



HAL
open science

Multilinguisation des systèmes de e-commerce traitant des énoncés spontanés en langue naturelle

Najeh Hajlaoui

► **To cite this version:**

Najeh Hajlaoui. Multilinguisation des systèmes de e-commerce traitant des énoncés spontanés en langue naturelle. Informatique [cs]. Université Joseph-Fourier - Grenoble I, 2008. Français. NNT : . tel-00337336

HAL Id: tel-00337336

<https://theses.hal.science/tel-00337336>

Submitted on 6 Nov 2008

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

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

Université Joseph Fourier de Grenoble

N° attribué par la bibliothèque

/ / / / / / / / / / / / / /

THESE

pour obtenir le grade de

DOCTEUR DE L'UJF

Spécialité : "INFORMATIQUE : SYSTEMES D'INFORMATION"

préparée au laboratoire GETALP-LIG (CNRS-INPG-UJF-UPMF)

dans le cadre de l'Ecole Doctorale "Mathématiques, Sciences et Technologie de l'Information"

présentée et soutenue publiquement

par

M. Najeh HAJLAOUI

le 25 septembre 2008

MULTILINGUISATION DE SYSTEMES DE E-COMMERCE TRAITANT DES ÉNONCÉS SPONTANÉS EN LANGUE NATURELLE

JURY

M. Ahmed LBATH	Président
M. Christophe ROCHE	Rapporteur
M. Jean VERONIS	Rapporteur
Mme. Violaine PRINCE	Rapporteur
M. Michel FANTON	Examinateur
M. Emmanuel PLANAS	Examinateur
M. Christian BOITET	Directeur de thèse
M. Hervé BLANCHON	Co-directeur de thèse

Remerciements

Je voudrais présenter mes vifs et sincères remerciements et ma profonde gratitude à mon directeur de thèse, le Professeur Christian BOITET, pour m'avoir accueilli au sein de son équipe et guidé tout au long de mes travaux de recherche. Il a su me conseiller efficacement et m'orienter dans le bon chemin de la recherche et aussi de l'enseignement. Malgré son emploi de temps chargé, il s'est montré toujours extrêmement disponible et efficace. Le temps que j'ai passé avec lui m'a permis de découvrir en lui un véritable esprit à la fois critique et constructif. Plus qu'un directeur, Christian est et sera toujours pour moi le maître qu'il a été durant toutes ces années. Ses qualités humaines et professionnelles ainsi que son aide sont pour moi un modèle à suivre.

J'aimerais, d'autre part, remercier mon co-directeur de thèse, Hervé BLANCHON, qui m'a également beaucoup aidé, avec patience et efficacité : il m'a accompagné de sa bienveillance pendant toute la durée de notre collaboration. Nos conversations quotidiennes passionnées ont été autant d'occasions pour moi de bénéficier de son savoir et de son expérience.

Je tiens à remercier vivement mes rapporteurs de thèse, le Professeur Christophe ROCHE, responsable des conférences TOTh et dans lesquelles j'ai eu l'occasion de publier deux fois, le Professeur Jean VERONIS, et le professeur Violaine PRINCE, deux spécialistes du domaine du TALN.

Je remercie enfin Michel FANTON, maître de conférence à l'Institut National des Langues et Civilisations Orientales « INALCO » et Emmanuel PLANAS, maître de conférence à l'université de l'Ouest à Angers d'avoir accepté de participer au jury, et le Professeur Ahmed LBATH d'avoir accepté de présider le jury.

Je n'oublierai pas de remercier tous les membres du laboratoire LIG, spécialement les membres de l'équipe GETALP pour les nombreuses discussions que nous avons eues ensemble, et Daoud DAOUD pour m'avoir permis d'accéder au code de son extracteur de contenu, ce qui m'a permis de procéder à trois expériences différentes de « portage » de CATS d'arabe en français.

Finalement, je remercie tous ceux qui m'ont aidé à terminer cette thèse, et également ceux qui la liront et m'enverront leurs commentaires.

Résumé : Nous nous intéressons à la multilinguisation, ou « portage linguistique » (plus simple que la localisation) des services de gestion de contenu traitant des énoncés spontanés en langue naturelle, souvent bruités, mais contraints par la situation, et constituant toujours un « sous-langage » plus ou moins restreint.

Un service de ce type (soit App) utilise une représentation du contenu spécifique (RC-App) sur laquelle travaille le noyau fonctionnel. Le plus souvent, cette représentation est produite à partir de la langue « native » L1 par un extracteur de contenu (EC-App). Nous avons dégagé trois méthodes de portage possibles, et les avons illustrées par le portage en français d'une partie de CATS, un système de traitement de petites annonces en SMS (en arabe) déployé à Amman, ainsi que sur IMRS, un système de recherche de morceaux de musique dont l'interface native est en japonais et dont seule la RC est accessible. Il s'agit de : (1) localisation « interne », i.e. adaptation à L2 de l'EC donnant EC-App-L2 ; (2) localisation « externe », i.e. adaptation d'un EC existant pour L2 au domaine et à la représentation de contenu de App (EC-X-L2-App); (3) traduction des énoncés de L2 vers L1.

Le choix de la stratégie est contraint par la situation traductionnelle : types et niveau d'accès possibles (accès complet au code source, accès limité à la représentation interne, accès limité au dictionnaire, et aucun accès), ressources disponibles (dictionnaires, corpus), compétences langagières et linguistiques des personnes intervenant dans la multilinguisation des applications.

Les trois méthodes ont donné de bons résultats sur le portage d'arabe en français de la partie de CATS concernant l'occasion automobile. En localisation interne, la partie grammaticale a été très faiblement modifiée, ce qui prouve que, malgré la grande distance entre l'arabe et le français, ces deux sous-langages sont très proches l'un de l'autre, une nouvelle illustration de l'analyse de R. Kittredge. La localisation externe a été expérimentée sur CATS et sur IMRS en adaptant au nouveau domaine considéré l'extracteur de contenu du français écrit initialement par H. Blanchon pour le domaine du tourisme (projet CSTAR/Nespole!), puis en changeant de langue pour IMRS (anglais).

Enfin, le portage par TA statistique a également donné de très bonnes performances, et cela avec un corpus d'apprentissage très petit (moins de 10.000 mots) et un dictionnaire complet. Cela prouve que, dans le cas de sous-langages très petits, la TA statistique peut être de qualité suffisante en partant de corpus 100 à 500 fois moins grands que pour de la langue générale.

Mots-clés : énoncés spontanés, langue naturelle, e-commerce, portage linguistique, sous-langage, extraction de contenu, traduction statistique, corpus.

Abstract: We are interested in the multilinguization, or “linguistic porting” (simpler than localization) of management content services processing spontaneous utterances in natural language, often noisy, but constrained by the situation and constituting a restricted “sublanguage”.

Any service of this type (App) uses a specific content representation (CR-App) on which the functional kernel operates. Most often, this representation is produced from the “native” language L1 by a content extractor (CE-App). We identified three possible methods of porting and have illustrated them by porting to French a part of CATS, a system handling small ads in SMS (in Arabic), deployed in Amman, as well as IMRS, a music retrieval system, where the native natural language interface is in Japanese and only the CR is accessible. These are: (1) “internal localisation”, i.e. adaptation to L2 of the CE, giving CE-App-L2; (2) “external” localization, i.e. adaptation of an existing CE for L2 to the domain and to the App content representation (CE-X-L2-App); (3) translation of utterances from L2 to L1.

The choice of the strategy is constrained by the translational situation: type and level of possible access (complete access to the source code, access limited to the internal representation, access limited to the dictionary, and no access), available resources (dictionaries, corpus), competences in languages and linguistics of persons taking part in the multilinguisation of application.

The three methods gave good results on the Arabic to French porting of the CARS part of CATS. For internal localization, the grammatical part was very little modified, which proves that, despite the great distance between Arabic and French, these two sub-languages are very near one to another. This is a new illustration of R.Kittredge’s analysis. The external localization was experimented with CATS and with IMRS by adapting to the new domain the French content extractor written initially by H. Blanchon for the tourism domain (CSTAR/Nespole! project), and then by changing the language for IMRS (English).

Finally, porting by statistical MT gave also a very good performance, and that with a very small training corpus (less than 10 000 words) and a complete dictionary. This proves that, in the case of very small sub-languages, statistical MT may be of sufficient quality, starting from a corpus 100 to 500 smaller than for the general language.

Keywords: spontaneous utterances, natural language, e-commerce, linguistic porting, sub-language, content extraction, statistical translation, corpora.

Abréviations

CRL	: Content Representation Language
DeCo	: outil de DéConversion (LSPL diffusé par le centre UNL de Tokyo)
EC	: Extracteur de Contenu
EI	: Extracteur d'Informations
EnCo	: outil d'EnConversion (LSPL diffusé par le centre UNL de Tokyo)
ESLN	: Énoncés Spontanés en Langue Naturelle
IF	: Interchange Format
LN	: Langage Naturel
LSPL	: Langage Spécialisé pour la Programmation Linguistique
RI	: Recherche d'Informations
SL	: Sous-Langage
SMT	: Statistical Machine Translation
TA	: Traduction Automatique
TALN	: Traitement Automatique des Langues Naturelles
TAS	: Traduction Automatique Statistique
UNL	: Universal Networking Language
XML	: Extensible Markup Language

Table des matières

Abréviations	5
Table des matières	7
Table des figures	11
Liste des tableaux	15
Introduction.....	17
Partie A Multilinguïisation des services de e-commerce.....	19
Chapitre 1 Multilinguïisation des services de e-commerce traitant des énoncés spontanés en langue naturelle	21
Introduction.....	21
1.1 Situation.....	21
1.1.1 Types d'applications et de services visés.....	21
1.1.2 Besoin d'utiliser les interfaces libres.....	23
1.1.3 Importance des services multilingues	24
1.2 Difficultés du traitement d'énoncés spontanés	24
1.2.1 Intérêt	24
1.2.2 Peu de services disponibles.....	24
1.2.3 Analyse préliminaire	25
1.3 Raisons possibles.....	26
1.3.1 Difficultés inhérentes au traitement automatique d'ESLN.....	26
1.3.2 Difficultés du portage linguistique.....	27
1.3.3 Difficultés du portage vers un autre domaine.....	29
Chapitre 2 État de l'art.....	31
Introduction.....	31
2.1 Portée de l'étude	31
2.1.1 Applications de e-commerce envisageables	31
2.1.2 Liste des applications considérées.....	33
2.2 Examen de quelques systèmes	34
2.2.1 NLSA	34
2.2.2 Pertinence Summarizer	41
2.2.3 CATS.....	43
2.2.4 IMRS	48
2.2.5 Récapitulation.....	52
2.3 Études méthodologiques antérieures	53
2.3.1 Génération multilingue : l'approche de MKBEEM	53
2.3.2 Études sur l'adaptation à une autre langue	55
2.3.3 Adaptation à un autre type de texte	61
2.3.4 Tableau récapitulatif.....	66
Chapitre 3 Démarche proposée	67
Introduction.....	67
3.1 Démarche	67
3.1.1 Choix de l'application	67
3.1.2 Méthodes de portage proposées.....	68

3.1.3	Ressources nécessaires	70
3.1.4	Techniques d'acquisition et de construction des ressources	72
3.1.5	Scénarios d'implémentation	74
3.1.6	Évaluation du portage	75
3.2	Portage par adaptation d'extracteur de contenu	75
3.2.1	Adaptation de l'extracteur de contenu de L1 à L2	75
3.2.2	Adaptation d'un extracteur de contenu de L2	76
3.2.3	Extracteur de contenu « universel »	76
3.3	Portage par TA	77
3.3.1	TA statistique	77
3.3.2	TA par analogie	78
Partie B Sous-langage, représentation de contenu et méthodes d'évaluation d'un portage linguistique		81
Chapitre 4 Caractérisation des sous-langages		83
Introduction		83
4.1	Définition d'un sous-langage naturel	83
4.1.1	Sous-langage selon Zellig Harris	84
4.1.2	Sous-langage selon Bross et autres	85
4.1.3	Sous-langage selon Kittredge et autres	85
4.2	Identification d'un sous-langage naturel	87
4.2.1	Phénomènes spécifiques dans un sous-langage	87
4.2.2	Comment caractériser un sous-langage	88
4.2.3	Autres caractérisations	90
4.3	Une taxonomie possible des sous-langages	94
4.3.1	Mesures possibles	94
4.3.2	Etude de quelques sous-langages	97
4.3.3	Inférence possible	101
Chapitre 5 Évaluation des représentations de contenu		103
Introduction		103
5.1	Représentations de contenu	103
5.1.1	Représentations classiques	104
5.1.2	Représentations spécifiques	107
5.2	Comparaison entre deux CRL-CATS	110
5.2.1	Algorithme de Wagner et Fischer	111
5.2.2	Algorithme de Selkow	113
5.3	Autre mesure	116
5.3.1	Arbres sémantiques	116
5.3.2	Algorithme de similarité entre arbres	117
5.3.3	Analyse critique	118
5.3.4	Simplicité d'un arbre	119
Conclusion		120
Chapitre 6 Évaluation d'un portage linguistique		121
Introduction		121
6.1	Langage d'entrée	121
6.1.1	Couverture	121
6.1.2	Mesure de l'efficacité	122
6.1.3	Coût	122
6.2	Représentation du contenu	122
6.2.1	Discussion générale	122
6.2.2	Application à CATS	123
6.3	Évaluation liée à la tâche	128
6.3.1	L'idéal	128

6.3.2	Avec des « cobayes »	128
6.3.3	Autre possibilité.....	128
6.3.4	En « magicien d’Oz ».....	129
	Conclusion	129
Partie C Implémentations et expériences de portage linguistique		131
Chapitre 7 Fabrication des données nécessaires pour le portage de CATS vers le français ..		133
	Introduction.....	133
7.1	Etude des données disponibles dans CATS et méthode de portage	133
7.1.1	Le corpus CATS	133
7.1.2	Dictionnaire.....	135
7.1.3	Vers le portage.....	136
7.2	Corpus.....	138
7.2.1	Par traduction.....	138
7.2.2	Étude de la traduction d’un corpus réel.....	138
7.2.3	Évaluation de cette hypothèse.....	139
7.3	Dictionnaire.....	144
7.3.1	Comparaison dictionnaire.....	145
7.3.2	Recouvrement du domaine	145
7.3.3	Conservation de l’utilisation de l’ontologie.....	146
Chapitre 8 Portage « interne ».....		149
	Introduction.....	149
8.1	Extracteur de contenu pour les SMS en arabe	149
8.1.1	Architecture générale de l’outil EnCo.....	149
8.1.2	Fonctionnement d’EnCo	150
8.1.3	Règles	151
8.1.4	Dictionnaire.....	152
8.2	Adaptation de l’extracteur de contenu.....	153
8.2.1	Exemple de traitement.....	153
8.2.2	Légère modification des règles.....	156
8.2.3	Exemple d’ambiguïté	156
8.3	Résultats et évaluations	157
8.3.1	Coût du portage interne	157
8.3.2	Mesure informationnelle	158
8.3.3	Évaluation générale	161
	Conclusion	163
Chapitre 9 Portage « externe »		165
	Introduction.....	165
9.1	Description de l’existant : EC pour le domaine du tourisme	165
9.1.1	Spécification de l’IF	165
9.1.2	Ce dont on dispose et ce qu’on souhaite	167
9.1.3	Méthode d’analyse dans Nespole!	167
9.2	Adaptation à CATS	170
9.2.1	Spécification de l’IF-CATS	170
9.2.2	Adaptation de l’extracteur.....	171
9.2.3	Traduction IF-CATS vers CRL-CATS	174
9.2.4	Évaluation et résultats	177
9.3	Adaptation à IMRS	183
9.3.1	Étude du sous-langage de IMRS	183
9.3.2	Méthode d’adaptation.....	184
9.3.3	Résultats et coût.....	185

Conclusion	186
Partie D Exploration d'une méthode de portage par TA	189
Chapitre 10 Architectures possibles pour la TA.....	191
Introduction.....	191
10.1 Architectures calculatoires envisageables.....	191
10.2 TA statistique à petits corpus pour de petits sous-langages.....	193
10.3 Démarche	194
10.3.1 Génération de corpus.....	194
10.3.2 Traduction Automatique Statistique : TAS	194
10.3.3 Évaluation	195
Chapitre 11 Production de corpus parallèles par déconversion	197
Introduction.....	197
11.1 Ressources et outils mis en oeuvre.....	197
11.1.1 Ressources.....	197
11.1.2 Outils	198
11.1.3 Ce que nous souhaitons.....	202
11.2 Déconversion à partir de CRL-CATS	202
11.2.1 Entrée	202
11.2.2 Sortie	203
11.2.3 Dictionnaire	203
11.2.4 Construction du déconvertisseur	204
11.3 Résultats et coûts	206
11.3.1 Résultats.....	206
11.3.2 Coût.....	207
11.3.3 Remarques sur la gestion des arguments	208
Chapitre 12 Portage par traduction statistique	209
Introduction.....	209
12.1 Travail préparatoire pour la traduction.....	209
12.1.1 Ressources utilisées.....	209
12.1.2 Alignement GIZA++.....	210
12.1.3 Analyse des résultats d'alignement	212
12.2 Mise en place du système.....	213
12.2.1 Modélisation du langage cible.....	214
12.2.2 Pharaoh en mode apprentissage.....	214
12.2.3 Pharaoh en mode développement.....	216
12.3 Résultats et évaluations	217
12.3.1 Évaluation de la traduction	217
12.3.2 Évaluation de l'extraction d'information.....	219
12.3.3 Évaluation générale.....	225
12.3.4 Coût de la méthode.....	227
Conclusion de la partie D.....	229
Conclusion et perspectives.....	230
Bibliographie	234
Annexes.....	241

Table des figures

Figure 1 : architecture générale des applications de traitement du contenu d'énoncés en LN	17
Figure 2 : exemple de traduction du résultat d'une recherche sur e-Bay (anglais vers français)	22
Figure 3 : exemple d'enchères de voiture	32
Figure 4 : recherche d'informations à l'aide des pages jaunes	33
Figure 5 : exemple de dialogue en anglais britannique	36
Figure 6 : architecture du système NLSA	37
Figure 7 : interface initiale pour utilisateur	38
Figure 8 : exemple de structure dans le format LG	39
Figure 9 : fragment du concept <Advanced-Graphic-Capability>	40
Figure 10 : Pertinence Summarizer en ligne	42
Figure 11 : exemple de texte traité par Pertinence Summarizer	42
Figure 12 : architecture du système Pertinence Summarizer	43
Figure 13 : exemple de SMS envoyé à CATS	44
Figure 14 : exemple de SMS envoyé par CATS	44
Figure 15 : exemples de SMS traités par CATS	44
Figure 16 : architecture du système CATS	45
Figure 17 : exemple de représentation CRL-CATS	46
Figure 18 : exemples de CRL-CATS transformés en requêtes SQL	46
Figure 19 : exemple d'entrées du dictionnaire CATS	47
Figure 20 : écran de choix des impressions (Kumamoto 2007)	49
Figure 21 : résultats trouvés par IMRS	50
Figure 22 : étapes pour générer les vecteurs requêtes à partir des énoncés	52
Figure 23 : exemple de texte annoté et aligné au niveau des mots	57
Figure 24 : portage d'un EC basé sur la TBL	59
Figure 25 : mode normal d'Amilcare	62
Figure 26 : exemple de texte d'entraînement	63
Figure 27 : exemple de généralisation à partir de la règle initiale	65
Figure 28 : généralisation possible de la règle de la figure précédente	65
Figure 29 : méthode de portage interne	68
Figure 30 : méthode de portage externe	69
Figure 31 : méthode de portage par TA	70
Figure 32 : langue générale, langue standard, et sous-langage	87
Figure 33 : exemple de calcul du TTR	91
Figure 34 : évolution de la taille de la grammaire générée en fonction de la taille du corpus (Lopez, Fay-Varnier et al. 1999)	93
Figure 35 : quelques classes et sous-classes pour le domaine de l'automobile dans CATS	95
Figure 36 : analyse du sous-langage de l'automobile dans CATS	98

Figure 37 : analyse du sous-langage de l'immobilier dans CATS	99
Figure 38 : analyse du sous-langage des bulletins météo.....	101
Figure 39 : exemple de représentation sous forme de liste de propriétés	104
Figure 40 : exemple de représentation vectorielle.....	104
Figure 41 : exemple de représentation sous forme de structure de traits avec disjonction.....	105
Figure 42 : exemple de représentation sous forme de graphe conceptuel.....	107
Figure 43 : exemple de représentation sous forme d'arbre.....	107
Figure 44 : exemple de représentation CRL-CATS (1)	108
Figure 45 : exemple de représentation CRL-CATS (2)	109
Figure 46 : exemple de représentation IF.....	109
Figure 47 : représentation d'un texte sous forme d'arbre.....	110
Figure 48 : algorithme de Wagner et Fischer.....	111
Figure 49 : calcul de trace à partir de l'algorithme de Wagner et Fischer.....	112
Figure 50 : algorithme de Wagner et Fisher (exemple 1)	112
Figure 51 : algorithme de Wagner et Fisher (exemple 2)	113
Figure 52 : algorithme de Selkow pour le calcul de la distance entre deux arbres ordonnés, orientés et étiquetés.....	114
Figure 53 : exemple de tri de deux CRL-CATS	114
Figure 54 : exemple détaillé de l'application de l'algorithme de Selkow.....	115
Figure 55 : schéma standard pour e-Learning (Bhavsar, Boley et al. 2004).....	116
Figure 56 : instance d'arbre générée à partir du schéma précédent.	117
Figure 57 : exemples d'arbre dans le domaine du e-learning.....	117
Figure 58 : exemple d'arbre dans le domaine du e-commerce	117
Figure 59 : formule de calcul de similarité entre deux arbres.....	118
Figure 60 : exemples d'arbres avec des poids d'arcs opposés.....	119
Figure 61 : exemple de calcul de similarité	119
Figure 62 : exemple de calcul de simplicité.....	120
Figure 63 : exemple de différences entre les CRL-CATS obtenues par deux extracteurs de contenu différents.....	125
Figure 64 : affectation des poids aux dix attributs de CRL-CATS	126
Figure 65 : exemple du corpus SMS_ar_orig	134
Figure 66 : exemple de deux SMS arabes en format CPXM.....	135
Figure 67 : exemple d'entrées du dictionnaire arabe	135
Figure 68 : exemple de SMS du corpus arabe	138
Figure 69 : exemple de traduction brute.....	138
Figure 70 : exemple de traduction littérale	138
Figure 71 : exemple de traduction fonctionnelle	138
Figure 72 : exemple de calcul de distance avec une représentation en suivi des modifications	140
Figure 73 : même exemple de calcul de distance avec une représentation sur 3 lignes.....	141
Figure 74 : exemples de traductions brutes, littérales et fonctionnelles.....	141
Figure 75 : mesures de distance d'édition	142
Figure 76 : mesures de similarité.....	143
Figure 77 : quelques exemples d'entrées dictionnaires secondaires.....	144

Figure 78 : convergence d'entrées dictionnairiques vers un même CW.....	145
Figure 79 : évolution de la taille du dictionnaire en fonction du corpus.....	146
Figure 80 : passage du sens à la signification en utilisant l'ontologie.....	147
Figure 81 : architecture de l'extracteur de contenu pour l'arabe	149
Figure 82 : structure d'EnCo, "A" indique la fenêtre d'analyse, "C" étiquette une fenêtre de Condition (Uchida and Zhu 1999).....	150
Figure 83 : fonctionnement d'EnCo.....	151
Figure 84 : syntaxe d'une règle en EnCo.....	151
Figure 85 : syntaxe d'une entrée dictionnairique	152
Figure 86 : configuration initiale d'EnCo (Daoud 2006).....	153
Figure 87 : configuration initiale de « recherche voiture »	153
Figure 88 : segmentation de « recherche voiture »	154
Figure 89 : configuration d'EnCo après un <i>shift right</i>	154
Figure 90 : suppression du BLANK et <i>shift right</i>	155
Figure 91 : création d'une relation dans le graphe des nœuds.....	156
Figure 92 : les sous-langages sont proches	156
Figure 93 : exemple d'ambiguïté (première interprétation)	157
Figure 94 : exemple d'ambiguïté (deuxième interprétation).....	157
Figure 95 : comparaison entre F-mesure (par rapport à la version originale).....	159
Figure 96 : comparaison entre F-mesures (par rapport à une version de référence)	161
Figure 97 : comparaison des scores des attributs les plus importants par rapport à la version originale sur le corpus « CorpusEvalFr1100SMS ».....	162
Figure 98 : extrait de la spécification de l'IF	166
Figure 99 : exemple d'extraction vers l'IF.....	167
Figure 100 : architecture du module d'analyse français → IF du second démonstrateur Nespole! (Blanchon 2004).	168
Figure 101 : exemple de mauvaise segmentation en SDU.....	171
Figure 102 : exemple de combinaison possible des arguments	173
Figure 103 : recherche et de traitement des arguments	174
Figure 104 : exemple de sortie CRL-CATS.....	174
Figure 105 : exemple de sortie IF-CATS	175
Figure 106 : transformation de IF-CATS vers CRL-CATS.....	175
Figure 107 : procédure permettant le passage de IF-CATS à CRL-CATS.....	176
Figure 108 : exemple du dictionnaire (IF-CATS, CRL-CATS)	176
Figure 109 : exemple de convergence d'arguments IF-CATS vers une seule propriété CRL-CATS	176
Figure 110 : exemple de sortie du traducteur IF-CATS.....	177
Figure 111 : comparaison entre F-mesures (par rapport à la version originale).....	180
Figure 112 : comparaison entres F-mesures (par rapport à une version de référence)	181
Figure 113 : comparaison des scores obtenus par portage externe par rapport à la version originale	182
Figure 114 : analyse du sous-langage de IMRS	184
Figure 115 : recherche et de traitement des arguments liés au domaine de la musique.....	184

Figure 116 : exemple de résultats pour le français	185
Figure 117 : exemple de résultats pour l'anglais	186
Figure 118 : exemple d'un SMS arabe et de sa représentation CRL-CATS.....	198
Figure 119 : modèle de calcul du langage DeCo	198
Figure 120 : exemple d'entrées dictionnairiques en DeCo	200
Figure 121 : exemple d'entrées dictionnairiques numériques	200
Figure 122 : architecture de notre cas d'utilisation de DeCo.....	202
Figure 123 : liste d'entrées candidates triées pour un exemple d'UW	203
Figure 124 : exemple de mauvaise sortie	204
Figure 125 : exemple de sortie correcte	205
Figure 126 : deuxième exemple de mauvaise sortie.....	206
Figure 127 : troisième exemple de mauvaise sortie	206
Figure 128 : quelques exemples de bonnes sorties de déconversion.....	206
Figure 129 : conversion des données du format texte vers snt	210
Figure 130 : exemple de corpus-ar et de corpus-fr	211
Figure 131 : fichier de configuration (config.default) pour GIZA++	212
Figure 132 : résultat d'alignement pour le corpus exemple	212
Figure 133 : exemple d'alignement correct	213
Figure 134 : exemple d'alignement d'un mot composé.....	213
Figure 135 : exemple de mauvais alignement.....	213
Figure 136 : architecture générale d'un système de traduction statistique	214
Figure 137 : fichier de configuration pour Pharaoh « pharaohh.ini »	215
Figure 138 : extrait du fichier phrase-table	216
Figure 139 : exécution du script <i>pharaoh.perl</i>	216
Figure 140 : exemple de résultat obtenu par traduction statistique (Pharaoh).....	217
Figure 141 : BLEU en fonction de la taille du corpus d'entraînement.....	218
Figure 142 : NIST en fonction de la taille du corpus d'entraînement.....	219
Figure 143 : quelques exemples de CRL-CATS obtenus par traduction (FR-AR) et par la version originale.....	221
Figure 144 : comparaison entre F-mesure (par rapport à la version originale)	222
Figure 145 : quelques exemples de CRL-CATS obtenus par traduction statistique (FR-AR) et par la version de référence.....	224
Figure 146 : comparaison entre F-mesures (par rapport à une version de référence)	225
Figure 147 : comparaison des scores des attributs les plus importants obtenus par la méthode par traduction par rapport à la version originale	226

Liste des tableaux

Tableau 1 : analyse des applications les plus importantes par pays, langue et type de service.....	26
Tableau 2 : liste de applications considérés	34
Tableau 3 : TTR (Type/Token Ratio) dans les sous-corpus de CATS.....	47
Tableau 4 : dix axes (définis par des paires) pour représenter les impressions musicales.....	48
Tableau 5 : extrait des règles d'interprétation contrôlant le passage entre les impressions et les vecteurs requêtes (tiré de (Kumamoto 2007))	51
Tableau 6 : tableau récapitulatif des principaux systèmes d'e-commerce étudiés.....	53
Tableau 7 : exemples de règles TBL (LOC=location, VEH=vehicule, VIC=victim)	58
Tableau 8 : performance du portage de l'EC basée sur le TBL.....	61
Tableau 9 : tableau récapitulatif des principales approches étudiées	66
Tableau 10 : types de langues et caractéristiques associées.....	87
Tableau 11 : mesure de la complexité lexicale	94
Tableau 12 : mesure de la difficulté lexicale	95
Tableau 13 : mesure de la longueur des énoncés.....	96
Tableau 14 : mesure de la complexité syntaxique.....	96
Tableau 15 : mesure de la convergence grammaticale.....	97
Tableau 16 : mesure de la convergence lexicale.....	97
Tableau 17 : exemples d'énoncés du sous-langage de l'automobile en Jordanie.....	98
Tableau 18 : exemples d'énoncés du sous-langage de l'immobilier à Amman.....	99
Tableau 19 : échantillon de bulletins météorologiques	100
Tableau 20 : normalisation d'énoncés de même contenu	103
Tableau 21 : exemple de calcul de distance nulle entre deux RC de CATS.....	115
Tableau 22 : exemple de calcul de distance non nulle entre deux RC de CATS	116
Tableau 23 : comparaison entre versions originale, cible, et référence	123
Tableau 24 : comparaison entre version cible et version originale	123
Tableau 25 : exemple d'application de la mesure informationnelle	124
Tableau 26 : exemple d'application de la mesure simple	126
Tableau 27 : exemple de calcul de distance nulle entre CRL-CATS	127
Tableau 28 : mesure de la satisfaction des utilisateurs.....	128
Tableau 29 : champs du corpus original SMS_ar_orig	134
Tableau 30 : évaluation de la durée des tâches pour le portage interne de CATS	158
Tableau 31 : comparaison entre les résultats d'extraction de contenu pour 200 SMS d'achat et de vente par rapport à la version originale	159
Tableau 32 : exemple de correction de la représentation CRL-CATS.....	160
Tableau 33 : comparaison entre les résultats d'extraction de contenu pour 200 SMS d'achat et de vente par rapport à une version de référence	160
Tableau 34 : scores du portage interne sur la totalité du corpus de 1100 SMS	161

Tableau 35 : comparaison des pourcentages de différences des scores d'extraction de contenu sur le corpus « CorpusEvalFr1100SMS »	162
Tableau 36 : résultats de la mesure fine sur le corpus « CorpusEvalFr1100SMS ».....	163
Tableau 37 : exemple de découpage en SDU	168
Tableau 38 : évaluation de la durée des tâches pour le portage externe	178
Tableau 39 : comparaison entre les résultats d'extraction de contenu pour 200 SMS obtenus par portage externe par rapport à la version originale	179
Tableau 40 : comparaison entre les résultats d'extraction de contenu obtenus par portage externe pour 200 SMS par rapport à une version de référence	180
Tableau 41 : scores trouvés pour le portage externe sur tout le corpus de 1100 SMS	181
Tableau 42 : comparaison des pourcentages de différences des scores d'extraction de contenu sur le total du corpus.....	182
Tableau 43 : résultats du portage externe par la mesure fine sur « CorpusEvalFr1100SMS »	183
Tableau 44 : dix axes (définis par des paires) pour représenter les impressions musicales en français et en anglais.....	185
Tableau 45 : syntaxe des règles de DeCo (1).....	201
Tableau 46 : syntaxe des règles de DeCo (2).....	201
Tableau 47 : syntaxe des règles de DeCo (3).....	201
Tableau 48 : niveaux de la sortie de l'outil DeCo	203
Tableau 49 : exemples de résultat de déconversion	207
Tableau 50 : durée des tâches de déconversion	208
Tableau 51 : taille des données d'entraînement.....	210
Tableau 52 : fragments du vocabulaires générés par le script plain2snt.....	211
Tableau 53 : fragments du format numérique généré par le script plain2snt.....	212
Tableau 54 : mesure de NIST et BLEU en fonction de la taille du corpus d'entraînement	218
Tableau 55 : comparaison entre les résultats d'extraction de contenu pour 200 SMS obtenus par traduction statistique par rapport à la version originale	221
Tableau 56 : comparaison entre les résultats d'extraction de contenu obtenus par TA statistique pour 200 SMS par rapport à une version de référence.....	224
Tableau 57 : scores trouvés pour la méthode par traduction sur le total du corpus de 1100 SMS...	225
Tableau 58 : comparaison des pourcentages de différences des scores obtenus par la méthode de traduction sur le total du corpus.....	226
Tableau 59 : résultats du portage par TA par la mesure fine sur « CorpusEvalFr1100SMS ».....	227
Tableau 60 : évaluation de la durée des tâches pour la méthode de portage externe.	228
Conclusion de la partie D	229
Tableau 61 : récapitulatif des résultats d'évaluation informationnelle sur le corpus « CorpusEvalFr200SMS »	231
Tableau 62 : récapitulatif des résultats d'évaluation générale sur « CorpusEvalFr1100SMS ».....	231
Tableau 63 : récapitulatif des coûts des trois méthodes de portage.....	231

Introduction

Une conséquence majeure résultant de la mondialisation que nous vivons aujourd'hui est l'importance croissante du multilinguisme, accentuée par l'utilisation d'Internet et le multimédia. Face à ce phénomène, les développeurs d'applications informatiques ne peuvent pas encore fournir des applications utilisables dans la langue maternelle de tous les utilisateurs. Les services et les applications sont rarement disponibles dans de nombreuses langues, et encore moins traitent des textes naturels rédigés de façon spontanée et pouvant contenir des erreurs, des abréviations, etc.

Dans ce cadre, nous nous intéressons à un problème plus précis, celui de la multilinguisation des applications de e-commerce traitant des énoncés spontanés en langue naturelle (LN). Ces applications extraient le « contenu pertinent » des énoncés en langue naturelle, le représentent dans un langage approprié (CRL), et traitent ensuite les objets obtenus.

L'extraction de contenu est rarement fondée sur une analyse complète des énoncés : on utilise le plus souvent des « grammaires locales » et un dictionnaire mettant en relation des termes du domaine et les symboles (concepts, attributs, relations) du CRL. L'architecture des applications qui nous intéressent est donc la suivante :

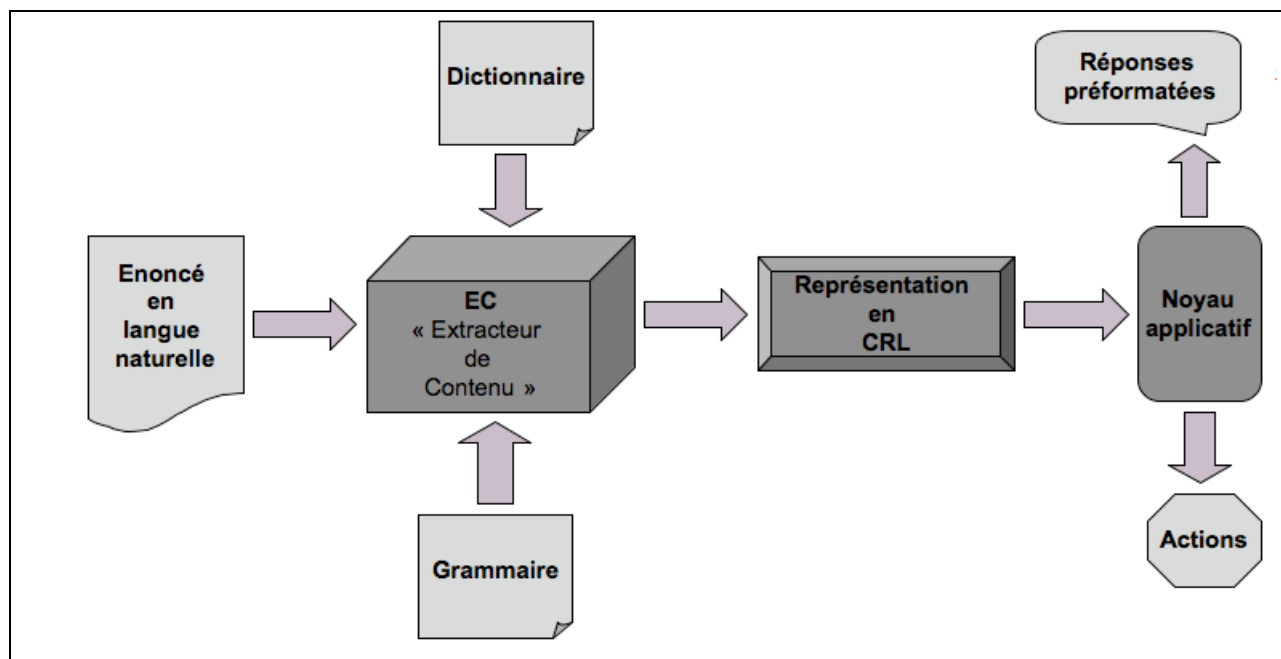


Figure 1 : architecture générale des applications de traitement du contenu d'énoncés en LN

La multilinguisation de telles applications est en fait un problème difficile, ce qui explique que très peu de services le proposent. La difficulté du problème dépend de deux facteurs relatifs à la situation traductionnelle :

- le niveau d'accès aux ressources des applications, pour lequel quatre situations se présentent : accès complet au code source, accès à la représentation interne des données manipulées uniquement, accès aux dictionnaires uniquement, et aucun accès.

- le niveau de compétence langagière et linguistique des intervenants pour la multilinguïisation des applications. Il peut être défini par rapport à la langue source ou par rapport aux compétences linguistiques de l'équipe qui veut faire la localisation (spécialiste ou non de TALN).

La multilinguïisation ou le « portage linguistique » dont nous parlons ici n'est pas nécessairement une « localisation ». Une localisation implique une adaptation à un autre contexte culturel. Par contre, un portage linguistique doit seulement permettre l'accès dans une autre langue à un service de e-commerce, tel qu'il est et où il est.

Notre but est de trouver, pour multilingualiser de telles applications, des solutions simples, peu coûteuses en termes de ressources et de temps, efficaces, applicables sur le terrain, adaptées à la situation traductionnelle, et en réutilisant ce qui est disponible comme outils, ressources et capacités langagières et linguistiques.

L'apport théorique principal de la thèse consiste en la démonstration ou la vérification des trois hypothèses suivantes :

- Bien que le traitement des énoncés spontanés soit difficile, le passage d'un sous-langage à son correspondant dans une autre langue est parfois très simple. C'est l'hypothèse de Kittredge (proximité grammaticale de deux sous-langages même si les deux langues-mères sont distantes).
- La traduction « usuelle », même de grande qualité, n'est pas bonne entre deux sous-langages. En conséquence, il faut corriger le résultat pour obtenir des traductions de même nature opératoire dans le sous-langage obtenu (spontanées, naturelles).
- La traduction statistique peut « bien marcher » pour l'extraction de contenu en se basant sur un corpus très petit mais formé de paires d'énoncés corrects (SL_1 , SL_2) et sur un dictionnaire complet.

L'apport pratique est de :

- proposer et expérimenter des méthodes de multilinguïisation.
- évaluer l'effort humain dans quelques cas représentatifs.
- définir et mettre en œuvre une méthode d'évaluation du résultat d'un portage linguistique.

La thèse est organisée en quatre parties. La première partie présente le contexte et la position du problème, la deuxième partie évalue a priori la difficulté d'un problème de portage linguistique, la troisième partie présente deux méthodes de portage par adaptation d'un extracteur de contenu « interne » et « externe », et la quatrième partie présente une troisième méthode de portage « frontal » par traduction des énoncés de la langue cible vers la langue « native ».

Partie A

Multilinguisation des services de e-commerce

Nous présentons dans cette partie une analyse de la situation actuelle, qui montre que, malgré l'intérêt de multilingualiser les services de gestion de contenu, très peu de développeurs le font. Nous détaillons les raisons de cet état de fait.

Nous faisons un état de l'art sur les systèmes et les prototypes qui ont été développés pour traiter ce type de textes spontanés et sur leur éventuel portage vers d'autres langues.

Finalement, nous présentons notre démarche scientifique, dans laquelle nous déterminons les notions et les difficultés principales, à savoir la construction d'un corpus réel fonctionnellement équivalent dans la nouvelle langue, et d'un dictionnaire. Nous définissons une méthode de travail apportant des solutions possibles suivant l'expérience passée, la situation présente, et la tâche concernée.

Chapitre 1 **Multilinguïisation des services de e-commerce**

traitant des énoncés spontanés en langue naturelle

Introduction

Nous allons d'abord préciser le type d'applications qui nous intéresse et les possibilités de les rendre disponibles dans d'autres langues que celle de départ.

Nous montrerons la spécificité du traitement du « sous-langage » particulier formé des énoncés spontanés en langue naturelle engendré par l'application en question et conclurons qu'on ne peut pas utiliser les outils faits pour le langage écrit « général » et « propre », basés sur un processus d'analyse, mais qu'il faut implémenter un processus d'extraction de contenu, guidé par la connaissance du domaine et de la tâche.

Nous présentons, dans la première section, une définition des applications qui nous intéressent dans cette étude, et l'illustrons par quelques exemples. Dans la deuxième section, nous montrons que, malgré l'importance croissante de ce type d'applications, très peu sont « multilinguïées », alors que le besoin de leur localisation semble bien réel. Dans la troisième section, nous essayons de faire une analyse des raisons possibles du manque de services traitant les énoncés spontanés en langue naturelle.

1.1 Situation

(Coupet, Buschbeck et al. 2005) montre une demande de plus en plus forte de services de gestion de contenu multilingue. En effet, de tels services sont toujours plus importants et plus efficaces s'ils sont disponibles dans la langue maternelle de chaque utilisateur.

1.1.1 Types d'applications et de services visés

Nous nous intéressons à un type particulier de service de gestion de contenu, ceux dans lesquels les conditions suivantes doivent être satisfaites : pas de traduction, interface « libre », besoin « naturel » de multilinguïisation.

1.1.1.1 Pas de la traduction

Dans cette étude, on ne s'intéresse pas aux applications commerciales qui ne font pas de traitement de contenu comme e-Bay. De telles applications sont nombreuses, comme celles qui font un stockage de la version originale des données dans une base de données, et qui pour le besoin multilingue, font un appel à des systèmes de traduction.

Dans d'autres applications, on trouve une génération multilingue de données qui sont déjà traduites en plusieurs langues comme la documentation technique ou les manuels d'utilisation de matériels techniques ou de logiciels informatiques. Par exemple, au Canada, les bulletins météo doivent être disponibles dans les deux langues française et anglaise, ils sont traduits grâce au système Météo (Chandioux, 1988).


Sur le Web, les offres d'emploi, les petites annonces de vente et d'achat sont disponibles par traduction en plusieurs langues. Par exemple, le site très connu e-Bay est disponible dans plus d'une

vingtaine de pays, le choix d'un pays affiche le site correspondant dans la langue du pays sélectionné. Si on élargit le critère de recherche en précisant le lieu de l'objet (en choisissant par exemple « *dans le monde entier* » au lieu de « *en France uniquement* »), un objet situé dans un autre pays peut être disponible par traduction dans la langue de navigation.

Caractéristiques de l'objet - Cars & Other Vehicles	
Manufacturer:	Fiat
Model:	Uno
Type:	Standard Car
Mileage:	75372
Number of Doors:	--
Reg. Date:	31 Août 1995
Previous Owners:	3
Transmission:	Manual
Fuel:	Diesel
Service History:	--
Safety Features:	Rear seat belts
Colour:	Red
Engine Size:	1,697 cc
Power:	--
Seats:	4
MOT Expiry:	--
Drive Side:	Right-hand drive
Road Tax:	--
Exterior Options:	--
V5 Doc.:	Present
Manufacturer's Warranty:	--
In-Car Audio Options:	AM/FM Stereo, Cassette Player
Interior/Comfort Options:	--

[Afficher les Caractéristiques de l'objet dans la langue locale.](#)

VERY GOOD RUNNER, STARTS FIRST TIME , NO RUST , LOW MILAGE ,DIESAL VERY GOOD ON FUEL.CRACK TO FRONT BUMPER AND RIGHT HAND INDICATER CASEING BROKE EASILY FIXED.GOOD TYRES NO MOT .YOU CAN NOT GO WRONG AT THIS PRICE.



Traduction fournie en tant que service,
son exactitude n'est pas garantie par e-Bay

Caractéristiques de l'objet - Voitures et autres véhicules	
Constructeur:	Fiat
Modèle:	Uno
Type:	Berline
Kilométrage:	75372
Nombre de portes:	--
Date of 1st Registration (DD/MM/YYYY):	31 Août 1995
Nombre de propriétaires précédents:	3
Boîte de Vitesse:	Manuelle
Carburant:	Diesel
Camet d'entretien disponible:	--
Equipements de Sécurité:	Ceintures de sécurité sur les sièges arrière
Couleur:	Rouge
Cylindrée (centimètres cubes):	1 697 cm3
Puissance (ch DIN):	--
Nombre de places:	4
Date du Contrôle Technique:	--
Volant à droite/Volant à gauche:	Volant à droite
Vignette:	--
Equipements Extérieurs:	--
Document d'inscription V5:	Present
Date expiration garantie constructeur:	--
Equipements Audio & Navigation:	Autoradio Ecran DVD, Autoradio K7
Equipements de Confort:	--

Cette traduction est fournie en tant que service. eBay ne sauraît garantir son exactitude.

[Afficher les Caractéristiques de l'objet dans la langue d'origine.](#)

VERY GOOD RUNNER, STARTS FIRST TIME , NO RUST , LOW MILAGE ,DIESAL VERY GOOD ON FUEL.CRACK TO FRONT BUMPER AND RIGHT HAND INDICATER CASEING BROKE EASILY FIXED.GOOD TYRES NO MOT .YOU CAN NOT GO WRONG AT THIS PRICE.

Figure 2 : exemple de traduction du résultat d'une recherche sur e-Bay (anglais vers français)

La Figure 2 est un exemple de résultat d'une petite annonce faite manuellement¹ à l'aide des menus de l'interface e-Bay, qui a été regénéré par traduction en français². Nous avons cherché sur le site français de e-Bay correspondant au pays « France » une voiture « Fiat » en précisant quelques

¹ L'extraction de contenu ici est manuelle (faite par l'utilisateur qui soumet son annonce à l'aide des menus).

² Ce n'est pas l'annonce qui est traduite mais ce sont les caractéristiques de l'objet (parfois mal traduites comme « AM/FM Stereo » traduite par « Autoradio Ecran DVD »).

critères à l'aide des menus de l'interface du site. Pour avoir plus de choix, nous avons élargi notre recherche au monde entier. L'objet de la Figure 2 qui se situe au Royaume-uni a été trouvé.

1.1.1.2 Caractéristique : traitement du contenu

Nous nous intéressons dans cette étude aux applications qui donnent de la valeur ajoutée par le traitement du contenu. C'est la caractéristique principale que nous imposons dans le choix de notre application. Un traitement de contenu peut avoir plusieurs catégories (extraction de contenu, résumé, catégorisation, recherche d'information, etc).

Bien sûr, nous ne négligeons pas les applications qui font de la gestion de données après leur traitement par enregistrement dans une base de données ou d'autres structures de données.

1.1.1.3 Principaux types d'application et de services

Les principaux types d'applications et de services pertinents sont :

- Catégorisation de documents divers (dépêches de l'AFP « Agence France-Presse », brèves des différentes bourses, messages de clients à un serveur de SAV (Service Après Vente)).
- Extraction d'informations pour nourrir ou consulter une base de données (exemple : petites annonces ciblées, FAQ « Foire Aux Questions » intelligentes, indexation ciblée à un domaine/métier jusqu'aux hotlines automatisées, rapports d'incidents par des utilisateurs).
- Traitement du Web pour des études d'opinion par exemple.
- Traitement du Web pour la catégorisation de courriels, le résumé automatique, etc.
- Recherche d'informations.
- Commerce : achat et vente de produits, assistance à la vente, etc.

1.1.2 Besoin d'utiliser les interfaces libres

Malgré le progrès et la multiplication des méthodes de TALN, elles sont très peu ou presque pas utilisées dans le genre d'applications décrites précédemment. Ce sont les menus et les formulaires qui sont plutôt utilisés pour répondre à des besoins commerciaux, qu'il s'agisse de catégorisation, de commerce, etc. La navigation et la recherche par mots-clés pilotées par des menus dans la plupart des sites commerciaux ont pourtant des limitations importantes : elles tendent à accabler et frustrer des utilisateurs avec des interactions prolongées et rigides. L'intérêt de l'utilisateur pour un site particulier diminue exponentiellement avec l'augmentation du nombre de clics de souris (Huberman, Pirolli et al. 1998). Par conséquent, le raccourcissement du chemin d'interaction pour obtenir des informations utiles devient important.

Beaucoup de sites de e-commerce essaient de résoudre ce problème en fournissant des possibilités de recherche par mots-clés. Cependant, les moteurs de recherche par mots-clés exigent habituellement que les utilisateurs connaissent le jargon spécifique du domaine. La recherche par mots-clés ne permet pas à des utilisateurs de décrire avec précision leurs intentions, et d'une manière primordiale, il leur manque les bons mots et les bonnes expressions de recherche.

Les utilisateurs frustrés et incapables de suivre des étapes de remplissage de formulaires et de recherche sont obligés d'abandonner ce mode d'interaction pour passer à une communication directe avec le propriétaire du service, soit en téléphonant soit en se déplaçant. Cela affaiblit le caractère autonome du service proposé censé fonctionner sans l'intervention d'une personne, et coûte très cher en temps pour répondre aux clients.

En conséquence, l'utilisation des interfaces libres dans lesquelles l'utilisateur peut exprimer sa requête (recherche, mécontentement, description d'un problème technique...) d'une façon très spontanée et non contrôlée peut dépasser les limites et les difficultés rencontrées dans l'utilisation des mots-clés et des formulaires.

1.1.3 Importance des services multilingues

Un service multilingue possède plusieurs avantages, il peut être déployé dans de nombreux pays. Par exemple, la société CISCO (<http://www.cisco.fr/>) utilise des systèmes de traduction pour traduire (mal) sa documentation. Il en résulte que les centres d'appel sont surchargés et coûtent très cher.

Le site numéro 1 de petite annonces en France (<http://www.paruvendu.fr/>) est disponible en une seule langue (français). Cela empêche un grand nombre d'étrangers ne parlant pas le français d'utiliser ce service assez connu pour répondre à leurs besoins quotidiens.

Les deux exemples précédents montrent bien la nécessité de services multilingues : bien sûr, la multilinguisation est nécessaire dans des pays multilingues (Canada, Inde, US), mais aussi dans des pays monolingues (France) à cause du tourisme et de la nouvelle mobilité.

1.2 Difficultés du traitement d'énoncés spontanés

1.2.1 Intérêt

L'intérêt d'un service de traitement d'énoncés spontanés se résume en différents points :

- Caractère naturel de l'interaction.
- Possibilité pour les utilisateurs de s'exprimer plus librement que ce que permettent les formulaires et les menus, à caractère modal et figé.
- Possibilité (plus récente) de construire et de faire évoluer la modélisation du domaine d'un service (thésaurus, ontologie, métier, etc).

Un service de traitement d'énoncés spontanés peut répondre à une lacune technique et ergonomique dans le domaine du e-commerce surtout, essentiellement aux problèmes soulevés par la navigation par mots-clés pilotées par des menus ou des formulaires. Des études ont été faites pour évaluer le système NLSA³ « Natural Language Sales Assistant » (Chai, Budzikowska et al. 2001). Elles montrent le succès des systèmes basés sur le dialogue dans un environnement de commerce électronique.

La première de ces études a comparé la première version du prototype à un système basé sur les menus entièrement développé. (Chai, Lin et al. 2000) ont montré qu'en comparant le système NLSA à la navigation à l'aide des menus pour trouver des produits, le nombre de clics (un clic est compté quand l'utilisateur clique sur un bouton « submit », un bouton « radio » ou un lien, etc) a été significativement réduit de 63.5 % et le temps passé pour trouver le produit a été significativement réduit de 33.3 %. Cette étude montre aussi que les utilisateurs préfèrent la navigation basée sur le dialogue à celle basée sur les menus (79% contre 21%).

1.2.2 Peu de services disponibles

Nous avons interrogé plusieurs moteurs de recherche (Google, Altavista, Tiscali...) pour trouver des services traitant des énoncés spontanés en langue naturelle (ESLN). Bien que nous ayons soumis de nombreuses requêtes⁴, nous avons obtenu relativement peu de résultats positifs et nous avons

³ Nous décrivons ce système, basé sur le dialogue à travers le Web et déployé par IBM, dans le chapitre suivant.

⁴ Localizing natural language message processors, localization NLP free text, localization NLP interfaces, multilingual customer care service web, multilingual customer message processing, multilingual customer messages products, multilingual customer messages tools, multilingual customer relationship processing, multilingual customer support products, multilingual NLP e-commerce, multilingual NLP interfaces, multilingual online sales customer support, multilingual online sales NLP customer support, multilingual written customer support, product for conversational Interface in e-commerce, categorizing natural language message, handling natural language messages in business, Natural Language Conversational Interface in Online Sales, natural language customer service, natural language customer service, natural language messages

obtenu très peu de renseignements sur le fonctionnement interne des services trouvés et leur multilinguisation.

Il semble donc qu'il y ait encore peu de tels services, ainsi que d'études théoriques sur le sujet.

1.2.3 Analyse préliminaire

On trouve très peu de services qui traitent des énoncés spontanés en langue naturelle et cela même dans les pays les plus développés, pour les langues les plus utilisées et pour les domaines les plus fréquents.

Par exemple, aux États-Unis, le système NLSA traite des énoncés spontanés, pour lesquels on ne contrôle ni la langue utilisée ni le dialogue (il n'est pas modal). Ce sont des énoncés spontanés pour chercher des produits, par exemple « *I'm looking for a piano*⁵ ». Au Japon, le système IMRS (Impression-based Music-Retrieval System) (Kumamoto 2007) de recherche de morceaux de musique traite des énoncés spontanés, .

Pour l'anglais, on a trouvé un peu plus d'applications et/ou d'approches construites pour traiter des textes anglais, mais ce ne sont pas forcément des textes formés d'énoncés spontanés. Il y a des applications, comme NLSA qui traite des textes anglais ou Amilcare (Ciravegna 2001a), un système adaptatif d'extraction d'information traitant des textes normaux en anglais et qui ne sont pas des textes d'énoncés spontanés.

Pour le français, le projet MKBEEM (Multilingual Knowledge-Based European Electronic Marketplace, 01/02/2000 – 30/11/2002)⁶ est un projet qui n'a pas abouti à des résultats exploités. Il était censé construire une plate-forme de commerce électronique multilingue et pluriculturelle et offrir des interfaces homme-machine en langue naturelle traitant des énoncés spontanés. Dans ce projet, l'usage d'ontologies était censé permettre de résoudre les ambiguïtés linguistiques et de trouver le service le plus adapté à la demande de l'utilisateur. Par exemple les requêtes comme « *nous aimerions aller à Montpellier mais pas en TGV* » et « *nous aimerions aller à Montpellier en voiture* » ne montrent aucune différence syntaxique ni sémantique (à part la négation et le moyen de transport). C'est à l'ontologie d'identifier un service « *réservation de train* » dans le premier cas et le service « *location voiture* » dans le deuxième.

Concernant le résumé automatique, les approches retenues pour la construction de systèmes de résumé automatique sont différentes et variées. La plupart d'entre elles sont, par ordre chronologique, des réalisations universitaires (Lehman 1996), d'entreprise nationale (EDF) en collaboration avec une équipe universitaire (CAMS, Centre d'Analyses Mathématiques en Sciences Sociales, thèse de Jawad Berri, 96, Université Paris 4) ou enfin de quelques entreprises privées dont les systèmes restent peu connus. Il existe aussi des modules de résumé rattachés à des moteurs de recherche professionnels, ou intégrés à des applications grand public (synthèse automatique de MSWord97).

Le produit Pertinence Summarizer (Lehman 1996) est basé sur une technologie propriétaire appelée KENIA (Knowledge Extraction and Notification Architecture). C'est un outil de résumé automatique qui procède par des méthodes statistiques. Cet outil utilise des méthodes « aveugles » car

in e-commerce, Natural Language operational Interface in Online Sales, NLP of customer complaints, NLP of customer complaints multilingual, NLP of customer messages multilingual, NLP of customer reports, NLP of multilingual customer messages, product for conversational Interface in e-commerce, traitement de messages en langue naturelle d'utilisateur, Natural Language Interface database, etc.

⁵ Il s'agit de chercher un ordinateur de type Notebook

⁶ <http://www.mkbeem.com/> : site Web du projet MKBEEM visité en 2006.

purement statistiques sans aucune participation analytique d'ordre linguistique. Pertinence Summarizer a la particularité d'analyser et de traiter un grand nombre de langues.

Pour l'extraction d'information, nous avons trouvé le système CATS (Classifieds Ads Through SMS) (Daoud 2005). C'est une plate-forme d'achat et de vente de biens (automobile, immobilier...) déployée par Fastlink, le plus gros opérateur de téléphonie mobile en Jordanie. CATS est un système qui traite des énoncés spontanés en arabe sous forme de SMS.

Le Tableau 1 résume l'ensemble des applications trouvées par pays, langue et types de service.

Applications	Pays	Langues	Type de service
NLSA	Etats Unis	Anglais	Extraction d'information
Amilcare	Grande-Bretagne	Anglais	Extraction d'information
MKBEEM	France	Français	Désambiguïsation
Pertinence Summarizer	France	Français, allemand, anglais, arabe, chinois, coréen, espagnol, grec, italien, japonais, néerlandais, norvégien, portugais, russe.	Résumé automatique
CATS	Jordanie	Arabe	Extraction d'information
IMRS	Japon	Japonais	Extraction d'information

Tableau 1 : analyse des applications les plus importantes par pays, langue et type de service

1.3 Raisons possibles

Peu de services traitent donc les énoncés spontanés en langue naturelle, bien que cela semble très souhaitable. Les raisons possibles en sont à notre avis que :

- réaliser un système de traitement de contenu pour des énoncés spontanés en langue naturelle est très difficile.
- la multilinguïsation d'un tel système pose de gros problèmes.

Nous montrerons dans la troisième partie, la validité de cette hypothèse sur un cas réel de multilinguïsation de tels services.

1.3.1 Difficultés inhérentes au traitement automatique d'ESLN

Plusieurs types de difficultés sont liés au traitement automatique des énoncés spontanés en langue naturelle, à savoir : des difficultés similaires à celles du traitement de l'oral qui peut contenir du bruit, des difficultés liées à l'utilisation d'un formalisme spécifique de représentation de contenu, et des difficultés liées à la nécessité de garantir un minimum de qualité des résultats des traitements.

1.3.1.1 Problèmes similaires à ceux de l'oral

Les problèmes qui persistent, dans le cas d'un traitement des ESLN, sont très similaires à ceux rencontrés dans le traitement de l'oral (Boitet 1990). En effet, on observe :

- l'utilisation de grammaire « non standard » (ordre des mots, réponses et formes familières).
- une abondance d'erreurs (fautes de frappe, d'orthographe).

- l'utilisation de conventions typographiques propres au contexte (émoticons et abréviations en tchat ou SMS).

Assez souvent, on est dans un « sous-langage » relativement éloigné de la « langue générale », comme les petites annonces ou les alarmes/avertissements (trafic routier, catastrophes naturelles). En conséquence, les outils existants faits pour la « langue générale » ne marchent pas (trop couvrants), qu'il s'agisse d'outils de TA (Traduction Automatique) ou d'EC (Extraction de Contenu). Une approche spécifique à chaque sous langage s'impose alors.

En conclusion, on ne peut pas utiliser des outils faits pour du langage écrit « général » et « propre » et il faut « traduire » les énoncés spontanés en langue naturelle dans un formalisme de « représentation de contenu » et, en général, chaque application possède son propre formalisme.

1.3.1.2 Utilisation d'un formalisme spécifique « représentation de contenu »

Les applications traitant les énoncés spontanés en langue naturelle utilisent une représentation de contenu. On trouve plusieurs formes de représentation de contenu avec différents niveaux de « puissance expressive » comme (voir Chapitre 5) :

- des listes de propriétés <attributs, valeurs>,
- des structures de traits typés,
- des expressions logiques (Prolog),
- des expressions logico-fonctionnelles,
- des objets (classes (méthodes, attributs), instances),
- des descripteurs liés à une ontologie.

Assez souvent, une application possède un formalisme très spécifique. En conséquence, l'adaptation ou le portage d'un extracteur de contenu d'une application à l'autre, même pour la même langue, nécessite une traduction d'un formalisme à l'autre. Par exemple, le système CATS utilise une représentation spécifique proche des listes de propriétés *Propriété = couleur {objet = saloon, valeur = bleu}* pour le domaine des achats et ventes de voitures d'occasion.

1.3.1.3 Nécessité de garantir un niveau minimum de qualité

Si l'on adapte ou porte un extracteur de contenu d'une application à une autre, même pour la même langue, il faut pouvoir garantir un niveau minimum de qualité :

- Exactitude/complétude de l'EC.
- Pertinence des réponses produites (traitement + gestion).
- Adéquation linguistique des réponses, si on répond en langue naturelle et pas sous forme de tableaux.

Le portage d'une application d'une langue à une autre doit être défini et mesuré par un pourcentage de portage par rapport à la version originale, et si c'est possible par rapport à une version de référence.

1.3.2 Difficultés du portage linguistique

1.3.2.1 Comment créer un corpus en sous-langage cible ?

La donnée primaire à fournir pour un portage linguistique d'une application traitant les énoncés spontanés est un corpus de développement relatif à la même tâche, mais en langue cible (un minimum pour évaluer, et pour développer). Cela n'est pas toujours évident à trouver. Il faut travailler le plus souvent par adaptation/simulation/imagination.

Nous nous bornons au portage pour éliminer le travail spécifique de l'adaptation, qui peut être lourd. Par exemple, CATS traite des SMS en arabe sur l'immobilier à Amman, mais l'adapter au français pour la France suppose de choisir des noms de ville en France, des noms de lieu en France, des noms de rues en France, etc. C'est un travail d'adaptation qui diffère d'un autre travail d'adaptation pour le français en Belgique, en Suisse, au Canada, ou en Afrique.

Pour le corpus, le problème se réduit donc, dans le cas du portage, à construire un corpus cible « naturel » correspondant à la situation de l'application telle quelle. Par exemple, on voudra disposer d'un corpus de SMS en français que des francophones pourraient envoyer à Amman sur de l'occasion automobile ou de l'immobilier. Par définition, un tel corpus n'existe pas encore quand on commence.

Une première idée serait de prendre un corpus en langue cible d'une application existante, par exemple tirer de Paruvendu.fr ou de e-Bay.fr un corpus de petites annonces « Web » concernant l'immobilier à Grenoble. Mais alors il faudrait non seulement l'adapter au « style SMS », mais surtout le « localiser » à Amman, travail lourd qu'on cherche justement à éviter en se limitant au portage.

La deuxième possibilité, que nous avons retenue, est de traduire et réviser un corpus existant puis de l'adapter au sous-langage. Pour ce qui est de la traduction, les systèmes de traduction automatique ne sont pas disponibles pour tous les couples de langues (ex : arabe-français). De plus, ils ne sont utiles que pour des textes « propres » et ne marchent pas pour les énoncés spontanés. Dans ce cas, ils peuvent juste servir de « dictionnaire en contexte ».

En conséquence, dans le cas d'un portage linguistique, on est confronté à un vrai problème. Ce problème est doublement posé, car il faut aussi trouver un « corpus d'évaluation » qui permet l'évaluation du portage.

La question de la taille se pose aussi : quelles tailles de corpus sont-elles suffisantes pour le développement et l'évaluation d'un portage ?

1.3.2.2 Comment construire l'extracteur de contenu ?

La réalisation d'un nouvel EC (Extracteur de Contenu) peut se faire par plusieurs méthodes.

La première solution est d'adapter l'EC existant de L1 à L2, mais cela n'est possible que si

- les développeurs acceptent d'ouvrir leur code ou leur boîte à outils (BàO) à des collaborateurs nécessairement éphémères,
- ce code / BàO est assez facile à maîtriser,
- les ressources ne sont pas trop lourdes à créer (en particulier le dictionnaire s'il existe),
- la maintenance peut ensuite se faire à coût faible, par des collaborations épisodiques.

Cette méthode d'adaptation interne de l'EC natif nécessite bien sûr une formation de l'équipe de localisation aux outils et aux méthodes utilisées.

La deuxième solution consisterait, pour une société voulant offrir des services de multilinguisation/portage, à implémenter un EC générique et à l'adapter à chaque situation (langues, sous-langage, domaine, langage de représentation de contenu, tâche, contraintes). On verra dans la partie suivante que cela semble très difficile à envisager actuellement.

Dans beaucoup de contextes, les « multilinguiseurs » n'auront donc pas accès à l'EC de l'application, ni à un EC « universel ». Une troisième solution pourrait alors être de rechercher et d'adapter un EC existant et disponible, soit pour la langue L2, pour un domaine différent ou une tâche différente, soit pour le même domaine et la même tâche, mais pour une autre langue (différente de L2).

1.3.2.3 Peut-on n'utiliser que l'extracteur de contenu de l'application ?

On se demande si l'on peut éviter le problème précédent avec une autre idée qui consiste à traduire tout simplement « à la volée » les énoncés spontanés d'une langue L2 vers la langue originale L1 de

l'application à localiser. Dans ce cas là, il faut bien sûr étudier le dictionnaire de L1 pour la situation/domaine visé.

Cela n'est a priori pas possible par adaptation de systèmes de TA « experts », généralistes⁷. C'est peut-être possible avec des méthodes « empiriques » (TA statistiques, TA par analogie, TA par l'exemple utilisant des corpus annotés et de l'apprentissage).

On pourrait aussi développer rapidement des systèmes de TA spécialisés aux sous-langages visés, mais l'approche « experte » (procédurale ou par règles) suppose de disposer de collaborateurs eux-mêmes experts dans les couples de langues visés, et en programmation linguistique.

L'idée de réaliser un système de Traduction Automatique pour les énoncés spontanés est donc envisageable, mais plusieurs questions se posent : quel type de système ? Quelle taille de ressources doit-on avoir ?

1.3.3 Difficultés du portage vers un autre domaine

Dans le cas d'une migration d'un domaine vers un autre (par exemple tourisme vers automobile), nous sommes confronté aux mêmes types de problèmes et de difficultés que ceux posés dans le cas d'un portage linguistique. À savoir la construction d'un corpus cible d'une taille suffisante pour le développement et l'évaluation du portage de l'application vers un nouveau domaine en restant dans la même langue ou en changeant de langue.

Plusieurs questions se posent aussi concernant la construction de l'extracteur de contenu pour le nouveau domaine. Peut-on adapter l'extracteur de contenu du domaine précédent ? Cela n'est possible que si les mêmes conditions annoncées précédemment (dans le cas de l'adaptation de l'EC existant de L1 à L2) sont assurées. En particulier, cela nécessite un accès total à l'application et la maîtrise du fonctionnement interne de l'application et de sa spécification, pour pouvoir définir les parties qui doivent être adaptées, modifiées, enrichies, supprimées, etc.

Plusieurs difficultés semblables à celles du portage linguistique se posent dans le cas d'un changement de domaine, mais cela semble réalisable une fois que les difficultés rencontrées sont résolues.

⁷ Ce serait possible s'il s'agissait d'un système de TA spécialisé à un sous-langage très proche, mais il n'y en a pas de disponibles.

Chapitre 2 État de l'art

Introduction

Nous limiterons notre état de l'art aux systèmes d'e-commerce traitant des énoncés spontanés en langue naturelle et à leur portage vers d'autres langues et/ou domaines, tels que les systèmes qui gèrent les petites annonces d'achat et de vente sur le Web.

Les travaux de portage et de multilinguisation des systèmes d'e-commerce cités nous serviront à mieux cerner le problème, et à voir comment, en général, on pourrait évaluer le coût d'un portage linguistique en fonction de tous les paramètres de la « situation traductionnelle ».

Dans la première section, nous définissons la portée de notre étude. Dans la deuxième section, nous présentons quelques systèmes d'e-commerce en nous concentrant sur leur portage vers d'autres langues. Dans la troisième section, nous présentons quelques approches plus théoriques qui peuvent s'appliquer à notre problème.

2.1 Portée de l'étude

Nous partons d'un cas particulier qui est le système CATS, et nous cherchons à élargir notre étude à tous les systèmes similaires au sens où ils traitent des énoncés spontanés en commerce électronique.

Nous précisons d'abord la terminologie que nous emploierons.

- Portage linguistique : il s'agit de détailler et analyser la méthode de passage d'une langue à l'autre ; il ne s'agit ni de génération multilingue, ni de localisation.
- Localisation : portage vers une autre langue et adaptation à un autre contexte similaire (tâche, domaine, lieu).
- Systèmes de e-commerce : systèmes utilisés pour faire du commerce électronique utilisant divers media (Web, dialogue, SMS...), que ce soit en B2B ou en B2C. C'est donc le grand public qui est visé dans ce type d'application.
- Énoncés spontanés en langue naturelle (ESLN) : les énoncés à traiter sont rédigés d'une façon naturelle ; ce sont par exemple des annonces qui contiennent du bruit, des erreurs, des fautes... ou les feedbacks des utilisateurs. A contrario, ce ne sont pas les annonces de la presse, car elles sont corrigées et éditées.
- Ecrit/oral : on traite ici l'écrit, car, jusqu'à présent, il n'existe pas de systèmes de e-commerce pour l'oral disponibles commercialement, il y a seulement des projets de recherche comme Nespole! (NESPOLE! 2000).

2.1.1 Applications de e-commerce envisageables

Les applications envisageables sont a priori nombreuses, mais si nous limitons notre recherche aux applications de e-commerce traitant obligatoirement des données spontanées, la liste devient moins importante :

- Petites annonces : des sites comme e-Bay.fr, bonjour.fr et autres existent sur le Web, mais l'opération d'extraction de contenu est réalisée manuellement par l'utilisateur qui choisit des items dans des menus à partir d'une interface contrôlée.

Par exemple, Comintoo est un moteur de recherche pour l'immobilier lancé début février 2007 (<http://comintoo.com/>). L'idée est de regrouper tout les sites sur l'immobilier en France dans un seul site basé sur la communication en langage naturel. Le langage utilisé n'est pas spontané, il est contrôlé par des menus déroulants et des textes gardés en mémoire. Il s'agit de contrôler la saisie et en même temps d'aider à la rédaction du texte par l'affichage d'exemples de recherche. Ensuite, les annonces sont affichées dans leur site original par redirection à l'aide des liens. Les annonces sont parfois disponibles dans d'autres langues, mais il s'agit d'une génération multilingue. La même chose est prévue pour le domaine de l'automobile.

Nous avons trouvé un traitement d'énoncés dans le cas du système CATS, basé sur le SMS comme media et recevant des énoncés spontanés en arabe.

- Enchères : il en existe surtout sur le Web, notamment Advise, e-Bay... Il n'y a pas de traitement des énoncés spontanés, et, pour le cas multilingue, il y a appel à des systèmes de TA comme Systran. Ces applications font de la recherche « exacte » par critères, sans inférence.

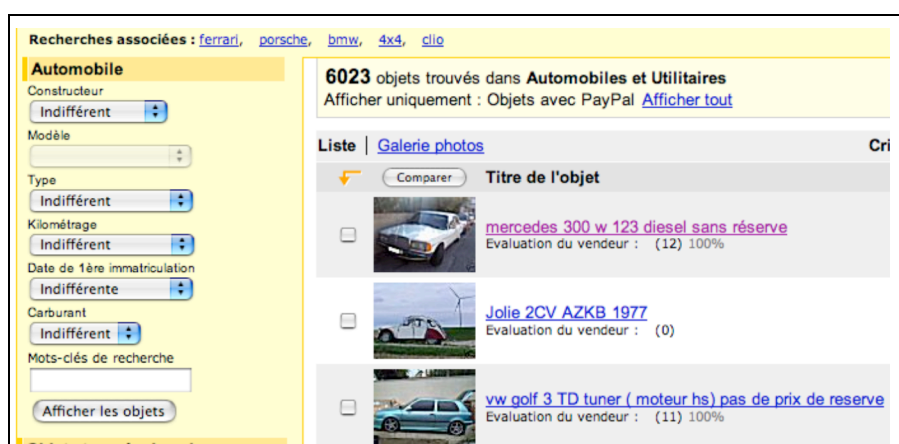


Figure 3 : exemple d'enchères de voiture

- Rapports d'incidents par des utilisateurs : on a trouvé un cas d'approche d'extraction d'information, pour lequel il y a eu portage d'une langue à une autre dans le domaine des accidents d'avion (Riloff, Schafer et al. 2002).
- Traitement du Web de type catégorisation de courriels, résumé automatique, etc. : ce type de service existe. La compagnie 5ca (<http://www.5ca.com/>) fait la gestion des communications basées essentiellement sur l'IP (téléphone, email, tchat, sms, etc). 5ca utilise des méthodes semi-automatiques pour la catégorisation des courriels et fait appel à des systèmes de traduction automatique pour des cas multilingues (manuels de produits...). Pertinence Summarizer, un système de résumé automatique, est aussi utilisé pour la gestion des courriels, de sites Web, etc. Nous avons trouvé des services proposés par DATOPS (<http://www.datops.com/>) pour les études d'opinion et l'identification d'informations-clés.
- Recherche d'information : il y a des projets comme MIETTA « Multilingual Information Extraction for Tourism and Travel Assistance ». MIETTA est un projet européen qui intègre la recherche d'information avec l'extraction et le traitement d'information (Buitelaar, Netter et al. 1998).

Les pages jaunes permettent aussi la recherche d'information, mais en contrôlant la saisie par un affichage d'exemples de saisie, comme le montre la figure suivante.

The image shows a search interface with three sections, each with a text input field, a 'Guide' link, and an example:

- Quoi ?**: Activity (ex: restaurant, médecin)
- Qui ?**: Nom : (ex: Hôtel de la Poste, Dupont)
- Où ?**: Localité (ex: paris 9, 37100), Adresse (ex: 7 av de la Cristallerie), Département ou région (ex: 92, idf)

A red 'Rechercher' button is located at the bottom center.

Figure 4 : recherche d'informations à l'aide des pages jaunes

- Extraction d'information : nous avons trouvé des systèmes ; un des plus importants est une application adaptative de type « Amilcare », c'est une application adaptable d'extraction d'information.
- Assistance à la vente : le système NLSA, une application déployée par IBM.

Dans cette étude, on ne s'intéresse qu'aux applications commerciales dans lesquelles il y a un traitement du contenu d'énoncés spontanés. En fait, ce traitement peut être automatique (CATS) ou manuel (e-Bay). L'essentiel est qu'il existe une représentation interne du contenu. Dans le cas d'e-Bay, le portage de la représentation interne du contenu est fait par appel à de la TA experte (Systran ou Reverso), et le portage des annonces elles-mêmes n'est pas fait (elles sont gardées telles quelles), mais on pourrait le faire.

Tenant compte de ce filtrage, la liste des types de services trouvés devient un peu moins longue. Voici la liste retenue :

- Petites annonces,
- Enchères,
- Rapports d'incidents,
- Traitement du Web,
- Recherche d'information,
- Assistance à la vente.

Les éléments de la liste retenue peuvent répondre partiellement à notre étude ; il est donc important de procéder à une analyse détaillée des systèmes et des approches correspondants.

2.1.2 Liste des applications considérées

Dans ce qui suit, nous dressons la liste des applications retenues pour les domaines envisagés. Le Tableau 2 présente la liste des applications classées par ordre d'importance.

Application	Note
NLSA (Naturel Language Sales Assistant)	+
Pertinence Summarizer	+
CATS (Classified Ads through SMS)	+
IMRS (Impression-based Music-Retrieval System)	+
MKBEEM (Multilingual Knowledge Based European Electronic Marketplace)	-
Projection interlingue d'extraction d'information	+
Amilcare	-
e-Bay	-

Tableau 2 : liste de applications considérés

+ : application très importante pour notre étude.

- : application moins importante pour notre étude.

Nous classons la liste des applications considérées en deux catégories : celles qui peuvent être considérées comme des systèmes et celles qui peuvent être considérées comme des approches. Ainsi, nous adoptons pour chaque système de e-commerce analysé le type d'examen suivant :

- Brève présentation
- Type des énoncés traités
- Architecture et ressources (pour traiter les énoncés)
- Portage linguistique (nous essayons d'analyser la méthode de portage s'il a été réalisé, sinon nous essayons de détecter les composants qui dépendent de la langue).

Nous adoptons pour chaque approche analysée le type d'examen suivant :

- Brève présentation
- Type des énoncés traités
- Méthode et ressources (pour traiter les énoncés)
- Domaine d'application

2.2 Examen de quelques systèmes

Nous appliquons la méthode d'examen présentée précédemment aux systèmes qui semblent répondre le mieux à nos critères. Dans certains cas, nous enrichissons l'examen par un simple rappel historique qui montre le besoin lié à la création du système en question.

2.2.1 NLSA

NLSA « Naturel Language Sales Assistant » (Chai, Budzikowska et al. 2001) est le résultat de recherches faites autour de la possibilité d'intégration du dialogue en langue naturelle dans les sites de e-commerce dans le cadre d'un produit d'IBM, déployé en Angleterre et qui facilite l'échange d'informations sur des produits informatiques.

2.2.1.1 Brève présentation

Le dialogue en langue naturelle a été employé pour des applications comme les centres d'appel (Carpenter and Chu-Carroll 1998), le routage de courriels (Walker, Fromer et al. 1998), la recherche d'information, l'accès aux bases de données (Ritchie 1995), et pour des opérations bancaires effectuées par téléphone (Zadrozny, Wolf et al. 1998).

L'objectif des recherches faites autour du système NLSA est l'intégration du dialogue en langue naturelle dans les environnements d'e-commerce et le progrès par rapport aux interfaces des sites Web en langue naturelle comme www.askjeeves.com, www.neuromedia.com, etc, qui fonctionnent en mode question-réponse et n'utilisent pas le dialogue.

Des études (Chai, Lin et al. 2000) ont montré qu'un tel moyen de communication basé essentiellement sur le dialogue permettrait de comprendre efficacement les requêtes, les besoins et les intentions des utilisateurs et par la suite de produire de l'aide, des conseils, et des recommandations aux utilisateurs.

NLSA est un système d'assistance à la vente basé sur le dialogue à travers le Web déployé en 2001 par IBM sur son site britannique. C'est un système qui aide les utilisateurs à trouver des informations sur des produits et des services dans des sites de e-commerce. Le système combine les technologies de traitement de la langue naturelle et de l'interaction homme-machine pour permettre une interaction plus rapide et plus intuitive avec les sites Web.

2.2.1.2 Type des énoncés traités

Comme le montre l'exemple de la Figure 5, le système NLSA traite des énoncés spontanés. Les données forment l'état du dialogue représenté par une combinaison de l'expression courante de l'utilisateur, de la réponse courante du système, et de l'historique de l'interaction.

Dans un exemple d'interaction, le système essaie de préciser les questions suivantes :

- L'utilisateur a-t-il indiqué des préférences de choix d'ordinateur ?
- Y a-t-il une ambiguïté détectée dans l'entrée de l'utilisateur ?
- Y a-t-il des concepts identifiés par l'analyseur dans l'expression courante ?
- Y a-t-il des contraintes contradictoires détectées ?
- Y a-t-il plus de trois produits recommandés à l'utilisateur dans l'interaction précédente ?

La combinaison des réponses à ces questions forme l'état du dialogue pour cette interaction particulière. L'histoire du dialogue est une représentation sous forme d'un arbre qui contient l'état du dialogue à tous les points d'interaction. En se fondant sur l'interprétation de l'entrée de l'utilisateur, le système forme alors la requête correspondant à l'information spécifique. Voici un exemple en anglais britannique.

U₀: I'm looking for a piano⁸.

S₀: I'll find you a notebook if you tell me what your notebook needs are.

U₁: I'm looking for a laptop less than 500 pounds.

S₁: Display: a table of most popular models
Reply: The price range for Thinkpads is between 799 and 2389 pounds. You may specify a value with the range or continue your search by specifying other features you are looking for.

U₂: I want something easy to carry and inexpensive.

S₂: Display: We have found several notebooks based on your search criteria: Ultraportability: models which weigh less than 3 kilograms, Low price: affordable models which sell for less than 1500 pounds, followed by a table of six products satisfying these criteria).
Reply: Do you need a large hard disk capacity for backup data or storing large files?

U₃: No, I want something top-of-the-line.

S₃: Display: We didn't find a product that matches your terms perfectly, but we have listed the closest match to your request. (Display two tables: one table satisfies the extreme-performance criteria with weights of products sorted from lightest to the heaviest; the other table satisfies the low price criteria with weights of products sorted from lightest to heaviest).

Reply: Battery life is important to frequent travelers. Is long battery life important to you?

...

Figure 5 : exemple de dialogue en anglais britannique

Dans une étude de cas (Chai, Budzikowska et al. 2001), des contradictions se sont produites pour plus de 40% des tours de parole dans lesquels le système propose des suggestions. Par conséquent, pour de tels systèmes, il est crucial que la source de conflit soit identifiée et que le système puisse fournir un retour.

2.2.1.3 Architecture et ressources

2.2.1.3.1 Architecture

Le système NLSA est basé sur une architecture client-serveur utilisant des servlets java et HTTP comme protocole d'échange entre le client et le serveur. Il inclut deux sous-systèmes comme le montre la Figure 6 : un sous-système pour la gestion des données et un sous-système d'interaction en ligne.

⁸ Piano est un ordinateur portable commercialisé par LG.

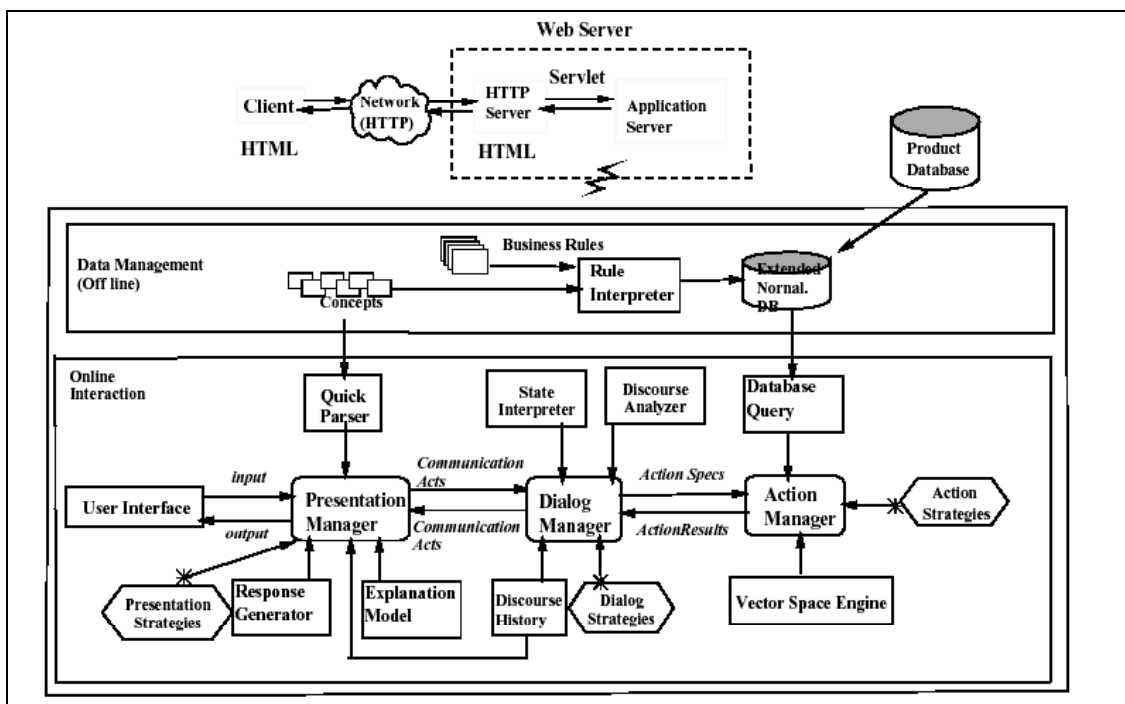


Figure 6 : architecture du système NLSA

2.2.1.3.2 Composants linguiciels et sémantiques

Les principaux composants linguiciels sont : un sous-système d'interaction en ligne et un sous-système de gestion des données.

❖ Sous-système d'interaction en ligne

L'architecture du sous-système d'interaction supporte un dialogue multimodal et des entrées de plusieurs canaux comme le clavier, l'entrée et la sortie d'un téléphone ou d'un microphone, la souris, etc.

Le sous-système d'interaction est composé de trois modules :

- Le module de Gestion de la Présentation (GP) qui utilise un analyseur pour transformer la requête de l'utilisateur en un format logique et l'envoyer au Gestionnaire de Dialogue (GD). Il reçoit aussi la réponse du système en provenance du gestionnaire de dialogue. Il se compose d'interfaces, un analyseur « Quick Parser », un générateur de réponse et un module d'explication.
- Le module de Gestion du Dialogue (GD) qui détermine et envoie/reçoit l'action au/du gestionnaire de l'action (GA).
- Le module de Gestion des Actions (GA) qui détermine le meilleur mécanisme pour l'exécution de l'action. Par exemple, ayant une spécification d'une action de recherche d'un produit, il décide quelle information source doit être cherchée et extraite.

L'un des principaux avantages de l'utilisation du dialogue est l'ajout de nouvelles informations à l'aide des composants de l'interface comme les boutons, les formulaires, les liens (voir Figure 7) et le dialogue en langue naturelle. Cela permet à l'utilisateur d'avoir aussi des exemples, des explications et de comprendre les limitations du système.

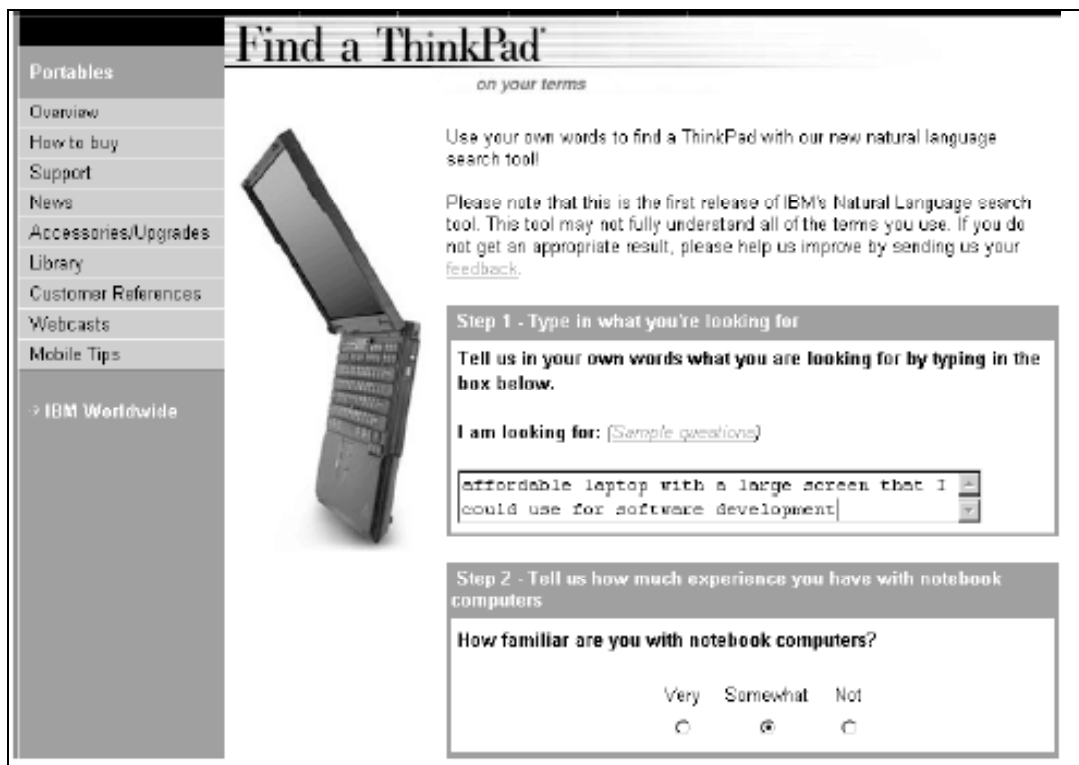


Figure 7 : interface initiale pour utilisateur

Un analyseur sémantique est utilisé pour identifier l'information sémantique à partir de l'entrée textuelle de l'utilisateur. Plus spécifiquement, en utilisant une collection d'expressions régulières et une grammaire, l'analyseur identifie les entités nommées spéciales (ex : CPU rapide, type de processeur, etc), la tête de chaque groupe nominal et les concepts pour le traitement ultérieur. L'analyseur est implémenté comme une cascade de transducteurs d'état finis. Il traduit des chaînes d'entrée en langage naturel en un message XML bien formé appelé « Logical Format » (LF), qui est transmis au GD (Gestionnaire de dialogue).

Par exemple, si l'entrée de l'utilisateur est *"nice travel companion for \$2000-3000 with at least 400Mhz Pentium processor"*, l'analyseur produira la forme suivante dans le format XML. C'est un exemple de message dans le format LG.

```
<LF>
<NP>
<ATTRIBUTE><SPECS>nice</SPECS></ATTRIBUTE>
<HEAD><SPECS CONCEPT="PORTABILITY"
SCALE="weight">travel companion</SPECS></HEAD>
<PP>
<PREP VALUE="for" />
<NP>
<ATTRIBUTE><SPECS
TYPE="BASE_PRICE">gt2000</SPECS></ATTRIBUTE>
<HEAD><SPECS TYPE="BASE_PRICE">1t3000</SPECS></HEAD>
</NP>
</PP>
<PP>
<PREP VALUE="with" />
<NP>
<ATTRIBUTE><SPECS
TYPE="CPU_SPEED">gt400</SPECS></ATTRIBUTE>
<ATTRIBUTE><SPECS
TYPE="CPU_BRAND">Pentium</SPECS></ATTRIBUTE>
<HEAD><SPECS>processor</SPECS></HEAD>
</NP>
</PP>
</NP>
</LF>
```

Le format LG est reçu par le gestionnaire de dialogue qui se charge de l'interaction ultérieure comme demander plus d'information ou une confirmation.

L'analyseur vise à produire une analyse linguistique complète (un arbre de constituants enrichi par des décorations) ; il annote la forme de surface avec une description de contenu. La représentation est donc liée à la langue source et assez complexe. Une analyse possible de la structure logique du format LG peut être la suivante (voir Figure 8). C'est un arbre dont la racine est LF « Logical Format » et dont les nœuds sont les concepts détectés et peuvent avoir des attributs.

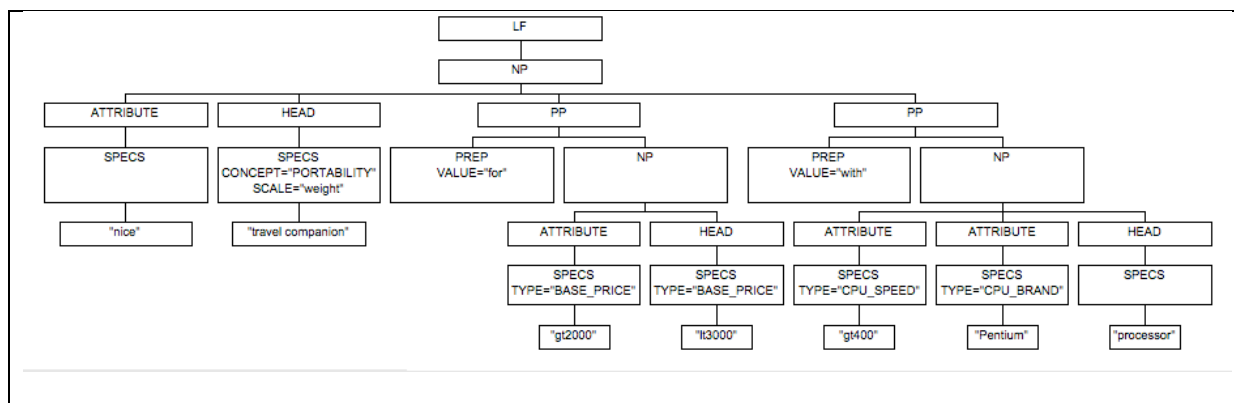


Figure 8 : exemple de structure dans le format LG

❖ Sous-système de gestion des données

Le sous-système de gestion des données contient les données spécifiques à l'application et au domaine, les informations et les connaissances, et traite les données, les connaissances et la maintenance (outils et processus).

Dans ce sous-système, la connaissance du domaine est représentée par les concepts et les règles du niveau « business », qui traitent respectivement les besoins du client et les demandes du « business ». Comme dans la plupart des autres systèmes, l'ensemble du vocabulaire est représenté par des concepts qui indiquent les intentions des utilisateurs et leurs intérêts dans le marché des ordinateurs. Trois types de concepts sont définis :

- Les concepts d'entités qui représentent des choses et des substances (produits, accessoires, etc.).
- Les concepts d'attributs qui représentent les caractéristiques des produits, comme « FAST », « LIGHT », « HIGH-PERFORMANCE » etc.
- Les concepts objectifs qui représentent les rôles et les fonctions des produits recherchés comme « SMALL-BUSINESS », « TRAVEL », etc.

Chaque concept contient une liste de mots et/ou de phrases définissant ce concept particulier. Un format XML est utilisé pour organiser et gérer ces concepts, comme le montre l'exemple de la Figure 9. Chaque concept a différents attributs pour indiquer d'autres informations. L'analyseur « Quick Parser » est utilisé pour identifier les concepts appropriés dans l'entrée de l'utilisateur.

Les règles de la couche « business » font le lien entre les concepts et les spécifications du « business ». Ces spécifications sont les valeurs et les termes utilisés par les utilisateurs pour décrire leurs marchandises. Par exemple, dans le domaine de la vente d'ordinateurs, les spécifications peuvent être : le modèle d'ordinateur, la vitesse du processeur, la capacité de la mémoire, la taille du disque dur, le prix, etc. En terme de ces spécifications, les règles de la couche « business » donnent des définitions aux concepts du point de vue du « business », de la même façon que ces modèles sont placés sur le marché.


```

</ENTRY><ENTRY NORMAL_FORM = "AGP", SCALE = 0,SUPERLATIVE = 0>
<NOVICE>Is 3-D graphics performance, which speeds online gaming and graphic
design, important to you?</NOVICE>
<EXPERT>Are you interested in having an Advanced Graphic Port for improved 3-D
graphics performance? </EXPERT>
<NL_FORM>Advanced Graphics Port</NL_FORM>
<DEF><![CDATA[hd_min >= 6.4 & resolution is 1024x768]]></DEF>
<WORD>agp</WORD>
<WORD>graphics</WORD>
<WORD>game</WORD>
<WORD>3-D</WORD>
<WORD>multimedia</WORD>
<WORD>...</WORD>
</ENTRY>

```

Figure 9 : fragment du concept <Advanced-Graphic-Capability>

Les règles du business ont la forme suivante :

```

CONCEPT ::= eval(ATTR) = CONST_VALUE &
eval(ATTR) in RANGE &
not(eval(ATTR))

```

Par exemple, la règle suivante interprète le concept « Advanced-Graphic-Capability » en tant que la combinaison : disque dur de 6.4 gigaoctets minimum et résolution de 1024x768.

```

<Advance-Graphics-Capability> ::=
eval(hard_disk_min) > 6.4 GB &
eval(resolution) = 1024x768

```

Pendant la phase d'interaction en ligne, les règles de « business » sont utilisées pour traduire des concepts identifiés en des spécifications du business. Ainsi, la requête appropriée peut être produite en se basant sur ces spécifications.

2.2.1.3.3 Corpus

Pour démarrer NLSA et étudier le marché, il y a eu besoin d'un corpus de développement et de test. Afin de l'obtenir, les premiers participants ont fourni les trois requêtes suivantes :

“What kind of notebook computer are you looking for?”,
“What features are important to you?”, et
“What do you plan to use this notebook computer for?”

En appliquant des techniques statistiques (modèles de n-grammes) et une grammaire du groupe nominal à une collection de réponses en langue naturelle, une extraction de mots-clés significatifs et d'expressions a été réalisée. Les expressions visées expriment des intentions d'utilisateur et des intérêts pour des achats d'ordinateurs. Ensuite, les participants ont choisi les 10 premiers termes techniques classés comme les plus pertinents parmi 90 termes. Cette étude a permis de mieux formuler les réponses du système. Au cours d'une période de 30 jours, 705 réponses ont été reçues. A partir de ces réponses en langue naturelle, 195 mots-clés ont été appris et ont été ajoutés au vocabulaire du système déployé.

La grande majorité des requêtes des utilisateurs (85%) étaient formée de phrases courtes (5,31 mots en moyenne) constituées de groupes nominaux attachés à une liste de mots clés.

2.2.1.4 Portage linguistique

Le système NLSA a été le premier système basé sur le dialogue dans le domaine du e-commerce, c'est un système qui dépasse les limites posées par les systèmes basés sur les menus, car il permet aux utilisateurs de s'exprimer en anglais.

Le système NLSA n'a pour l'instant pas été porté vers d'autres langues. Ce portage nécessiterait les deux opérations suivantes :

- Portage des ressources externes et internes utilisées vers la nouvelle langue désirée.
- Adaptation de la méthode d'extraction d'information utilisant un analyseur sémantique pour identifier l'information sémantique et syntaxique à partir de l'entrée textuelle en langage naturel spontané d'un utilisateur.

NLSA est très intéressant par sa structure et sa puissance, on voit qu'il est conçu pour être localisé, mais on ne sait pas si sa localisation rencontrerait des problèmes dans des cas pratiques.

Nous avons trouvé très peu d'informations concernant la taille des composants du système NLSA.

Par contre, du point de vue de la complexité, les composants du système sont faciles à exploiter. Par exemple, l'analyseur utilise plusieurs ressources externes, et est adaptable à de nouveaux domaines d'application. En effet, chacune de ces ressources est configurable pour s'adapter à différents domaines. L'adaptation à de nouveaux domaines exigera seulement la mise à jour de ces ressources sans modification du code.

2.2.2 Pertinence Summarizer

Pertinence Summarizer (Lehman 1996) est un logiciel multilingue de résumé automatique de textes. C'est un outil qui a été primé par l'ANVAR, l'Agence française de l'innovation. Il permet de résumer des documents écrits dans les langues suivantes : français, allemand, anglais, arabe, chinois, coréen, espagnol, grec, italien, japonais, néerlandais, norvégien, portugais, et russe.

Bien que ce système ne traite clairement pas des énoncés spontanés, mais du texte « réfléchi » (Blanchon 2004), le nombre impressionnant de langues dans lesquelles il a été porté le rend a priori très intéressant pour notre étude.

2.2.2.1 Scénario

Un utilisateur fournit un document à résumer, il peut sélectionner un domaine particulier ou entrer ses propres mots, groupes de mots ou expressions pour personnaliser le résumé pour ses propres centres d'intérêt. Chaque terme trouvé dans le texte est accompagné du nombre de ses occurrences. Tous les termes sont présentés sous forme de liens hypertextes permettant de naviguer dans le texte source et/ou son résumé (en allant d'une occurrence à l'autre à l'aide de la touche de tabulation). Une fonctionnalité ajoutée récemment permet d'entrer des mots, groupes de mots ou expressions « parasites » pour provoquer la suppression des phrases les contenant.

Il permet d'accéder à l'information essentielle du texte via une extraction paramétrée des phrases les plus significatives à travers une interface conviviale : les phrases les plus informatives sont représentées sous la forme d'un dégradé de couleurs allant du bleu foncé, pour les phrases les plus importantes, au bleu très clair, pour les informations de moindre importance. La production du résumé automatique se fait en ligne en sélectionnant un taux de contraction allant de 100% à 1% du texte source.

Pertinence Summarizer applique un ensemble de règles s'appuyant sur l'analyse linguistique du discours pour extraire l'information pertinente du texte. Ces règles sont spécialisées par domaine. Ainsi, il est possible d'accéder rapidement aux seules informations liées à des besoins particuliers.

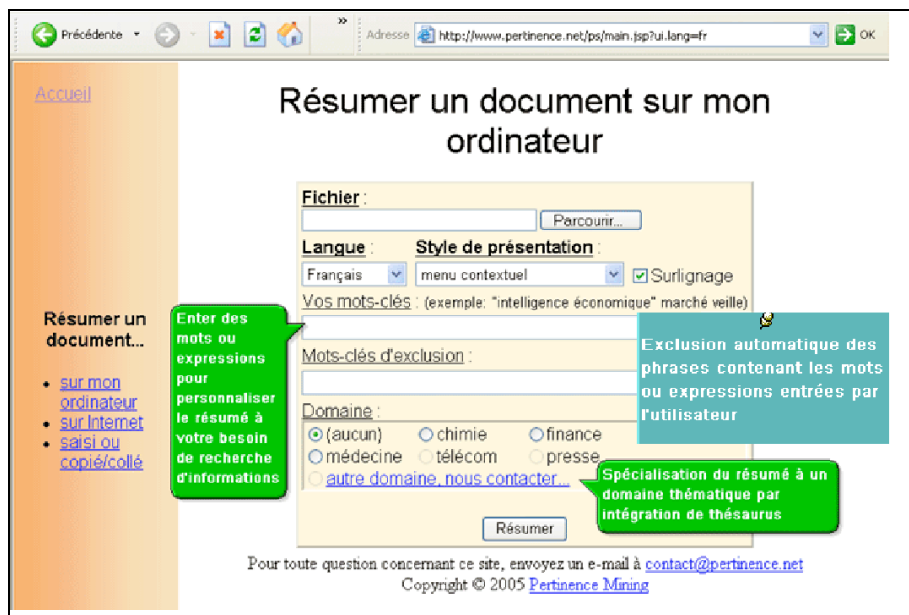


Figure 10 : Pertinence Summarizer en ligne

2.2.2.2 Type des énoncés traités

Pertinence Summarizer traite des textes de type scientifique et technique. L'écran de la Figure 11 montre un exemple de texte sur la grippe aviaire. Comme on l'a annoncé plus haut, ce ne sont pas des textes d'énoncés spontanés.

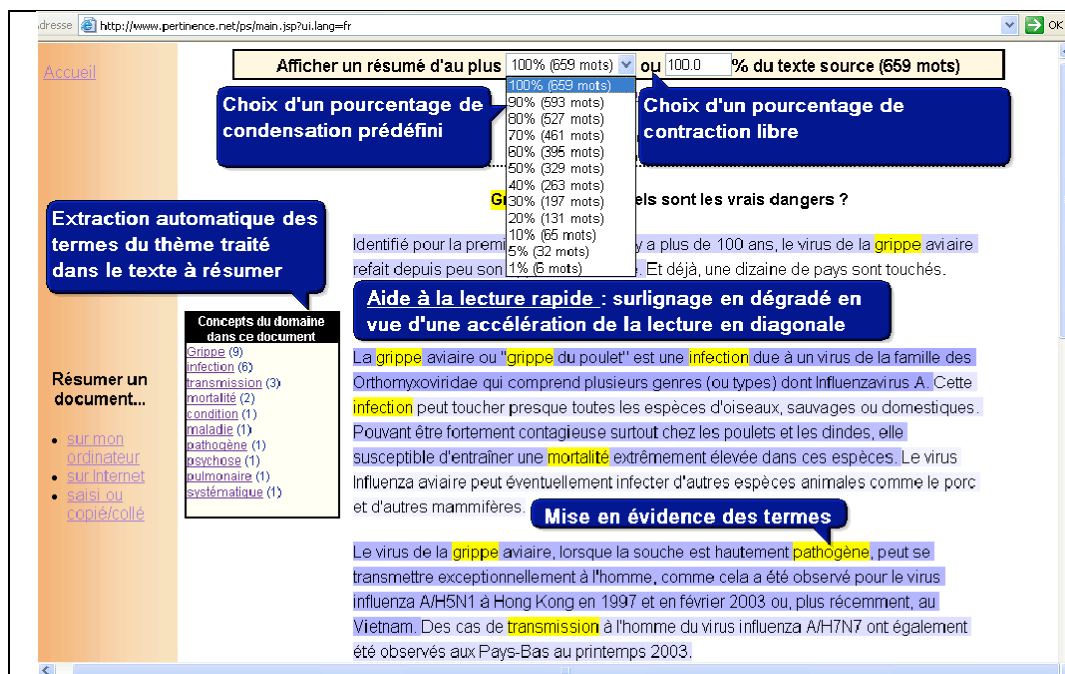


Figure 11 : exemple de texte traité par Pertinence Summarizer

2.2.2.3 Architecture et ressources

2.2.2.3.1 Architecture et composants

La Figure 12 représente l'architecture du système Pertinence Summarizer. Un texte dans l'une des langues citées précédemment est placé en entrée et sera analysé. Les phrases pertinentes seront extraites, avec la possibilité d'appel à des fonctionnalités d'aide à la lecture et à la cohérence.

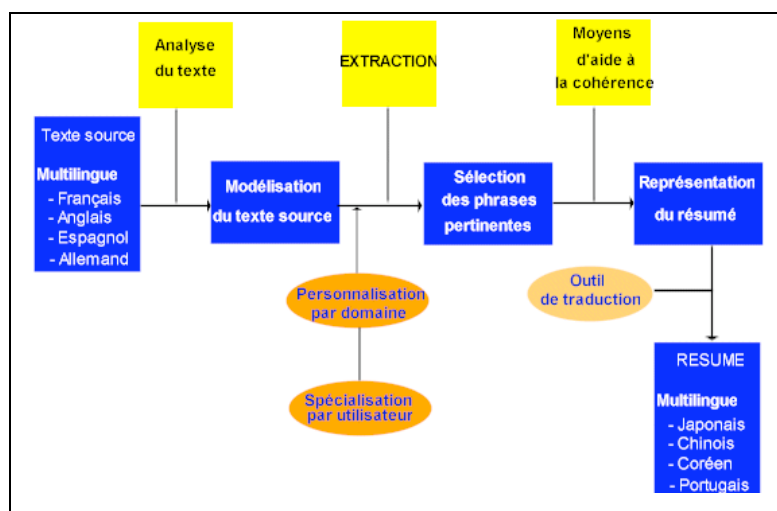


Figure 12 : architecture du système Pertinence Summarizer

Le résumé obtenu peut être présenté dans d'autres langues par appel à des outils de TA, mais bien sûr la qualité n'est pas garantie... et, s'il s'agit de système de TA généraliste, on peut être sûr que ce n'est pas bon en général.

Pertinence est un produit qui a été commercialisé. Malgré les efforts de contact direct avec le propriétaire de ce produit et pour des raisons commerciales, nous n'avons pas pu avoir plus d'informations sur la taille et la complexité de ce produit.

2.2.2.3.2 Corpus

Nous n'avons pas non plus pu obtenir plus de détail sur les corpus utilisés pour le développement. Mais on sait que Pertinence Summarizer résume automatiquement des textes scientifiques et techniques de différents formats : DOC (Word), PDF, RTF, HTML (.html, .htm, .asp, .php, etc), ASCII (.txt), WML, XML, PPT et XLS.

2.2.2.4 Portage linguistique

Pertinence Summarizer a été porté d'une langue à une autre en utilisant des connaissances linguistiques, il s'agissait de traduire les marqueurs sémantiques pour passer d'une langue à l'autre. Nous n'avons pas pu avoir plus d'informations sur les détails du portage, ni sur son coût.

2.2.3 CATS

CATS est un système de petites annonces d'achat et de vente basé sur l'utilisation des SMS en arabe (Daoud 2006), il est déployé en Jordanie par la société FastLink.

2.2.3.1 Brève présentation

CATS permet à des acheteurs de prendre contact avec des vendeurs et réciproquement. N'importe quelle personne se trouvant en Jordanie qui veut vendre une voiture, peut envoyer un SMS au système en indiquant sa proposition de vente en précisant quelques attributs comme le modèle, l'année, etc.

Une autre personne, intéressée par l'achat d'une voiture, peut simplement envoyer une requête au système. En cas de correspondance, le système CATS les met en contact. Ce service est disponible dans les domaines de l'automobile (Cars), de l'immobilier (Real Estate) et de l'emploi (Jobs).

Un vrai utilisateur a envoyé à partir de son téléphone mobile le SMS suivant décrivant sa requête d'achat :

6/1/2005 11:32:20 AM Message In : 962795028727,3750 الى 3500 والسعر ما بين 97 و المطلوب سيارة هوندا موديل

Figure 13 : exemple de SMS envoyé à CATS

Notre traduction :

6/1/2005 11 :32 :20 AM Message reçu : 962795028727 : recherche voiture Honda, modèle 97, prix entre 3500 et 3750

CATS envoie la réponse de la Figure 14.

Honda:
07999855 Civic
07956843 Accord

Figure 14 : exemple de SMS envoyé par CATS

Les requêtes et/ou propositions SMS sont envoyées à un numéro spécial et sont enregistrées automatiquement dans un corpus de données, précédées de la date, de l'heure et du numéro de téléphone. Leur texte est analysé et le résultat de l'extracteur de contenu est enregistré dans une base de données.

Une réponse est envoyée automatiquement à l'expéditeur d'un SMS contenant une requête, seulement en cas de correspondance exacte ou approchée de la demande avec l'une des propositions. Cette réponse permet de contacter la personne qui a envoyé la proposition (grâce au numéro de téléphone).

En cas d'échec, le système mémorise la demande et envoie une réponse dès qu'une nouvelle proposition le permet, dans un certain délai.

2.2.3.2 Type des énoncés traités

Voici quelques exemples typiques de SMS reçu par CATS.

9/3/2005 12:52:40 PM Message In : +962796770392,+962796770392, ش الفلل, 190 م الواحد جديده
9/3/2005 12:56:11 PM 90050
9/3/2005 12:56:11 PM Message In : +962796322201,Kia rio 2003 silver 1500cc
9/3/2005 12:57:28 PM 90050
9/3/2005 12:57:28 PM Message In : +962795202973,منظفه الروضه ناعور, 230 م
9/3/2005 1:04:36 PM 90050
9/3/2005 1:04:36 PM Message In : 962796322201,2003, لبيع كيا ريو
9/3/2005 1:05:36 PM 90050
9/3/2005 1:05:36 PM Message In : +962795769711,

Figure 15 : exemples de SMS traités par CATS

Comme le montre la Figure 15, les données envoyées à CATS contiennent accidentellement des lignes vides, des morceaux d'annonces incomplètes, des mots étrangers, ou bien sont complètement en anglais. Ce sont des énoncés spontanés rédigés de façon très naturelle.

Les raisons d'avoir un tel mélange de textes non propres sont nombreuses : erreurs humaines naturelles (envoi d'un message incomplet en appuyant sur la touche « OK » involontairement, SMS rédigés par des personnes résidant en Jordanie mais parlant l'anglais et pas l'arabe, et n'hésitant pas à envoyer des annonces en anglais pour leurs besoins commerciaux).

2.2.3.3 Architecture et ressources

2.2.3.3.1 Architecture

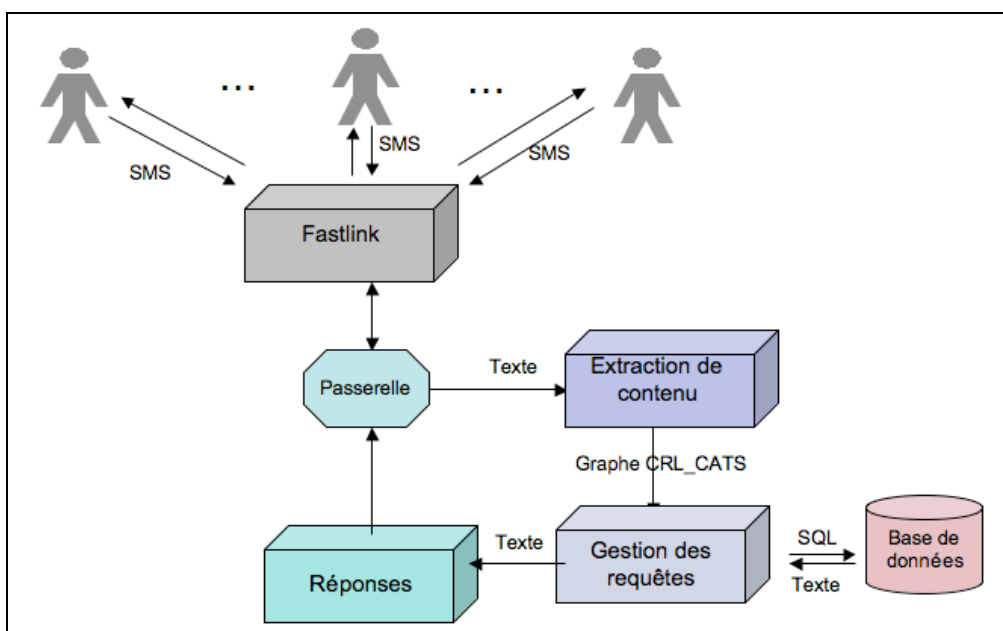


Figure 16 : architecture du système CATS

La Figure 16 montre l'architecture du système CATS : un vendeur ou un acheteur envoie un SMS, le texte de ce SMS est analysé par un extracteur d'information et est représenté dans le format intermédiaire CRL-CATS (Content Representation Language). Cette représentation est traduite en requête SQL et envoyée vers une base de données.

Cette traduction comporte un module d'inférence, qui explicite les informations implicites sûres (comme le montre l'exemple de la Figure 18, la marque [Mitsubishi] implique que le pays est le [Japon]).

2.2.3.3.2 Composants linguistiques et sémantiques

❖ Description

Le système CATS met en œuvre deux composants principaux, un extracteur de contenu et un composant de gestion de requêtes (le noyau fonctionnel) :

▪ Un Extracteur de Contenu « EC »

L'extracteur de contenu reçoit les SMS et traduit (« enconvertit ») leur texte vers la représentation CRL-CATS en utilisant un dictionnaire spécifique du domaine et une grammaire « dynamique » écrits dans le langage EnCo (Uchida and Zhu 1999).

Comme le montre l'exemple de la Figure 17, CRL-CATS ressemble à une représentation de type *Propriété {objet, valeur}*.

```

2000 رينو ميغان م 2000
; لبيع رينو ميغان م 2000
; A vendre Renault Mégane m 2000
[S]
sal(saloon:00, sale:00)
mak(saloon:00, RENAULT(country<France, country<europe):07)
mod(saloon:00, Megane(country<France, country<europe, make<RENAULT):0C)
yea(saloon:00, 2000:0K)
[/S]

```

Figure 17 : exemple de représentation CRL-CATS

Pour la propriété type (*make*), l'objet est une voiture (*saloon*) et la valeur est égale à `RENAULT(country<France, country<europe)`.

Pour la propriété modèle (*mod*), l'objet est une voiture (*saloon*) et la valeur est égale à `Megane(country<France, country<europe, make<RENAULT)`. Pour la propriété année (*yea*), l'objet est une voiture (*saloon*) et la valeur est égale à « 2000 ».

▪ Un composant de gestion de requêtes

Le gestionnaire de requêtes (QM : Query Manager) permet de convertir la représentation CRL-CATS vers une requête SQL (de sélection pour l'achat et/ou d'insertion pour la vente). Il manipule aussi les situations dans lesquelles aucune réponse n'a été trouvée.

SMS	CRL-CATS	SQL
<p>مطلوب ميتسوبيش لانسر</p> <p>« Recherche une Mitsubishi Lancer »</p>	<pre>wan(saloon:00, wanted:00) mak(saloon:00, MITSUBISHI(country<japan):06) mod(saloon:00, Lancer(country<japan, make<MITSUBISHI):0G)</pre>	<pre>select MsgContact from Cars where make ='mitsubishi' and model ='lancer' and maincat ='saloon'</pre>
<p>للبيع لانسر 99 سعر 5000 دينار</p> <p>« A vendre une Lancer 99 prix 5000 dinars »</p>	<pre>sal(saloon:00, sale:00) mod(saloon:00, Lancer(country<japan, make<MITSUBISHI):06) yea(saloon:00, 99:0I) pri(saloon :00, 5000:0L)</pre>	<pre>Insert into cars (maincat, model, year, price, country, make) Values('saloon', 'lancer', '99', '5000', 'japan', 'mitsubishi')</pre>

Figure 18 : exemples de CRL-CATS transformés en requêtes SQL

❖ Taille et complexité

L'extracteur de contenu utilisé dans le système CATS nécessite un dictionnaire et des règles.

Le dictionnaire utilisé contient 30.000 entrées, dont 20.000 ont été générées automatiquement (grâce à un répertoire de variantes et de fautes d'orthographe fréquentes). Il relie les mots et les expressions arabes d'un domaine spécifique aux concepts de ce domaine, en gardant les caractéristiques sémantiques, syntaxiques et morphologiques utilisées dans l'analyse des SMS arabes.

La structure des entrées inclut des abréviations, différentes écritures pour la même entrée, différentes formes orthographiques, et d'autres formes de jargon utilisé dans le sous-langage en question.

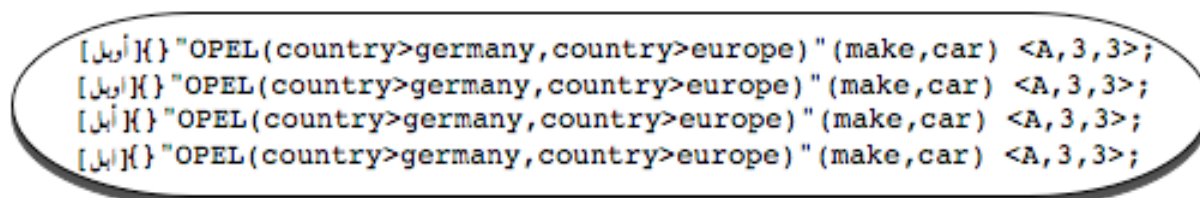


Figure 19 : exemple d'entrées du dictionnaire CATS

Les 710 règles utilisées dans le système CATS permettent l'analyse du texte des annonces envoyées au système.

2.2.3.3.3 Corpus

Le corpus actuel comprend 12000 annonces. Il grossit tous les jours mais cet accroissement évolue suivant le nombre de publicités commerciales effectuées dans les journaux et les médias à propos du système CATS. Ce corpus n'est accessible qu'au développeur (D. Daoud).

Le corpus de développement a été « inventé » à partir de la collecte de petites annonces sur le Web. Il contient 1817 SMS (16047 mots au total).

Pour mesurer la complexité lexicale du corpus produit par CATS, (Daoud 2006) a utilisé comme mesure principale le TTR (Type/Token Ratio), le rapport entre le nombre des types⁹ représentant les différents mots et le nombre total des mots d'un texte donné. D'après lui, plus ce rapport est faible, plus le langage est spécialisé. Le tableau Tableau 3 résume les différentes valeurs trouvées. Cependant, comme le TTR décroît quand la taille du corpus croît, il faut comparer les TTR pour des corpus de même taille.

Domaine	Nombre de SMS	Longueur moyenne (mots)	Types	Tokens	TTR (Types/Token)
Automobile	771	9	1181	5875	0,201
Immobilier	641	12,5	1441	6182	0,233
Emploi	174	14,61	921	2452	0,375
Divers	231	6,42	1099	1538	0,714

Tableau 3 : TTR (Type/Token Ratio) dans les sous-corpus de CATS.

2.2.3.4 Portage linguistique

Le système CATS n'avait pas été porté vers d'autres langues au début de notre étude, mais c'est un bon exemple d'application traitant des énoncés spontanés en langue naturelle, et un très bon candidat à des expériences de portage. Pour aborder concrètement le problème général du portage linguistique d'applications de e-commerce traitant des énoncés spontanés, nous avons donc entrepris de porter CATS, dans un premier temps vers le français, puis si possible vers l'anglais et d'autres langues européennes.

⁹ Un « type » est la forme de base dans le dictionnaire, i.e. un lemme en français, et en général une racine consonantique trilitère ou quadrilitère en arabe.

2.2.4 IMRS

Le système IMRS est un bon exemple de service Web traitant des énoncés en LN pour la gestion de contenu.

2.2.4.1 Brève présentation

IMRS « Impression-based Music-Retrieval System » (Kumamoto 2007) est un système de recherche de musique basé sur un dialogue en langue naturelle en japonais.

Le service visé consiste à permettre à un utilisateur de rechercher des morceaux de musique à partir d'impressions subjectives (dansant, calme, rythme plus rapide, etc.). Les informations de ce type sont généralement entrées en utilisant l'une des trois méthodes suivantes : (i) les utilisateurs sélectionnent un ou plusieurs mots à partir d'une liste présentée (Tsuji, Hoshi et al. 1997), (ii) les utilisateurs sélectionnent un ou plusieurs axes, chacun défini par un mot (une « impression »), à partir d'une liste présentée, et donnent un score à chacun de ces mots à l'aide d'un curseur (Sato, Ogawa et al. 2000), (iii) les utilisateurs sélectionnent un ou plusieurs axes, chacun défini par une paire d'impressions (antonymes) à partir d'une liste présentée (voir Tableau 4), et évaluent leur besoin selon chacun des axes à l'aide d'un curseur placé entre les 2 extrémités de l'intervalle associé à la paire (Ikezoe, Kajikawa et al. 2001).

Score N°	Paire de mots d'impressions
1	Quiet – Noisy
2	Calm - Agitated
3	Refreshing - Depressing
4	Bright - Dark
5	Solemn - Flippant
6	Leisurely - Restricted
7	Pretty - Unattractive
8	Happy - Sad
9	Relax - Arouse
10	The mind is restored – The mind is vulnerable

Tableau 4 : dix axes (définis par des paires) pour représenter les impressions musicales

Impression-based Music-Retrieval System

You are requested to select one or more impression scales from the ten scales listed below to input impressions, and check a radio button in each of the selected scales.

<input checked="" type="radio"/> Very quiet	<input type="radio"/> Quiet	<input type="radio"/> A little quiet	<input type="radio"/> Average	<input type="radio"/> A little noisy	<input type="radio"/> Noisy	<input type="radio"/> Very noisy
<input type="radio"/> Very calm	<input checked="" type="radio"/> Calm	<input type="radio"/> A little calm	<input type="radio"/> Average	<input type="radio"/> A little agitated	<input type="radio"/> Agitated	<input type="radio"/> Very agitated
<input type="radio"/> Very refresh	<input type="radio"/> Refresh	<input checked="" type="radio"/> A little refresh	<input type="radio"/> Average	<input type="radio"/> A little depressing	<input type="radio"/> Depressing	<input type="radio"/> Very depressing
<input type="radio"/> Very bright	<input type="radio"/> Bright	<input type="radio"/> A little bright	<input checked="" type="radio"/> Average	<input type="radio"/> A little dark	<input type="radio"/> Dark	<input type="radio"/> Very dark
<input type="radio"/> Very solemn	<input type="radio"/> Solemn	<input type="radio"/> A little solemn	<input type="radio"/> Average	<input checked="" type="radio"/> A little flippant	<input type="radio"/> Flippant	<input type="radio"/> Very flippant
<input type="radio"/> Very leisurely	<input type="radio"/> Leisurely	<input type="radio"/> A little leisurely	<input type="radio"/> Average	<input type="radio"/> A little restricted	<input checked="" type="radio"/> Restricted	<input type="radio"/> Very restricted
<input type="radio"/> Very pretty	<input type="radio"/> Pretty	<input type="radio"/> A little pretty	<input type="radio"/> Average	<input type="radio"/> A little unattractive	<input checked="" type="radio"/> Unattractive	<input type="radio"/> Very unattractive
<input type="radio"/> Very happy	<input type="radio"/> Happy	<input type="radio"/> A little happy	<input type="radio"/> Average	<input type="radio"/> A little sad	<input checked="" type="radio"/> Sad	<input type="radio"/> Very sad
<input type="radio"/> Calm down very much	<input type="radio"/> Calm down	<input type="radio"/> Calm down a little	<input type="radio"/> Average	<input type="radio"/> Arouse a little	<input checked="" type="radio"/> Arouse	<input type="radio"/> Arouse very much
<input type="radio"/> The mind is healed very much	<input type="radio"/> The mind is healed	<input type="radio"/> The mind is healed a little	<input type="radio"/> Average	<input type="radio"/> The mind is vulnerable a little	<input checked="" type="radio"/> The mind is vulnerable	<input type="radio"/> The mind is vulnerable very much

Figure 20 : écran de choix des impressions (Kumamoto 2007)

Avec l'augmentation du nombre de mots représentant les impressions, ces approches augmentent le coût en terme de temps et de travail nécessaire pour saisir les impressions. Mais si l'ensemble de mots est trop limité, les utilisateurs ne peuvent souvent pas représenter exactement leurs propres impressions. Une interface en langue naturelle est donc une solution qui permet aux utilisateurs d'entrer leurs impressions sans limitation du vocabulaire.

Pour répondre à cette dernière solution, (Kumamoto 2004) et (Kumamoto and Ohta 2003) ont développé les premières interfaces qui permettent d'interpréter 164 mots d'impressions, comme « *happy* » et « *sad* » et 119 niveaux de modifications, comme « *a little* » et « *comparatively* » et de générer un vecteur de requête pour retrouver la musique correspondante. Par contre, ces interfaces ne permettent pas de modifier les vecteurs requêtes utilisés récemment : les utilisateurs sont obligés d'entrer une nouvelle phrase et en conséquence une nouvelle requête est formée, au lieu de modifier l'ancienne, récemment utilisée. Une nouvelle version du système de Kumamoto a été créée. C'est un système de dialogue en langue naturelle qui dépasse les précédentes limitations et permet de générer ou modifier les vecteurs de requêtes pour la recherche de musique.

La Figure 20 est un exemple d'écran qui montre les différents choix possibles des impressions. Le résultat trouvé pour cette requête est l'ensemble des fichiers audio représentés dans l'écran de la Figure 21.

Retrieval Results

Retrieval Conditions

Quiet	⇔ Noisy	Very Quiet
Calm	⇔ Agitated	Calm
Refresh	⇔ Depressing	A Little Refresh
Bright	⇔ Dark	Average
Solemn	⇔ Flippant	A Little Flippant
Leisurely	⇔ Restricted	Restricted
Pretty	⇔ Unattractive	Unattractive
Happy	⇔ Sad	Sad
Calm Down	⇔ Arouse	Arouse
The Mind Is Healed	⇔ The Mind Is Vulnerable	The Mind Is Vulnerable

(Copyright Notice for MIDI Data)

	MIDI File Name	Distance	File Size (byte)		MIDI File Name	Distance	File Size (byte)
1	kleine_romanze_grt	284.26	3903	11	loirei_mb	338.93	2521
2	nocturne_sat5_pi	293.56	5416	12	gnossienne2_pi	340.36	3809
3	bm4	309.50	2371	13	flower	341.54	5200
4	caresse_pi	310.06	3414	14	traumerei_mb	341.93	4357
5	catbedrale_engloutie_pi	317.39	16312	15	kanon_pachelbel_pi	344.66	17463
6	bm3	320.76	4769	16	pavane_belle_pi	346.78	2355
7	snow0	323.24	9574	17	pavane_defunte_pi	347.12	17195
8	n_season	333.81	8463	18	rondonderry	351.85	13182
9	snow3	336.16	16427	19	reverie_du_pauvre_pi	352.06	3241
10	melodie_pi	337.29	2833	20	berceuse_grg_pi	352.08	8370

[Go Back](#)

Figure 21 : résultats trouvés par IMRS

2.2.4.2 Type des énoncés traités

Les énoncés traités par le système sont des phrases ou fragments de phrase spontanés qui sont transformés en vecteurs de requête. Chaque vecteur est composé de dix composants présentés dans le Tableau 4. La valeur d'un composant est un nombre réel entre 1 et 7. Le symbole particulier « *nil* » veut dire « *don't care* ».

Par exemple, l'axe numéro 8 « *Happy – Sad* » est caractérisé par sept valeurs intermédiaires : « *very happy* », « *happy* », « *a little happy* », « *medium* », « *a little sad* », « *sad* », et « *very sad* » qui correspondent respectivement aux valeurs 7.0, 6.0, 5.0, 4.0, 3.0, 2.0, et 1.0. Une requête pour trouver un morceau de musique qui donne une heureuse impression (*happy*) correspondra ainsi au vecteur à dix dimensions suivant :

```
(nil nil nil nil nil nil nil 6,0 nil nil)
```

De la même façon, pour trouver un morceau de musique clair (axe 4) et triste (axe 8), le vecteur à générer est :

(nil nil nil 2,0 nil nil nil 2,0 nil nil)

Des nuances pourront être apportées, soit par le contexte de la demande (« *somewhat quite sad* » → 1,5 ?), soit par réglage (apprentissage) en fonction des choix de l'utilisateur.

La génération des vecteurs de requêtes est faite à l'aide de règles d'interprétation à partir de 164 formes d'impression (mots). Le Tableau 5 est un extrait des règles d'interprétation.

Scale No.	1	2	3	4	5	6	7	8	9	10
Pitiful	nil	Nil	2.48	2.13	nil	nil	nil	1.75	nil	nil
not pitiful	nil	Nil	nil	5.41	nil	nil	nil	5.37	nil	nil
Classic	5.42	5.56	nil	3.47	5.57	nil	5.51	nil	5.06	5.09
Gentle	5.49	5.79	5.62	5.27	nil	5.62	6.01	5.10	5.85	6.16
Powerful	2.13	Nil	nil	nil	nil	nil	nil	nil	2.38	nil

Tableau 5 : extrait des règles d'interprétation contrôlant le passage entre les impressions et les vecteurs requêtes (tiré de (Kumamoto 2007))

Dans IMRS, la représentation interne se présente donc comme une liste de couples <attribut, valeur>, où le nom d'un attribut est pris dans une langue naturelle (l'anglais dans le prototype démontré, mais on passe à une autre langue par une simple table de correspondance), et où les valeurs sont scalaires.

Pour l'évaluation, on calcule une distance entre le vecteur requête généré par le système et le vecteur barycentre (représentant la moyenne) des trois morceaux musicaux les plus semblables à la requête, i.e les plus proches au sens de la distance cartésienne.

Nous nous intéressons dans ce qui suit à l'extraction des (mots-clés correspondant aux) impressions à partir des textes spontanés saisis par les utilisateurs. On commence par une analyse morphologique, puis on extrait le contenu, un peu comme dans CATS, mais avec des outils différents.

2.2.4.3 Architecture et ressources

Nous présentons dans ce qui suit l'architecture d'IMRS ainsi que ses composants linguistiques et sémantiques.

2.2.4.3.1 Architecture

Comme le montre la Figure 22, les traitements du système prototype se composent des étapes suivantes : saisie d'un énoncé exprimant la musique recherchée, analyse morphologique de l'énoncé, extraction des mots-clés exprimant les impressions, calcul d'un vecteur requête à partir de chaque mot-clé et de ses modificateurs, et construction du vecteur requête à dix dimensions par combinaison linéaire et normalisation du résultat.

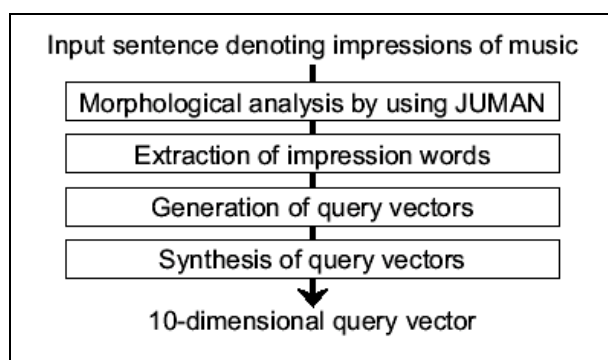


Figure 22 : étapes pour générer les vecteurs requêtes à partir des énoncés

2.2.4.3.2 Composants linguistiques et sémantiques

❖ Analyseur morphologique

L'analyseur morphologique utilisé dans ce travail pour décomposer les phrases japonaises en mots puis les lemmatiser et analyser est JUMAN (Kurohashi and Nagao 1999). Cet analyseur s'appuie sur un dictionnaire (celui de EDR). Pour chaque mot reconnu, JUMAN produit un ou plusieurs résultats, chacun composé d'un lemme (entrée dictionnaire), avec sa partie du discours (nom, verbe, adjectif, etc), et son paradigme de flexion (pour les verbes et les adjectifs).

❖ Extracteur de mots-clés représentant les impressions

L'étape d'analyse morphologique donne en sortie des mots annotés à partir desquels sont extraites les informations importantes et nécessaires pour la construction des vecteurs requêtes telles que les impressions, les expressions comparatives, les mots négatifs. Si des mots négatifs sont extraits, ils deviennent une paire composée de la négation et du mot représentant l'impression. Par exemple, la paire « *not-pretty* » est extraite à partir de l'énoncé « *I don't want a pretty one* ».

2.2.4.3.3 Corpus

La base de données contient 160 morceaux de musique classique qui durent entre 30 secondes et quelques minutes. Tous les morceaux sont en libre accès à l'adresse <http://nocturne.vis.ne.jp/>.

Les auteurs n'ont pas diffusé leur corpus (énoncés japonais), ni donné d'indications chiffrées. On n'a que les quelques exemples présentés dans leurs articles. À partir de ces exemples, nous devons donc créer un corpus de développement « naturel » dans chaque langue cible, en l'imaginant.

2.2.4.4 Portage linguistique

Le système présenté par (Kumamoto 2007) est un prototype testé pour des énoncés spontanés en langue japonaise. Il n'a pas été porté vers d'autres langues. Les composants dépendant de la langue sont l'analyseur morphologique et l'extracteur des mots-clés représentant les impressions. L'intérêt de ce système pour notre étude est que sa représentation interne du contenu semble accessible (ce sont les vecteurs vus plus haut).

2.2.5 Récapitulation

Le Tableau 6 résume les principales informations concernant les différents systèmes d'e-commerce détaillés précédemment et classés par ordre d'importance. CATS et IMRS sont les meilleurs systèmes qui correspondent totalement à notre thème de recherche : **systèmes d'e-commerce utilisant les énoncés spontanés** (comme les petites annonces ou les rapports d'utilisateurs sur des incidents, etc.). De plus, nous pouvons accéder à tout le code de CATS, et aux représentations internes de IMRS.

Dans d'autres cas, les énoncés sont plutôt « semi-spontanés », comme dans le cas du système NLSA qui traite aussi des annonces moins spontanées comme « *je cherche un ordinateur Pentium 4 disque dur 60 Giga* ». Elles ne contiennent pas des erreurs, des abréviations et des transcriptions multiples de noms, etc., comme c'est le cas dans CATS.

Systeme	Énoncés	Interface	Domaine	Langue initiale	Approche TALN	Portage linguistique
CATS (D.Daoud)	Spontanés	SMS	Occasion auto, immobilier, emploi, divers	Arabe	Sémantique	Non
IMRS (T.Kumamoto)	Spontanés	Dialogue non contrôlé	Recherche de morceaux de musique	Japonais	Morphologique et sémantique	Non
NLSA (IBM)	Semi-spontanés	Dialogue basé sur le Web	Ordinateurs	Anglais	Sémantique et syntaxique	Non
e-Bay	Spontanés	Contrôlée par des menus	Occasion auto, immobilier, informatique ...	Anglais	Manuelle	Site disponible dans plus de 20 langues
Pertinence Summarizer (A.Lehman)	Réfléchis (textes scientifiques et techniques)	Documents, Web, email	Scientifique	Français	Sémantique	Arabe, allemand, anglais, arabe, chinois, coréen, espagnol, grec, italien, japonais, néerlandais, norvégien, portugais, russe

Tableau 6 : tableau récapitulatif des principaux systèmes d'e-commerce étudiés

2.3 Études méthodologiques antérieures

Nous présentons maintenant quelques précédentes approches du portage ou plus généralement de la multilinguisation par approche de génération multilingue (cas du projet multilingue MKBEEM), par adaptation à une autre langue (cas d'une projection interlingue d'extraction d'information) et par adaptation à un autre type de texte (adaptation de type Amilcare).

2.3.1 Génération multilingue : l'approche de MKBEEM

Le projet MKBEEM « Multilingual Knowledge-Based European Electronic Marketplace » (<http://www.mkbeem.com/>) a démarré en février 2000 et a fini en fin novembre 2002.

2.3.1.1 Brève présentation

MKBEEM visait à réaliser une plate-forme de commerce électronique multilingue et pluriculturelle censée offrir des interfaces homme-machine en langue naturelle. Le consortium MKBEEM était coordonné par France Télécom (F), ses partenaires industriels étaient Ellos-Redoute

(Fin), SNCF (F), Schlumberger Sema (E) et FIDAL (F), et ses partenaires académiques étaient UPM « Universidad Politécnica de Madrid » (E), NTUA « National Technical University of Athens » (GR), CNRS (F) et TRCF « Technical Research Centre of Finland » (FIN).

Le projet MKBEEM a couplé le raisonnement sur la connaissance ("KRR") et le traitement automatique des langues naturelles (TALN) afin d'offrir des services de médiation de commerce électronique multilingue (finnois, français, espagnol et anglais).

Ses objectifs de services étaient :

- Maintenance semi-automatisée de catalogues de produits et de services multilingues ;
- Interprétation des requêtes en langue naturelle (ex : *on voudrait arriver à Paris de Lannion ce soir*) ;
- Transactions multilingues.

Ses objectifs techniques étaient :

- Développer une librairie de composants intelligents (Web sémantique) pour le commerce électronique multilingue ;
- Montrer l'efficacité de solutions génériques éprouvées par des utilisateurs français, anglais, espagnols et finnois pour les domaines du tourisme et de l'achat en ligne de vêtements.

Une séparation stricte entre les données spécifiques à une langue (lexiques, grammaires) ou à un pays (monnaie, mesures (cm, pouces)) et les connaissances du domaine (ontologies du domaine métier du prêt-à-porter ou du domaine des voyages) est un principe fondamental de l'approche MKBEEM.

MKBEEM est un projet qui n'a permis de produire qu'un prototype interne et n'aboutissant pas à des résultats commercialisés et publics. Nous n'avons donc pas pu trouver d'informations concernant la taille et la complexité du travail. Il n'y a pas eu d'application ni de déploiement de la technique d'extraction d'information utilisée, ni de portage vers d'autres langues.

2.3.1.2 Type des énoncés traités

Dans ce travail, nous distinguons deux types de données :

Le premier type concerne les données spontanées, c'est à dire l'ensemble des requêtes en langue naturelle émises par les utilisateurs. Par exemple, un utilisateur a donné la requête suivante en demandant plusieurs services : "*J'aimerais visiter Versailles et réserver un hôtel à Paris*". Dans ce cas, MKBEEM doit identifier deux services : réservation de train et réservation d'hôtel.

Le deuxième type concerne des données non spontanées comme les catalogues. Ce sont des textes multilingues de description des produits obtenus par traduction automatique.

Nous n'avons pas pu avoir un accès aux rapports internes de ce projet pour avoir plus d'informations sur la taille des données utilisées.

2.3.1.3 Méthodes et ressources

L'approche technique est censée permettre de :

- Établir la correspondance entre une requête d'utilisateur en langue naturelle et la représentation « ontologique » du contenu de cette requête ;
- Créer des services composites en utilisant les services ou les produits fournis par le(s) catalogue(s) du ou des fournisseurs d'offres (par ex. si l'utilisateur demande deux services différents à la fois dans une seule requête) ;
- Établir la correspondance entre la représentation ontologique du contenu de la requête de l'utilisateur et les ontologies des offres de services ;

- Établir la correspondance entre la description en langue naturelle d'un produit et sa représentation ontologique.

Les ontologies peuvent être exploitées pour la classification et l'indexation des produits ou des services dans les catalogues, aussi bien que pour faciliter l'interprétation et l'inférence d'information pertinente en rapport avec la requête de l'utilisateur. Les composants essentiels sont le catalogue multilingue et le traitement des requêtes de l'utilisateur.

2.3.1.3.1 Catalogue multilingue

Dans le cadre du projet MKBEEM, un scénario de catalogue multilingue a été réalisé : ce sous-système contrôle les descriptions des nouveaux produits rajoutés au catalogue et extrait les caractéristiques (indexation automatique) des nouveaux produits afin de proposer une ou des catégories possibles dans l'arborescence existante des produits. Ce but est atteint avec l'aide des ontologies des produits. Ensuite, la traduction automatique produit les textes de description des produits dans les autres langues.

2.3.1.3.2 Traitement des requêtes

Les catalogues disponibles peuvent être explorés avec des requêtes en langue naturelle. Après l'identification de la langue, la requête est analysée et transformée dans une représentation ontologique indépendante de la langue (extraction de contenu). Cette représentation est ensuite utilisée afin d'identifier le service demandé par l'utilisateur (ex. horaires de trains, location de voiture, réservations d'hôtel/chalet/appartements, achat de vêtements). Une fois qu'un ou plusieurs services ont été identifiés, MKBEEM récupère les paramètres obligatoires et facultatifs du service (par exemple date et heure de départ, lieu d'arrivée pour une réservation de train, taille et couleur pour l'achat d'un vêtement) avant de lancer une recherche dans la base de données du fournisseur de contenu.

Les requêtes sont analysées et les propriétés extraites sont liées aux propriétés enregistrées dans la base d'ontologie. Les liens qui forment la représentation sémantique entre la requête lancée par l'utilisateur et un produit sont basés sur des graphes.

2.3.1.4 Domaine d'application

Dans ce projet, l'usage des ontologies devrait permettre de résoudre les ambiguïtés linguistiques et de trouver le service le plus adapté à la demande de l'utilisateur. Soit par exemple, les requêtes comme « *On voudrait arriver à Paris de Lannion ce soir* » et « *on voudrait aller de Lannion à Paris ce soir* ». Avec le premier énoncé, l'utilisateur attend une correspondance qui arrive à Paris le soir, il doit donc partir bien avant, dans le deuxième cas il veut partir le soir (donc arriver pendant la nuit).

2.3.2 Études sur l'adaptation à une autre langue

Il s'agit ici d'adapter à une nouvelle langue une méthode existante d'extraction d'information. Il s'agit ici de passer d'une langue à une autre sans utiliser un pivot, c'est donc une « projection translingue » (terme peu exact, mais calqué de l'anglais, équivalent à « portage linguistique ») qui correspond à nos objectifs.

L'extraction d'information est une application importante pour le traitement du langage naturel. (Riloff, Schafer et al. 2002) montrent que les systèmes d'extraction d'information qui dépendent des outils d'analyse et des dictionnaires spécialisés et spécifiques à une langue ne sont pas facilement portables à de nouvelles langues.

2.3.2.1 Brève présentation

Nous commençons par un brève historique des systèmes d'extraction d'informations et leur portage linguistique.

2.3.2.1.1 Historique

Ces dernières années, il y a eu une foison de projets autour de l'idée du « portage linguistique ». (Walker, Fromer et al. 1998) ont développé des modèles de portage linguistique pour des étiquettes de parties du discours, de groupes nominaux, d'entités nommées, et d'analyse morphologique (lemmatisation) pour quatre langues. (Ngai and Florian 2001) ont développé les modèles pour porter des analyseurs de dépendance de l'anglais vers le chinois. Il y a eu également un gros travail sur le portage linguistique et le développement des ontologies et des WordNets pour différentes langues (Atserias, Climent et al. 1997).

Les systèmes d'extraction d'information sont coûteux à construire parce qu'ils exigent des textes, des outils d'analyse, et des dictionnaires spécialisés pour chaque domaine d'application et pour chaque langue naturelle. Une solution possible consiste à porter les ressources linguistiques, les outils, et la connaissance du domaine depuis des langues riches en de telles ressources vers des langues pauvres en ces mêmes ressources.

Étant donné un système d'extraction d'information pour une langue source (par exemple, anglais), il est possible de porter ses annotations aux textes qui leur correspondent dans une autre langue (par exemple, français), et d'apprendre automatiquement les règles d'extraction d'information pour la nouvelle langue. Nous présentons dans ce qui suit, un cas de portage (anglais vers français) d'un tel système pour le domaine d'accidents d'avions.

2.3.2.1.2 Portage linguistique

Il y a plusieurs stratégies pour porter les annotations anglaises d'extraction d'information vers la langue française, comme l'utilisation de l'apprentissage pour créer des règles françaises d'extraction des annotations françaises.

La plupart des systèmes d'extraction d'information emploient un certain modèle d'extraction pour identifier et extraire l'information appropriée. Plusieurs techniques ont été développées pour produire automatiquement des modèles d'extraction pour un nouveau domaine : PALKA (Kim and Moldovan 1993), AutoSlog (Riloff 1993), CRYSTAL (Soderland, Fisher et al. 1995), RAPIER (Califf 1998), SRV (Freitag 1998), meta-bootstrapping (Riloff and Jones 1999), et ExDisco (Yangarber, Grishman et al. 2000).

Pour ce travail, AutoSlog-TS (Riloff 1996) a été utilisé pour produire un modèle d'extraction d'information pour le domaine des accidents d'avions. AutoSlog-TS est un dérivé d'AutoSlog qui produit automatiquement des modèles d'extraction en recueillant des statistiques d'un corpus de textes appropriés (dans le domaine) et de textes inappropriés (hors domaine).

Chaque modèle d'extraction est une expression régulière (sur des primitives linguistiques) qui peut extraire des groupes nominaux à partir d'une des trois fonctions syntaxiques : objet, objet direct, ou objet d'une préposition. Par exemple, les modèles suivants ont pu extraire des véhicules impliqués dans un accident d'avion: “<subject> crashed”, “hijacked <direct-object>”, et “wreckage <np>”.

2.3.2.2 Type des énoncés traités

Dans ce travail, les participants se sont concentrés sur le domaine des accidents d'avion pour extraire la description de l'aéronef impliqué dans l'accident, la description des victimes de l'accident, et la description de l'endroit de l'accident.

AutoSlog-TS est initialisé en utilisant des textes relatant des accidents d'avion qui ont déjà eu lieu. De nouveaux textes sont utilisés pour tester le système. Un humain doit choisir la liste de textes qui peuvent servir comme modèles pour l'extraction d'information.

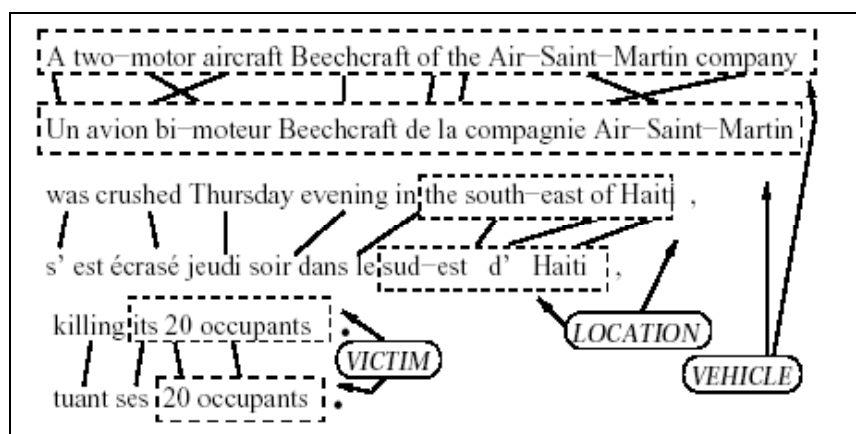


Figure 23 : exemple de texte annoté et aligné au niveau des mots

2.3.2.3 Méthode et ressources

2.3.2.3.1 Méthode

La méthode proposée par (Riloff, Schafer et al. 2002) est basée sur celle de (Yarowsky, Ngai et al. 2001). Elle utilise un système de traduction automatique commercial. La méthode consiste à appliquer les étapes suivantes :

❖ Traduction par système de TA

Un système de TA commercial est utilisé pour produire un corpus parallèle artificiel. Si les erreurs de TA présentent un problème substantiel, on peut recourir à de gros corpus bilingues, comme le corpus canadien « Hansard ».

❖ Alignement au niveau des phrases

Cette tâche était triviale car chaque phrase possède un numéro préservé par le système de TA.

❖ Alignement au niveau des mots

Cette tâche est effectuée en utilisant le système Giza++ (Och and Ney 2000).

❖ Transfert des annotations

Il s'agit de porter des annotations anglaises d'extraction d'information à l'aide du mécanisme décrit dans (Yarowsky, Ngai et al. 2001) rapprochant des paires de phrases annotées, comme le montre la Figure 23.

❖ Apprentissage d'un extracteur de contenu pour la nouvelle langue

Afin d'apprendre les règles d'extraction d'information pour le français, l'outil TBL (*TBL : Transformation-Based Learning*) (Brill 1995) a été adopté. Cet outil est bien adapté à cette tâche, parce qu'il emploie des modèles de règles comme base pour l'apprentissage, et que ces modèles peuvent être facilement réutilisés pour de nouveaux textes français.

Dans le Tableau 7, les règles 1-3 sont des exemples de règles lexicales basées sur des N-grammes.

Rule Condition	Rule Effect
1. $w_1 = \text{crashed}$ $w_2 = \text{in}$	w_3 is LOC
2. $w_1 = \text{wreckage}$ $w_2 = \text{of}$	w_3 is VEH
3. $w_1 = \text{injuring}$	w_2 is VIC
4. $w_1 = \text{NOUN}$ $w_2 = \text{crashed}$	w_1 is VEH
5. $w_1 = \text{VERB}$ $w_2 = \text{down}$ $w_3 = \text{in}$	w_4 is LOC
6. $w_1 = \epsilon$ $w_2 = \text{ART}$ $w_{4-7} = \text{crashed}$	w_3 is VEH
7. $w_2 = \text{COMMA}$ $w_3 = \text{which}$ $w_4 = \text{crashed}$	w_1 is VEH
8. $w_1 = \text{in}$ $w_2 = \text{LOCATION}$ $w_3 = \text{NOUN}$	w_3 is LOC
9. $w_1 = \text{VERB}$ $w_2 = \text{VICTIM}$ $w_3 = \text{NOUN}$	w_3 is VIC
10. $w_1 = \text{ART}$ $w_2 = \text{VEHICLE}$	w_2 is VEH

Tableau 7 : exemples de règles TBL
(LOC=location, VEH=vehicle, VIC=victim)

La règle 5 relie un trigramme verbal comme « *went down in* », « *shot down in* », et « *came down in* » à un circonstant de lieu.

Dans ce travail, l'absence d'un analyseur syntaxique a été compensée par des modèles de règles tels que des modèles de règles de capture du sujet. Par exemple :

- la règle 6 recherche un article au début d'une phrase et le mot "*crashed*" un peu après, et en déduit que l'article appartient à un groupe nominal dont la tête (w_3) est un véhicule.
- la règle 7 cherche les trois marques "*COMMA which crashed*" et en déduit que le mot précédant la virgule est un véhicule.

Des modèles de règles d'enrichissement des mots ont été conçus aussi pour regarder les mots qui ont déjà été marqués et pour prolonger les frontières de l'annotation afin de couvrir un groupe nominal complet. Par exemple :

- les règles 8 et 9 étendent les annotations « *victim* » et « *location* » vers la droite,
- la règle 10 étend l'annotation « *vehicle* » vers la gauche.

Le principe du TBL se base sur l'utilisation de plusieurs processus possibles et le choix de celui qui donne le meilleur résultat. La Figure 24 montre les différents processus possibles pour appliquer le TBL. Les données des deux corpus anglais et français sont divisées en deux sous-ensembles : non-annotées (« *plain* ») et annotées (« *antd* » ou « *Tst* ») (marqués en blanc dans la figure) et leurs images obtenues par traduction (marquées en noir dans la figure).

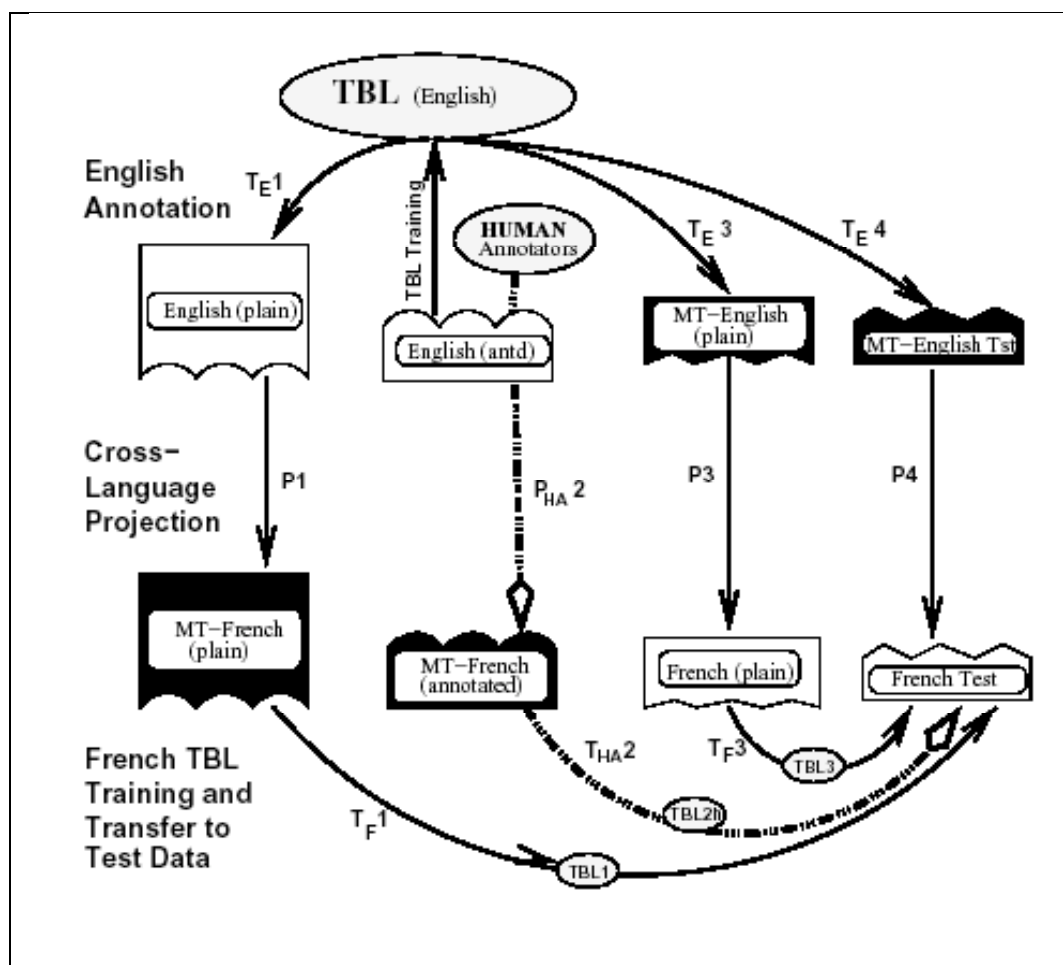


Figure 24 : portage d'un EC basé sur la TBL

Nous détaillons par exemple la voie $T_{E1} \rightarrow P1 \rightarrow T_{F1}$:

Entrée	=	corpus anglais d'entraînement (données anglaises annotées), corpus anglais de développement (données anglaises non-annotées), et corpus de test (données françaises non-annotées)
Processus	=	$T_{E1} \rightarrow P1 \rightarrow T_{F1}$
Sortie	=	corpus français (image du corpus anglais de développement), et corpus français de test annoté (données françaises annotées)

L'outil TBL est entraîné pour l'anglais sur un ensemble de 140 K mots de données annotées et les règles de TBL apprises sont appliquées sur le sous-corpus anglais non-annoté. Les annotations seront portées par la suite, via le système d'alignement au niveau des mots Giza++, sur l'image en français « *MT-French* ». Ensuite, TBL est entraîné sur les annotations de « *MT-French* », éventuellement post-éditées, et les règles apprises sont appliquées aux données de test du français.

Un autre chemin plus direct est $T_{E4} \rightarrow P4 \rightarrow \text{French-Test}$, dans lequel l'outil TBL adapté à l'anglais, est appliqué immédiatement sur le résultat de l'alignement au niveau des mots de la traduction « *MT-English* » des données test du français. Les annotations de « *MT-English* » seront traduites directement en les données de test du français.

Entrée	=	corpus français de test (données françaises non-annotées)
Processus	=	$T_{E4} \rightarrow P4 \rightarrow \text{French-Test}$
Sortie	=	corpus anglais de test annoté (image du corpus français de test), et corpus français de test annoté par projection

2.3.2.3.2 Ressources

En ce qui concerne l'apprentissage de l'extracteur de contenu, une expérimentation a été faite à partir de deux types de données : les annotations automatiques et les annotations humaines. 56 modèles de règles indépendants ont été définis.

Les corpus de développement utilisés dans ces expériences ont été extraits à partir de textes en anglais et en français tirés de journaux concernant l'aéronautique. Ils ont été créés automatiquement en recherchant les articles qui contiennent des mots-clés relatifs aux accidents d'avion. Deux corpus dans les deux langues ont été extraits à partir de ces journaux. Le corpus anglais contient approximativement 420.000 mots, et le corpus français contient environ 150.000 mots.

Pour chaque langue, des humains ont fait l'annotation à l'aide d'étiquettes de style SGML. Ces étiquettes indiquent l'emplacement de l'accident, l'aéronef impliqué dans l'accident, et des informations sur les victimes (morts, blessés, survivants). Un tiers du corpus anglais et la moitié du corpus français ont été annotés.

2.3.2.4 Domaine d'application

L'application d'extraction d'information, conçue initialement pour l'anglais, a été portée vers le français en suivant les étapes détaillées précédemment. Une évaluation a été faite par les auteurs. En ce qui concerne l'annotation, le nombre de coïncidences exactes (*exact-word-match*¹⁰), alignement au niveau du mot, varie de 16 à 31 % pour le français et de 24 à 27 % pour l'anglais.

Ces chiffres relativement bas suggèrent que le critère est trop strict. Le nombre d'accords (*Exact-NP-match*¹¹), alignement au niveau segment, était beaucoup plus élevé, de 43-54 % pour le français et de 51-59 % pour l'anglais.

L'approche (*exact-word-match*) est très rigoureuse parce que les annotateurs ne sont parfois pas d'accord sur certains termes (par exemple, "*Boeing 727*" contre "*Nouveau Boeing 727*"). Avec l'utilisation de la mesure (*Exact-NP-match*), ces deux termes se correspondent.

Une évaluation des différents processus utilisés est résumée dans le Tableau 8.

La meilleure performance est celle du processus $T_{E4} \rightarrow P4 \rightarrow \text{French-Test}$, dans lequel il y a deux étapes seulement. La F-mesure du meilleur processus pour le français monolingue est de 45%, elle est donc inférieure de 9% à celle de l'anglais monolingue (54%).

Sachant que les systèmes de TA ne sont pas disponibles pour tous les couples de langues ou sont très mauvais, l'application de cette technique à d'autres couples de langues distantes reste une question ouverte et les difficultés qui peuvent être rencontrées dépendent des couples de langues choisis.

¹⁰ On vérifie si deux mots ont la même étiquette.

¹¹ On vérifie si deux segments ont la même étiquette.

Projection et chemin d'entraînement	Précision	Rappel	F-mesure
T _{E1} : application du TBL _E à <i>English-plain</i> P ₁ : projection sur MT-French (<i>English-Plain</i>) T _{F1} : entraînement du TBL & application sur FrTest	0,69	0,24	0,36
Utilisation des annotations humaines à partir de Eng-Antd P _{ha2} : projection sur MT-French (<i>English-Antd</i>) T _{ha2} : entraînement TBL & application à FrTest	0,56	0,29	0,39
T _{E3} : application du TBL _E à MT-Eng (<i>FrenchPlain</i>) P ₃ : projection sur French-Plain T _{F3} : entraînement du TBL & application sur FrTest	0,49	0,34	0,40
T _{E4} : application du TBL _E à MT-Eng (<i>FrenchTest</i>) P ₄ : projection directe sur French-Test	0,49	0,41	0,45

Tableau 8 : performance du portage de l'EC basée sur le TBL

2.3.3 Adaptation à un autre type de texte

Dans le cadre des campagnes MUC de compréhension de messages, plusieurs technologies d'adaptation ont été développées (Humphreys, Gaizauskas et al. 1998), et un grand nombre d'évaluations comparatives ont été publiées par la DARPA (MUC7 1998).

2.3.3.1 Brève présentation

Avant de passer à la présentation d'une approche adaptative de type Amilcare (Ciravegna 2001a), nous présentons un brève historique sur les possibilités d'adaptation des applications à d'autres types de textes.

2.3.3.1.1 Historique

Selon (Miller, Crystal et al. 1998), une technologie d'adaptation qui demande une annotation linguistique n'est pas souhaitable dans le contexte de l'adaptation des applications à d'autres types de textes. De plus, la quantité de texte annoté nécessaire pour l'apprentissage de l'annotation linguistique doit être réduite afin de réduire au minimum la lourdeur de cette tâche (coût minimum, réduction du temps de développement/maintenance).

Dans le domaine du TALN, adapter une application d'extraction d'information à un nouveau type de texte a été considéré généralement comme une question de portage vers différents types de texte libre¹² (Cardie 1997). L'utilisation de l'extraction d'information pour le Web sémantique nécessite une extension du concept de « type de texte ». Selon (Cardie 1997), il est difficile ou même inefficace d'appliquer les méthodologies basées sur la linguistique et utilisées pour des textes libres à des textes fortement structurés tels que les pages Web produites par des bases de données. Ces méthodes ne sont pas capables de faire face à la variété des éléments extralinguistiques (étiquettes XML, documents formatés, et langue stéréotypée) qui sont employées dans de tels documents.

L'intérêt de l'approche adaptative que nous allons présenter est de bénéficier des technologies et des expériences qui ont été faites précédemment pour construire une application adaptable à différents types de textes et qui fait face à la variété extralinguistique.

¹² Texte non structuré

Nous décrivons, dans ce qui suit, une application d'extraction d'information pour l'annotation dans le domaine du Web sémantique, qui a été développée pour répondre à un critère de portabilité et d'adaptation.

2.3.3.1.2 Présentation

Amilcare (Ciravegna 2001a) est un système d'extraction d'information conçu comme support pour l'annotation des documents dans le cadre du Web sémantique. Il est basé sur une méthodologie adaptative tenant compte des besoins et en particulier :

- adaptation à différents types d'utilisateurs : de l'utilisateur naïf à l'utilisateur expert en matière d'extraction d'information ;
- adaptation à différents types de textes (y compris un mélange de types de textes) ;
- adaptation à différents environnements d'annotation existants en perturbant le moins possible la tâche habituelle de l'annotation.

La Figure 25 montre le mode de fonctionnement normal d'Amilcare. Un algorithme d'apprentissage induit les règles qui permettent l'extraction d'information. Ces règles sont apprises par la généralisation d'un ensemble d'annotations trouvées dans un corpus d'entraînement. Ainsi, le système apprend comment reproduire de telles annotations.

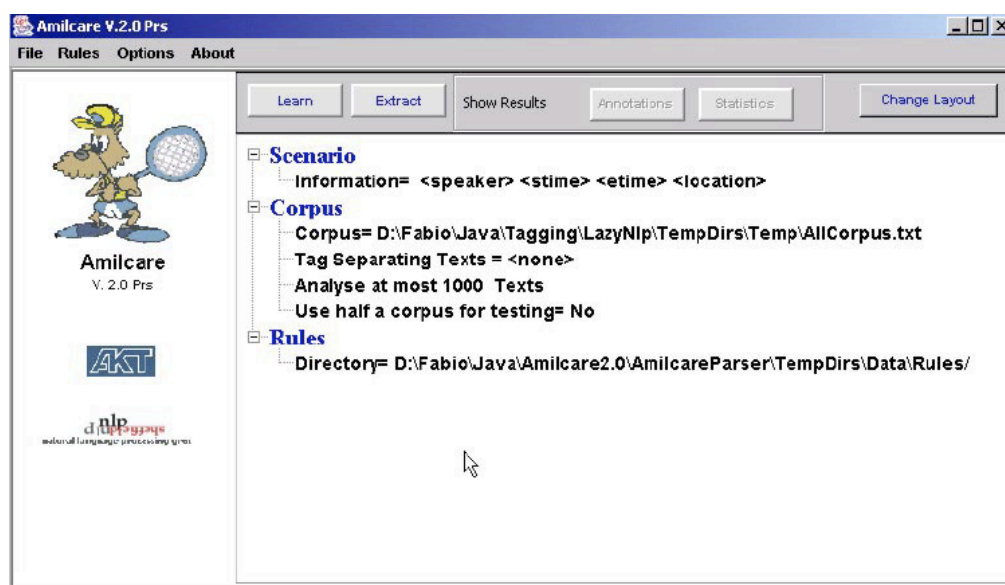


Figure 25 : mode normal d'Amilcare

2.3.3.2 Type des énoncés traités

Afin de porter et/ou de maintenir un système d'extraction d'information, on peut être confronté à quatre types de problèmes :

- l'adapter à de nouveaux domaines d'information, ce qui nécessite d'implémenter (ou modifier) le lexique, les bases de connaissances, etc., et de produire de nouveaux modèles afin de couvrir le nouveau domaine ;
- l'adapter à des sous-langages (voir la définition dans 4.1), ce qui nécessite de modifier la grammaire et le lexique afin de traiter les constructions linguistiques spécifiques et typiques du sous-langage ;
- l'adapter à différents genres de textes spécifiques qui peuvent combiner les deux précédents points : résumé médical, article scientifique, rapports de police peuvent avoir leurs propre lexique, grammaire, discours, structure ;

- l'adapter à différents formats de documents : les documents basés sur le Web peuvent varier de documents XML structurés à des textes non-structurés et/ou non-formatés. Chaque format de document peut donner lieu à différentes contraintes en termes d'analyse de langue.

Comme le montre l'exemple de la Figure 26, les données traitées par Amilcare sont des documents qui présentent un contenu textuel réfléchi.

```
<XML version="1.0" encoding="UTF-8">
Title: seminar announcement:
Speaker: <speaker>Steven Skiena</speaker>
Department of Computer Science
State University of New York, Stony Brook
Date: Tuesday, January 14
Time: <stime>1:30 pm</stime>
Place: <location>Wean Hall 4634</location>
Topic: Algorithms for Square Roots of Graphs

The kth power of a graph  $G = (V, E)$ , written  $G^k$ ,
is defined to be the graph having  $V$  as its vertex set,
with vertices  $u, v$  adjacent in  $G^k$  if and only if
there exists a path of length at most  $k$  between them.

Similarly, graph  $H$  has a  $k$ th root  $G$  if  $G^k = H$ .

Powers of graphs have several interesting properties.
For example, the square of any biconnected graph and the
cube of any connected graph is hamiltonian.
In this talk, we present algorithms to find a square
root  $G$  of a graph  $G^2$  for some special classes of graphs.
This is joint work with Yaw-Ling Lin.

Host: Dana Scott

Appointments can be made through Lydia DeFilippo, x3063,
lydia@cs.cmu.edu.
<speaker>Steve Skiena</speaker> will be at CMU Monday,
January 13, and Tuesday, February 14.
</XML>
```

Figure 26 : exemple de texte d'entraînement

2.3.3.3 Méthode et ressources

Amilcare peut fonctionner en trois modes : mode d'apprentissage (*training mode*), mode de test (*test mode*) et mode de production (*production mode*).

- Le mode d'apprentissage est employé pour induire des règles. L'entrée se compose d'un scénario et d'un corpus d'apprentissage annoté avec l'information à extraire.

La sortie de la phase d'apprentissage est un ensemble de règles capables de reproduire l'annotation sur des textes du même type.

- Le mode test est utilisé pour tester les règles induites sur un nouveau corpus étiqueté (non vu en mode apprentissage), cela permet de voir comment le système fonctionne pour une application spécifique. Lors de son exécution en mode test, Amilcare enlève tout d'abord toutes les annotations du corpus, ensuite, il réannote le corpus en utilisant les règles induites. Enfin, les résultats sont automatiquement comparés aux annotations originales et sont

présentés à l'utilisateur. La sortie de la phase de test est le corpus réannoté par le système et un ensemble de statistiques d'exactitude sur le corpus de test : le rappel, la précision et les détails sur les erreurs du système.

- Le mode de production est utilisé dans une application réelle. Amilcare annote les documents fournis. Si un utilisateur est disponible pour réviser ses résultats, l'administrateur utilise ses corrections pour recommencer l'apprentissage.

Les modes d'apprentissage/test/production peuvent être intercalés pour produire une annotation basée sur l'apprentissage. L'architecture par défaut d'Amilcare contient une ouverture vers les systèmes d'extraction d'information qui réalisent la segmentation, l'étiquetage par les parties du discours, et l'identification des entités nommées.

Amilcare utilise l'algorithme LP² (Ciravegna 2001b), algorithme d'extraction d'information adaptatif, classé parmi les systèmes d'induction (« *Wrapper Induction System* », WIS). Amilcare se base sur deux composants essentiels : un composant d'induction et un composant de généralisation.

❖ **Composant d'induction**

L'algorithme LP² induit deux types de règles symboliques en deux étapes :

- règles qui insèrent des annotations dans les textes ;
- règles qui corrigent les erreurs et les imprécisions dans les annotations fournies par l'étape (1).

Des règles sont apprises par la généralisation d'un ensemble d'exemples marqués par des balises XML dans un corpus d'apprentissage. Une règle d'étiquetage se compose d'une partie gauche, contenant un modèle de conditions sur une séquence de mots connectés, et d'une partie droite qui est une action insérant une balise XML dans le texte. Chaque règle insère une balise simple, par exemple `</speaker>`.

Pour chaque exemple, l'algorithme construit une règle initiale, généralise cette règle, et garde les k meilleures généralisations de la règle initiale.

L'étiquetage produit par les règles appliquées au corpus peut conduire à quelques imprécisions dans la détection des frontières d'expressions. Une erreur typique est par exemple "at `<time> 4 </time> pm`", où le "pm" devrait faire partie de l'expression de temps. L'algorithme d'induction utilisé pour les meilleures règles d'étiquetage est également employé pour des règles de décalage : identification, généralisation, test et sélection.

Si, comme cela arrive souvent, le système produit une balise ouvrante (par exemple `<speaker>`) sans la balise fermante correspondante (`</speaker>`), on l'enlève.

❖ **Composant de généralisation**

Le modèle de règle initial relie des conditions sur des chaînes de mots. Ce type de règle est typique des systèmes d'induction (« *wrappers* »). Ce modèle de règle est intégré avec des conditions additionnelles sur les résultats du pré-processeur linguistique. Cela signifie que des conditions sont imposées non seulement aux chaînes (word="companies"), mais également aux lemmes ("company"), les utilisateurs définissent leurs listes de classes à partir d'un dictionnaire.

Condition	Additional Knowledge				Action
	Word	Lemma	LexCat	Case	
the	The	det	low		
seminar	Seminar	noun	low		
at	At	prep	low		
4		digit	low		<time>
pm		noun	low	timeid	
will	Will	verb	low		

Figure 27 : exemple de généralisation à partir de la règle initiale

L'efficacité de la règle initiale est comparée à celle d'un ensemble des règles plus générales obtenues par un algorithme de recherche spécifique qui commence à modéliser l'exemple d'annotation avec la règle la plus générale et le spécialise en ajoutant des contraintes. Le processus de spécialisation s'arrête quand les règles produites ont atteint une exactitude satisfaisante et quand aucune nouvelle spécialisation n'améliore la précision. De cette façon, les règles qui incorporent différents niveaux de connaissance linguistique sont mises en concurrence directe, et seule la meilleure combinaison de règles survit.

Condition					Action
Word	Lemma	LexCat	Case	SemCat	Tag
	At				
		digit			<time>
				timeid	

Figure 28 : généralisation possible de la règle de la figure précédente

2.3.3.4 Domaine d'application

L'algorithme LP² a été utilisé dans un système industriel appelé « *LearningPinocchio* » sur lequel nous avons trouvé quelques informations sur le coût en termes de temps développement.

Le système « *LearningPinocchio* » a été déployé dans plusieurs applications industrielles telles que l'extraction d'information à partir de curriculum vitae professionnels écrits en anglais et trouvés sur le Web. Il classe les informations trouvées par thème et alimente une base de données.

Le temps mis pour développer une telle application a été estimé à partir d'un exemple concret « moyen » : il a fallu 24 heures pour la définition et la révision d'un scénario (un scénario est raffiné par étiquetage de quelques textes de différentes manières), puis 10 heures environ pour étiqueter 250 textes. L'induction des règles a pris 72 heures sur une machine à 450MHz, avec une fenêtre de taille $w = 4$. Enfin, la validation de résultats du système a nécessité 4 heures.

Amilcare a été porté vers plusieurs domaines et pour plusieurs types de textes, dans la même langue. Étant financé par la DARPA, ce système n'a pas été porté vers d'autres langues. Pour le faire, il faudrait:

- porter les ressources externes et internes utilisées vers la nouvelle langue désirée.
- réécrire le pré-processeur linguistique, seul module dépendant de la langue, le reste du système étant indépendant de la langue.

Le coût serait alors la somme du coût de développement du pré-processeur et du coût estimé dans le paragraphe précédent.

L'approche répond à l'exigence de faire face à différents types de textes comprenant les types mélangés, parce qu'ils peuvent contenir différentes parties de texte dans lesquelles différentes informations sont localisées, en utilisant différentes stratégies (linguistiques). Mais l'adaptation de cette approche à des cas de sous-langage spécialisés tels que les textes des petites annonces spontanées n'a pas été étudiée dans cette approche.

Pour nous, l'intérêt de cette approche est l'idée d'adaptation d'un extracteur d'information à un nouveau domaine en restant dans la même langue. Amilcare prouve qu'il est parfaitement possible de passer à de nouveaux domaines pour un traitement de textes « réfléchis ».

2.3.4 Tableau récapitulatif

Le Tableau 9 résume les principales informations concernant les différentes approches détaillées précédemment et classées par ordre d'importance. L'approche du portage linguistique est la plus importante car il s'agit d'un cas de portage de l'anglais vers le français.

Approches	Données (type de SL)	Interface	Domaine	Langue initiale	Approche TALN	Portage linguistique
Adaptation à une autre langue : portage linguistique¹³ (E.Riloff)	Articles sur des accidents d'avion	Contrôlée	Accident d'avion	Anglais	Apprentissage	Français
Adaptation à un autre type de texte : Amilcare (F.Ciravegna)	Réfléchies	Contrôlée	Web sémantique	Anglais	Apprentissage	Non
Génération multilingue : MKBEEM (FT R&D &al)	Spontanées	Web	Tourisme	Finnois, français, espagnol et anglais	Syntaxe basée sur WEBTRAN	Non

Tableau 9 : tableau récapitulatif des principales approches étudiées

¹³ “interlingual projection”, terme moins adéquat

Chapitre 3 Démarche proposée

Introduction

L'état de l'art présenté dans le chapitre précédent montre que le type d'application que l'on trouve et qui traitent les énoncés spontanés en langue naturelle est celui du domaine de l'extraction d'information. Dans ce chapitre, nous proposons une démarche pour effectuer la multilinguïstation de telles applications.

Supposons qu'on veuille créer une société de « portage linguistique ». Elle proposera à ses clients potentiels un service de portage d'une application App disponible en langue L1 vers une ou plusieurs autres langues L2, L3...

Dans la première partie de ce chapitre, nous discutons des solutions possibles pour trouver un corpus et un dictionnaire cibles, besoins nécessaires et communs pour toute opération de portage linguistique d'application d'extraction d'information vers une nouvelle langue Li. Nous proposons ensuite deux méthodes de portage : (1) adaptation de l'extracteur de contenu, et (2) portage direct par TA.

3.1 Démarche

Les principales étapes d'un portage linguistique sont le choix de l'application à porter et l'étude de sa situation traductionnelle, l'étude du sous-langage traité, la détermination du ou des chemins de portage, la préparation des données nécessaires, l'implémentation du portage, et finalement l'évaluation du portage.

3.1.1 Choix de l'application

Nous avons indiqué au début de ce mémoire que la multilinguïstation d'une application de e-commerce traitant des énoncés spontanés en langue naturelle dépend principalement des deux points suivants relatifs à la situation traductionnelle :

- le niveau d'accès aux ressources des applications, pour lequel quatre cas se présentent : accès complet au code source, accès à la représentation interne uniquement, accès au dictionnaire uniquement, et aucun accès.
- le niveau de compétences langagières et linguistiques des intervenants pour la multilinguïstation des applications, qui peut être défini par rapport à la langue source ou par rapport aux compétences linguistiques de l'équipe qui veut faire la localisation (spécialiste ou non de TALN et de tout ou partie des langues cibles).

Nous choisirons une application de préférence déployée pour laquelle les deux points présentés ci-dessus sont assurés, ce qui permettra alors de tester les différentes méthodes de portage que nous proposons.

Notre choix de l'application à porter dépendra aussi du type des énoncés traités, puisque nous nous intéressons au traitement des énoncés spontanés. Ces énoncés appartiennent généralement à un sous-langage bien déterminé, une notion très importante que nous détaillerons dans le Chapitre 4.

3.1.2 Méthodes de portage proposées

Le choix de l'application, l'étude de la situation traductionnelle, l'étude du sous-langage, le choix du couple de langue et notre savoir faire dans le domaine sont les facteurs permettant de déterminer la ou les méthodes de portage possibles pour chaque application.

Trois méthodes sont a priori possibles : portage interne, portage externe et portage par traduction.

3.1.2.1 Portage interne

La première méthode que nous proposons, et que nous détaillerons dans la section 3.2.1, consiste à adapter à la nouvelle langue L2, l'extracteur de contenu de l'application prévu pour la langue originale L1. C'est ce que certaines études vues dans le Chapitre 2 appellent "internlingual projection". La Figure 29 résume cette méthode de « portage interne ».

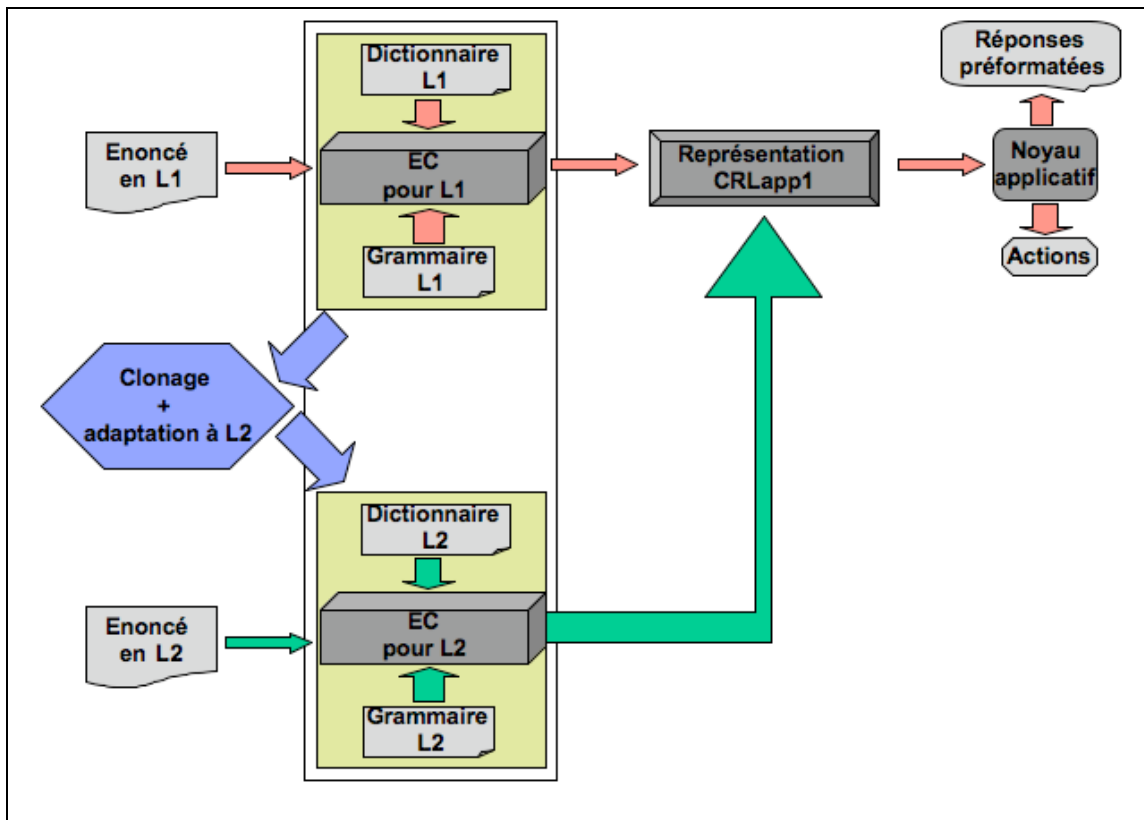


Figure 29 : méthode de portage interne

Par exemple, pour porter le système CATS décrit dans le 2.2.3 de l'arabe vers le français, nous avons bien accès à l'EC-CATS (Extracteur de Contenu de CATS) existant pour l'arabe, nous pouvons donc appliquer cette méthode en localisant et adaptant au français cet extracteur. Nous verrons l'expérimentation de cette méthode dans le Chapitre 8.

3.1.2.2 Portage externe

Nous proposons une deuxième méthode que nous détaillerons dans la section 3.2.2, et qui consiste à adapter un extracteur de contenu existant, traitant la même langue cible L2, mais pour un autre domaine et/ou une autre tâche, et à traduire le résultat obtenu vers la représentation de contenu de l'extracteur de contenu original. La Figure 30 résume cette méthode de « portage externe ».

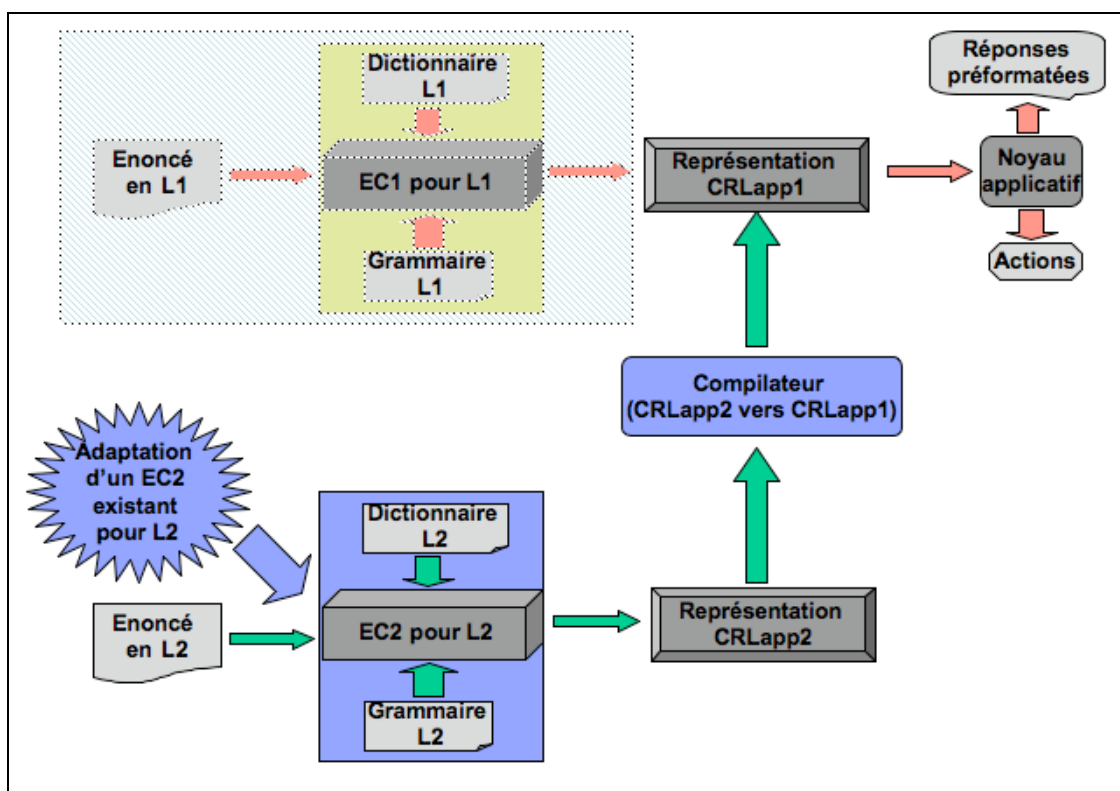


Figure 30 : méthode de portage externe

Par exemple, dans le cas du même système CATS, on peut effectuer une adaptation et une greffe d'un EC externe (EC-ext) d'un autre système traitant la langue cible, visée. Concrètement, nous avons adapté à CATS un EC du français existant, celui du projet Nespole!¹⁴ « *NEgotiating through SPOken Language in E-commerce* » écrit par (Blanchon 2004). Afin d'aboutir au même format que le format CRL-CATS généré par la version originale de CATS, nous avons construit un traducteur de format qui permet de traduire le format IF-CATS (Interchange Format) généré par cet EC vers le format CRL-CATS. Nous expérimenterons cette méthode dans le Chapitre 9.

Imaginons qu'on souhaite adapter au français le système IMRS présenté dans le paragraphe 2.2.4. La tâche est différente du portage de CATS. En effet,

- nous ne pouvons pas adapter l'extracteur de contenu de l'application, car nous n'y avons pas accès, et, si nous l'avions, nous devrions faire faire le travail par quelqu'un connaissant le japonais et le français,
- nous n'avons pas de corpus parallèle français-japonais adapté, et ne pouvons pas en créer.

Il nous resterait donc comme seule possibilité de créer un extracteur de contenu produisant un vecteur requête à partir d'un énoncé en français.

Pour créer cet extracteur, deux solutions peuvent être envisagées :

- Portage externe : adapter un extracteur de contenu existant pour le français à ce nouveau domaine, par exemple le précédent, comme pour CATS, et traduire la représentation obtenue (IFmusique) vers un vecteur de dimension 10. Nous présenterons ce travail dans la section 9.3.

¹⁴ <http://nespole.itc.it> : site Web du projet Nespole !, visité en 2006.

- Création : s'inspirer de l'approche décrite dans les articles de Kumamoto, et la mettre en œuvre sur le français (et éventuellement d'autres langues que nous connaissons, comme l'arabe et l'anglais).

3.1.2.3 Portage par traduction

La troisième méthode possible et que nous détaillerons dans la section 3.3, consiste à traduire les énoncés de la nouvelle langue L2 vers la langue originale de l'application L1 et à utiliser l'extracteur de contenu de l'application originale. La Figure 31 décrit l'architecture de cette méthode de « portage par TA ».

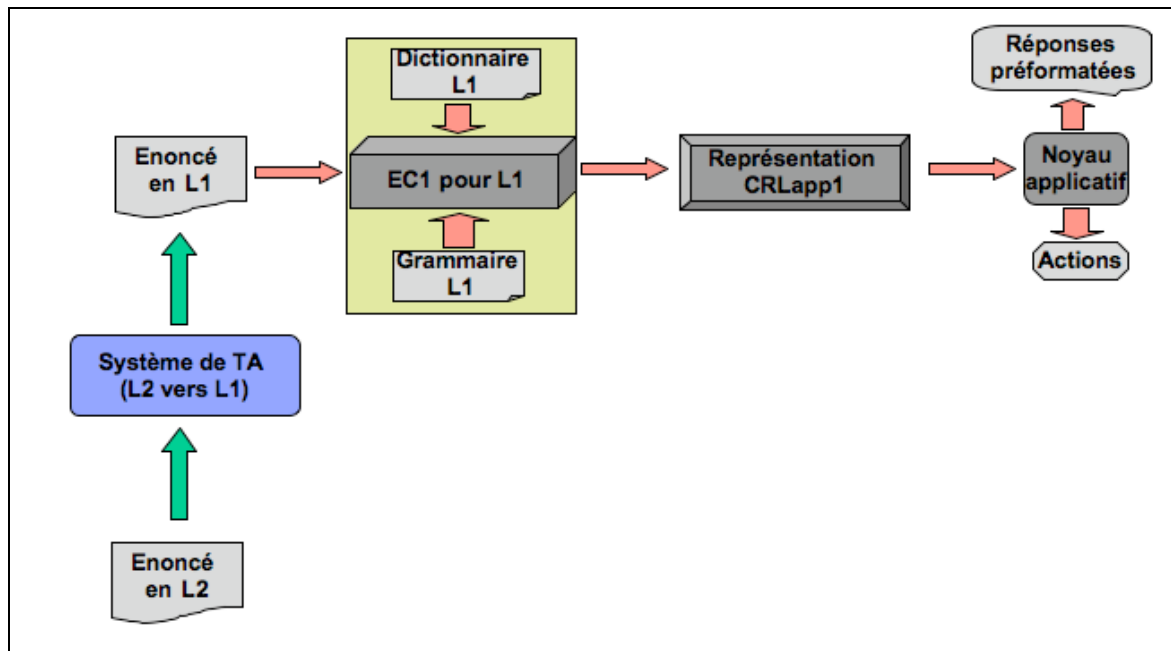


Figure 31 : méthode de portage par TA

Pour le cas du même système CATS, comme le décrit la Figure 31, on propose alors de traduire des énoncés de la langue L2 (L2 = français : langue source pour la traduction et cible pour le portage de CATS) vers la langue L1 (L1 = arabe : langue cible pour la traduction et source pour le portage de CATS) et d'utiliser l'extracteur de contenu original de l'application CATS. Nous présentons l'expérimentation de cette méthode dans le Chapitre 12.

Nous avons une difficulté commune pour toutes ces méthodes de portage, celle de trouver les données de développement (corpus et dictionnaire). Nous présentons dans ce qui suit une solution possible pour préparer ces données.

3.1.3 Ressources nécessaires

Nous présentons ici les ressources nécessaires pour chacune des méthodes de portage présentées dans la section précédente.

Les deux méthodes de portage interne et externe nécessitent comme ressources un corpus et un dictionnaire pour la nouvelle langue L2, qui soient fonctionnellement équivalents aux ressources de départ (corpus et dictionnaire pour L1).

La troisième méthode nécessite essentiellement un corpus parallèle bilingue (L2,L1) contenant la même qualité de données, et éventuellement un dictionnaire (L2,L1) pour améliorer le résultat de la traduction obtenue. A priori, le corpus parallèle nécessaire pour la TA « experte » est petit, mais celui nécessaire pour la TA « statistique » est très grand. Nous verrons cependant que, dans le cas d'un sous-langage restreint, on peut aussi se contenter d'un petit corpus parallèle.

Il est clair qu'il y a un besoin commun (corpus et dictionnaire) à préparer pour toutes les méthodes de portage envisageables. Nous commençons par définir les propriétés du corpus nécessaire.

3.1.3.1 Caractéristiques et définition d'un corpus

Selon (Nelson 1982) et (Nelson 1992), la définition d'un corpus est la suivante :

« A collection of texts assumed to be representative of a given language, dialect, or other subset of a language to be used for linguistic analysis ».

Cette définition peut être traduite de la façon suivante :

« Une collection de textes supposés être représentatifs d'une langue donnée, d'un dialecte, ou de tout autre sous-ensemble d'une langue pour être utilisée dans l'analyse linguistique ».

Un corpus est considéré comme étant un ensemble de documents (textes, images, vidéos, etc.), regroupés dans une optique précise. On peut utiliser des corpus dans plusieurs domaines : études littéraires, linguistiques, scientifiques, etc.

En littérature, un corpus est un regroupement de textes ayant une visée commune. En science, les corpus sont des outils indispensables et précieux en traitement automatique du langage naturel. Ils permettent en effet d'extraire un ensemble d'informations utiles pour des traitements statistiques.

D'un point de vue informatif, ils permettent de construire des ensembles de fréquences de n-grammes.

D'un point de vue méthodologique, ils apportent une objectivité nécessaire à la validation scientifique en traitement automatique du langage naturel. L'information n'est plus empirique, elle est vérifiée par le corpus. Il est donc possible de s'appuyer sur des corpus pour formuler et vérifier des hypothèses scientifiques.

Plusieurs caractéristiques entrent en jeu pour la création d'un corpus bien formé telles que :

- **la taille** : le corpus doit évidemment atteindre une taille critique pour permettre des traitements statistiques fiables. Il est impossible d'extraire des informations fiables à partir d'un corpus trop petit.
- **la langue du corpus** : un corpus bien formé doit nécessairement couvrir un seul langage, et une seule déclinaison de cette langue. Il existe par exemple de subtiles différences entre le français de France et le français parlé en Belgique. Il ne sera donc pas possible de tirer des conclusions fiables à partir d'un corpus franco-belge sur le français de France, ni sur le français de Belgique.
- **l'évolution des textes avec le temps** : le temps joue un rôle important dans l'évolution d'une langue : le français parlé aujourd'hui ne ressemble pas au français parlé il y a 200 ans ni, de façon plus subtile, au français parlé il y a 10 ans, à cause notamment des néologismes. C'est un phénomène à prendre en compte pour toutes les langues vivantes. Un corpus ne doit donc pas contenir des textes rédigés à des intervalles de temps trop larges.
- **le registre** : il ne faut pas mélanger des registres différents, un corpus construit à partir de textes scientifiques ne peut pas être utilisé pour extraire des informations sur les textes vulgarisés, et un corpus mélangeant des textes scientifiques et vulgarisés ne permettra de tirer aucune conclusion sur ces deux registres.

3.1.3.2 Caractéristiques et définition d'un corpus réel, naturel

Nous n'avons pas trouvé de définition claire d'un corpus réel. Nous proposons alors une définition basée sur celle de (Nelson 1982) :

« *Un corpus naturel (réel) est une collection de textes supposés être représentatifs d'une langue donnée, d'un dialecte, ou de tout autre sous-ensemble d'une langue construit d'une façon spontanée et naturelle pour être utilisé dans différents objectifs (commercial, linguistique, statistique...)* ».

Les caractéristiques citées précédemment restent applicables pour un corpus naturel, à savoir la taille, la langue, l'évolution des textes avec le temps et le registre.

3.1.3.3 Différences et particularités des données réelles et spontanées

Nous voudrions éclaircir ici la différence entre un corpus normal tel qu'il est défini par (Nelson 1982) et un corpus spontané ou réel. En effet, la différence est surtout au niveau de la façon de construire le corpus : un corpus normal peut être fabriqué automatiquement à partir de textes, de journaux ou autres supports ; par contre, un corpus spontané doit être fabriqué d'une façon naturelle et peut contenir des erreurs, des fautes, des abréviations.

Contrairement à un corpus normal, les données d'un corpus réel ne doivent pas être éditées, mais les outils de traitement doivent être adaptés à ce type de données non propres.

Il faut aussi préciser que les corpus réels ne sont pas des corpus spécialisés. En effet, un corpus spécialisé est un type de corpus normal. Pearson décrit plusieurs types de corpus selon les applications possibles (Pearson 1998). Par exemple, dans les travaux de l'équipe ÉCLECTIK (équipe de recherche en combinatoire lexicale, terminologie et informatique à Montréal), on aura affaire surtout à des corpus dits « spécialisés » ou « de spécialité », à savoir, des corpus contenant des textes traitant d'un sujet lié à un domaine de la connaissance comme la médecine, le droit ou l'informatique. Pearson, cependant, ne fait pas une distinction de base entre les corpus ouverts (auxquels on ajoute constamment de nouveaux textes) et fermés (qui restent « stables »). Les corpus ouverts, bien que plus à jour, impliquent un entretien constant et méticuleux.

3.1.4 Techniques d'acquisition et de construction des ressources

Nous décrivons dans ce qui suit quelques techniques possibles d'acquisition et de construction d'un corpus et d'un dictionnaire fonctionnellement équivalents dans la nouvelle langue L2 ainsi que la construction d'un corpus bilingue parallèle (L2, L1).

3.1.4.1 Corpus fonctionnellement équivalent dans L2

Dans le cas d'un portage linguistique, trouver un corpus fonctionnellement équivalent dans L2 est nécessaire pour faire d'abord le portage lui-même et pour l'évaluer ensuite.

Pour répondre à un tel besoin, deux options se présentent : traduction d'un corpus existant ou collecte d'un nouveau corpus.

Pour la première option, il s'agit de traduire et réviser un corpus de la langue L1 vers la langue L2, puis l'adapter au sous-langage supposé correspondre dans L2. Cela suppose bien sûr :

- qu'il y ait un corpus pour la langue L1 ;
- qu'on ait accès à ce corpus.

La traduction du corpus de L1 vers L2 est nécessaire pour effectuer le portage linguistique, car elle permet de créer une version initiale du corpus de développement dans L2, mais les systèmes de traduction automatique ne sont pas disponibles pour tous les couples de langues. Ils ne sont utiles que pour des textes « propres » et ne marchent pas pour les énoncés spontanés qui peuvent contenir des erreurs, des fautes, des abréviations.

On peut alors passer par une traduction humaine. Un nouveau problème se pose, c'est de savoir quel type de personne peut traduire le corpus source. Suffit-il d'une personne qui parle la langue L2, ou faut-il une personne dont la langue L2 est la langue maternelle ? Nous ne pouvons pas donner de

réponse immédiate à cette question, mais nous devons expérimenter et tester. Pour cela, nous devons émettre une hypothèse qui puisse confirmer ou infirmer notre supposition. Nous détaillerons ce travail dans la section 7.2.3.

Pour la deuxième option (collecte d'un nouveau corpus), il s'agit de collecter un corpus dans L2 dans le domaine visé, puis de le « localiser » (dans la même langue) au domaine, à la tâche, au lieu et aux contraintes visées. Par exemple, on pourrait collecter un corpus de petites annonces Web sur l'immobilier à Grenoble, et l'adapter (en français) à l'immobilier à Amman ainsi qu'au style et à la contrainte de longueur des SMS. Cette méthode est la seule à pouvoir être mise en œuvre dans le cas où on n'a pas accès au corpus de l'application à porter, ou bien dans le cas où on ne connaît pas la langue du corpus à porter. L'inconvénient de cette approche est qu'on ne traite pas la même qualité de données du point de vue de la spontanéité et du point de vue du contenu. Par exemple, un corpus collecté sur le Web pour CATS peut contenir des critères de choix de véhicules que CATS ne gère pas comme le kilométrage.

Après avoir trouvé ou construit un corpus cible, le second problème est de déterminer la taille du corpus dont on a besoin pour effectuer le portage vers une langue L2.

Nous pouvons commencer par un corpus cible de petite taille, et augmenter peu à peu sa taille en fonction de l'augmentation de la taille du vocabulaire. Une fois que la taille du vocabulaire se stabilise¹⁵, on peut dire que la taille du corpus cible correspondant est la taille minimale nécessaire pour couvrir le sous-langage considéré.

3.1.4.2 Dictionnaire fonctionnellement équivalent dans L2

Le second problème qui se pose consiste à trouver un dictionnaire pour le sous-langage correspondant de L2.

Faut-il une traduction (automatique ou humaine), ou une adaptation du dictionnaire de L1 ? Dans les deux cas, cela suppose :

- qu'il y ait un dictionnaire pour la langue L1 ;
- qu'on ait accès à ce dictionnaire.

Dans le premier cas (traduction), nous devons répondre aux mêmes questions que dans le cas d'une traduction de corpus source.

Dans le deuxième cas (adaptation), nous pensons qu'un travail humain considérable est indispensable. Une personne parlant la langue L2 peut faire l'adaptation en question, à condition qu'elle soit aidée par le type et la variante lexicale et orthographique des données du corpus cible de L2 (supposé créé avant le dictionnaire cible).

De même que pour les corpus, une deuxième approche peut être de collecter un dictionnaire du domaine en L2, et de l'adapter au contexte de l'application (par exemple, en introduisant des erreurs, et des abréviations).

¹⁵ Dans le cas d'un sous-langage très restreint, le vocabulaire peut se stabiliser très rapidement.

3.1.4.3 Corpus parallèle (L2,L1)

L'obtention manuelle d'un corpus parallèle (L2,L1) semble être très difficile et coûteuse. Une façon possible pour obtenir un tel corpus est de « déconvertir¹⁶ » vers L2 les représentations internes correspondant au corpus disponible L1. Cela suppose d'avoir :

- un corpus en L1,
- la représentation interne (CRL) correspondante de ce corpus L1,
- un environnement (contenant par exemple un LSPL adéquat) de déconversion,
- des compétences en déconversion.

Grâce à la proximité des sous-langages, le travail de déconversion peut être effectué très rapidement si on maîtrise l'environnement de déconversion et la langue cible en question.

Il reste à effectuer une post-édition manuelle visant à obtenir un corpus « naturel », « pseudo-spontané ». Nous présentons les détails d'expérimentation de ce travail dans le Chapitre 11.

3.1.5 Scénarios d'implémentation

Cette étape dépend de la méthode de portage choisie : portage interne ou externe de l'extracteur de contenu, ou portage par TA.

3.1.5.1 Implémentation d'un portage interne

Le scénario d'implémentation d'un portage interne est le suivant :

1. Identifier la partie ou les composants de l'application qui dépendent de la langue. Généralement, il s'agit d'un seul composant qui est l'extracteur de contenu.
2. Cloner la partie identifiée.
3. Adapter le code (les règles et le dictionnaire dans le cas d'un EC utilisant un LSPL) pour extraire le contenu des énoncés de la nouvelle langue L2.

3.1.5.2 Implémentation d'un portage externe

Le scénario d'implémentation d'un portage externe est le suivant :

1. Identifier la partie ou les composants de l'application originale qui dépendent de la langue.
2. Cloner la partie non identifiée (la partie indépendante de la langue).
3. Greffer un extracteur de contenu existant pour L2 à la partie non identifiée.
4. Adapter le code du nouvel extracteur de contenu de L2 à la tâche de l'application originale.
5. Traduire le résultat obtenu dans la représentation interne générée par l'extracteur de contenu adapté vers la représentation interne de l'application originale.

3.1.5.3 Implémentation d'un portage par TA

Le scénario d'implémentation d'un portage par TA est le suivant :

1. Construire un système de TA (L2 vers L1).
2. Traduire les énoncés de L2 vers L1.
3. Utiliser l'extracteur de contenu de l'application originale.

¹⁶ déconvertir : c'est le terme employé quand il s'agit d'une traduction à partir d'une représentation interne vers une langue, pour l'inverse, on dit « convertir ». Il ne s'agit pas d'analyse ou de génération, car deux « espaces lexicaux » sont en jeu, et pas un seul.

3.1.6 Évaluation du portage

Nous proposons de distinguer deux types d'évaluation : évaluation de la performance de l'application portée et évaluation du coût du portage.

3.1.6.1 Évaluation de l'application portée

Plusieurs méthodes sont envisageables pour évaluer le portage d'une application d'extraction d'information traitant des énoncés spontanés en langue naturelle. Dans le Chapitre 6, nous détaillons trois niveaux d'évaluation, à savoir :

- le langage d'entrée en mesurant sa couverture, son efficacité et son coût,
- la représentation de contenu en mesurant l'acceptabilité des résultats d'extraction de contenu,
- la performance liée à la tâche en mesurant la réalité et l'efficacité du portage sur le terrain.

Nous définirons un seuil objectif à partir duquel nous pourrions dire que le portage est effectué et donne une version initiale de l'application en question pour la nouvelle langue L2. Bien évidemment, les résultats d'extraction peuvent s'améliorer une fois que l'application a commencé à être déployée sur le terrain. À ce moment là, on pourra mesurer le degré de satisfaction des utilisateurs et définir un seuil subjectif.

3.1.6.2 Évaluation de coût (humain)

Nous proposons une mesure de coût du portage vers une nouvelle langue L2, qui dépend :

- de la méthode de portage utilisée,
- du niveau d'accès aux ressources de l'application à porter,
- de la disponibilité des ressources pour le sous-langage considéré dans la nouvelle langue visée,
- des compétences langagières et linguistiques utilisées.

Nous mesurons le coût en terme de temps de l'opération de portage (en hommes/mois) pour différentes compétences (hommes/mois de linguiste, hommes/mois d'informaticien, hommes/mois de linguiste informaticien). Les résultats de cette évaluation sont comparables à condition d'avoir les mêmes situations de portage (ressources, niveaux d'accès, ressources et compétences...).

3.2 Portage par adaptation d'extracteur de contenu

Plusieurs possibilités d'adaptation de l'extracteur de contenu peuvent être envisagées telles que : (1) l'adaptation de l'extracteur de l'application à L2, (2) l'adaptation d'un extracteur existant pour la langue L2 pour la nouvelle tâche, (3) l'adaptation d'un extracteur de contenu universel. Cette dernière option fait l'objet de recherches (voir section 3.2.3) mais aucun extracteur de contenu universel n'était disponible quand nous avons commencé ce travail.

3.2.1 Adaptation de l'extracteur de contenu de L1 à L2

Des études précédentes (Kittredge and Lehrberger 1982a) montrent que deux sous-langages équivalents dans deux langues différentes sont proches l'un de l'autre, même si leurs deux langues mères sont éloignées. On devrait donc pouvoir, à faible coût (hors construction du dictionnaire) adapter à L2 l'extracteur de contenu prévu initialement pour la langue L1. Concrètement, il s'agit :

- d'adapter le dictionnaire source à L2 ;
- d'adapter les règles et/ou les programmes (grammaire) sources à L2.

Pour cela, il faut que l'on ait :

1. un accès complet au code source de l'extracteur de contenu,

2. des compétences dans le langage de programmation en question (celui dans lequel est programmé l'extracteur de contenu),
3. des compétences dans la langue source et dans la langue cible.

Nous présentons une mise en œuvre de cette méthode dans le Chapitre 8.

3.2.2 Adaptation d'un extracteur de contenu de L2

Une autre idée intéressante consiste à exploiter un extracteur de contenu EC2 disponible pour L2. Deux cas de figure se présentent : (1) EC2 est prévu pour un domaine différent, (2) EC2 est prévu pour le même domaine.

3.2.2.1 EC2 traite un domaine différent

Dans le cas où EC2 est prévu pour un domaine différent, trois sous-tâches sont indispensables :

1. Travail dictionnaire : adaptation et enrichissement du dictionnaire utilisé dans EC2 au domaine de l'application.
2. Travail grammatical/procédural : adaptation des procédures, des fonctions, et des grammaires programmées dans EC2 en tenant compte des modifications faites dans la partie dictionnaire.
3. Traduction vers la représentation interne : compilation (traduction) du résultat de EC2 vers la représentation interne de l'application originale.

Nous présentons une mise en œuvre de cette méthode dans le Chapitre 9.

3.2.2.2 EC2 traite le même domaine

Dans le cas où l'on dispose d'un extracteur de contenu pour L2 pour le même domaine, le travail est moins lourd et moins coûteux, car il suffit de passer directement à la troisième sous-tâche qui consiste à traduire de la nouvelle représentation de contenu vers la représentation de contenu originale.

3.2.3 Extracteur de contenu « universel »

On peut aussi envisager de trouver ou construire un extracteur de contenu universel, qui serait paramétrable par :

- un sous-langage (lexique, grammaire),
- un domaine (dictionnaire),
- une CRL (représentation interne de contenu).

Nous n'avons pas trouvé d'extracteur de contenu universel paramétrable. Par contre, réutiliser des outils d'extraction de contenu et essayer de généraliser la tâche d'une façon générique peut être réalisable. Au Canada, dans le laboratoire RALI «Recherche Appliquée en Linguistique Informatique», plusieurs pistes ont été explorées, qui ont abouti à différents prototypes comme :

- Un outil de question-réponse : Quantum (Plamondon 2002), est un système de question-réponse développé dans le cadre du projet Merkure. Le but est d'identifier, dans une base documentaire, une réponse factuelle à une question courte et explicite (par ex. *Quel est le nom de jeune fille d'Hillary Clinton ?*). Le système effectue une analyse de la question pour en déterminer le focus et une recherche d'information à partir des termes de la question pour identifier les passages de la base documentaire les plus susceptibles de contenir la réponse. Ces passages sont ensuite analysés pour extraire des entités sémantiques particulières (noms de personnes, dates,...). En fonction du type de réponse recherchée (date, lieu,...), l'entité la plus probable est extraite et est retournée comme réponse à l'utilisateur.
- Un outil de traitement de courriels : (Bélanger 2003) a proposé d'adapter les outils et les techniques utilisées dans les systèmes de question-réponse comme Quantum au traitement de

courriels. En effet, l'analyse de courriels du corpus BCE (Bell Canada Entreprise) montre qu'une bonne partie d'entre eux sont des questions posées à l'entreprise.

- Un outil d'extraction d'information : (Boufaden 2004) a travaillé sur l'extraction d'information à partir de transcriptions manuelles de conversations téléphoniques recueillies dans un contexte de recherche-sauvetage.

Tous ces travaux peuvent être des candidats comme point de départ vers l'objectif de construire un extracteur de contenu universel. Cela semble être une tâche assez délicate à réaliser et demander des efforts considérables.

3.3 Portage par TA

La troisième méthode de portage consiste à traduire les énoncés de la nouvelle langue L2 vers L1. Elle nécessite la réalisation d'un système de Traduction Automatique (Boitet 2003) pour les énoncés spontanés. Cela n'est pas impossible, mais plusieurs questions se posent : quel type de système choisir ? et de quelles ressources doit-on disposer ?

Pour le type de système, deux pistes de l'approche « empirique » s'offrent à nous, la TA statistique (voir Chapitre 12) et la TA analogique. Les systèmes « empiriques », construits à partir de corpus parallèles plus ou moins annotés, permettent a priori à une équipe réduite d'effectuer un portage vers des langues qu'elle ne connaît pas.

3.3.1 TA statistique

Pour le cas des systèmes statistiques, deux problèmes principaux se posent :

- l'estimation de la taille du corpus bilingue nécessaire en fonction de la « proximité » estimée entre L2 et L1,
- la construction d'un tel corpus.

Les deux problèmes précédents semblent être délicats mais nous devons expérimenter pour pouvoir tirer des conclusions.

3.3.1.1 Taille du corpus

La détermination, à l'avance, de la taille du corpus nécessaire pour pouvoir traduire, avec des systèmes statistiques, des énoncés spontanés est impossible, mais plusieurs éléments que nous détaillons dans la section 10.2 nous motivent à expérimenter cette piste, comme la proximité des sous-langages. Nous pouvons alors commencer l'expérience avec une taille minimale de corpus.

Nous citons le texte suivant tiré d'un rapport à Christian Boitet sur la TA (Boitet 2007) qui montre que les systèmes de TA peuvent marcher dans le cas d'un sous-langage.

La TA statistique "de base" (modèles de Brown à IBM vers 1980) est directe, de chaîne à chaîne, sans même d'analyse ou de génération morphologique.

Par contre, Language Weaver (1999—) utilise des modèles intermédiaires, et apprend toutes les correspondances de façon non supervisée, à partir d'un gigantesque corpus aligné. Pour obtenir des traductions assez fidèles et fluides, K. Knight a dit à CICLING-05 que, pour la TA de textes généraux, variés et difficiles, comme des débats parlementaires ou des journaux, il faut un corpus de 50 M mots (200.000 pages), et plus récemment Ph. Koehn a parlé de 200 M mots (800.000 pages). Ces corpus leur sont fournis gratuitement par leurs clients, et ils se limitent à les aligner puis à construire un système. Mais peu de clients ont des corpus aussi gros, pour une raison simple : leur construction a demandé autant d'heures de travail que de pages, ce qui est considérable.

La société dit pouvoir construire un système de TA "virtuellement d'un jour à l'autre"... mais n'en a construit apparemment que 3 ou 4 depuis sa création. En effet, si le corpus n'est pas assez grand

(comme dans le cas de leur prototype anglais-hindi), les résultats sont du charabia incompréhensible, et presque inutilisable en tant que prétraduction pour un traducteur. Il s'agit donc, curieusement, d'une approche qui ne semble marcher que dans des "niches" assez rares, un peu comme l'approche par règles et optimisation à un sous-langage des systèmes TAUM-météo pour les bulletins météorologiques et ALTFLASH de NTT pour les "brèves" (market flash reports) du Nikkei.

En 2005, Google a mis sur son site deux systèmes de TA statistique en version bêta (chinois-anglais et arabe-anglais). Tous les autres couples de langues sont fournis par Systran, qui a eu un contrat de Google plus gros en 2006 qu'en 2005. Pour des textes journalistiques, les résultats sont assez satisfaisants, car les corpus d'entraînement sont de ce type. Par contre, ça ne donne pas grand chose de compréhensible sur des textes techniques.

3.3.1.2 Expérimenter et mesurer

Afin de trouver quelle est la taille minimale nécessaire pour construire un traducteur statistique du sous-langage L2, nous pouvons commencer l'expérimentation avec une taille minimale et mesurer les résultats trouvés, augmenter la taille du corpus au fur et mesure et voir l'évolution des résultats.

Notre objectif ici n'est pas d'obtenir une excellente qualité de traduction destinée à un humain, mais d'utiliser la traduction statistique pour l'extraction de contenu. Autrement dit, le résultat de traduction sera destinée à une machine (extracteur de contenu).

De la même façon, nous utilisons les mêmes mesures d'évaluation du portage après extraction de contenu à partir des résultats produits en L1 (après traduction de L2 vers L1).

3.3.2 TA par analogie

L'avantage de la réalisation d'un système de TA par analogie (Lepage 2006) (Denoual 2006) est le même que celui de la TA statistique : elle ne demande pas de compétences linguistiques. De plus, cette méthode ne requiert aucun prétraitement, alors qu'il faut en général un segmenteur et souvent un analyseur morphologique pour la TA statistique. Mais il n'existe pas encore de boîte à outils comparable à ce qu'on trouve pour la TA statistique.

3.3.2.1 Taille du corpus

Les expériences effectuées par (Lepage 2005) et (Denoual 2006) montrent que la taille du corpus doit être importante. Dans ces travaux, le corpus BTEC (Basic Travel Expression Corpus) constitué de 163.000 phrases de tourisme alignées en japonais-anglais, d'une longueur moyenne de 6,5 mots (~ 1 M mots) en anglais a été utilisé. Dans une évaluation basée sur la méthode BLEU (Blanchon 2004) et utilisant un ensemble de 510 phrases, ce système (Denoual 2006) a obtenu le troisième rang parmi six systèmes candidats.

(Lepage 2005) a évalué ce que cela donne avec un corpus plus petit, comme celui distribué pour le développement lors de la campagne IWSLT-05. Le résultat est que 40000 phrases (3 M mots dans le BTEC) semblent insuffisantes, car la « densité d'analogies » est trop faible.

Nous n'avons malheureusement pas eu accès à une boîte à outil nous permettant de mettre en œuvre une expérience en TA analogique.

Conclusion de la partie A

Nous avons montré dans le premier chapitre que la multilinguisation des services de e-commerce traitant des énoncés spontanés en langue naturelle est un problème important mais difficile. Très peu de services le font et cela dépend la situation traductionnelle. Nous avons cherché à trouver les raisons possibles qui font que très peu de services traitent de tels types d'énoncés.

Nous avons présenté, dans le deuxième chapitre, une recherche d'applications traitant des énoncés spontanés en langue naturelle. Il s'agit d'une recherche d'applications de gestion de contenu qui sont généralement orientées et conçues pour traiter un sous-langage lié à un domaine particulier.

Nous avons trouvé quelques applications qui correspondent à notre domaine d'étude. Ces applications sont orientées plutôt vers le domaine de l'extraction d'information. Nous avons analysé ces applications en détail. Nous sommes intéressés par une première expérimentation sur le système CATS, une vraie application déployée qui traite des SMS en arabe, et une deuxième sur le système IMRS.

Nous avons présenté, dans le troisième chapitre, une analyse des approches possibles pour porter de telles applications à de nouvelles langues, dans laquelle nous essayons de répondre aux différentes difficultés que nous pouvons rencontrer lors de la réalisation d'un portage linguistique de telles applications.

Partie B

Sous-langage, représentation de contenu et méthodes d'évaluation d'un portage linguistique

Dans le premier chapitre, nous étudions la notion de sous-langage, puisque la plupart des textes traités par le genre d'applications auxquelles on s'intéresse sont des énoncés appartenant un à sous-langage particulier, une notion sur laquelle plusieurs chercheurs ont travaillé et ont divergé sur sa définition et son identification. Un sous-langage est-il vraiment un sous-ensemble de la langue standard ? Nous présentons une description des précédents travaux liés à cette notion, et définissons une taxonomie pour identifier et définir un sous-langage.

Dans les applications qui nous intéressent, le passage par une représentation de contenu est presque obligatoire pour pouvoir normaliser différents énoncés qui portent le même sens. Nous parlerons dans le deuxième chapitre de cette partie des différents types de représentations de contenu et de la possibilité de comparer deux instances d'une même représentation.

Finalement, dans le troisième chapitre de cette partie, nous présenterons les différentes possibilités d'évaluation d'un portage linguistique, à savoir une évaluation du sous-langage traité, une évaluation de la représentation de contenu, et une évaluation de performance liée à la tâche.

Chapitre 4 Caractérisation des sous-langages

Introduction

La plupart des applications décrites dans le deuxième chapitre traitent des textes de sous-langages. Mais comment peut-on savoir si une application traite ou non un sous-langage, et si oui quelles sont ses caractéristiques ? On souhaite proposer une méthode ou une taxonomie qui nous permette dans des cas pratiques de juger si on a affaire à un sous-langage.

Dans la première section, nous présentons plusieurs définitions qui ont été proposées pour la notion de sous-langage. Dans la deuxième section, nous dressons un état de l'art des méthodes d'indentification d'un sous-langage. Finalement, nous proposons une taxonomie applicable dans des cas pratiques pour étudier et préciser les caractéristiques d'un sous-langage.

4.1 Définition d'un sous-langage naturel

Dans un cours sur la traduction automatique, Christian Boitet présente ainsi les origines de l'intérêt pour les sous-langages en TALN : « *les chercheurs du groupe TAUM à l'UdM (Université de Montréal) furent les premiers à se rendre compte de la relative facilité de construction de certains systèmes de TALN, et de leur grande qualité. Après avoir connu une telle « bonne surprise » avec la traduction de bulletins météo (système TAUM-météo, 1976)¹⁷, le groupe TAUM a cherché (en vain d'ailleurs¹⁸) d'autres sous-langages aussi « faciles » pour la méthode employée (programmation « experte » reposant sur une étude précise du sous-langage en question et sur la mise en œuvre d'heuristiques adaptées). Cela conduisit les linguistes du groupe TAUM (surtout R. Kittredge et J. Lehrberger) à approfondir cette notion, introduite par Z. Harris en 1968, pour la rendre « opérationnelle »* ».

Kittredge et d'autres ont montré l'importance de la notion de sous-langage dans le traitement des textes d'un langage naturel modifié ou simplifié par l'utilisation de restrictions lexicales, syntaxiques ou sémantiques spécifiques (Kittredge and Lehrberger 1982a), (Grishman and Kittredge 1986), (Slocum 1986), (Biber 1993), (Sekine 1994).

Dans son article intitulé « A New direction for Sublanguage NLP », (Sekine 1994) montre que la restriction (explicite ou implicite) à des sous-langages « assez restreints » conduit au succès : on arrive à construire des systèmes très performants avec un investissement très raisonnable en temps humain de spécialistes et en ressources de calcul (temps, place). Il cite, par exemple, le cas du système TAUM-METEO (Isabelle 1984), (Chandioux 1988), un système de traduction automatique qui marche

¹⁷ Ce système, construit par le groupe TAUM de l'UDM en 1975-76, fut mis en service opérationnel à l'environnement Canada le 24 mai 1977 par la société J. Chandioux Conseils. Il traduit environ 20 M mots/an d'anglais en français et 10 M mots/an dans l'autre sens, avec une qualité liée à la tâche de plus de 97 % (moins de 3 opérations d'édition pour 100 mots traduits).

¹⁸ NTT a trouvé une application de ce type, la traduction en anglais des brèves du Nikkei (bourse de Tokyo), et développé pour cela le système ALTFlash, totalement automatique, de grande qualité, et "bimoteur" (système à patrons avec en secours une version spécialisée du système général ALT/JE).

extrêmement bien pour le sous-langage des bulletins météo (mais pas pour ceux des situations ou des avertissements météo !).

Selon lui, il reste deux grands problèmes non encore résolus : (1) la définition automatique et l'identification dynamique du ou des sous-langages d'un texte, et (2) l'acquisition automatique de la connaissance linguistique pour les sous-langages.

Nous étudions, dans ce qui suit, les principales définitions des sous-langages.

4.1.1 Sous-langage selon Zellig Harris

Plusieurs définitions pour le terme « sous-langage » ont été données. Il semble que la première a été proposée par Zellig Harris (Harris 1968) :

“Certain proper subsets of the sentences of a language may be closed under some or all of the operations defined in the language, and thus constitute a sublanguage of it.”

Autrement dit, un sous-langage est un sous-ensemble strict de phrases d'une langue qui peut être fermé pour tout ou partie des opérations définies dans cette langue.

Cette définition semble à première vue incorrecte et trop vague, car les phrases d'un « sous-langage » ne sont souvent pas des phrases (correctes) de la « langue standard », dont on suppose que parle un linguiste, et alors on ne pourrait pas parler de « sous-ensemble » au sens usuel. Comme Z. Harris savait parfaitement ce qu'est un sous-ensemble d'un ensemble, nous sommes conduits à admettre cette extension de la notion usuelle de langue.

Définitions

Nous appellerons « langue standard » l'ensemble des énoncés d'une communauté linguistique formés d'une façon « correcte » par rapport à la grammaire et au vocabulaire usuels, tels qu'enseignés dans les cours de langue.

Nous appellerons « langue générale » l'union d'une langue standard et de toutes ses variantes (jargons, langues de spécialité, parlons régionaux, langages « techniques », et langages « secrétés » par des contextes socioprofessionnels).

D'autre part, il faudrait préciser les opérations dont on parle¹⁹. Par exemple, les phrases suivantes sont toutes déduites de la première par une opération de la langue (Denos 1994).

« Ce théorème apporte la solution au problème des valeurs aux bornes.

C'est ce théorème qui apporte la solution au problème des valeurs aux bornes.

Ce que ce théorème apporte est la solution au problème des valeurs aux bornes.

La solution au problème des valeurs aux bornes est apportée par ce théorème.

Ce théorème apporte-t-il la solution au problème des valeurs aux bornes ?

Ce théorème n'apporte pas la solution au problème des valeurs aux bornes.»

Dans sa définition d'un sous-langage, Harris ne dit pas de quelles opérations il parle. Mais il propose ensuite une définition « inductive » plus précise : un sous-langage SL est le plus petit ensemble contenant une base B et fermé (stable) par un ensemble de règles R.

SL = <B, R>, où

- la base B est un ensemble « noyau » d'énoncés ou schémas d'énoncés observés ;

¹⁹ Par exemple, mise au passif, à l'impersonnel, à l'interrogatif, ou simplement à un autre temps ou un autre mode.

- les règles R sont des règles de transformation comme la passivation, l'extraposition, l'interrogation, la mise au passif, à l'impersonnel, à l'interrogatif, ou simplement à un autre temps ou un autre mode, etc.

Ou bien un énoncé du sous-langage est dans le « noyau », ou bien il résulte d'un énoncé du sous-langage par une transformation de R. Par exemple, si *“The enzyme activated the process.”* est dans le sous-langage, et si la passivation est une des transformations permises, *“The process was activated by the enzyme.”* le sera aussi.

Cette définition est difficilement utilisable en pratique, car elle ne fournit pas de moyen opérationnel pour identifier le noyau et les règles caractérisant un sous-langage observé.

En se concentrant principalement sur les sous-langages scientifiques et techniques, Harris en arrive à considérer un sous-langage comme un sous-ensemble de la langue générale qui peut ne pas être contenu dans la langue standard. Il y a des phrases qui sont acceptables dans le sous-langage même si elles sont « incorrectes » dans la langue standard, et réciproquement. Par exemple, il est acceptable, dans un article de biochimie, de dire *“The polypeptides were washed in hydrochloric acid”* et pas *“hydrochloric acid was washed in polypeptides”*.

Le sous-ensemble est fermé pour certaines opérations syntaxiques de la langue générale. Cela signifie que, si une phrase est acceptable dans le sous-langage, ses transformations syntaxiques le seront aussi.

Harris a ainsi introduit en linguistique une caractérisation précise d'un sous-langage, et cela par analogie avec un sous-système en mathématiques, qui peut être un système indépendant. Un sous-langage est alors un sous-système ou un système indépendant de la langue standard.

4.1.2 Sous-langage selon Bross et autres

Une deuxième définition donnée par Bross et autres (Bross, Shapiro et al. 1972) est la suivante :

“Informally, we can define a sublanguage as the language used by a particular community of speakers, say, those concerned with a particular subject matter or those engaged in a specialized occupation.”

Autrement dit, un sous-langage est l'ensemble des énoncés susceptibles d'être prononcés par une communauté (de communication) en un certains temps et en certains lieux.

Cette définition est observationnelle et expérimentale, et prend directement en compte un contexte d'usage particulier. C'est celle qui a été utilisée dans le projet TAUM-METEO (1972-1973), puis pour des manuels de maintenance d'avions dans le cadre du projet TAUM-AVIATION (1974-1981) et du PN-TAO (Projet National de TAO, 1982-87) en France.

4.1.3 Sous-langage selon Kittredge et autres

(Grishman and Kittredge 1986) et (Deville 1989) définissent aussi un sous-langage comme étant une forme spécialisée d'une langue naturelle employée dans un domaine ou un thème particulier.

À titre d'exemples de sous-langages, on peut citer les bulletins METEO, les manuels techniques d'un avion, les articles scientifiques concernant la pharmacologie, les rapports de radiologie, et les annonces immobilières ou autres.

Un sous-langage est caractérisé par un vocabulaire spécialisé, une sémantique restreinte, et dans beaucoup de cas une syntaxe spécialisée (voir Tableau 10). Ainsi, les prépositions et articles normalement obligatoires peuvent être omis. Exemple : *« trappe visite réservoir avant gauche », « vent fort lac Saint-Jean », « orienté objet ».*

(Kittredge and Lehrberger 1982a) ont quelques affirmations communes :

- “*A sublanguage of a natural language L is part of L.*”

Dans ce critère, il y a une ambiguïté sur l'utilisation du terme “*natural language L*”. On ne sait pas s'il s'agit de toute la langue, ou de la langue standard. De plus, quand on parle de « partie », on ne sait pas si on parle de sous-ensemble ou d'autre chose. Ici, il s'agit manifestement de sous-ensemble, et on considère alors un sous-langage comme l'ensemble des phrases « identifiables » comme phrases de l'ensemble d'énoncés défini plus haut comme la « langue générale ».

Il s'agit alors de définir l'identifiabilité, qui n'est pas l'acceptabilité, car cette dernière dépend du type de locuteur. En effet, il est très fréquent que les énoncés « identifiés » comme appartenant à un « sous-langage » soient inacceptables dans la langue standard généralement admise ainsi que dans d'autres sous-langages.

- “*A sublanguage is identified with a particular semantic domain.*”

Ce critère veut dire qu'un sous-langage apparaît dans des textes limités à un domaine spécifique. Par contre, Harris a utilisé le terme sous-langage d'une façon plus générale, dans laquelle il n'y a pas d'indication de limitation du domaine sémantique.

- “*A sublanguage of a natural language L grows in a natural way through the use of L, albeit in special circumstances.*”

La troisième caractéristique admise est donc qu'un sous-langage ne peut pas être fabriqué, contrairement à une langue artificielle, mais se développe avec l'utilisation de la langue dans un contexte particulier.

- “*What we refer to as sublanguage texts usually contains some material that does not belong to the sublanguage proper.*”

Le matériau en question peut se présenter comme des phrases standard mélangées avec des phrases du sous-langage proprement dit, ou comme des syntagmes non standard incorporés dans des phrases ou des expressions du langage standard. Par exemple, dans un texte sur la géométrie, on peut trouver :

“*(1) Pythagoras proved the theorem that now bears his name in the sixth century B.C. He was able to show that the area of the square on the hypotenuse of a right triangle equals the sum of the areas of the squares on the other two sides.*”

La première phrase n'est pas une phrase de la géométrie mais de l'histoire. La deuxième phrase est une phrase de la géométrie (« *the area of the square on the hypotenuse* ») incorporée sous (« *He was able to show that* »). Cela divise un livre sur la géométrie en deux types d'énoncés selon les propriétés linguistiques, aussi bien syntaxiques que sémantiques. C'est un phénomène général qu'on trouve dans des textes réels restreints à des domaines sémantiques. Typiquement, un texte scientifique se compose d'un mélange de discours dans un domaine particulier et de métadiscours sur le sujet.

Plusieurs affirmations sont donc partagées par Grishman, Kittredge et Lehrberger. Ils sont d'accord sur la définition d'un sous-langage comme étant la forme spécialisée d'une langue naturelle employée dans un domaine ou un thème particulier.

Kittredge a donné une définition informelle mais assez précise, qui résume bien les précédentes :

Un sous-langage est un sous-ensemble d'une langue ayant les propriétés suivantes :

- le sous-ensemble fait référence à un domaine particulier ou à une famille de domaines liés,
- l'ensemble des phrases et des textes du sous-système reflète l'usage d'une communauté de personnes ayant en commun des connaissances élaborées du domaine,
- le sous-système a les propriétés fondamentales d'un système linguistique : consistance, complétude, économie d'expression, préférences lexicales et stylistiques.
- le sous-système est maximal par rapport au domaine (il n'y en a pas de plus grand qui possède ces propriétés).

Le Tableau 10 résume les trois différents types de langue : langue générale, langue standard, et sous-langage.

Type de langue / Niveau considéré	Langue générale	Langue standard	Sous-langage
Textes	Corrects + incorrects	Corrects	Corrects + incorrects
Syntaxe	Usuelle + spécialisée	Usuelle	Spécialisée
Vocabulaire	Usuel + spécialisé	Usuel	Spécialisé

Tableau 10 : types de langues et caractéristiques associées

Comme le montre la Figure 32, nous acceptons donc l'idée qu'une langue générale est l'union de la langue standard et de tous ses sous-langages, et par conséquent, aussi l'idée d'Harris, selon laquelle un sous-langage est un sous-ensemble de la langue générale.

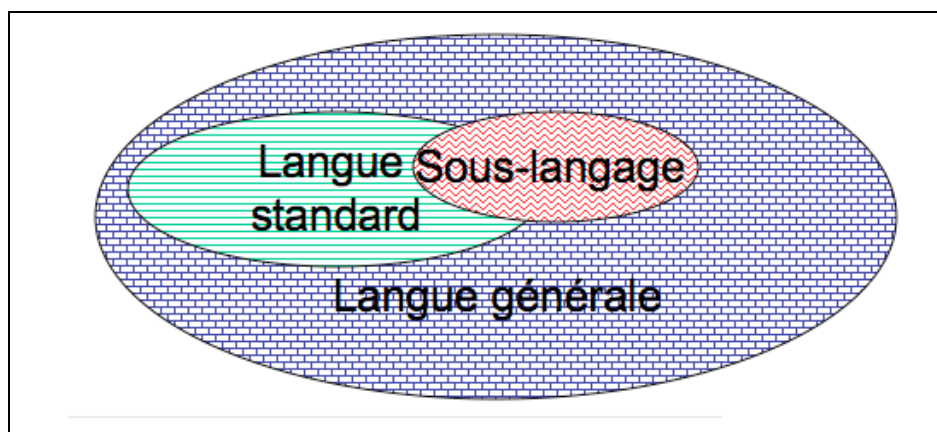


Figure 32 : langue générale, langue standard, et sous-langage

4.2 Identification d'un sous-langage naturel

Dans les systèmes de TALN cités par (Sekine 1994), la détection du domaine et la définition de sous-langage sont faites par un humain. Cependant, quelques critères sont utilisés pour caractériser un sous-langage.

L'intervention humaine n'est pas possible dans les cas où un système doit traiter une grande variété de genres de textes et de domaines. Ainsi, la tâche doit être automatisée. Nous commençons par présenter les phénomènes qui caractérisent un sous-langage.

4.2.1 Phénomènes spécifiques dans un sous-langage

Ces phénomènes concernent les classes sémantiques et grammaticales de mots, les modèles de discours et les sous-classes de mots.

4.2.1.1 Les classes de mots

(Kittredge 1978; Kittredge 1993) utilise une analyse distributionnelle du corpus. Cela permet de définir un ensemble de classes de mots pour le sous-langage, fondées sur les similarités d'occurrences. Ainsi, chaque notion grammaticale et sémantique est représentée par une classe de mots qui peut avoir des sous-classes. Par exemple, dans le domaine de la biochimie, les mots qui désignent des ions chimiques appartiennent tous à une même classe.

Kittredge définit alors des relations entre un verbe et les classes autorisées en tenant compte de la sémantique associée au sous-langage en question. Par exemple, dans le sous-langage associé à la biochimie, on peut dire « *La cellule assimile le potassium* », mais on ne peut pas dire « *Le potassium assimile la cellule* ».

4.2.1.2 Les modèles de discours

(Lehrberger 1982) utilise le comportement des noms dans le sous-langage traité pour définir des relations entre ces noms. Généralement, les modèles de phrase s'expriment en terme de combinaisons observées de classes de mots. Ainsi, chaque relation R définit deux sous-ensembles de noms, appelés domaine « *domain* » et image « *range* ». Par exemple, la relation *INSTR* contient les paires (*installation, kit*) et (*access, component*), et la relation BUT^{-1} ($x BUT^{-1} y$ si et seulement si $y BUT x$) contient les paires (*installation, kit*) et (*access, component*), (*safety, pipe*).

Ces informations aident à la désambiguïsation relative au parenthésage des subordinations dans un groupe nominal (Denos 1994).

Le même principe peut être appliqué aux adjectifs, et aux arguments des verbes. Dans d'autres cas de sous-langages, des modèles de discours sont définis pour accompagner les parties du texte et constituent l'enchaînement entre les phrases du sous-langage proprement dit.

4.2.1.3 Les sous-classes

Dans certains cas, on a besoin d'utiliser des sous-classes (Denos 1994).

Si « *installation* » est l'image pour la relation *BUT* de « *kit* » dans le groupe « *installation kit* », cela ne signifie pas pour autant que tous les noms qui vont apparaître à la droite de « *installation* » lui seront liés par cette même relation. Par exemple, dans « *installation difficulty* », « *installation* » n'est pas le but de « *difficulty* ».

C'est pourquoi on est amené à définir des sous-classes, pour donner l'ensemble des mots avec lesquels « *installation* » est dans la relation *BUT*, et à sa droite, dans le sous-langage. On ajoute cette connaissance dans le système de traduction en indiquant pour chaque mot du dictionnaire quelles relations il entretient avec quelles sous-classes de mots.

On obtient un dictionnaire sensiblement plus compliqué qu'un simple lexique. Ces connaissances, qui reviennent en fait à attribuer aux noms des propriétés sur leurs arguments « droite » et « gauche », sont l'équivalent des connaissances classiquement spécifiées pour les différents arguments des verbes.

Alors que, dans beaucoup de sous-langages, les « classes » sont fondées sur des relations d'ordre général (par exemple *Fonction de, Partie de, Sujet de, Objet de*), au contraire, les sous-classes sont très spécifiques à chaque sous-langage.

4.2.2 Comment caractériser un sous-langage

Nous présentons un état de l'art sur les précédents travaux de caractérisation d'un sous-langage en nous focalisant maintenant sur le travail (Kittredge 1982b) sur la TA des sous-langages. Kittredge a procédé à une analyse détaillée en proposant les mesures suivantes : la complexité de sous-langage (taille et complexité du lexique, structure des phrases, éléments de liens entre phrases), et l'homogénéité du corpus traité.

4.2.2.1 Complexité d'un sous-langage

La complexité d'un sous-langage est liée au lexique et à la syntaxe utilisés. En matière de sous-langage, la taille du lexique seule ne suffit pas à caractériser la complexité. C'est le cas dans les recettes de cuisine, où le lexique est extrêmement riche à cause de la variété des ingrédients. Mais les mots utilisés pour décrire ces ingrédients sont répartis sur un nombre très limité de sous-classes, car

syntactiquement ils jouent le même rôle. Dans ce cas, mesurer la complexité du sous-langage en utilisant la taille du lexique sans la pondérer conduit à surestimer cette complexité.

Pour aboutir à une mesure plus pertinente, on peut compter le nombre de catégories et de sous-catégories morpho-syntaxiques nécessaires pour spécifier correctement la grammaire de ce sous-langage. Pour chaque item du lexique, on peut aussi calculer la complexité moyenne de sa description en terme de ces catégories. Par exemple, la description d'un nom est simple, contrairement à celle d'un prédicat, pour lequel il faut spécifier le sujet, l'objet et éventuellement les autres compléments fortement régis.

La structure des phrases peut aussi être utilisée pour estimer la complexité grammaticale, mais deux problèmes se posent.

Premièrement, il est difficile de trouver des critères d'évaluation identiques, voire même semblables, pour un sous-langage de style télégraphique et pour un sous-langage dont les constructions syntaxiques sont proches de celles de la langue standard. En effet, les deux grammaires associées sont très différentes.

Deuxièmement, souvent, on ne peut pas définir de critères de comparaison généraux, car les catégories principales du sous-langage sont très spécifiques. Par exemple, « condition météo » est une catégorie pour le sous-langage des bulletins météo. Elle reflète une catégorie plus sémantique que syntaxique, car on peut y trouver des noms, des adjectifs, ou des gérondifs (« pluie », « nuageux », « se réchauffant »,...). Il se trouve que cette catégorie (condition météo) est pourtant très pertinente dans le sous-langage en question. Il est alors nécessaire d'identifier les motifs grammaticaux de la langue standard qui sont pertinents pour comparer des sous-langages. Kittredge propose pour cela d'évaluer quantitativement le temps dominant (présent, futur, passé simple...), le type de modalité (probabilité, obligation), les types d'interrogatif et d'impératif, le type de subordination, etc.

(Kittredge 1978) s'est aussi intéressé à l'organisation du texte, qui peut être un élément révélateur de sa complexité. Il ne s'agit pas de l'enchaînement des types de phrases entre elles, car dans ce cas, un sous-langage tel que celui des bulletins météo serait évalué de manière erronée puisqu'il n'y a pas de surprise dans la présentation de ses textes. En effet, ils sont présentés selon une structure très régulière et selon un schéma préétabli, avec des items obligatoires et d'autres facultatifs. Il s'agit de mesurer les *liens de cohésion*, à savoir les liens lexicaux et les liens grammaticaux. Ce sont des liens concrets entre les phrases, au niveau local, par exemple, ceux créés par les pronoms, les comparatifs et superlatifs, les adverbes conjonctifs (« néanmoins », « au contraire »,...), la répétition lexicale, la synonymie, l'hyponymie²⁰.

4.2.2.2 Homogénéité d'un sous-langage

(Kittredge 1978) dit aussi que les textes d'un sous-langage doivent présenter une certaine homogénéité du point de vue de l'emploi des mots. Les classes de mots sont établies à partir des similarités de leur emploi. Il faut donc un corpus « représentatif » qui permette d'établir toutes les catégories nécessaires au sous-langage, toutes les séquences de catégories, et les sous-catégories.

Cela amène à parler de la *convergence* d'un sous-langage. Un sous-langage converge plus ou moins vite selon qu'il faut un corpus plus ou moins volumineux pour que le nombre de nouveaux motifs rencontrés dans une nouvelle phrase soit très faible. Pour fixer les idées, la probabilité d'apparition d'un nouvel élément lors de l'ajout d'une phrase au corpus doit être inférieure à 0,01 % (Kittredge 1978) (un mot nouveau ou une construction nouvelle tous les 10000 mots, soit 40 pages standard).

²⁰ A est hyponyme de B si A appartient à une sous-classe sémantique de B, par exemple : « bouteille » est hyponyme de « récipient ».

Généralement, un sous-langage est un système fermé (clos) et deux types de convergence peuvent être définis : *convergence grammaticale* et *convergence lexicale*. Dans certains cas, les deux types de convergence ne sont pas assurés. Ainsi, un sous-langage peut converger grammaticalement mais pas lexicalement, pour diverses raisons. Par exemple, le sous-langage des avertissements météo converge grammaticalement d'une façon rapide, mais ne converge pas lexicalement, car :

- de nouveaux noms propres, surtout des noms de lieu, peuvent apparaître lors d'un phénomène météorologique localisé à un endroit non cité d'habitude ;
- de nouveaux verbes et noms « hors météo » relatifs aux événements provoqués par un phénomène météo inhabituel (lignes électriques coupées, ponts bloqués...) peuvent être utilisés.

Pour ce qui est des noms de lieu, il y a convergence pour le sous-langage des bulletins météo, car on peut recenser à l'avance tous les noms de lieu risquant d'apparaître dans ces bulletins et ne se traduisant pas à l'identique. Au total, le vocabulaire du système METEO ne comporte qu'environ 3000 entrées.

4.2.3 Autres caractérisations

Il faut aussi mentionner deux autres méthodes d'identification d'un sous-langage qui ont été proposées, mais que nous ne considérons pas comme de bonnes mesures.

4.2.3.1 TTR (Type to Token Ratio)

Le TTR est égal au rapport entre le nombre des types²¹ de mots (lemmes différents) et le nombre total des mots (i.e., d'occurrences) d'un texte donné. Cette mesure est utilisée dans plusieurs domaines et types de documents (WordList, corpus LOB et Brown...). (Daoud 2006) dit par exemple que, plus ce rapport est faible, plus le texte est spécialisé. Cependant, cette mesure est mal définie et doit être améliorée. En effet, on peut faire diminuer le TTR arbitrairement en augmentant la taille du texte considéré. Nous montrons dans ce qui suit pourquoi et comment on peut améliorer cette mesure.

Exemple : si un texte comporte 1 000 mots, on dit qu'il comporte 1 000 « tokens ». Cependant, un bon nombre de ces mots vont se répéter, et il n'y aura peut-être que 400 mots différents dans le texte. Si on convient pour cet exemple que les « types » sont les différents mots, le rapport entre les types et les tokens serait de 40 %.

Le tableau suivant est un exemple détaillé du calcul du TTR tiré d'un travail sur le désordre de la communication humaine²².

²¹ Un « type » peut correspondre la forme de base dans le dictionnaire, i.e. un lemme en français comme c'est le cas dans CATS.

²² <http://www.ux1.eiu.edu/~cfrma/>, site web visité en 2006.

Enoncés	Nombre de mots	Nombre de lemmes nouveaux
/I/, /need/, /that/	3	3
/he/, /eating/, /that/	3	2
/got/, /my/, /dolls/	3	3
/what/, /that/, /is/	3	2
/I/, /'m/, /gonna/, /eat/, /that/	5	3
/that/, /my/, /doll/	3	0
/he/, /not/, /go/	3	2
/That/, /'s/, /my/, /car/	4	1
/he/, /going/	2	0
/he/, /'s/, /not/, /eating/	4	0
Total	33	16
TTR	0,48	
Décision	Normal	

Figure 33 : exemple de calcul du TTR

Remarquons ici que les mots « eat » et « eating », « is » et « 's »... sont comptés comme ayant le même lemme.

Le ratio type/token est évidemment une fonction décroissante de la longueur du texte (ou du corpus de texte) étudié. Un article de 1.000 mots peut avoir un TTR de 40 %, un texte plus court peut avoir un ratio de 70 % et un texte de 4 millions de mots, quant à lui, peut avoir un ratio de 2 %, et ainsi de suite.

Cependant, WordList par exemple calcule cela de manière différente, en utilisant la méthode WordSmith : le ratio type/token standardisé (également appelé STTR) est calculé tous les N mots à mesure que WordList parcourt chaque fichier texte.

Le ratio est calculé pour les N premiers mots, puis recalculé pour les N suivants, et ainsi de suite jusqu'à la fin du texte ou du corpus. Par défaut N = 1.000. De cette façon, on obtient un ratio type/token basé sur des blocs de texte consécutifs de N mots. Les textes comprenant moins de N mots obtiennent un ratio standardisé de 0.

Dans la méthode WordSmith, plusieurs paramètres peuvent être fixés :

- Le nombre N qui doit être compris entre 100 et 20000.
- Une liste de mots à exclure (stop list) : il peut s'agir par exemple de mots très courants sans grand intérêt comme *le*, *de*, *était*, *est*, qu'on souhaite exclure d'une liste de mots ou d'une analyse de mots-clés.
- Les formes différentes : il peut être utile d'enregistrer plusieurs entrées en même temps : par exemple *want*, *wants*, *wanting*, *wanted*, qui font partie du même lemme.
- Les nombres : on peut décider d'inclure ou d'exclure les nombres des listes de mots.
- Les traits d'union : il s'agit du caractère utilisé pour séparer les mots. L'élément « *auto-formation* » peut être considéré comme étant deux mots ou un seul mot, selon le choix de paramétrage.

La figure suivante montre la différence des résultats en utilisant le TTR et le STTR. En effet, le TTR varie d'une façon très rapide (10.78 ; 23.64 ; 21,82 et 16.10) par rapport au STTR (37.15 ; 36.37 ; 37.37 et 35.08).

TTR et STTR sont des mesures relativement grossières, même si beaucoup d'auteurs disent qu'elles donnent des informations sur la « densité lexicale ». Le critère STTR ne permet en effet pas de calculer et de voir la croissance du vocabulaire dans un langage, et donc ne permet pas à lui seul de déterminer si on a affaire à un sous-langage plus ou moins spécialisé. Ainsi, si le vocabulaire spécialisé change tous les N ou $K \cdot N$ mots, le STTR restera le même, alors que le vocabulaire total croîtra.

N	0	1	2	3
text file	overall	31.5.98.txt	alex salmond 15.3.98.txt	4.6.98.tx
file size	151,474	23,489	20,216	37,53
tokens (running words) in text	22,992	3,485	3,089	5,73
tokens used for word list	22,992	3,485	3,089	5,73
types (distinct words)	2,479	824	674	92
type/token ratio (TTR)	10.78	23.64	21.82	16.10
standardised TTR	37.15	36.37	37.37	35.08
standardised TTR std.dev.	58.95	49.00	46.95	53.5
standardised TTR basis	1,000	1,000	1,000	1,000
mean word length (in characters)	4.49	4.52	4.55	4.4
word length std.dev.	2.42	2.48	2.58	2.3
sentences	928	147	132	23
mean (in words)	24.78	23.71	23.40	24.2
std dev	19.73	20.38	15.28	18.7

Figure : exemple de calcul du TTR et STTR

4.2.3.2 Taille de la grammaire en fonction de la taille des données

Une autre méthode qui a été proposée pour identifier un sous-langage consiste à mesurer la taille de la grammaire générée automatiquement ou créée manuellement en fonction de la taille du corpus. S'il semble qu'on se rapproche d'une asymptote horizontale, cette asymptote représente en un sens la complexité syntaxique du texte étudié. Si elle est assez basse (petite taille de la grammaire à la limite), on peut en inférer qu'on a affaire à un sous-langage de « typologie » restreinte (Lopez, Fay-Varnier et al. 1999).

Cependant, on peut tout de suite remarquer que cette méthode est absolument inapplicable en pratique, dans la mesure où il n'existe, pour aucune langue, une grammaire la couvrant totalement, ni pour la langue standard, ni a fortiori pour la langue générale. Cela constitue une première raison pour rejeter cette méthode.

Dans les travaux sur le système EGAL²³ « Extraction de Grammaire d'Arbres Lexicalisés » (Lopez, Fay-Varnier et al. 1999) dont l'objectif final était de concevoir des interfaces vocales Homme-Machine génériques et portables permettant la communication en langue spontanée, Lopez et autres ont réalisé des tests systématiques sur l'évolution des données générées.

Ils ont proposé la méthode suivante. Dans un premier temps, on fait des tirages aléatoires d'énoncés à partir d'un corpus, et on construit automatiquement la grammaire associée à partir de la grammaire générale, supposée connue. Ensuite, une fois la sous-grammaire d'application obtenue, un modèle d'analyse permet de la mettre à l'épreuve sur un corpus de test, en utilisant plusieurs algorithmes et stratégies. Ces tests permettent en outre d'obtenir des informations sur la représentativité du corpus initial utilisé pour décrire le sous-langage. On étudie l'évolution de la taille de la grammaire, donnée par le nombre total de liens vers un *schème*²⁴ (Candito 1999) en fonction de l'augmentation du nombre des énoncés pris en compte.

La Figure 34, dans laquelle les abscisses représentent la taille du corpus et les ordonnées la taille de la grammaire générée, illustre le résultat de la méthode proposée. D'après les auteurs, elle montre que la courbe représentant l'évolution en question atteint une asymptote. Mais, à y regarder de plus près, cette figure ne prouve absolument pas cela. Rien n'empêche en effet la courbe de continuer à croître, même faiblement, et à dépasser n'importe quelle taille, aussi grande soit-elle, après une augmentation suffisante de la taille du corpus. Il est par exemple possible que des phénomènes grammaticaux nouveaux apparaissent de temps en temps, sans que d'autres disparaissent. Il faudrait vraiment beaucoup plus que 800 phrases pour qu'on puisse se persuader de l'existence d'une asymptote. Cela constitue une deuxième raison pour rejeter cette méthode.

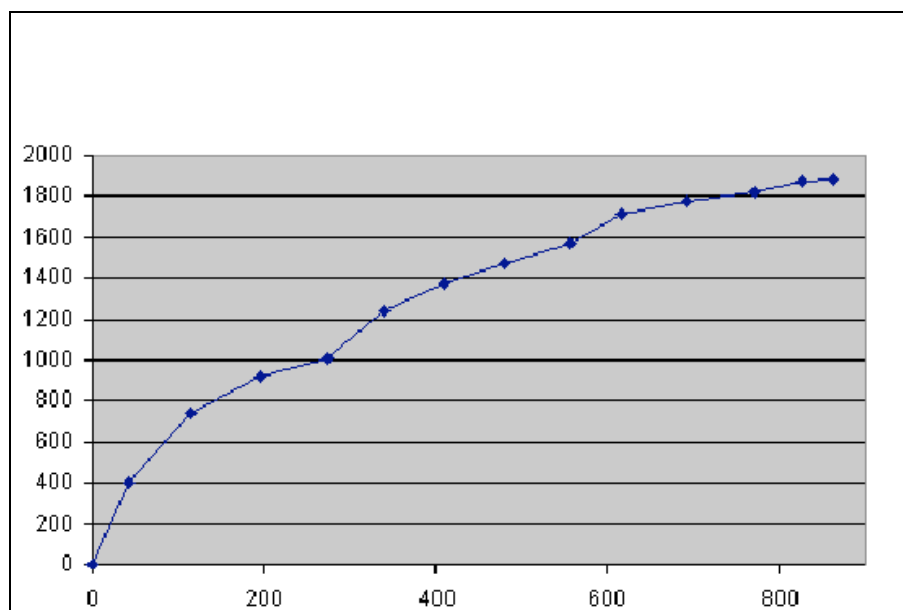


Figure 34 : évolution de la taille de la grammaire générée en fonction de la taille du corpus (Lopez, Fay-Varnier et al. 1999)

Pour les deux raisons détaillées précédemment, nous jugeons que cette méthode n'est pas bonne pour identifier un sous-langage.

²³ EGAL est un système capable d'extraire de façon semi automatique un sous-langage applicatif à partir d'une grammaire générale et de corpus.

²⁴ Un schème est un motif d'arbre élémentaire non instancié.

4.3 Une taxonomie possible des sous-langages

Étant confronté à une tâche de portage linguistique d'applications de e-commerce traitant des énoncés spontanés appartenant à des sous-langages, et cherchant à déterminer les méthodes de portage les plus adaptées, et le coût du portage, nous devons d'abord voir si nous avons affaire à un sous-langage ou non, et si oui, quelles sont ses caractéristiques.

Nous ne pouvons pas utiliser la mesure de la taille de la grammaire générée à partir de la grammaire générale en fonction de la taille du corpus utilisée (comme c'est le cas dans les travaux du système EGAL), car cela nécessite bien sûr d'avoir la grammaire générale, chose qu'on n'a jamais.

Nous caractériserons donc un sous-langage par *observation* et par *estimation* d'un certain nombre de paramètres objectifs. Ensuite, nous donnerons des « diagrammes radar » pour quelques exemples de sous-langages. Enfin, nous présenterons les inférences possibles suite à l'application de cette taxonomie.

4.3.1 Mesures possibles

On peut améliorer la mesure du TTR en évaluant essentiellement deux critères, la complexité lexicale et la complexité syntaxique.

4.3.1.1 Complexité lexicale

Nous mesurons la complexité lexicale par (1) le nombre total de lemmes, (2) le nombre total de lexèmes (total des nombres de formes orthographiques d'un lemme²⁵). Le quotient du deuxième nombre par le premier donne une information sur le coefficient d'expansion dictionnaire, la taille du vocabulaire et sa croissance.

Voici certaines valeurs typiques (chacune illustrée par un exemple) :

500	Nombre de lemmes pour l'occasion automobile dans CATS
2000	Nombre de lemmes pour l'immobilier à Amman dans CATS
3000	Nombre de lemmes dans le système METEO de traduction de bulletins météo
10.000	Nombre de termes techniques spécialisés en aéronautique pour un avionneur ²⁶
20.000	Nombre de tous les termes techniques en aéronautique pour un avionneur
50.000	Nombre de lexies dans un dictionnaire d'usage comme le FEM (français-anglais-malais)
100.000	Nombre de termes dans un système de TA généraliste de base

Tableau 11 : mesure de la complexité lexicale

Par exemple, dans CATS (voir section 2.2.3.3.3) et pour l'ensemble de ses domaines (automobile, immobilier, emplois, divers), le nombre de lemmes est d'environ 10.000 et celui des lexèmes est d'environ 30.000, ce qui donne un coefficient d'expansion égal à 3. Concernant le domaine de l'automobile, la taille du dictionnaire est de l'ordre de 2000 entrées et varie très faiblement dans le temps. En effet, ce n'est qu'après une longue période qu'il y a un nouveau modèle de voiture qui apparaît. Par contre, le dictionnaire du domaine de l'immobilier est beaucoup plus grand et sa taille varie plus rapidement, car non seulement on ajoute des noms de villes et de rues pour étendre son

²⁵ Par exemple, CLEF a deux formes (standard) « clef » et « clé », et ALPHA-ROMÉO peut avoir « Alpha-Roméo » et « Alfa-Roméo » (avec erreur typique du sous-langage).

²⁶ Dans les années 70-80, le terme « avionneur » désignait surtout un « équipement d'avion ». Depuis, il semble surtout désigner un conducteur d'avions.

usage, mais aussi on y ajoute des noms de nouvelles constructions qui apparaissent fréquemment (ex : « immeuble Total »).

Cela nous conduit à définir la notion de mise à jour du vocabulaire, qui dépend du domaine traité. Pour celui de l'automobile, un entretien annuel est donc nécessaire. Par contre, pour l'immobilier, un entretien doit être fait chaque fois que l'on ajoute une ville.

Nous classons le vocabulaire traité en classes et sous-classes sémantiques en regardant la variation de la taille du dictionnaire en fonction de la taille du corpus.

Par exemple, dans CATS et pour le domaine de l'automobile, on peut trouver les classes et les sous-classes sémantiques suivantes :

Classe verbe = { vendre, acheter, échanger... }
Classe vehicule = { voiture, bus,... }
Classe marque = { Mercedes, Renault, Peugeot... }
Classe couleur = { noire, bleue... }
Classe année = { 96, 95... }
Sous-classe année = { avant, après... }

Figure 35 : quelques classes et sous-classes pour le domaine de l'automobile dans CATS

Nous pouvons avoir un autre type d'information qui concerne la difficulté lexicale. Nous utilisons l'échelle suivante :

Facile	Le vocabulaire n'est pas à traduire.
Plus ou moins translittérés	Présence de mots ou d'expressions étrangers (d'une autre langue mais saisis dans la même langue), ce qui augmente le coefficient d'expansion
Difficile à comprendre	Vocabulaire très spécifique
Difficile à traduire	Difficile à comprendre, ou difficulté à trouver l'équivalent dans la nouvelle langue

Tableau 12 : mesure de la difficulté lexicale

4.3.1.2 Complexité syntaxique

4.3.1.2.1 Longueur des énoncés

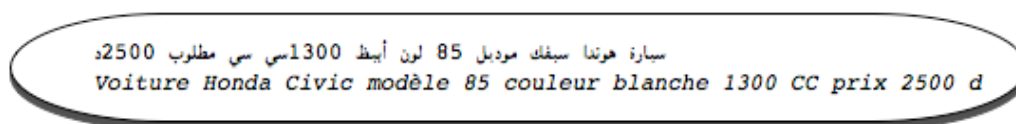
Nous mesurons la complexité syntaxique par la longueur des phrases utilisées. Nous calculons la longueur des phrases par bloc (par exemple de 100 phrases). Cette information donne une idée sur le type des phrases utilisées.

Nous utiliserons les valeurs de longueur suivantes :

Très courte	=< 6,5 mots/phrased (comme le BTEC en anglais, des phrases de voyage)
Courte]6,5..8] mots/phrased
Assez courte]8..10] mots/phrased (comme le domaine de l'automobile dans CATS)
Moyenne]10..15] mots/phrased
Peu longue]15..20] mots/phrased
Assez longue]20..25] mots/phrased
Longue]25..35] mots/phrased
Très longue]35..60] mots/phrased
Extrêmement longue	> 60 mots/phrased

Tableau 13 : mesure de la longueur des énoncés

Par exemple, dans CATS, la longueur moyenne des phrases est de l'ordre de 9 mots pour le domaine de l'automobile, et de 12,5 mots pour le domaine de l'immobilier. Ce sont des phrases simples et parfois même pas des phrases, mais des suites de groupes nominaux sans verbe, comme le montre l'exemple suivant.



4.3.1.2.2 Complexité grammaticale

Nous examinons aussi la structure et l'organisation des énoncés. Les phrases peuvent être simples, composées, imbriquées... Il peut y avoir ou non des constructions de type extraposition, anaphore, ellipse, et de la phraséologie spécifique (outre le vocabulaire).

Nous prendrons comme points de repère pour la complexité syntaxique :

Très simple	Liste de groupes élémentaires
Simple	Pas plus d'un niveau de subordination, pas d'imbrication/enchâssement central
Moyennement complexe	Possibilité d'imbrications, coordinations et énumérations, anaphore intraphrastique
Complexe	Possibilité d'anaphore extraphrastique, d'ellipse, d'extraposition, etc.

Tableau 14 : mesure de la complexité syntaxique

4.3.1.2.3 Convergence grammaticale

La convergence grammaticale est inversement proportionnelle au nombre de lignes de la grammaire utilisée. Nous constatons les cas de convergence suivants :

Grammaire très peu convergente	Pas de sous-langage
Grammaire peu convergente	Mélange de sous-langage et de la langue standard
Grammaire rapidement convergente	Gros sous-langage, pas de langue standard, cas d'un manuel d'entretien
Grammaire très rapidement convergente	Petit sous-langage, cas de MÉTÉO, AltFlash (traducteur de NTT J→E pour les « brèves » du Nikkei –la bourse de Tokyo), CATS

Tableau 15 : mesure de la convergence grammaticale

4.3.1.2.4 Convergence lexicale

La convergence lexicale est liée au nombre de nouveaux motifs ou mots qui apparaissent dans le temps. Nous utilisons les mesures suivantes en précisant qu'il s'agit ici de page standard (250 mots) :

Lexique très peu convergent	1 mot/page après 5000 pages, cas de l'encyclopédie EOLSS (http://www.eolss.net/)
Lexique peu convergent	1 mot/10 pages après 5000 pages
Lexique rapidement convergent	1 mot/100 pages après 5000 pages
Lexique très rapidement convergent	1 mot/1000 pages après 5000 pages

Tableau 16 : mesure de la convergence lexicale

4.3.2 Etude de quelques sous-langages

4.3.2.1 L'automobile en Jordanie

Dans CATS et pour le domaine fermé de l'automobile, la taille du vocabulaire utilisé est d'environ 638 entrées principales, il s'agit d'un vocabulaire facile et plus ou moins translittéré. On peut avoir des phrases très simples et très courtes (zone sombre dans la Figure 36), la convergence grammaticale est très rapide, mais la convergence lexicale est moins rapide, à cause du fait que de nouveaux motifs (comme les nouveaux modèles de voitures) peuvent apparaître après un certain temps. On peut aussi avoir des phrases simples et courtes (zone claire dans la Figure 36).

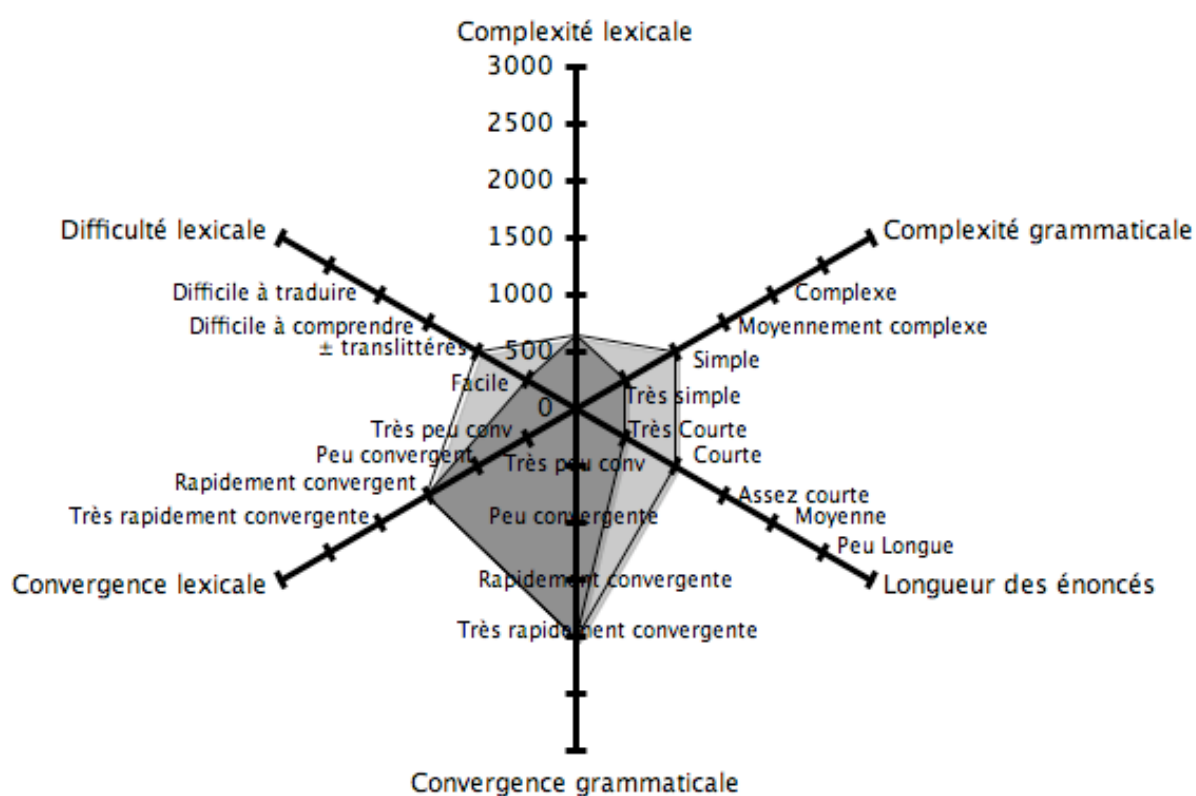


Figure 36 : analyse du sous-langage de l'automobile dans CATS

Nous présentons quelques exemples d'énoncés dans le Tableau 17.

Énoncé arabe	Traduction française	Nombre de mots pour l'énoncé arabe	Nombre de caractères pour l'énoncé arabe
مطلوب سيارة هونداى موديل 97 والسعر ما بين 3500 الى 3750	Recherche voiture Honda, modèle 97, prix entre 3500 et 3750	11	55
مطلوب سيارة سبور	Recherche voiture sport	3	16
اريد سيارة مرسيدس موديل 82 لون ابيض	Je veux une voiture Mercedes modèle 82 couleur blanche	7	35

Tableau 17 : exemples d'énoncés du sous-langage de l'automobile en Jordanie

4.3.2.2 L'immobilier à Amman

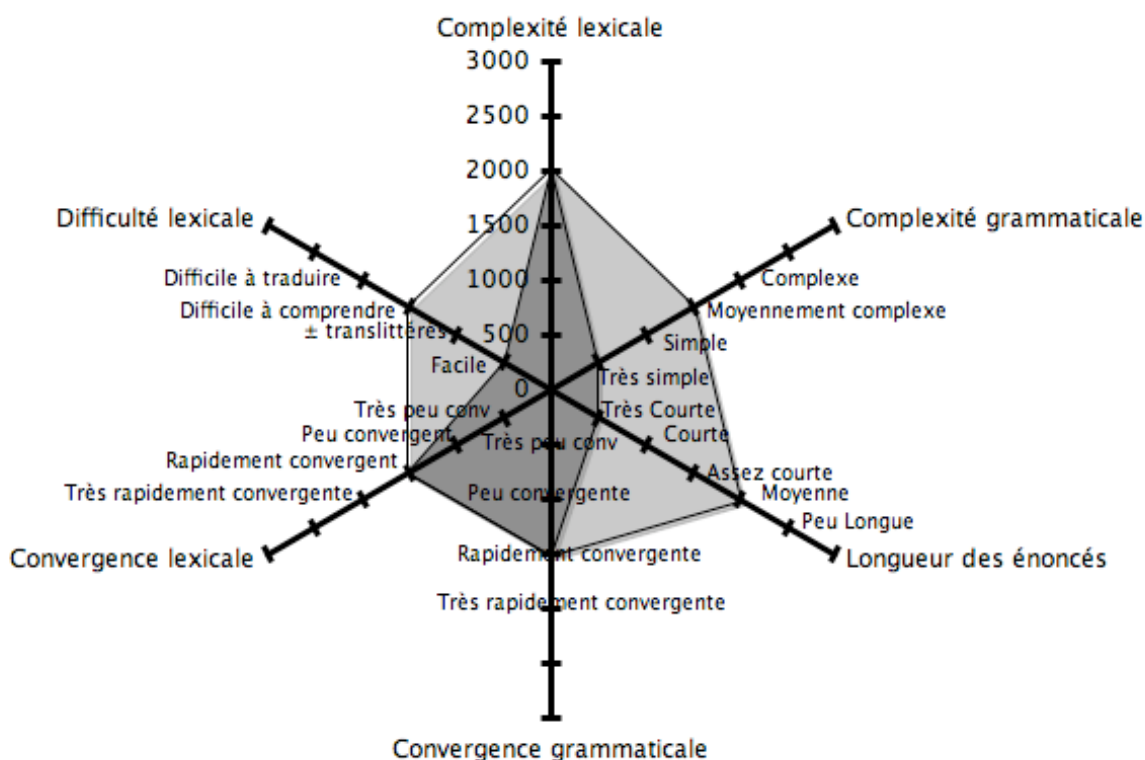


Figure 37 : analyse du sous-langage de l'immobilier dans CATS

Pour le domaine de l'immobilier, on trouve des caractéristiques qui rendent le portage un peu plus difficile. La taille du vocabulaire utilisé est d'environ 2000 entrées principales, et il s'agit d'un vocabulaire parfois difficile à comprendre (voir Tableau 18).

Énoncé arabe	Traduction française	Nombre de mots pour l'énoncé arabe	Nombre de caractères pour l'énoncé arabe
شقق مفروشة لايجار 2 نوم صالة صالون مطبخ حمام الشمساني بسعر مناسب	A louer appartements meublés 2 chb, salon, cuisine, 2 sdb, ECHAMSSINI, TB prix	9	62
شقة مستقلة بثمن جيد	Appartement isolé à TB prix	4	19
عماره للبيع بخلدا 3 طوابق بسعر 170 الف	À vendre, bâtiment de 3 étages à KALDA, 170000 dinars	7	37

Tableau 18 : exemples d'énoncés du sous-langage de l'immobilier à Amman

On peut avoir des phrases très simples et très courtes (zone sombre dans la Figure 37) la convergence grammaticale est rapide, mais le lexique est peu convergent à cause des nouveaux noms

de lieu qui peuvent apparaître, en petit nombre mais régulièrement si on reste au même endroit (Amman en l'occurrence), ou en grand nombre, chaque fois qu'on change d'endroit²⁷. On peut aussi avoir des phrases moyennement compliquées et moyennement longues (zone claire dans la Figure 37).

Le Tableau 18 présente quelques exemples d'énoncés.

4.3.2.3 La météo

Le style du sous-langage des bulletins météorologiques est télégraphique. Nous présentons un échantillon en anglais et sa correspondance en français dans le Tableau 19.

Énoncé anglais	Énoncé français	Nombre de mots pour l'énoncé anglais	Nombre de caractères pour l'énoncé anglais
Showers or snowshowers	Averses de pluie ou de neige	3	33
Thunderstorms likely	Risque d'orages	2	22
Afternoon showers tomorrow	Averses demain après-midi	3	26
Cloudy clearing overnight	Nuageux se dégageant durant la nuit	3	25
Rain developing	Pluie commençant	2	25
Clouding over in the morning	Se couvrant dans la journée	5	28

Tableau 19 : échantillon de bulletins météorologiques

Le vocabulaire de ce sous-langage comporte 3000 mots liés à un domaine fermé. Comme le montre la zone sombre de la Figure 38, on peut avoir des énoncés très simples et très courts avec une convergence grammaticale rapide. La convergence du lexique est assez lente, pour les raisons données précédemment au sujet des avertissements météo (voir 4.2.2.2), et qui sont liées à l'apparition de nouveaux noms de lieu. Mais, dans le cas de METEO, on considère que tout mot inconnu est un nom de lieu qui se traduit par lui-même, et on ne l'intègre pas au dictionnaire. La convergence « pratique » du lexique est donc en fait très rapide. On peut aussi avoir dans ce sous-langage des phrases complexes et moyennement longues.

²⁷ Dans ce second cas, il ne s'agit plus de portage, mais de localisation.

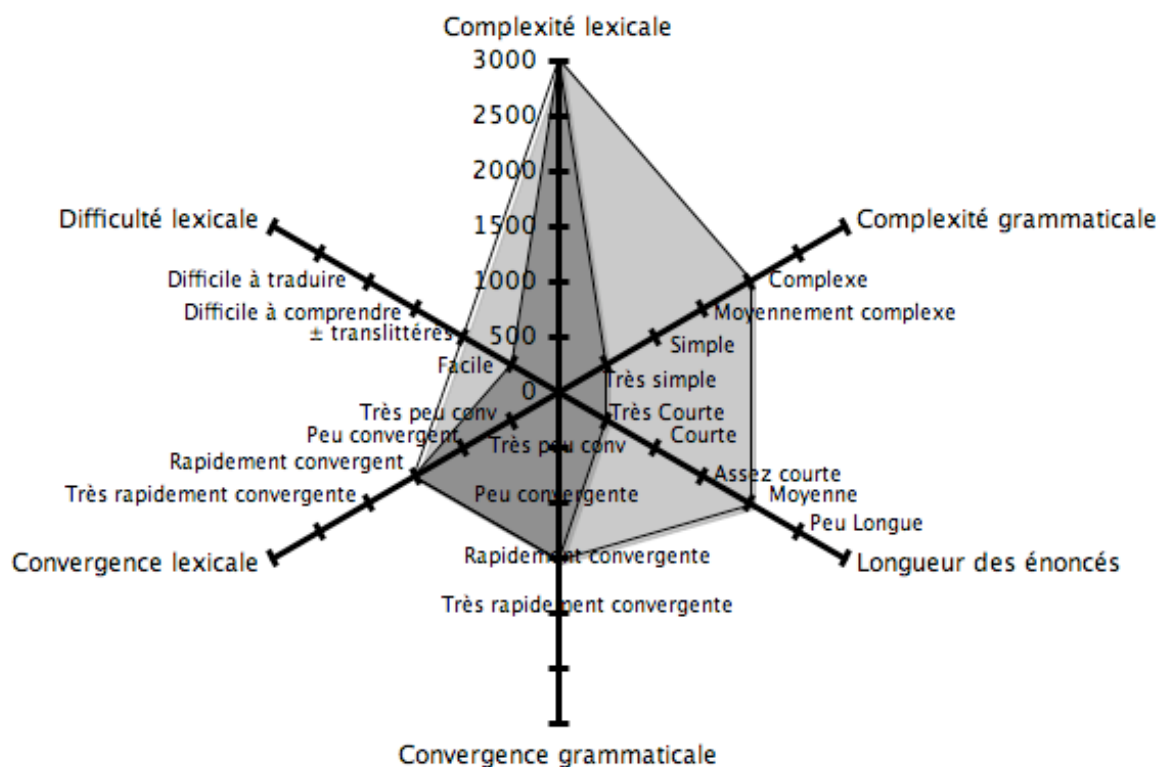


Figure 38 : analyse du sous-langage des bulletins météo

D'autre part, l'analyse dans METEO est plus sémantique que syntaxique, les grammaires extraient le contenu de chaque phrase avant de la traduire. La technique employée est celle des « grammaires sémantiques ». Du fait du style télégraphique, on a peu de repères syntaxiques (ponctuation, conjonction, articles, prépositions) pour procéder à l'analyse. La sémantique du petit monde des bulletins météo est facile à cerner, et se résume à la formule suivante : une « *condition météo* » (limitée dans le temps et dans l'espace) « *début* » ou « *se maintient* », ou « *cesse* », ou « *fait place à une autre* ».

4.3.3 Inférence possible

L'ensemble de ces mesures nous aident à choisir et/ou préciser :

- la méthode de portage ou de localisation adéquate,
- le coût associé à un certain type d'application (recherche d'information, traduction automatique, résumé, catégorisation...),
- la durée de la tâche.

Rappelons que les précédents points énumérés dépendent aussi d'autres facteurs et paramètres étudiés dans le premier chapitre, qui précisent la situation traductionnelle, à savoir le niveau d'accès permis et les compétences linguistiques et langagières disponibles.

On peut établir une hiérarchie ensembliste en étudiant des sous-langages pour chaque cas d'application ou pour chaque compagnie, suivant le contenu existant de chaque application à porter et les fonctionnalités existantes et souhaitées. Lors d'un raffinement des énoncés, on n'obtient pas nécessairement une partition, car certains énoncés peuvent être à la fois dans deux « sous-sous-langages ».

Par exemple, le sous-langage des documents de Microsoft peut être décomposé en deux sous-langages, celui des aides en ligne et celui des manuels de référence.

De même, les énoncés traités par le système CATS seront subdivisés en quatre catégories de sous-langages : occasion automobile, immobilier, emplois, divers. Une étude de leur multilinguïsation, du portage linguistique et de la localisation résumée dans la grille de l'Annexe I.

Nous remplissons cette classification au fur et mesure du traitement de chaque nouveau type de sous-langage en appliquant notre taxonomie.

Chapitre 5 Évaluation des représentations de contenu

Introduction

Les systèmes de e-commerce utilisent différentes représentations de contenu. Les représentations de contenu possibles sont soit classiques (vecteurs, listes de propriétés, structures de traits...), soit spécifiques (comme celle de CATS (Daoud 2006), proches de graphes sémantiques à la Sowa (Sowa 1984) ou de graphes UNL (Uchida and Zhu 2003)).

Nous nous intéressons ici à l'évaluation d'un portage linguistique au niveau de la représentation de contenu (RC) obtenue à partir d'un énoncé dans une nouvelle langue. Pour cela, nous devons comparer cette RC avec celle obtenue à partir d'un énoncé équivalent dans la langue native de l'application, et éventuellement avec une RC de référence idéale.

Nous présentons les types de représentation de contenu intéressants dans notre cas. Nous décrivons quelques algorithmes de comparaison (calcul de distances et/ou de similarités) entre deux instances de représentations de contenu. Enfin, nous décrivons un cas d'utilisation expérimentale de la comparaison entre arbres pondérés par des poids dans un système de e-commerce. Une idée fructueuse est de passer d'un type de RC à un autre, pour pouvoir utiliser des algorithmes de comparaison plus performants. C'est le cas de CRL-CATS (Daoud 2006), dont les RC (pseudo-graphes UNL) peuvent être transformées en structures de traits et comparées sous cette forme.

5.1 Représentations de contenu

Voici d'abord un exemple de l'utilisation du niveau « contenu » dans CATS (Tableau 20). L'idée est de transformer tous les SMS synonymes (du point de vue de l'application) en une même RC.

Énoncés	Représentation de contenu CRL-CATS
Recherche Kia année 2000 couleur noire.	<pre>[S] wan(saloon:00, wanted:00) mak(saloon:00, KIA(country>korea):0A) yea(saloon:00, 2000:0L) col(saloon:00, black:0Y) [/S]</pre>
Recherche voiture Kia année 2000 couleur noire.	
Recherche Kia modèle 2000 couleur noire.	
Recherche Kia modèle 2000 noire.	
Je cherche une voiture Kia année 2000 noire.	
Je suis à la recherche d'une voiture Kia modèle 2000 noire.	

Tableau 20 : normalisation d'énoncés de même contenu

Il existe plusieurs types de représentations de contenu comme :

- Liste de propriétés (couples <attribut, valeur>),
- Vecteurs (k attributs à valeurs numériques),
- Structures de traits (typés ou non), attributs sur les arcs, valeurs élémentaires sur les feuilles,
- Graphes conceptuels ou sémantiques (exemple : graphe de Sowa),

- Arbres (attributs et valeurs sur les nœuds),
- CRL-CATS (pseudo-graphe UNL),
- Représentation IF « Interchange Format ».

Nous détaillons d'abord les représentations de contenu classiques. Ensuite, nous présentons quelques représentations spécifiques qui ont été utilisées comme « pivot » en traduction automatique comme la représentation IF ou sont dérivées d'une représentation pivot comme CRL-CATS qui est dérivée du pivot UNL.

5.1.1 Représentations classiques

5.1.1.1 Liste de propriétés

Une liste de propriétés est une structure de données qui regroupe des informations sous forme de couples (attribut, valeur) associés à un objet. Dans l'exemple de la Figure 39, les couples (*sal*, *sale*), (*make*, *HONDA*), ... sont associés à l'objet *saloon*.

Ainsi, l'exemple « *A vendre Honda modèle 1997 boîte automatique climatisée fermeture centrale prix 7750 dinars* » peut être représenté sous la forme de la liste de propriétés de la Figure 39.

```
saloon (
  (sal          sale)
  (make        HONDA)
  (country     Japan)
  (yea        1997)
  (gear       automatic)
  (air-condition yes)
  (center lock yes)
  (price      7750)
)
```

Figure 39 : exemple de représentation sous forme de liste de propriétés

5.1.1.2 Représentation vectorielle

Les vecteurs dont il est question ici sont des listes ordonnées de nombres, de longueur fixe. Ainsi, nous représentons le même exemple « *A vendre Honda modèle 1997 boîte automatique climatisée fermeture centrale prix 7750 dinars* » par la représentation vectorielle de la Figure 40 .

Sal	Year	Price	Make	Country	Center lock	Automatic gear	Air condition
1	1997	7750	2	2	1	1	1

Figure 40 : exemple de représentation vectorielle

5.1.1.3 Structure de traits

La représentation par structures de traits est fortement présente dans le domaine linguistique (phonologie, morpho-syntaxe, lexicque, ...). Elle forme un mécanisme générique pour la représentation d'informations linguistiques.

Les structures de traits décrivent chaque élément d'une phrase en énumérant ses caractéristiques linguistiques, syntaxiques ou sémantiques, sous la forme de listes de couples trait-valeur. Un trait peut avoir pour valeur un ou plusieurs couples trait-valeur. Une structure de traits peut être réentrante.

La définition très complète (<http://www.loria.fr/~azim/FS/presentation.html>) est la suivante :

Une structure de traits est un ensemble $[T_1, \dots, T_n]$ où chaque T_i est :

- une paire ($t : n$) où t est le nom du trait et n est la valeur associée au trait ;
- une boîte de coindexation, qui permet de faire partager la même valeur à deux traits ;

- un ensemble de structures de traits appelé disjonction généralisée.

La valeur v d'un trait peut, quant à elle, être :

- atomique ;
- un ensemble de valeurs appelé disjonction sur valeurs ;
- une structure de traits.

L'unification de deux structures de traits consiste à combiner leurs informations et permet de vérifier leur compatibilité. Si les deux structures de traits ont les mêmes valeurs pour des traits identiques, alors elles peuvent s'unifier. Si deux traits identiques n'ont pas la même valeur dans les deux structures de traits, alors l'unification échoue. Si un trait n'est présent que dans une des deux structures, cela n'influe pas sur l'unification, qui échoue dans le cas de valeurs contradictoires pour un même trait, mais pas dans le cas d'absence d'un trait.

L'exemple précédent peut être représenté sous forme de la structure de traits de la Figure 41.

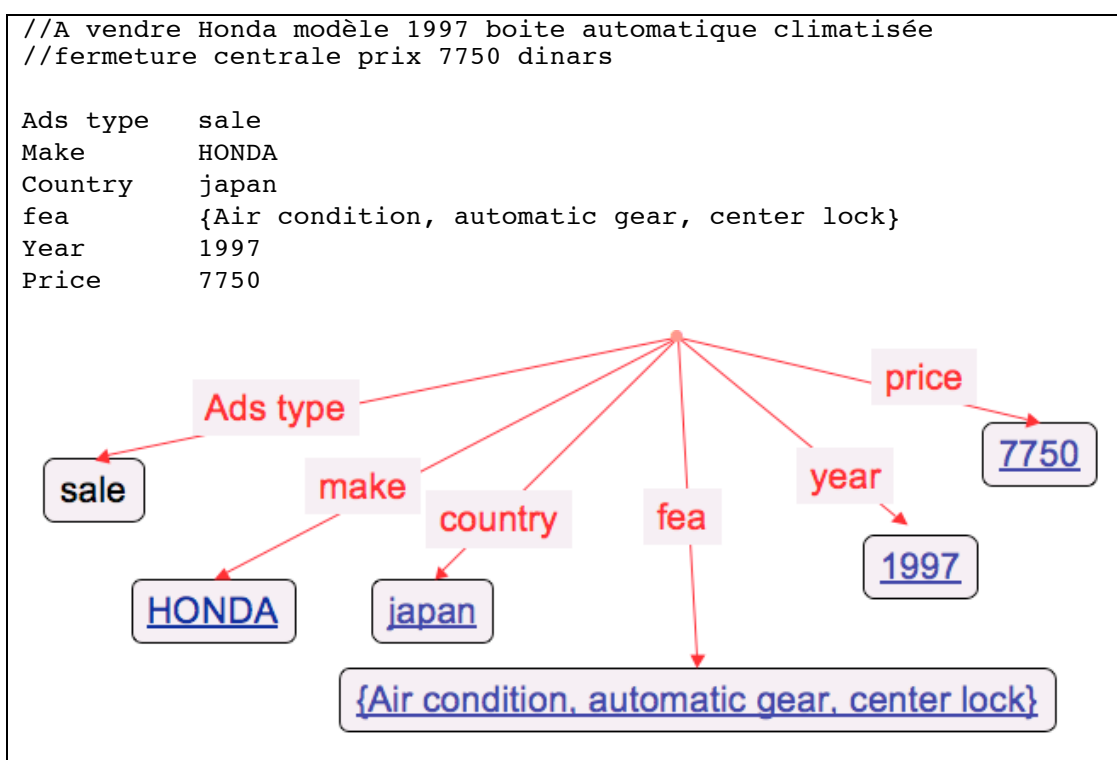


Figure 41 : exemple de représentation sous forme de structure de traits avec disjonction

5.1.1.4 Graphe conceptuel

La théorie des graphes conceptuels a été développée par (Sowa 1984), c'est un formalisme général fondé sur la logique pour représenter les connaissances.

Les graphes conceptuels ont été conçus pour représenter la sémantique du langage naturel ; ils ont évolué pour devenir des systèmes complets au sens de la logique du premier ordre. De façon générale, un graphe conceptuel est défini comme un graphe qui a deux sortes de nœuds :

- les nœuds concepts qui représentent des entités, des attributs, des états, des événements...
- les nœuds relations conceptuelles qui symbolisent les liens qui existent entre deux concepts.

[CONCEPT] (RELATION) [CONCEPT]

Les concepts sont représentés graphiquement par des boîtes ou entre crochets [Concept], et les nœuds relations conceptuelles sont représentés entre parenthèses avec un seul arc entrant et un seul arc sortant (RELATION CONCEPTUELLE) ou par des cercles.

Un arc entrant relie un concept à une relation conceptuelle, et un arc sortant relie une relation conceptuelle à un concept. Ainsi, un graphe conceptuel est :

- orienté,
- fini : tout graphe dans une mémoire d'un ordinateur ne peut avoir qu'un nombre fini de nœuds,
- connexe : si deux parties n'étaient pas connectées entre elles on aurait deux graphes conceptuels,
- bipartite : il ne possède que deux sortes de nœuds (les concepts et les relations conceptuelles) chaque arc reliant une sorte de nœud à l'autre sorte de nœud.

Selon (Sowa 1984), un graphe conceptuel se représente comme suit :

$GC = (S, A)$ un graphe conceptuel est constitué de sommets et d'arcs.

$S = (C, R)$ les sommets sont partitionnés en deux ensembles : C et R

$C = \{c1, \dots, cn\}$ ensemble de nœuds concepts

$R = \{r1, \dots, rm\}$ ensemble de relations conceptuelles

$A = (AE, AS)$ les arcs sont partitionnés en deux ensembles : AE et AS

$AE = \{e1, \dots, em\}$ où $e_i = \langle c_j, r_i \rangle$ avec $i \in [1 \dots m]$ et $j \in [1 \dots n]$ arcs entrants

$AS = \{s1, \dots, sm\}$ où $s_i = \langle r_i, c_k \rangle$ avec $i \in [1 \dots m]$ et $k \in [1 \dots n]$ arcs sortants.

Une représentation sous forme de graphe conceptuel du même exemple peut être de la forme montrée par la Figure 42.

```
//A vendre Honda modèle 1997 boîte automatique climatisée fermeture
//centrale prix 7750 dinars
```

```
[sale] (ads type) [saloon]
[Saloon] (Make) [Honda]
[Honda] (Country) [japan]
[Saloon] (Year) [1997]
[Saloon] (Price) [7750]
[Saloon] (have) [Gear]
[Gear] (is) [automatic]
[Saloon] (Lock) [Center lock]
[Saloon] (Air) [air condition]
```

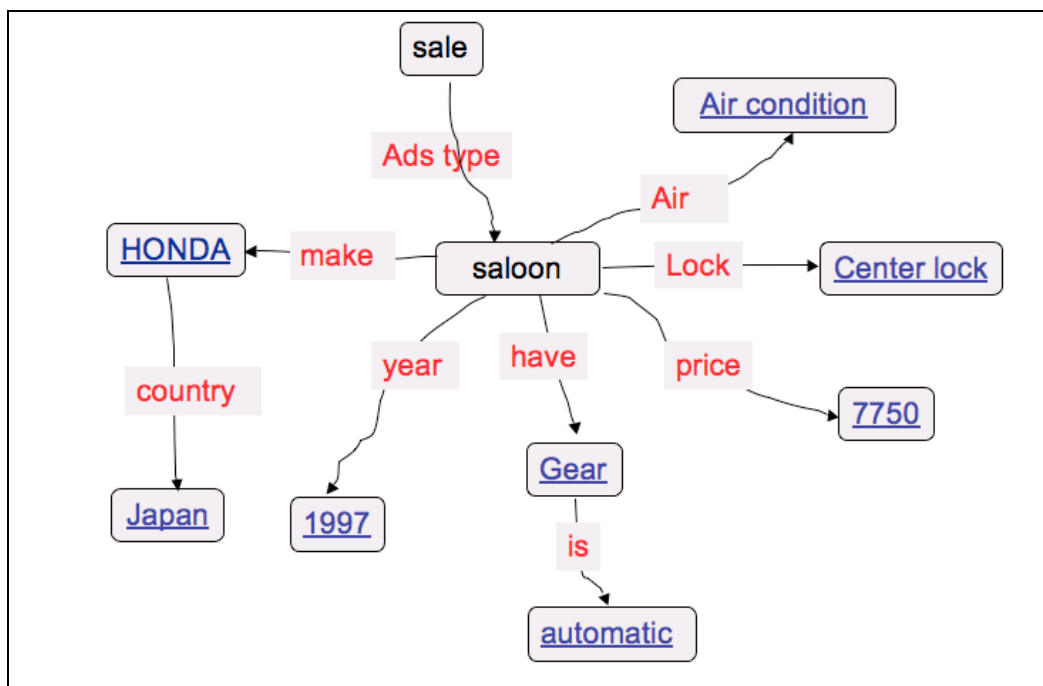


Figure 42 : exemple de représentation sous forme de graphe conceptuel

5.1.1.5 Arbres étiquetés (orientés non ordonnés)

Nous pouvons aussi représenter l'exemple précédent sous forme de l'arbre étiqueté de la Figure 43.

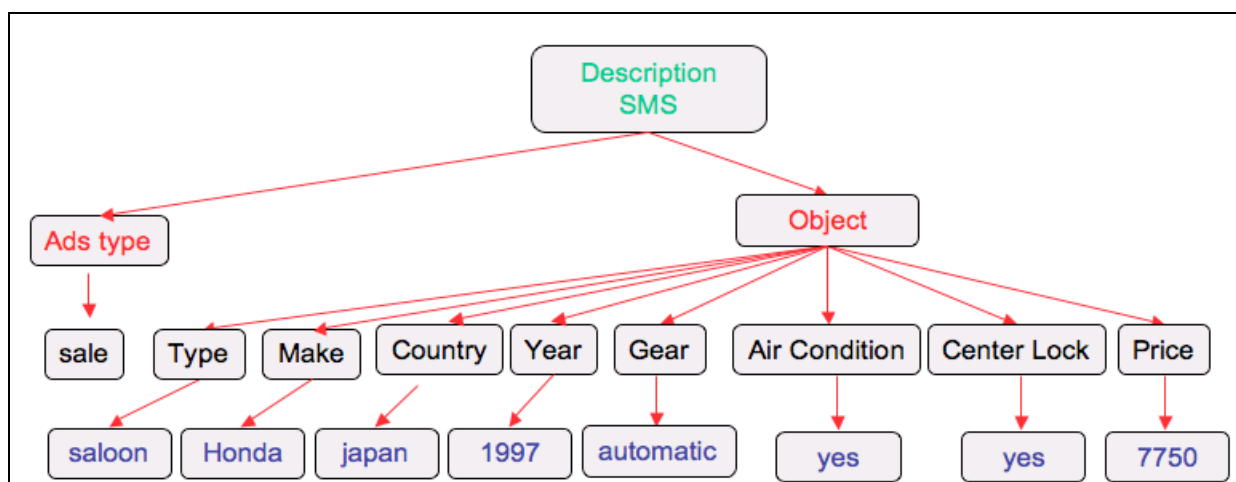


Figure 43 : exemple de représentation sous forme d'arbre

5.1.2 Représentations spécifiques

5.1.2.1 Définition d'un pivot

En TA, un « pivot sémantique » est une représentation logico-sémantique profonde qui représente un énoncé par un ensemble de prédicats munis de leurs arguments et reliés entre eux par des méta-prédicats (Vauquois, Veillon et al. 1969).

D'après (Boitet 2007), un « pivot » en TA est une représentation intermédiaire par laquelle passe toute traduction d'une langue vers une autre. C'est donc une notion purement architecturale,

rien n'est dit sur la nature dudit pivot. On veut seulement que le système, s'il est multilingue, ait une structure « en étoile ».

Nous pouvons considérer que les RC sont des « pivots sémantiques » dans les applications de e-commerce, une fois multilinguées. Ainsi, le résultat de la traduction ne dépend pas de la langue source. Ce pivot est « hybride », parce que d'un côté il a ses propres descriptions grammaticales universelles mais d'un autre côté il emploie le vocabulaire d'une langue naturelle (plus précisément les « unités lexicales » ou familles dérivationnelles) pour exprimer les concepts. Donc, dans l'application à la traduction automatique, il y a seulement un transfert lexical entre deux langues naturelles, réalisé en consultant un dictionnaire bilingue.

Dans le cas d'un dialogue entre un système de e-commerce et un utilisateur, il est même possible qu'une RC représente l'information ou la requête précise par une partie du dialogue en question.

5.1.2.2 CRL-CATS

CATS (Daoud 2006) utilise une représentation de contenu appelée ici « CRL-CATS ». Elle est syntaxiquement semblable à UNL, mais ne correspond pas du tout à la représentation UNL standard, liée à une expression linguistique en anglais (même si elle en est une représentation profonde). Certains exemples simples (voir Figure 17) pourraient faire croire qu'une RC en CRL-CATS peut se ramener à une liste de propriétés. Mais comme le montre la Figure 44, sa définition comme liste de triplets *propriété (objet, valeur)* fait qu'il peut y avoir deux attributs (propriétés) de même nom (par exemple *fea*) pour un même objet du domaine (par exemple *saloon*) ayant des valeurs différentes. Il faudrait donc comprendre ici *fea* comme un attribut avec une valeur ensembliste.

Dans l'exemple suivant, l'ensemble des valeurs de l'attribut *fea* est {*automatic gear, air condition, center lock*}. Moyennant cela, on peut donc transformer une RC CRL-CATS quelconque en une liste de propriétés.

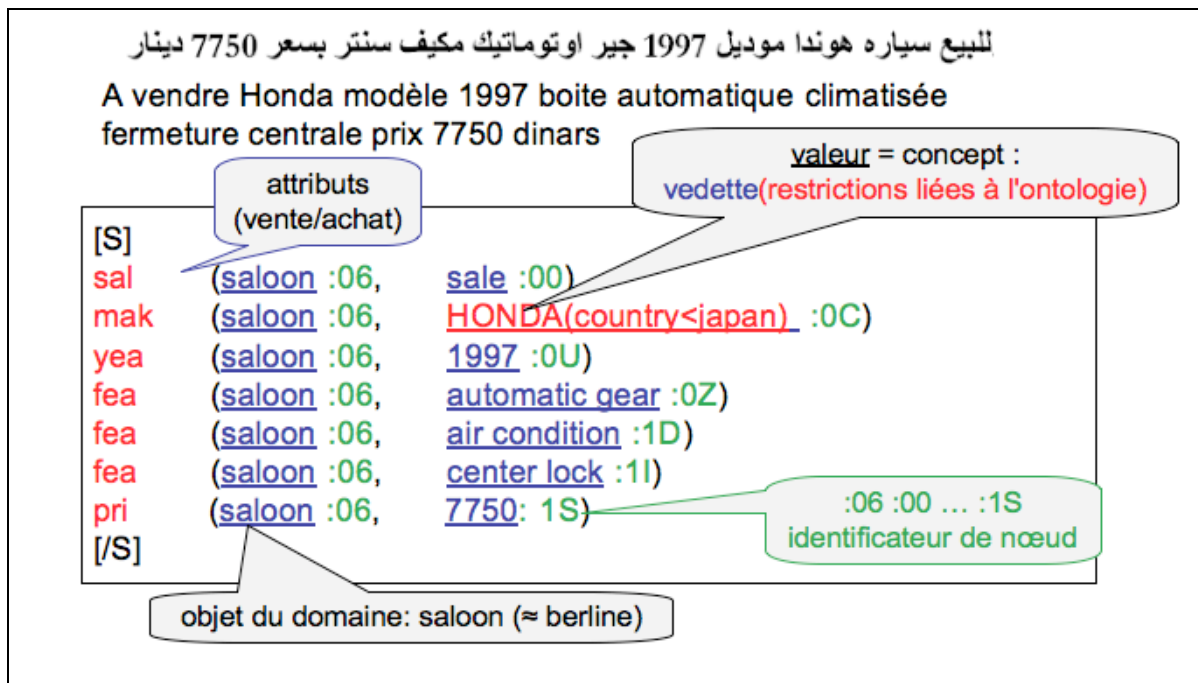


Figure 44 : exemple de représentation CRL-CATS (1)

L'écriture linéaire RC en CRL-CATS est composée du texte de la phrase en entrée placé en commentaire, suivi du symbole [S] indiquant le début l'analyse de la phrase et fini par [/S], indiquant la fin de l'analyse de la phrase en question.

La représentation précédente peut être dessinée sous forme de graphe, comme le montre la Figure 45.

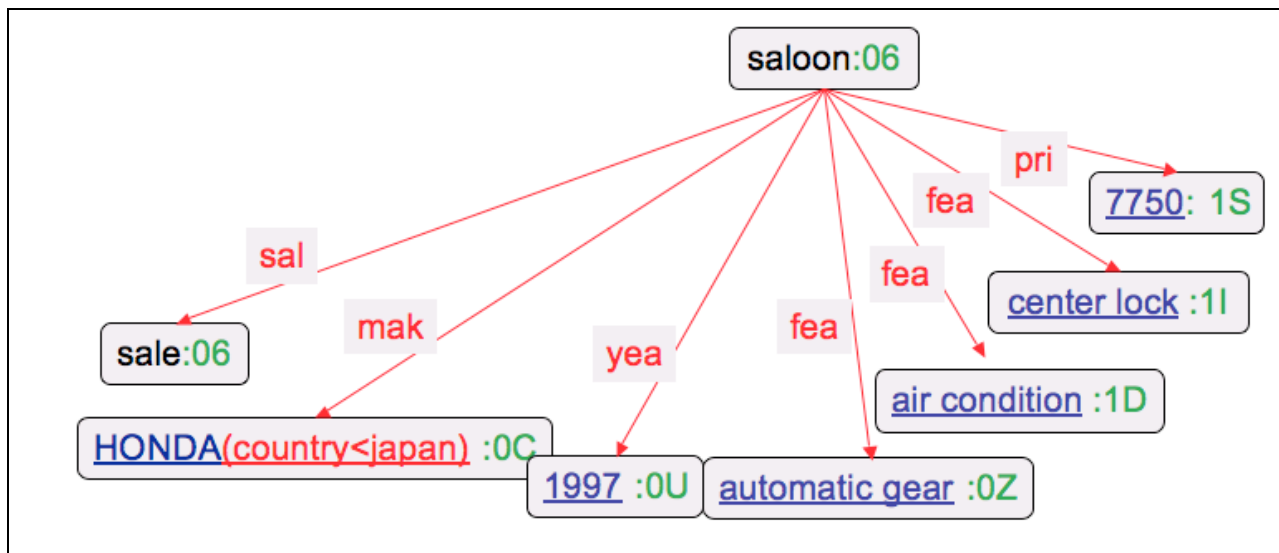


Figure 45 : exemple de représentation CRL-CATS (2)

5.1.2.3 IF « Interchange Format »

L'IF est le langage pivot utilisé dans les projets C-STAR et Nespole! (Blanchon 2004), avec de petites modifications dans chaque projet. Proposé par Hans-Ulrich Block de Siemens, l'IF a été adopté par le projet C-STAR II en mai 1997, et a été développé pendant les deux années suivantes, pour la démonstration internationale et quintilingue (anglais, allemand, coréen, français, et italien) du 22/07/1999. Il a ensuite été exploité et très souvent modifié par le projet européen Nespole! (Besacier, Blanchon et al. 2001).

L'IF est un pivot sémantico-pragmatique utilisé pour des domaines restreints. Dans le projet Nespole!, le passage par le pivot IF utilise une méthode basée sur les automates de reconnaissance de séquences pertinentes. Nous détaillons la spécification de cette représentation dans la section 9.1.1.

La représentation en IF de l'exemple précédent, adaptée au domaine de l'automobile peut, être la suivante :

```
; A vendre Honda modèle 1997 boîte automatique climatisée
fermeture centrale prix 7750 dinars
```

```
IF = {a:give-information+concept(action=e_sell, vehicle-spec=(car, vehicle-
make=Honda, vehicle-nationality=japan, vehicle-age=1997, vehicle-
gear=automatic, vehicle-air-condition=air_condition, vehicle-
center_lock=center_lock, vehicle-price=7750))}
```

Figure 46 : exemple de représentation IF

Pour l'évaluation du portage linguistique de CATS, nous aurons besoin de comparer deux représentations CRL-CATS. Nous cherchons, dans ce qui suit, à appliquer ou à adapter des algorithmes de comparaison entre arbres. Pour cela, regardons les possibilités de calcul de distance et/ou de similarité entre deux représentations sous forme d'arbre.

5.2 Comparaison entre deux CRL-CATS

Les techniques de mesure de similarité entre arbres sont utiles pour plusieurs applications comme la reconnaissance des formes, l'analyse et le traitement de l'image, le TALN (Kamat 1996) et la bioinformatique. Les précédents travaux ont traité des arbres étiquetés ordonnés (Wang, Shapiro et al. 1998; Shasha, Wang et al. 2001) ou non ordonnés (Shasha, Wang et al. 1994).

Les opérations incluant l'insertion, la suppression, et la substitution (Lu 1979) avec l'affectation de coûts sont définies pour transformer un arbre en un autre et pour définir une distance ou une similarité. Pour la correspondance locale entre arbres (Liu and Geiger 1999), des opérations comme *merge*, *cut*, et *merge-and-cut* incluant des coûts ont également été définies pour trouver la meilleure correspondance approximative et le coût associé. La distance de Hamming est également utilisée dans certaines approches (Schindler, Rothlauf et al. 2002) pour calculer la distance entre deux arbres.

Comme le montre la Figure 47, un document peut être représenté non seulement sous forme de chaînes de caractères dans un fichier mais aussi sous forme d'une organisation hiérarchique de textes et de nœuds. Les parties d'un document sont des nœuds étiquetés et tous les nœuds terminaux sont des textes.

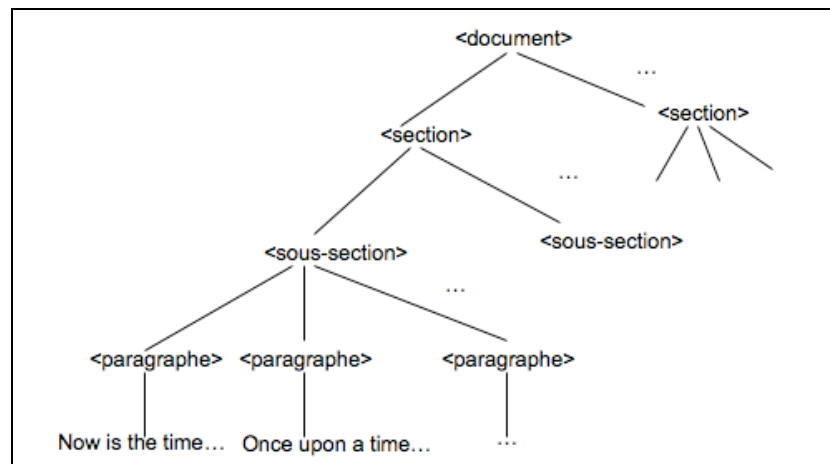


Figure 47 : représentation d'un texte sous forme d'arbre

Trouver le minimum de différences entre deux documents représentés par deux arbres, est un problème similaire à celui de trouver les différences entre chaînes de caractères (Wagner and Fischer 1974) ou entre arbres (Tai 1979).

La définition d'une distance d'édition entre deux arbres A1 et A2 est le coût minimal d'une suite de transformations permettant de transformer A1 en A2. On considère les transformations suivantes :

- Insertion d'arbre : ajouter un sous-arbre entier (qui peut être un nœud terminal) à l'arbre courant.
- Suppression d'arbre : supprimer un sous-arbre entier (qui peut être un nœud terminal) de l'arbre courant.
- Insertion de nœud : ajouter un nœud interne à l'arbre courant.
- Suppression de nœud : supprimer un nœud interne de l'arbre courant.
- Remplacement : changer l'étiquette du nœud de l'arbre courant par l'étiquette du nœud dans l'arbre cible.
- Échange de sous-arbres : échange de sous-arbres enracinés adjacents dans l'arbre courant.

À chaque opération est affecté un coût variable. Par exemple, le coût de la suppression d'un nœud interne pourra dépendre du nombre de fils de ce dernier.

Nous détaillons, dans ce qui suit, quelques algorithmes de calcul de distance entre arbres. Ils sont basés sur l'algorithme de mesure de distance entre chaînes de caractères que nous rappelons dans le paragraphe suivant.

5.2.1 Algorithme de Wagner et Fischer

Wagner et Fischer (Wagner and Fischer 1974) ont cherché à résoudre deux problèmes : la détermination du coût minimum d'opérations et la production d'une liste d'opérations qui permettent de transformer une chaîne de caractères originale S en une chaîne cible S' en utilisant les opérations d'insertion, de suppression, et de substitution d'un caractère.

Les opérations d'édition sont notées de la façon suivante (la notation est ici généralisée d'avance aux arbres étiquetés pour l'utiliser dans la section 5.2.2) :

$S_i \rightarrow S_j'$: substitution du $i^{\text{ième}}$ caractère dans S par le $j^{\text{ième}}$ caractère dans S' .

ϵ : une chaîne vide ou un arbre vide.

$a \rightarrow \epsilon$: suppression du caractère a .

$\epsilon \rightarrow b$: insertion du caractère b .

$\gamma(a \rightarrow b)$: coût de l'opération d'édition $a \rightarrow b$.

n : longueur de la chaîne originale S , ou nombre total de nœuds dans l'arbre original T

m : longueur de la chaîne cible S' , ou nombre total de nœuds dans l'arbre cible T'

D : matrice des distances d'édition entre les préfixes des chaînes S et S' ou entre les préfixes des arbres T et T' (écrits linéairement).

$D[i,j]$: distance d'édition cumulée entre $S_{1\dots i}$ et $S'_{1\dots j}$ ou entre $T_{1\dots i}$ et $T'_{1\dots j}$

$||$: opérateur de concaténation

L'algorithme de Wagner et Fischer est le suivant :

```

D[0,0] := 0
Pour i de 1 à n faire
  D[i,0] := D[i-1,0] +  $\gamma(S_i \rightarrow \epsilon)$  //distance entre les préfixes de S et  $\epsilon$ 
Fin faire
Pour j de 1 à m faire
  D[0,j] := D[0,j-1] +  $\gamma(\epsilon \rightarrow S'_j)$  //distance entre les préfixes de S' et  $\epsilon$ 
Fin faire
Pour i de 1 à n faire //calcul des D[i,j] pour i,j>0 par
  Pour j de 1 à m faire // programmation dynamique
    D[i,j] := min (D[i-1,j-1] +  $\gamma(S_i \rightarrow S'_j)$ ,
                  D[i-1,j] +  $\gamma(S_i \rightarrow \epsilon)$ ,
                  D[i,j-1] +  $\gamma(\epsilon \rightarrow S'_j)$ )
  Fin faire
Fin faire

```

Figure 48 : algorithme de Wagner et Fischer

La distance d'édition minimale entre S et S' est égale à $D[n,m]$.

L'algorithme de calcul de trace est le suivant :


```

i := n ; j := m ; L := nul ;
Tant que i > 0 et j > 0 faire
  Si D[i,j] = D[i-1,j] +  $\gamma$ (Si $\rightarrow$  $\epsilon$ ) alors
    L := L || (Si $\rightarrow$  $\epsilon$ ) ; // Suppression
    i := i-1 ;
  Fin Si
  Sinon si D[i,j] = D[i,j-1] +  $\gamma$ ( $\epsilon$  $\rightarrow$ S'j) alors
    L := L || ( $\epsilon$  $\rightarrow$ S'j) ; // Insertion
    j := j-1 ;
  Fin si
  Sinon
    L := L || (Si $\rightarrow$ S'j) ; // Substitution
    i := i-1 ;
    j := j-1 ;
  Fin faire

```

Figure 49 : calcul de trace à partir de l'algorithme de Wagner et Fischer

Dans le cadre du TER de (Pineau 2004), nous avons implémenté cet algorithme, ainsi qu'une extension et plusieurs visualisations de la trace. L'extension consiste à calculer une distance mixte, combinaison linéaire de la distance en caractères et de la distance en mots. Pour cette dernière, les mots $m_1 \dots m_k$ sont traités comme des caractères, avec $\gamma(m \rightarrow \epsilon) = D[m, \epsilon]$; $\gamma(\epsilon \rightarrow m) = D[\epsilon, m]$; $\gamma(m \rightarrow m') = D[m, m']$.

La visualisation de la trace est analogue à celle du mode « suivi de modifications » de Word.

Exemple 1 : mode « suivi de modifications » du calcul de distance entre SMS1 et SMS2

SMS1 = Recherche voiture TOYOTA Cresseda modèle 95

SMS2 = Recherche TOYOTA Cresseda mod 95

Représentation en suivi des modifications

Représentation : Recherche ~~voiture~~ TOYOTA Cresseda **modèle** 95

Préférence

	Coût
Insertion	1
Suppression :	1
Changement :	1
Egalité :	0

Coût uniforme

distance de : 11

Figure 50 : algorithme de Wagner et Fisher (exemple 1)

Exemple 2 : mode « 3 lignes synchronisées » du calcul de distance entre SMS3 et SMS4

SMS3 = Recherche voiture MITSHUBITSHI modèle Lancer

SMS4 = Recherche MITSHUBITSHI Lancer

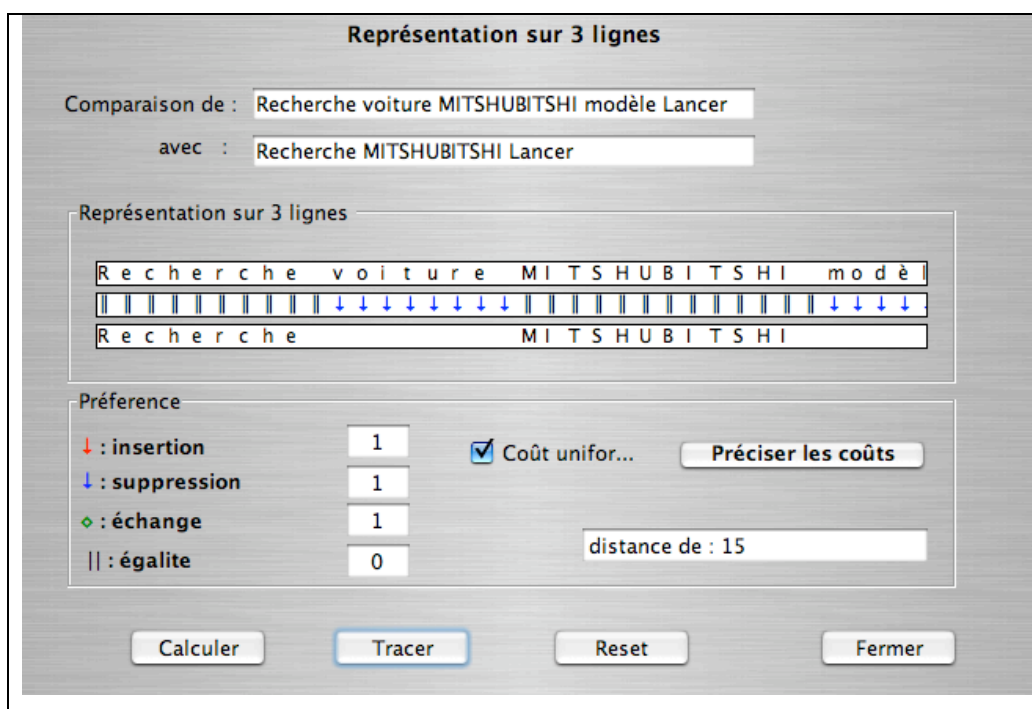


Figure 51 : algorithme de Wagner et Fisher (exemple 2)

5.2.2 Algorithme de Selkow

(Selkow 1977) a étendu la distance entre chaînes de caractères pour trouver une distance d'édition entre arbres. Sa solution utilise les opérations basiques de changement d'étiquette, de suppression et d'insertion. La suppression et l'insertion de nœuds peuvent être faites, sauf dans le cas des nœuds non-terminaux. Les notations utilisées sont les suivantes.

c : nombre des sous-arbre de racine (T). On note $T = r(T_1, \dots, T_c)$, avec $r = \text{racine}(T)$. Si $c = 0$, alors T est réduit à un nœud.

c' : idem pour T' , en notant $T' = r'(T'_1, \dots, T'_{c'})$.

T_i : le $i^{\text{ème}}$ sous-arbre de père $r = \text{racine}(T)$ (compté à partir de la gauche)

$\lambda(T)$: étiquette (racine (T)) = étiquette (r)

$\text{ins}(T'_x)$: coût d'insertion du sous-arbre de T' enraciné en x

$\text{del}(T_x)$: coût de suppression du sous-arbre de T enraciné en x

CoûtType : type prédéfini (probablement entier)

$D[i, j]$: distance d'édition cumulée entre $T_{1 \dots i}$ et $T'_{1 \dots j}$

```

edit(T,T') : CoûtType
D[0,0] :=  $\gamma(\lambda(T) \rightarrow \lambda(T'))$ 
Pour i de 1 à c faire // parcourir les sous-arbres de T fils de r
  D[i,0] := D[i-1,0] + del(Ti)
Fin faire
Pour j de 1 à c' faire // parcourir les sous-arbres de T' fils de r'
  D[0,j] := D[0,j-1] + ins(T'j)
Fin faire
Pour i de 1 à c faire // parcourir les sous-arbres de T fils de r
  Pour j de 1 à c' faire // parcourir les sous-arbres de T' fils de r'
    D[i,j] := min (D[i-1,j-1] + edit(Ti,T'j), // appel récursif
                  D[i-1,j] + del(Ti),
                  D[i,j-1] + ins(T'j))
  Fin faire
Fin faire
Retourner (D[c,c'])

```

Figure 52 : algorithme de Selkow pour le calcul de la distance entre deux arbres ordonnés, orientés et étiquetés

Nous avons implémenté l'algorithme de Selkow dans le cadre du TER de (Zair 2008), et d'une comparaison de deux représentations CRL-CATS (voir section 6.2.2.3). Puisque cet algorithme traite seulement des arbres étiquetés, nous avons calculé la distance entre deux représentations CRL-CATS en intégrant les noms des attributs dans les étiquettes des nœuds (fils). Nous avons alors un attribut et sa valeur dans le même nœud attaché à la racine.

Nous avons trié les nœuds de gauche à droite par les noms d'attribut puis par sa valeur. Autrement dit, nous avons trié les nœuds de gauche à droite suivant le maximum commun en terme d'attributs et valeurs. La Figure 53 montre un exemple de tri de deux CRL-CATS.

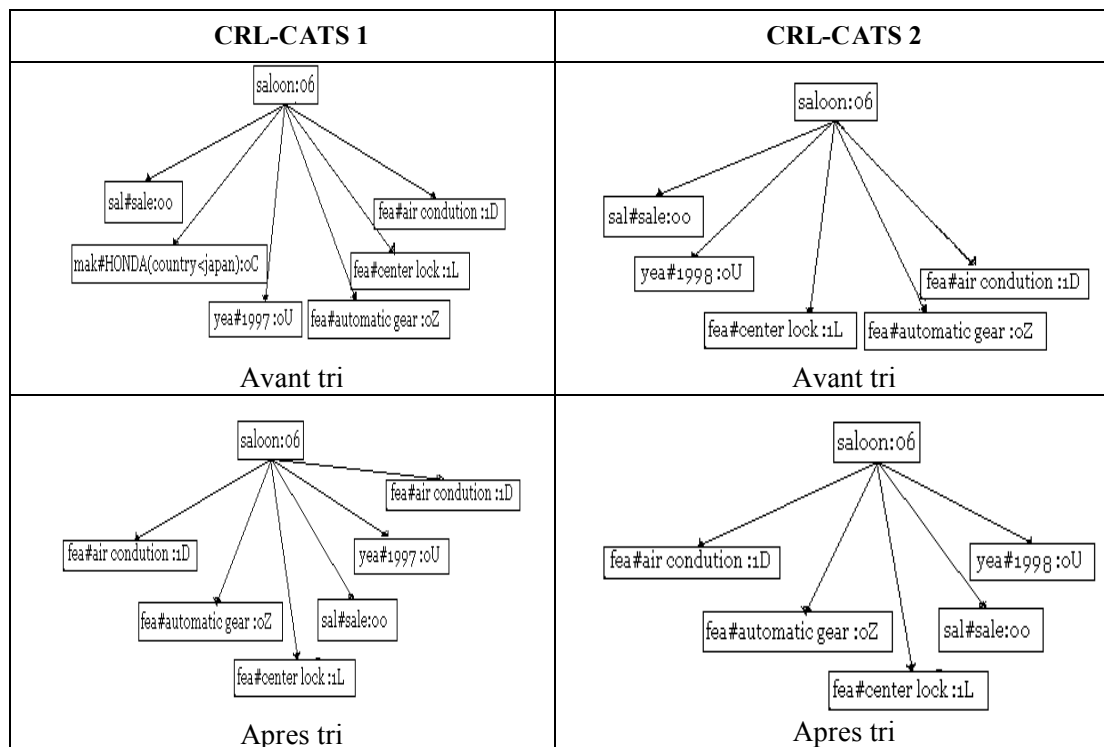


Figure 53 : exemple de tri de deux CRL-CATS

L'algorithme utilise trois opérations (suppression, insertion, substitution). Nous avons pris un coût égal à 1 pour chacune des opérations, et nous utilisons une fonction *edit* qui compare deux nœuds n_1 et n_2 .

Soit : $n_1 = \text{attribut}_1\#\text{valeur}_1$ et $n_2 = \text{attribut}_2\#\text{valeur}_2$

Nous utilisons les règles de calcul suivantes :

Si $\text{attribut}_1 = \text{attribut}_2$ et $\text{valeur}_1 = \text{valeur}_2$ alors $\text{edit} = 0$

Si $\text{attribut}_1 = \text{attribut}_2$ et $\text{valeur}_1 \neq \text{valeur}_2$ alors $\text{edit} = \frac{1}{2}$

Si $\text{attribut}_1 \neq \text{attribut}_2$ et $\text{valeur}_1 \neq \text{valeur}_2$ alors $\text{edit} = 1$

Si $\text{attribut}_1 \neq \text{attribut}_2$ et $\text{valeur}_1 = \text{valeur}_2$ alors $\text{edit} = \frac{1}{2}$

La Figure 54 détaille le calcul de distance entre les deux CRL-CATS de la Figure 53.

	sal#sale	fea#air condition	fea#automatic gear	fea#center lock	yea#1997	mak#HONDA(country>japan)
sal#sale	0	1	2	3	4	5
fea#air condition	1	0	1	2	3	4
fea#automatic gear	2	1	0	1	2	3
fea#center lock	3	2	1	0	1	2
yea#1998	4	3	2	1	0.5	1.5

— insertion
— suppression
— substitution

$$\text{Min} [2 + \text{coût}(\text{ins}), 0.5 + \text{coût}(\text{supp}), 1 + \text{edit}(\text{yea}\#1998, \text{mak}\#HONDA(\text{country}\>\text{japan}))]$$

Figure 54 : exemple détaillé de l'application de l'algorithme de Selkow

Nous illustrons ce calcul par d'autres exemples de CRL-CATS, résultats de deux extracteurs de contenu (le premier pour l'arabe et le deuxième pour le français) :

مطلوب سيارة تويوتا كريسيدا موديل 95	;Recherche voiture TOYOTA Cresseda modèle 95																
[S] wan(saloon:06, wanted:00) mak(saloon:06, TOYOTA(country>japan):0C) yea(saloon:06, 95:0X) [/S]	[S] wan(saloon:0A, wanted:00) mak(saloon:0A, TOYOTA(country>japan):0I) yea(saloon:0A, 95:15) [/S]																
	<table border="1"> <thead> <tr> <th></th> <th>mak#TOYOTA(country>japan)</th> <th>wan#wanted</th> <th>yea#95</th> </tr> </thead> <tbody> <tr> <th>mak#TOYOTA(country>japan)</th> <td>0</td> <td>1</td> <td>2</td> </tr> <tr> <th>wan#wanted</th> <td>1</td> <td>0</td> <td>1</td> </tr> <tr> <th>yea#95</th> <td>2</td> <td>1</td> <td>0</td> </tr> </tbody> </table>		mak#TOYOTA(country>japan)	wan#wanted	yea#95	mak#TOYOTA(country>japan)	0	1	2	wan#wanted	1	0	1	yea#95	2	1	0
	mak#TOYOTA(country>japan)	wan#wanted	yea#95														
mak#TOYOTA(country>japan)	0	1	2														
wan#wanted	1	0	1														
yea#95	2	1	0														
Distance = 0																	

Tableau 21 : exemple de calcul de distance nulle entre deux RC de CATS

Si les deux CRL-CATS (du Tableau 21) n'étaient pas dans le même ordre (absence de tri), par exemple si mak#TOYOTA(country>japan) vient après yea#95 (venant du SMS français), l'algorithme donnerait une distance non nulle.

مطلوب سيارة ميتسوبيشي لانسر	;Recherche voiture MITSHUBITSHI Lancer												
[S] wan(saloon:06, wanted:00) mak(saloon:06, MITSUBISHI(country>japan):0C) mod(saloon:06, Lancer(country>japan,make>MITSUBISHI): 0N) [/S]	[S] wan(saloon:0A, wanted:00) mod(saloon:0A, Lancer(country>japan,make>MITSUBISHI): 0V) [/S]												
	<table border="1"> <tr> <td></td> <td>mod#Lancer(country>japan,make>MITSUBISHI)</td> <td>wan#wanted</td> </tr> <tr> <td>mod#Lancer(country>japan,make>MITSUBISHI)</td> <td>0</td> <td>1</td> </tr> <tr> <td>wan#wanted</td> <td>1</td> <td>0</td> </tr> <tr> <td>mak#MITSUBISHI(country>japan)</td> <td>2</td> <td>1</td> </tr> </table>		mod#Lancer(country>japan,make>MITSUBISHI)	wan#wanted	mod#Lancer(country>japan,make>MITSUBISHI)	0	1	wan#wanted	1	0	mak#MITSUBISHI(country>japan)	2	1
	mod#Lancer(country>japan,make>MITSUBISHI)	wan#wanted											
mod#Lancer(country>japan,make>MITSUBISHI)	0	1											
wan#wanted	1	0											
mak#MITSUBISHI(country>japan)	2	1											
Distance = 1													

Tableau 22 : exemple de calcul de distance non nulle entre deux RC de CATS

5.3 Autre mesure

Nous présentons un algorithme de calcul de la similarité entre arbres pour l'appariement des agents dans les environnements de commerce électronique (Bhavsar, Boley et al. 2004). Cette mesure a été développée pour des raisons de différences de représentations internes par rapport à celles qui la précèdent. Les descriptions des produits et des services offerts par les agents vendeurs et acheteurs sont représentés par des arbres à nœuds étiquetés, et à arcs étiquetés et pondérés. Nous décrivons un algorithme de calcul de similarité entre de tels arbres, utilisé pour l'appariement sémantique dans une place de marché virtuelle.

5.3.1 Arbres sémantiques

La description précise des produits ou services est traduite par l'affectation des poids aux attributs de l'arbre sémantique correspondant suivant leurs importances.

Comme le montre la figure suivante, un schéma standard, spécifique à chaque application (ici, c'est e-Learning), est utilisé pour décrire les besoins.

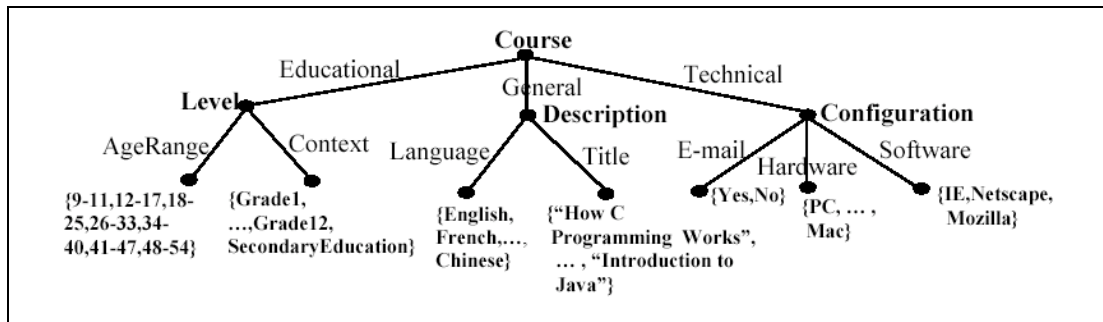


Figure 55 : schéma standard pour e-Learning (Bhavsar, Boley et al. 2004)

Voici une instance possible. Les poids des arcs adjacents à un même nœud sont normalisés (somme = 1). La présentation est un arbre ordonné selon l'ordre lexicographique des étiquettes des arcs adjacents à un même nœud.

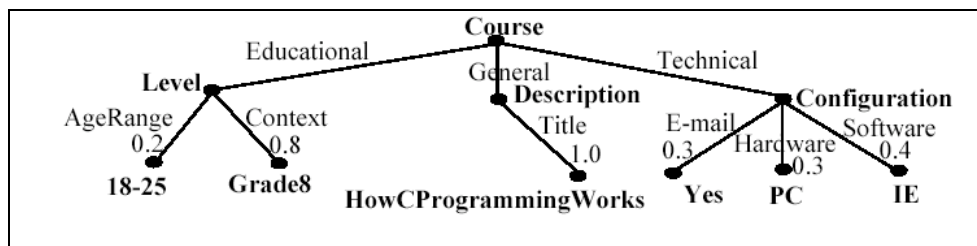


Figure 56 : instance d'arbre générée à partir du schéma précédent.

Deux instances (envoyées par exemple par un acheteur et un vendeur) donnent lieu à un calcul de similarité.

Diverses représentations de ces arbres sont possibles. Pour simplifier l'algorithme de calcul de similarité, une forme normalisée a été utilisée :

- les arcs sont étiquetés dans l'ordre alphabétique appliqué de gauche à droite,
- la somme des poids des arcs d'un même niveau et dans n'importe quel sous-arbre est égale à 1,
- les étiquettes des arcs adjacents à un même nœud doivent être uniques,

Voici quelques exemples (Marsh, Ghorbani et al. 2003).

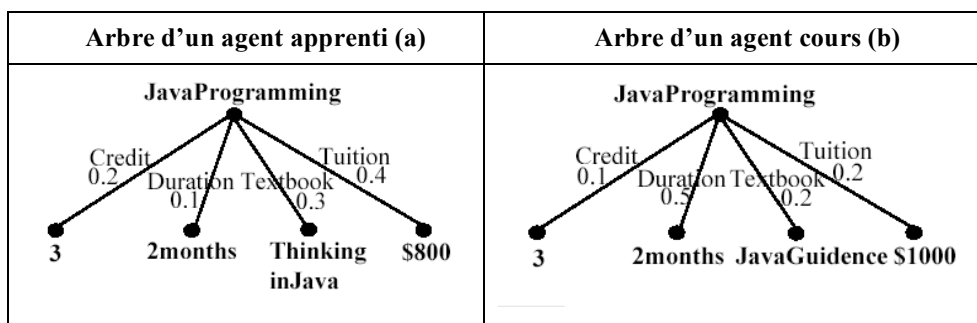


Figure 57 : exemples d'arbre dans le domaine du e-learning

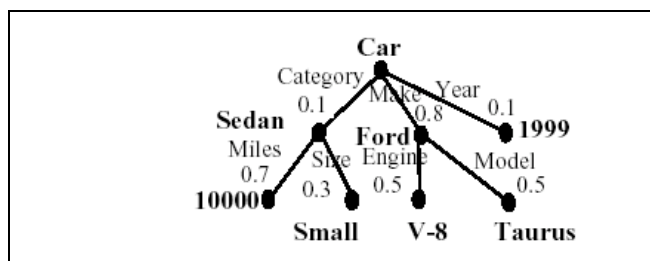


Figure 58 : exemple d'arbre dans le domaine du e-commerce

L'arbre (a) de la Figure 57 représente les préférences d'un apprenant, il cherche un cours de « programmation java » dont la durée lui importe très peu (poids = 0,1).

5.3.2 Algorithme de similarité entre arbres

De manière générale, l'algorithme parcourt deux arbres placés en entrée de haut en bas (racine-feuilles), puis calcule leur similarité de bas en haut. Si les racines de deux (sous-) arbres non vides ont

la même étiquette, leur similarité est calculée par un parcours récursif de haut en bas (racine-feuilles) à travers les sous-arbres accessibles à chaque niveau via les étiquettes identiques des arcs.

La récursivité se termine par deux (sous-) arbres (racine-feuille) qui sont des nœuds terminaux (nœuds-feuilles) ou des arbres vides, qui ont la valeur 1 comme similarité si leurs étiquettes sont égales, et 0 sinon. Le parcours de haut en bas avec un calcul de bas en haut est récursivement employé pour chaque paire de sous-arbres.

5.3.3 Analyse critique

(Bhavsar, Boley et al. 2004) ne donnent pas de définition claire, avec une formule simple qui permette de calculer la similarité entre deux arbres sémantiques. Nous avons trouvé des définitions formelles assez complexes suivies des résultats de calculs, mais aucune information sur le détail du calcul. Ajoutons une fausse utilisation des termes comme « arbre vide » qui veut dire ici un arbre possédant un seul nœud, au lieu d'un arbre qui possède zéro nœud.

Nous avons formulé alors une définition mathématique de la similarité entre deux arbres, en l'étendant à la similarité entre deux forêts, et en utilisant les notations suivantes.

ϕ : arbre vide

φ, Ψ, λ et $\varphi', \Psi', \lambda'$: des forêts

$\varphi = n(\Psi)$: arbre de racine n dominant la forêt Ψ

On compare deux forêts φ , et φ' on notera : si $\varphi \neq \phi$, $\varphi = n(\Psi), \lambda$ et si $\varphi' \neq \phi$, $\varphi' = n'(\Psi'), \lambda'$

$l(n)$: étiquette de l'arc adjacent au nœud n (venant de son père)

$e(n)$: étiquette du nœud n

$S(\varphi, \varphi')$: similarité entre les forêts φ et φ'

σ : arbre $\rightarrow [0,1]$: simplicité d'un arbre

T : arbre \times arbre $\rightarrow [0,1]$: fonction qui calcule la proximité de deux arbres

$$S(\phi, \phi) = 0$$

$$S((n(\Psi), \lambda), \phi) = \sigma(n(\Psi)) + S(\lambda, \phi)$$

$$S(\phi, (n'(\Psi'), \lambda')) = \sigma(n'(\Psi')) + S(\phi, \lambda')$$

$$S(((n(\Psi), \lambda), (n'(\Psi'), \lambda')) = \begin{cases} T(n(\Psi), n'(\Psi')) + S(\lambda, \lambda') & \text{si } l(n) = l(n') \\ \sigma(n(\Psi)) + S(\lambda, (n'(\Psi'), \lambda')) & \text{si } l(n) < l(n') \\ \sigma(n'(\Psi')) + S((n(\Psi), \lambda), \lambda') & \text{si } l(n) > l(n') \end{cases}$$

$$T(n(\Psi), n'(\Psi')) = \text{Proximité}(e(n), e(n')) * S(\Psi, \Psi') * (w(n) + w(n')) / 2$$

$$\text{Proximité}(e(n), e(n')) = \begin{cases} 1 & \text{si } e(n) = e(n') \\ 0 & \text{si } e(n) \neq e(n') \end{cases}$$

$\sigma(n(\Psi))$: simplicité d'un arbre, que nous détaillons plus loin

Figure 59 : formule de calcul de similarité entre deux arbres

Exemples (cas où $l(n) = l(n')$) :

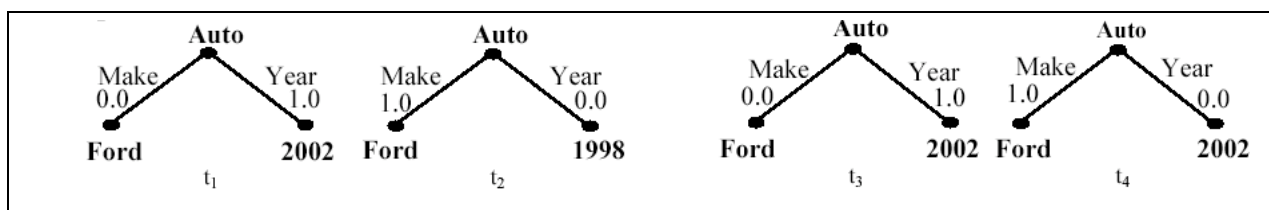


Figure 60 : exemples d'arbres avec des poids d'arcs opposés

$$S(t_1, t_2) = 1.0 * (0.0 + 1.0) / 2 + 0.0 * (1.0 + 0.0) / 2 = 0.5$$

$$S(t_3, t_4) = 1.0 * (0.0 + 1.0) / 2 + 1.0 * (1.0 + 0.0) / 2 = 1.0$$

Figure 61 : exemple de calcul de similarité

Les arbres t_1 et t_2 ont les mêmes étiquettes « Ford » sur les nœuds d'arrivée des branches « Make », la valeur de multiplication est donc égale à 1. Par contre, ils n'ont pas les mêmes étiquettes pour la branche « Year », et la valeur de multiplication est alors égale à 0.

Les arbres t_3 et t_4 ont les mêmes étiquettes de nœuds, la valeur de multiplication est égale à 1 pour les deux branches « Make » et « Year ».

5.3.4 Simplicité d'un arbre

L'algorithme de calcul de la simplicité d'un arbre est dérivé directement de la définition récursive suivante. Par définition, la simplicité d'un arbre à un seul nœud est égale à 1.

On utilisera les notations suivantes, reprises de (Marsh, Ghorbani et al. 2003).

$\delta(T)$: valeur de la simplicité d'un seul arbre T

D_I : index de dégradation de la profondeur

D_F : facteur de dégradation de la profondeur

d : profondeur du nœud terminal

m : degré du nœud racine de l'arbre T qui n'est pas terminal

w_j : poids du $j^{\text{ème}}$ arc en dessous du nœud racine de l'arbre T

T_j : sous-arbre accédé par le $j^{\text{ème}}$ arc partant de la racine (T)

$$\delta(T) = \begin{cases} D_I \cdot (D_F)^d & \text{si } T \text{ est un nœud terminal} \\ \frac{1}{M} \sum_{j=1}^m w_j \cdot \delta(T_j) & \text{sinon} \end{cases}$$

Lorsqu'un arbre grossit verticalement et horizontalement de façon non bornée, les valeurs de m et de d tendent vers l'infini. Toutefois, $(D_F)^d$ et $1/m$ tendent vers 0. Comme w_j et (T_j) sont des valeurs comprises entre 0 et 1, la simplicité tend vers 0.

$$\lim_{\substack{m \rightarrow \infty \\ d \rightarrow \infty}} \delta(T) = 0$$

La Figure 62 montre un exemple de calcul de simplicité.

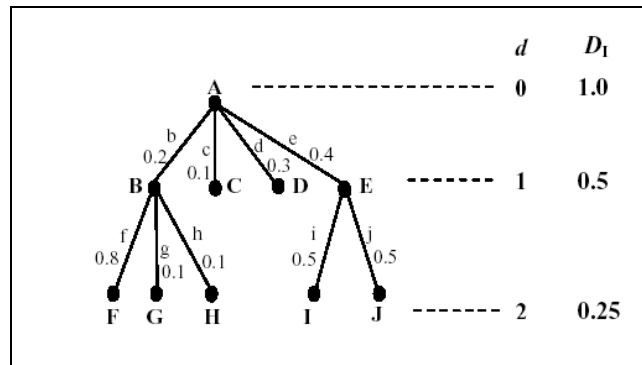


Figure 62 : exemple de calcul de simplicité

$$\delta(T) = \frac{1}{4}(0.2 * \delta(T_B) + 0.1 * \delta(T_C) + 0.3 * \delta(T_D) + 0.4 * \delta(T_E))$$

$$\delta(T_B) = \frac{1}{3}(0.8 * \delta(T_F) + 0.1 * \delta(T_G) + 0.1 * \delta(T_H))$$

$$\delta(T_E) = \frac{1}{2}(0.5 * \delta(T_I) + 0.5 * \delta(T_J))$$

La valeur de la simplicité pour les sous-arbres (nœuds-terminaux) T_F , T_G , T_H , T_I et T_J est celle de D_1 et égale à 0,25. Toutefois la simplicité de $T_B = 0,25/3$ et celle de $T_E = 0,25/2$. Finalement, nous obtenons une valeur de simplicité de l'arbre T égale à 0,06667.

Conclusion

Nous avons classé les représentations de contenu en deux types : des représentations de contenu classiques (vecteurs, listes de propriétés, structures de traits...), et des représentations de contenu spécifiques (CRL-CATS, IF-CATS). Nous utiliserons ces deux dernières représentations dans les prochains chapitres.

Nous avons cherché à trouver quelques algorithmes de calcul de distance et/ou de similarité entre deux instances d'une même représentation de contenu. Nous avons choisi d'adopter l'algorithme de Selkow pour la comparaison entre deux CRL-CATS.

Chapitre 6 Évaluation d'un portage linguistique

Introduction

L'idée ici est d'évaluer l'opération globale de portage en évaluant la performance du système porté et le coût du portage. Pour le coût, c'est facile *a posteriori*, et l'intéressant serait d'arriver à le faire *a priori*.

Dans ce chapitre, on s'intéresse d'abord aux mesures de performance, puis aux mesures de coût. On se propose de faire ces mesures par rapport à trois éléments :

- le langage d'entrée,
- la représentation de contenu obtenue,
- la tâche elle même.

6.1 Langage d'entrée

Trois points principaux liés au langage traité sont à mesurer, à savoir la couverture, l'efficacité et le coût.

6.1.1 Couverture

Il s'agit principalement de mesurer la couverture lexicale pour le langage en question et de mesurer la couverture grammaticale dans le cas d'un sous-langage. Ces deux notions sont liées respectivement aux notions de convergence lexicale et de convergence grammaticale (voir 4.3).

Nous mesurons la couverture lexicale du domaine par l'évolution de la taille du dictionnaire par rapport à la taille du corpus. La taille du dictionnaire peut d'abord augmenter fortement avec à la taille du corpus puis atteindre une certaine stabilité. Cette stabilité se traduit par la suite par l'obtention de bons résultats d'extraction de contenu (si la couverture grammaticale est atteinte bien sûr). Il suffit donc d'examiner le corpus et de voir si tous les mots et/ou expressions de mots constituant ses énoncés sont représentés dans le dictionnaire sous forme d'entrées principales (formes standard) ou secondaires (formes équivalentes, éventuellement « bruitées »).

Par exemple, dans CATS, l'ensemble des mots et des expressions constituant les énoncés « Recherche voiture astra rouge » et « Recherche achat Laguna » sont représentés dans le fragment du dictionnaire suivant :

```
[rouge]{} "red" (color) <F,1,1>;
[voiture]{} "saloon" (vech) <F,1,1>;
[Recherche]{} "wanted" (want) <F,1,1>;
[Recherche achat]{} "wanted" (want) <F,1,1>;
[Laguna]{} "Laguna(country>France, country>europe, make>RENAULT)" (model, car)
<F,3,3>;
[astra]{} "Astra(country>germany, country>europe, make>OPEL)" (model, car)
<F,3,3>;
```

Nous mesurons la couverture grammaticale du domaine par l'évolution des résultats d'extraction de contenu obtenus en fonction de la taille du dictionnaire et de la taille du corpus.

Dans le cas d'un sous-langage et si la couverture lexicale est atteinte, la taille de la grammaire peut atteindre une stabilité. Cette stabilité se traduit par de bons résultats d'extraction de contenu. Nous utilisons une mesure observationnelle pour évaluer la couverture grammaticale. Cette mesure doit être confirmée par une mesure des résultats du portage au niveau de la représentation de contenu que nous détaillerons dans la section 6.2.

Notons qu'il n'est pas toujours possible de mesurer ou de modifier la couverture grammaticale. Par exemple, dans le cas d'un portage par traduction statistique, nous utiliserons des outils fermés (voir Chapitre 12) sur lesquels nous ne pouvons pas intervenir.

6.1.2 Mesure de l'efficacité

On mesure l'efficacité d'un portage principalement par le temps de réponse aux requêtes. Cette mesure n'a pas autant d'importance que la mesure précédente (la couverture) car généralement il s'agit d'un temps de calcul très faible et la différence de temps de calcul entre la version originale de l'application et la version cible obtenue par portage est négligeable.

6.1.3 Coût

On peut mesurer le coût lié au langage suite à une opération de portage linguistique. Cela dépend de la méthode utilisée pour effectuer le portage.

Il s'agit principalement de mesurer le coût de construction d'un corpus et d'un dictionnaire (besoin commun pour la plupart des méthodes).

Mais aussi, dans certaines méthodes comme le portage interne (Chapitre 8) et externe (Chapitre 9), il s'agit d'ajouter le coût de construction ou d'adaptation de la grammaire au coût de construction du corpus et du dictionnaire.

6.2 Représentation du contenu

L'évaluation au niveau de la représentation du contenu dépend de la situation traductionnelle de l'application, et plusieurs cas se présentent. Nous faisons une étude générale illustrée par une étude de cas sur l'application CATS.

6.2.1 Discussion générale

Dans le cas général, deux situations se présentent dans l'évaluation au niveau de la représentation du contenu. Une comparaison par rapport aux représentations de contenu « natives » produites par l'application originale et une comparaison par rapport à une version « de référence ».

Les deux méthodes nécessitent une évaluation basée sur le même corpus d'évaluation. Nous devons alors porter par traduction et/ou révision les mêmes données d'évaluation de la version originale vers la nouvelle langue. Cela suppose bien sûr qu'on ait accès :

- au corpus d'évaluation utilisé dans la version originale,

- aux résultats d'évaluation de la version originale (ou qu'on puisse évaluer l'application originale).

Deux cas se présentent : présence et absence d'une version de référence.

6.2.1.1 Présence d'une version de référence

Nous pouvons obtenir une version de référence en révisant manuellement les résultats donnés par la version originale, puis comparer les résultats du portage par rapport à cette version de référence.

En principe, les résultats trouvés par la version cible (candidate) seront moins bons que ceux de la version originale, et encore moins bons que ceux de la version de référence. Mais on peut améliorer aussi l'opération d'extraction de contenu dans la version cible.

Nous calculons la distance de chacune des versions originale et cible par rapport à la version de référence. Le Tableau 23 résume les situations possibles.

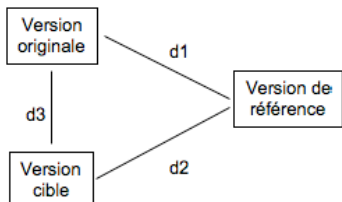
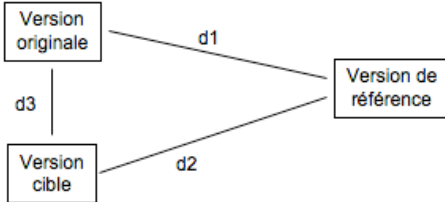
	
<p>Si $d2 < d1$ alors le portage améliore le résultat</p> <p>Si $d2 > d1$ alors le portage affaiblit le résultat</p>	<p>Si $d2$ et $d1$ sont très grandes alors la version originale et la version cible sont mauvaises toutes les deux</p>

Tableau 23 : comparaison entre versions originale, cible, et référence

6.2.1.2 Absence d'une version de référence

Dans le cas de l'absence d'une version de référence, nous pouvons considérer la version originale de l'application comme étant une version de référence. Cela nous laisse supposer que les résultats de cette version sont corrects. Ainsi, nous calculons la distance entre la version originale et la version cible (Tableau 24).

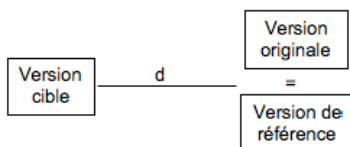

<p>d = distance entre la version cible et la version originale considérée comme version de référence</p>

Tableau 24 : comparaison entre version cible et version originale

6.2.2 Application à CATS

Nous utiliserons principalement trois mesures pour évaluer les résultats d'extraction au niveau de la représentation de contenu CRL-CATS : une première « mesure informationnelle » basée sur un corpus de taille limitée, une deuxième mesure générale basée sur tout le corpus (obtenu par déconversion, voir Chapitre 11), qu'on qualifie de « mesure simple », et une troisième mesure générale basée aussi sur la totalité du corpus, qu'on qualifie de « mesure fine ».

6.2.2.1 Mesure informationnelle

(Daoud 2006), a utilisé les mesures connues de rappel et de précision pour évaluer la version originale (arabe) de son système. Sur un corpus limité à 200 SMS (100 SMS d'achat et 100 SMS de vente) que nous appelons « CorpusEvalAr200SMS », il se base sur un jugement humain pour calculer ces mesures. De la même façon, et sur le même corpus que nous appelons « CorpusEvalFr200SMS », obtenu par traduction dans la nouvelle langue L2 (français), nous utilisons ces mesures pour évaluer la nouvelle version de CATS, obtenue par chacune des méthodes de portage.

Nous calculons le rappel R, la précision P et la F-mesure F pour chacune des propriétés les plus importantes (*action de vente ou d'achat, marque, modèle, année, prix*), avec les notations suivantes :

P = Nombre d'entités correctes identifiées par le système / Nombre total d'entités identifiées par le système ;

(P = 0 si nombre total d'entités identifiées par le système = 0)

R = Nombre d'entités correctes identifiées par le système / Nombre d'entités identifiées par l'humain ;

(R = 0 si nombre d'entités identifiées par l'humain = 0)

F = $2 * P * R / (P + R)$

Par convention F = 0 si P = R = 0

Le Tableau 25 présente un exemple d'application de cette mesure.

	Version originale (arabe)				Version cible (française)			
Énoncé	البيع مرسيدس 1985م.				;A vendre Mercedes m 1985.			
CRL-CATS	[S] sal(saloon:00, sale:00) mak(saloon:00, MERCEDES(country>germany, country>europe):06) yea(saloon:00, 1985:1G) [/S]				[S] sal(saloon:00, sale:00) mak(saloon:00, MERCEDES(country>germany, country>europe):09) mod(saloon:00, 1985:16) [/S]			
Propriété	sal	Mak	mod	yea	sal	mak	mod	yea
Nbre d'entité correctes identifiées par le système	1	1	0	1	1	1	0	0
Nbre total d'entité identifiées par le système	1	1	0	1	1	1	1	0
Nbre d'entité identifiées par l'humain	1	1	0	1	1	1	0	1
Précision (P)	1	1	0	1	1	1	0	0
Rappel (R)	1	1	0	1	1	1	0	0
F-mesure (F)	1	1	0	1	1	1	0	0

Tableau 25 : exemple d'application de la mesure informationnelle

La deuxième façon d'évaluer la nouvelle version (française) consiste à comparer ces résultats à une version de référence. Pour cela, on prend les CRL-CATS produites par la version arabe et on les corrige manuellement, ce qui donne des représentations de référence.

6.2.2.2 Mesure simple

Afin, d'évaluer la nouvelle version obtenue en L2 sur le total du corpus (également disponible en L2) et appelé « CorpusEvalFr1100SMS », il nous faut trouver une méthode automatique de mesure de la similarité entre deux représentations de contenu de CATS, celle qui est produite par la version originale du système, et celle qui est produite par la nouvelle version du système.

Trois particularités essentielles liées à la représentation CRL-CATS sont importantes à considérer.

La première est qu'un attribut peut avoir une valeur ensembliste, par exemple l'attribut *fea* dans la Figure 44.

La deuxième est que, quelle que soit la méthode d'extraction de contenu utilisée, et bien que toutes ces méthodes aboutissent finalement à la représentation CRL-CATS, les objets et leurs valeurs sont produits à partir de dictionnaires qui contiennent les mêmes données (lemmes = entrées principales du dictionnaire). Les données (en L2) absentes du dictionnaire ne pourront donc être détectées par aucun des extracteurs de contenu construits pour L2. Une fois que des données sont détectées, l'extracteur de contenu les affiche telles quelles sans aucune modification, sauf pour les deux attributs *pri* et *yea* qui peuvent avoir des sous-attributs ajoutés par un extracteur et pas par un autre.

Par exemple, un premier extracteur a détecté l'attribut *pri* suivant avec le sous-attribut ajouté *@ALF* qui veut dire *mille*, alors qu'un deuxième extracteur n'a pas détecté ce sous-attribut.

```
pri(saloon:00, 23:1R.@ALF)
```

De même pour le sous-attribut *@installment* qui veut dire « *paiement par tranches* ».

```
pri(saloon:00, 250:1H.@installment)
```

Nous avons constaté que l'ajout des sous-attributs pour les deux relations sémantiques *pri* et *yea* et dans les toutes les versions d'extraction de contenu est d'une fréquence très faible.

CRL-CATS obtenue par l'extracteur de contenu original EnCo pour l'arabe	CRL-CATS obtenue par l'extracteur de contenu adapté pour l'automobile IF-CATS
<pre>مطلوب سيارة اوبل فكترا بمحرك 1600 موديل 94 فما فوق. [S] wan(saloon:06, wanted:00) mak(saloon:06, OPEL(country>germany, country>europe):0C) mod(saloon:06, Vectra(country>germany, country>europe,make>OPEL):0H) mot(saloon:06, 1600:0T) yea(saloon:06, 94:14.@more) [/S]</pre>	<pre>; recherche voiture OPEL Vectra moteur 1600 modèle 94 ou plus S wan(saloon, wanted) mak(saloon, OPEL(country>germany, country>europe)) mod(saloon, Vectra(country>germany, country>europe,make>OPEL)) /S</pre>

Figure 63 : exemple de différences entre les CRL-CATS obtenues par deux extracteurs de contenu différents

La troisième particularité, comme le montre la Figure 63, est que les attributs (les étiquettes des arcs) ne sont pas pondérés par des poids.

Nous pouvons utiliser une mesure basée sur la combinaison d'une mesure linguistique intrinsèque en comparant les attributs (les propriétés), les objets du domaine et leurs valeurs détectés par chaque version d'extraction de contenu.

Les trois particularités précédentes nous conduisent à nous limiter à la mesure des attributs sémantiques détectés. En effet, dans cette application, l'opération d'extraction de contenu revient à construire ces attributs sémantiques entre les objets du domaine et leurs valeurs. Ainsi, dans une représentation CRL-CATS, nous calculons un score pour chaque attribut (*wan, sal, mak, mod, etc*), en affectant la valeur 1 si l'attribut est détecté et la valeur 0 sinon. Nous affectons des poids à valeurs fixes aux différents attributs traités suivant leur importance, de façon que leur somme soit égale à 1. Ainsi, nous multiplions le score obtenu pour chaque attribut par son poids correspondant.

Attribut	wan	sal	cou	yea	col	fea	pri	mak	mod	mot
Poids	1/9	1/9	1/12	1/9	1/12	1/12	1/9	1/9	1/9	1/12

Figure 64 : affectation des poids aux dix attributs de CRL-CATS

Ensuite, nous affectons un score global au corpus total, pour chaque extracteur de contenu, par rapport aux propriétés les plus importantes.

Le score d'un attribut x dans un corpus de n SMS est égal à (n est un entier) :

$$\text{Score}_{\text{corpus}(n \text{ SMS})}(\text{attribut}_x) = \sum_n \text{poids}(\text{attribut}_x) * \text{Detection}_{\text{SMS}}(\text{attribut}_x)$$

$$\text{Detection}_{\text{SMS}}(\text{attribut}_x) = 1 \text{ si } \text{attribut}_x \text{ est détecté et } 0 \text{ sinon}$$

Le score de m attributs dans un corpus de n SMS est égal à (m est un entier) :

$$\text{Score}_{\text{corpus}(n \text{ SMS})}(m \text{ attribut}) = \sum_m \text{Score}_{\text{corpus}(n \text{ SMS})}(\text{attribut}_m)$$

	Version originale (arabe)				Version cible (française)			
SMS	;للبيع مرسيدس م 1985.				;A vendre Mercedes m 1985.			
CRL-CATS	[S] sal(saloon:00, sale:00) mak(saloon:00, MERCEDES(country>germany, country>europe):06) yea(saloon:00, 1985:1G) [/S]				[S] sal(saloon:00, sale:00) mak(saloon:00, MERCEDES(country>germany, country>europe):09) mod(saloon:00, 1985:16) [/S]			
Attribut	sal	mak	mod	yea	Sal	mak	mod	yea
Poids	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9
Detection (attribut)	1	1	0	1	1	1	1	0
Score (attribut)	1/9	1/9	0	1/9	1/9	1/9	1/9	0
Score (sal, mak, mod, yea)	3/9				3/9			

Tableau 26 : exemple d'application de la mesure simple

Le Tableau 26 présente un exemple d'application de cette mesure.

L'objectif de cette mesure est d'évaluer l'extracteur de contenu en terme de quantité d'attributs détectés, sans tenir compte de la qualité de l'attribut détecté.

6.2.2.3 Mesure fine

Toujours dans le but d'effectuer une évaluation sur la totalité du corpus (obtenu par déconversion, voir Chapitre 11) nommé « CorpusEvalFr1100SMS », nous utiliserons ici une deuxième mesure automatique, plus fine que la précédente. Dans cette mesure, nous allons au delà de la détection des noms des attributs en comparant leurs valeurs par rapport à une version de référence. Nous supposons alors que la version originale produite par le système est une version de référence. Ainsi, nous calculons la distance entre la version de référence et la version candidate résultant d'une méthode de portage. Nous utilisons l'algorithme de calcul de distance d'édition entre arbres (Selkow 1977) présenté dans la section 5.2.2 du chapitre précédent.

Pour pouvoir comparer les résultats de cette mesure à d'autres, nous pouvons utiliser cette mesure sur le total des attributs ou nous limiter à un calcul de distance basé sur un certain ensemble d'attributs considérés comme les plus importants comme *sal*, *wan*, *mak*, *mod*, *pri*, *yea*.

Voici un exemple de calcul de distance :

Version originale (arabe)	Version cible (française)																				
2001 م 140 للبيع مرسيدس	;A vendre Mercedes 140 m 2001																				
[S] sal(saloon:00, sale:00) mak(saloon:00, MERCEDES(country>germany, country>europe):06) mod(saloon:00, 140:0D) yea(saloon:00, 2001:0J) [/S]	[S] sal(saloon:00, sale:00) mak(saloon:00, MERCEDES(country>germany, country>europe):09) mod(saloon:00, 140:0I) [/S]																				
	<table border="1"> <thead> <tr> <th></th> <th>mak#MERCEDES(country>germany,country>europe)</th> <th>mod#140</th> <th>sal#sale</th> </tr> </thead> <tbody> <tr> <td>mak#MERCEDES(country>germany,country>europe)</td> <td>0</td> <td>1</td> <td>2</td> </tr> <tr> <td>mod#140</td> <td>1</td> <td>0</td> <td>1</td> </tr> <tr> <td>sal#sale</td> <td>2</td> <td>1</td> <td>0</td> </tr> <tr> <td>yea#2001</td> <td>3</td> <td>2</td> <td>1</td> </tr> </tbody> </table>		mak#MERCEDES(country>germany,country>europe)	mod#140	sal#sale	mak#MERCEDES(country>germany,country>europe)	0	1	2	mod#140	1	0	1	sal#sale	2	1	0	yea#2001	3	2	1
	mak#MERCEDES(country>germany,country>europe)	mod#140	sal#sale																		
mak#MERCEDES(country>germany,country>europe)	0	1	2																		
mod#140	1	0	1																		
sal#sale	2	1	0																		
yea#2001	3	2	1																		
Distance = 1																					

Tableau 27 : exemple de calcul de distance nulle entre CRL-CATS

Nous passons à un calcul des mesures connues (rappel R, précision P, et F-mesure F) par simple appel à la formule suivante.

Soit

Ref : une CRL de référence (triée), avec $|Ref| = m$

Cand : une CRL candidate (triée), avec $|Cand| = n$

$K = |\{\text{attributs corrects dans Cand}\}|$: la longueur d'une sous chaîne maximale commune entre une CRL candidate et référence

Alors $K = |\text{sscm}(\text{Ref}, \text{Cand})|$ et $D(\text{Ref}, \text{Cand}) = m + n - 2K$

Donc $R = K/m$, $P = K/n$,

d'où $F = 1 - D(\text{Ref}, \text{Cand})/(m+n)$

Dans l'exemple précédent, $m = n = 4$ d'où $k = 4$ et $F = 1$

Nous calculons ensuite, la moyenne des différentes valeurs des rappels, des précisions et des F-mesures trouvées.

6.3 Évaluation liée à la tâche

Nous détaillons dans ce paragraphe quelques pistes d'évaluation de la performance liée à la tâche d'une version obtenue par un portage linguistique.

6.3.1 L'idéal

L'idéal est de réaliser une évaluation en réel avec de vrais utilisateurs (comme le fait Multicom, une équipe de recherche dans notre laboratoire LIG), de la même façon que l'application « native » est évaluée. Il s'agirait par exemple dans notre étude de cas sur CATS de tester la version française de CATS-Fr à Amman, en situation réelle.

En fait, ce type d'évaluation n'est pas facilement réalisable. Par exemple, déployer en situation réelle une version française de CATS nécessiterait d'abord une coopération avec une société de téléphonie et demande des accords à l'avance et des pré-tests. Ensuite, le déploiement d'un nouveau service de ce type doit démarrer par le passage de flashes de publicité (journaux, télévision, envoi de SMS...).

L'ensemble de ce travail d'installation et de publicité ainsi que d'autres tâches préparatoires demande un coût non négligeable et beaucoup de temps. Il n'est pas garanti non plus qu'on arrive à « enrôler » assez d'utilisateurs (potentiels) sur place.

6.3.2 Avec des « cobayes »

On peut imaginer une autre piste d'évaluation en « inventant » des utilisateurs (ici francophones), qui seraient des « cobayes » payés pour un travail de simulation. Il s'agit d'évaluer le degré de leur satisfaction par rapport au service de l'application proposé. Par exemple, est ce qu'un utilisateur, acheteur d'une voiture a bien reçu des propositions qui correspondent bien à sa requête.

Cette évaluation basée sur des utilisateurs « cobayes » coûterait aussi très cher en termes de prix et de temps (au moins trois mois de travail). En effet, nous estimons qu'il nous faut au moins :

- une dizaine de cobayes de type vendeur,
- une dizaine de cobayes de type acheteur,
- un total de 600 SMS envoyés par l'ensemble des utilisateurs.

Comme indique le Tableau 28, chaque utilisateur peut exprimer son degré de satisfaction en donnant une note comprise entre 0 et 20.

	Pas satisfait	Presque satisfait	Satisfait	Très satisfait
Notes	[0..5[[5..10[[10..15[[15..20]

Tableau 28 : mesure de la satisfaction des utilisateurs

6.3.3 Autre possibilité

Dans le cas où on n'a pas accès à l'application par le Web, ce qui est le cas de toutes les applications sur les téléphones mobiles, il faut alors avoir des abonnements.

On peut créer un site Web simulant l'application jusqu'à un certain point. L'inconvénient ici c'est que ce travail est assez coûteux, donc il faut de la ré-ingénierie d'outils, mais c'est moins fiable au niveau du contexte qui est différent (le sous-langage peut être modifié) et au niveau de l'application qui est différente de la réalité.

6.3.4 En « magicien d'Oz »

On peut se servir d'un humain pour jouer le rôle de l'application : en voyant une représentation de contenu, il répond/agit aux requêtes.

Ce n'est pas une possibilité pour évaluer un « portage linguistique », car si on doit utiliser un magicien d'Oz (MOz), c'est comme s'il n'y avait pas (encore) d'application.

Conclusion

Nous avons étudié les méthodes possibles pour évaluer un portage linguistique d'une application d'extraction de contenu à partir d'énoncés spontanés en langue naturelle. Nous avons détaillé trois possibilités d'évaluation portant sur le langage ou sous-langage traité, la représentation de contenu, et la performance liée à la tâche.

Nous avons montré comment on peut appliquer certaines méthodes d'évaluation (de la représentation de contenu par exemple) à l'application CATS, nous décrivons le portage de l'arabe vers le français par plusieurs méthodes dans les prochains chapitres.

Concernant l'évaluation liée à la tâche, nous avons décidé de faire une évaluation « de bout en bout » grâce à une simulation de CATS par un site Web. Ce travail a déjà commencé dans le cadre d'un TER (Zair 2008).

Partie C

Implémentations

et expériences de portage linguistique

Pour différentes raisons que nous détaillons dans cette partie, nous nous limitons à un portage linguistique (arabe vers français) de l'application de e-commerce CATS, pour laquelle les facteurs dont dépend la situation traductionnelle sont assurés.

Nous présentons, dans un premier chapitre, le travail de construction d'un dictionnaire et d'un corpus français fonctionnellement équivalents à ceux de l'arabe. Nous montrons qu'une simple traduction directe du corpus de départ ne donne pas un corpus de développement satisfaisant.

Nous détaillons dans le deuxième chapitre, la mise en œuvre de la première méthode de portage interne qui nécessite un accès complet au code et aux ressources de l'application. Elle consiste à adapter au français l'extracteur de contenu de l'application.

Enfin, le troisième chapitre de cette partie présente la mise en œuvre de la deuxième méthode de portage externe, qui demande un simple accès à la représentation interne de l'application. Elle consiste à adapter un extracteur de contenu existant pour la même langue (français) mais pour un domaine différent (tourisme) et à traduire les représentations internes produites vers le format CRL-CATS. Dans le cadre du portage d'une autre application de recherche de musique en langue japonaise, nous avons aussi adapté le même extracteur de contenu au domaine de la musique, pour les deux langues française et anglaise.

Chapitre 7 Fabrication des données nécessaires pour le portage de CATS vers le français

Introduction

Nous choisissons le système CATS comme application principale à porter vers le français pour plusieurs raisons :

- CATS traite des énoncés spontanés en langue naturelle.
- Nous avons un accès complet au code source et aux ressources de l'application, un niveau de compétences langagières suffisant par rapport à la langue source de l'application (arabe, langue maternelle) et à la langue cible (français, deuxième langue), et un niveau de compétence informatique suffisant par rapport à la technique et à l'outil d'extraction utilisé.
- L'existence d'une motivation réaliste : il est souhaitable de permettre l'accès dans plusieurs langues au service proposé par CATS en Jordanie (à Amman).

L'intérêt dans ce travail est double. D'abord, le portage de CATS vers le français (en Jordanie) donnerait la possibilité à des francophones vivant en Jordanie (à Amman) et disposant d'un mobile d'utiliser le système. Ensuite, ce travail constitue une réelle expérience qui nous permet de confirmer ou d'infirmer les hypothèses et les méthodes de portage proposées.

Dans la première partie, nous faisons une étude concernant les données de la version originale disponibles, à savoir les caractéristiques, la structure, et la taille de ces données. Dans la deuxième partie, nous présentons une méthode pour produire un corpus français fonctionnellement équivalent à celui de l'arabe. Dans la troisième partie, nous présentons la méthode mise en œuvre pour la préparation du dictionnaire, une tâche qui demande beaucoup d'imagination, et de simulation.

7.1 Etude des données disponibles dans CATS et méthode de portage

Après avoir présenté en détail le corpus et le dictionnaire de CATS, nous étudions la façon de les porter vers le français.

7.1.1 Le corpus CATS

Le corpus de développement SMS_ar_orig est constitué de 12372 SMS en arabe, de taille 448 Ko encodé en arabe (Windows).

7.1.1.1 Description du contenu

Le texte de ces annonces en SMS est un mélange entre des SMS pour l'achat ou la vente de voitures, de biens immobiliers et autres (voir section 2.2.3.3.3).

Il y a environ 23 % d'arabe jordanien, 5% de mots anglais importés, 70% d'arabe littéraire et 2% de mots inconnus. Ce mélange de données nécessite un traitement particulier. D'autres difficultés sont liées à la structure du corpus.

7.1.1.2 Structure du corpus CATS

Comme le montre l'exemple de la Figure 65, le corpus contient accidentellement des lignes vides, des annonces incomplètes (présence de la date et/ou de l'heure de l'annonce, mais absence du cœur de l'annonce), ou des annonces en anglais.

La dernière version du système CATS accepte encore de tels SMS. Les messages en anglais doivent être traduits manuellement en arabe avant d'être passés au système.

```

4/16/2005 8:00:27 AM   Message In : +962796635036,مطلوب شغه نوم واحد مع حمام ومطيخ وصالون في عبدون غير مفروش,
4/16/2005 9:05:56 AM   90050
4/16/2005 9:05:56 AM   Message In : +962795171114,مطلوب سيارة بسعر لايزيد عن ألفين دينار,
4/16/2005 9:16:48 AM   90050
4/16/2005 9:16:48 AM   Message In : +962796666670,wanted toyota corola model 93
4/16/2005 9:18:46 AM   90050
4/16/2005 9:18:46 AM   Message In : +962795912862,99 .مطلوب سيارة نوبونا كورولا
  
```

Figure 65 : exemple du corpus SMS_ar_orig

Les données du corpus CATS sont spontanées, réelles, avec des erreurs, des fautes, des blagues et il y a même parfois des messages exprimant des besoins hors sujet et hors domaines. L'ensemble de ces données forme un corpus naturel, évolutif²⁸ et assez intéressant. Son point fort est qu'il contient des données du monde réel.

7.1.1.3 Traitement particulier pour extraire le contenu

Les annonces du corpus sont constituées des champs suivants :

Champ	Date	Heure	Message	N° téléphone	Cœur du message
Exemples	9/3/2005	12:47:47 PM	Message	+962795731129,	مطلوب سيارة ميتسوبيشي لانسير
	9/9/2005	7:57:17 PM	Message	+962795314763,	

Tableau 29 : champs du corpus original SMS_ar_orig

Les cœurs de ces annonces sont précédés par une virgule, ce qui facilite le nettoyage. Par des programmes Perl, nous avons filtré ce corpus en supprimant les messages vides, les lignes vides et autres textes parfois incompréhensibles (ou hors sujet). Cela a réduit le nombre d'annonces de 12372 à 3017 (25%). Nous obtenons ainsi un nouveau corpus propre que nous appelons SMS_ar_extr_content.

Afin de faciliter la manipulation de ce corpus et son traitement sur d'autres plates-formes (UNIX, LINUX, MAC OS X), nous l'avons d'abord transformé vers un format texte en codage utf-8, puis dans un format XML (codage utf-8) appelé CPXM (Common Parallel eXample Markup), destiné à coder des fichiers multilingues, dont la dtd est détaillée dans l'Annexe II (Hajlaoui and Boitet 2004).

²⁸ Nous disposons de la version de décembre 2005.

```

<?xml version="1.0" standalone="no" ?>
<!DOCTYPE document SYSTEM "CPXM-v5.dtd">
<document>
<information documentname="CATS Corpus"
creation-date="** ** ** ** ** ** ** ** ** **"
modification-date=" septembre 12 JST 2005"
coding-set="UTF-8"
number-of-language="1"
number-of-polyphrases="3017"/>

<polyphrase polyphrase-id="0">
  <monolingual-component xml:lang="Arabe">
    <proposal proposal-id="1">
      نعم نعم نفعه اريد التفاصيل
    </proposal>
  </monolingual-component>
</polyphrase>
<polyphrase polyphrase-id="1">
  <monolingual-component xml:lang="Arabe">
    <proposal proposal-id="1">
      مطلوب نرا سياره نوينا بکب 4 باي 4 کيشه واحد
    </proposal>
  </monolingual-component>
</polyphrase>
</document>

```

Figure 66 : exemple de deux SMS arabes en format CPXM

7.1.2 Dictionnaire

L'extracteur de contenu utilisé dans le système CATS nécessite un dictionnaire et des règles. Le dictionnaire relie les mots et les expressions arabes d'un domaine spécifique aux concepts de CRL-CATS en conservant les caractéristiques sémantiques, syntaxiques et morphologiques utilisées dans l'analyse des SMS arabes. La structure des entrées inclut des abréviations, différentes formes orthographiques pour la même entrée, et d'autres formes de jargon utilisé dans les différents sous-langages.

```

[بي ام دبليو] "B.M.W(country>germany,country>europe)" (make,nmodel,car) <A,3,3>;
[بي ام دبليو] "B.M.W(country>germany,country>europe)" (make,nmodel,car) <A,3,3>;
[ام] "B.M.W(country>germany,country>europe)" (make,nmodel,car) <A,3,3>;
[بي ام] "B.M.W(country>germany,country>europe)" (make,nmodel,car) <A,3,3>;
[بي.ام.دبليو] "B.M.W(country>germany,country>europe)" (make,nmodel,car) <A,3,3>;
[بي.ام.دبليو] "B.M.W(country>germany,country>europe)" (make,nmodel,car) <A,3,3>;
[دبليو] "B.M.W(country>germany,country>europe)" (make,nmodel,car) <A,3,3>;
[بي.ام.] "B.M.W(country>germany,country>europe)" (make,nmodel,car) <A,3,3>;

```

Figure 67 : exemple d'entrées du dictionnaire arabe

Plusieurs entrées secondaires convergent vers un « Concept Word » (CW); il contient l'information sémantique nécessaire pour le processus d'extraction. Un CW possède un seul sens. Par contre, un lexème UNL, ou UW (Universal Word) utilisé dans UNL (Uchida, Zhu et al. 2005-2006) peut avoir plusieurs sens (Sérasset and Boitet 2000).

Chaque CW appartient à une catégorie sémantique. Par exemple, dans le domaine de l'automobile, le modèle, la couleur, le prix sont des catégories sémantiques différentes.

Taille : comme dit dans la section 4.3.1.1 le dictionnaire utilisé dans la version arabe a environ 30.000 entrées pour les deux domaines de l'automobile et de l'immobilier, dont 20.000 ont été générées automatiquement grâce à un répertoire de variantes et de fautes d'orthographe fréquentes.

Coefficient d'expansion : c'est le quotient du nombre total d'entrées (30.982) par le nombre d'entrées de base (10.828), ce qui donne une valeur égale à 3. Cela est dû aux erreurs, formes

diacritiques, et transcriptions multiples de noms étrangers etc. Ces différentes formes peuvent être écrites de plusieurs façons possibles dans une langue assez riche morphologiquement comme l'arabe.

L'expansion peut être faite d'une façon manuelle ou automatique. Dans ce cas, elle a été faite avec des algorithmes de normalisation (voir Figure 67).

Plusieurs entrées lexicales sont regroupées sous une même entrée principale (et partagent un même « concept dictionnaire » CW). Ces diverses variantes ont été collectées à partir du corpus CATS qui ne cesse pas d'augmenter. En effet, chaque fois qu'il y a un nouvel ensemble de mots inconnus pour l'extracteur de contenu, il sera ajouté dans le dictionnaire à la main. Il peut s'agir d'un nouveau modèle de voiture, d'une nouvelle transcription, ou d'une nouvelle orthographe.

Autrement dit, le corpus est enrichi directement par les nouveaux énoncés soumis par les utilisateurs, et le dictionnaire est aussi enrichi, mais indirectement, par ces mêmes énoncés.

7.1.3 Vers le portage

L'opération du portage du système CATS vers le français doit être précédée d'une étape de préparation de corpus et du dictionnaire. Cela peut se faire en deux étapes : création des données primaires²⁹ et création des données dérivées³⁰.

7.1.3.1 Éléments à prendre en compte

Plusieurs éléments sont à prendre en compte, à savoir les différences culturelles, l'utilisation d'un langage codé, et la présence ou l'absence d'un dictionnaire qui aide à la saisie des SMS (T9).

❖ Différences culturelles

Les différences culturelles se concrétisent par l'utilisation d'un dialecte pour chaque langue. Dans le cas de notre corpus arabe, quelques mots sont écrits en arabe comme ils sont prononcés en anglais, c'est une transcription en écriture arabe de la prononciation anglaise du mot. D'autre part, l'anglais est la deuxième langue en Jordanie, contrairement à d'autres pays arabes et francophones. Mais un phénomène similaire est observé dans d'autres pays, par exemple, en Tunisie, pays francophone, où on arabise des mots français et où on les emploie dans les discussions quotidiennes au lieu d'utiliser les équivalents arabes.

Par exemple, dans la plupart des SMS du corpus arabe, nous trouvons le mot « **موديل** » qui est une prononciation du mot anglais "model " mais ne signifie pas « modèle » en arabe.

❖ Utilisation d'un langage codé

Il existe des logiciels de rédaction des SMS pour le français et d'autres langues.

Un communiqué de presse (<http://www.languefrancaise.net/>) montre que :

les Français veulent pourtant défendre le bon langage, mais un vrai phénomène de société s'empare de la France : les SMS forment un langage codé riche en expressions familières ou abrégées du genre MDR (mort de rire), BI1 (bien) ou A2M1 (à demain) (Véronis and Neef 2006) (Fairon, Klein et al. 2006).

Parmi les personnes interrogées, 41 % disent utiliser toujours ou très souvent de tels termes codés lorsqu'elles envoient des messages SMS à leurs amis ou aux membres de leur famille. Cette proportion fait un bond spectaculaire chez les jeunes de 18 à 24 ans où elle passe à plus de 73 %.

²⁹ Les données primaires sont les entrées de base (une pour chaque CW) pour le dictionnaire et les messages types pour le corpus.

³⁰ Pour le dictionnaire, les données dérivées sont les entrées secondaires qui contiennent différentes formes pour le CW de l'entrée de base considérée. Pour le corpus, ce sont les autres formes d'un même énoncé.

❖ Absence de TextInput

T9 (Text Input), célèbre logiciel de saisie de texte utilisant l'alphabet ou les caractères est fréquemment utilisé pour certaines langues. Théoriquement l'utilisation du T9 réduit l'utilisation des abréviations dans la rédaction des SMS, mais tout dépend de la culture, des habitudes et du savoir faire des utilisateurs : certains ne savent pas utiliser le T9, d'autres n'aiment pas l'utilisation d'un dictionnaire, etc.

7.1.3.2 Restrictions sur le domaine et le langage

Pour notre expérience du portage de CATS, nous nous sommes restreints à un seul domaine et à une seule langue cible.

❖ Sur le domaine

D'après les mesures données dans le Tableau 3, le TTR pour le domaine de l'automobile (0,201) représente la valeur la plus faible par rapport aux autres domaines, ce qui veut dire qu'une grande partie du contenu du corpus CATS concerne ce domaine. Pour cette raison et afin de faciliter la tâche, nous avons décidé de nous limiter au domaine de l'automobile.

❖ Sur le langage

Nous nous sommes limité à un portage de CATS vers le français en Jordanie et ne l'avons pas « localisé » vers la France, à cause des différences culturelles, traditionnelles, commerciales... au niveau des critères de recherche d'une voiture. En effet, les Français s'intéressent à certaines options (diesel ou essence, première main ou non, option cuir, kilométrage...). Par contre, en Jordanie, ces options sont négligées par les acheteurs.

La spécification de CRL-CATS telle qu'elle est ne couvre pas tous les critères de choix d'une voiture en France. Pour cette raison et afin de garder la spécification originale sans introduire de modifications, nous avons décidé de porter CATS, c'est-à-dire de changer la langue, en conservant les critères de choix jordaniens.

7.1.3.3 Préparation des données

Il faut distinguer la préparation des données primaires et des données secondaires.

❖ Création des données primaires

Dans cette étape, il s'agit de créer et de préparer les données primaires (données de base) pour le domaine de l'automobile à savoir les entrées de base (les concepts) pour le dictionnaire et les messages types qu'on peut imaginer être envoyés par un Français en Jordanie avec des critères de choix jordaniens.

Cette opération peut être faite d'une façon manuelle et/ou automatique et une grande partie du travail peut être basée sur les données disponibles en arabe. Nous détaillerons les résultats de ce travail dans les sections 7.2 et 7.3.

❖ Création des données dérivées

Dans cette étape, il s'agit de créer les données dérivées (données secondaires) pour le domaine de l'automobile, à savoir les entrées secondaires qui doivent être connectées aux concepts de base du dictionnaire créés dans l'étape précédente. Dans ce cas, il est important de travailler par imagination, simulation, interrogation pour arriver à créer des erreurs, des abréviations, et des transcriptions multiples de noms étrangers que peut rédiger un Français.

De la même manière, et à partir des messages des données primaires, on essaie d'imaginer d'autres formes de SMS rédigés à la même manière d'un Français et aussi en changeant de critères de recherche ou de vente des voitures.

Nous pouvons améliorer le corpus obtenu en supprimant beaucoup de mots outils (articles, prépositions, etc), en introduisant des abréviations usuelles dans les petites annonces et en ajoutant des abréviations de type phonétique ("*ke*" pour "*que*", "*g*" pour "*j'ai*", etc) (Véronis and Neef 2006).

7.2 Corpus

Nous étudions dans ce qui suit la possibilité de fabriquer un corpus français fonctionnellement équivalent au corpus arabe par traduction. Nous montrons qu'une traduction directe d'un corpus réel ne donne pas un corpus réel.

7.2.1 Par traduction

L'idée de fabriquer un corpus français de développement à partir du corpus CATS en arabe en le traduisant en « SMS français spontanés » semble intéressante mais on est confronté à plusieurs types de traductions.

Dans la suite, nous prenons comme exemple le SMS original suivant :

6/1/2005 11:32:20 AM Message In : 962795028727,3750 الى والسعر ما بين 3500 ومطلوب سيارة هونداي موديل 97

Figure 68 : exemple de SMS du corpus arabe

Une traduction « brute » est une traduction qui est soit le résultat d'une traduction automatique soit le résultat produit par un non-Français. Dans les deux cas, le résultat n'aura pas le style d'un message produit par un Français, comme le montre la traduction suivante produite par un non-Français.

6/1/2005 11:32:20 AM Message reçu : 962795028727 : recherche d'une voiture Honda, modèle 97, et le prix entre 3500 et 3750.

Figure 69 : exemple de traduction brute

Une traduction littérale peut ne pas respecter la façon de dire d'un Français. Une traduction littérale du message précédent donne (« *une* » effacé, « *et* » remplacé par « *jusqu'à* ») :

6/1/2005 11:32:20 AM Message reçu : 962795028727 : recherche voiture Honda, modèle 97, et le prix entre 3500 jusqu'à 3750.

Figure 70 : exemple de traduction littérale

Une traduction fonctionnelle ou naturelle est une traduction correcte qui correspondrait à ce que dirait un Français d'une façon spontanée dans la même situation.

Pour le même SMS et si c'était un Français, il écrirait plutôt (en ajoutant l'unité de prix, nécessaire en français) :

6/1/2005 11:32:20 AM Message reçu : 962795028727 : recherche voiture Honda, modèle 97, prix entre 3500 et 3750 Dinars.

Figure 71 : exemple de traduction fonctionnelle

7.2.2 Étude de la traduction d'un corpus réel

Nous nous intéressons à l'étude de la traduction d'un corpus réel et précisément à la différence entre une traduction brute et une traduction fonctionnelle.

7.2.2.1 Différence entre une traduction brute et une traduction fonctionnelle

Afin d'obtenir un corpus fonctionnel en français, nous avons produit une première traduction, sachant que le français n'est pas notre langue maternelle. Cette traduction correspond bien à une traduction brute qui doit être révisée et corrigée par des Français (elle l'a été par des membres de notre équipe de recherche). Le résultat obtenu peut être considéré comme une traduction fonctionnelle.

Nous avons observé une énorme différence (du point de vue de la spontanéité et de la façon de parler) entre notre traduction brute et la traduction fonctionnelle, et ces différences peuvent jouer un rôle important dans la préparation du corpus souhaité et par conséquent dans l'évaluation du résultat obtenu.

L'ensemble des différences observées entre une traduction brute et une traduction fonctionnelle nous a permis de proposer l'hypothèse suivante :

Les traductions directes d'un corpus réel d'un sous-langage restreint constitué de phrases naturelles ne sont le plus souvent pas dans le sous-langage cible correspondant.

Nous montrons dans ce qui suit la validité de cette hypothèse et nous proposons une solution au problème suivant : *comment construire des corpus parallèles de SMS fonctionnellement équivalents dans des situations de e-commerce ?*

7.2.2.2 Arguments en faveur de l'hypothèse de « non-traductibilité directe des sous-langages d'énoncés spontanés »

La traduction directe d'un corpus réel de départ en arabe ne peut pas donner un résultat fonctionnel et jugé réel, que cette traduction soit faite par un système automatique (qui n'existe pas dans le cas du couple (arabe, français)), soit par un humain qui parle le français.

Cela est dû aux mêmes raisons de différences culturelles décrites précédemment et à d'autres causes comme l'utilisation d'un jargon spécifique aux domaines et/ou langages.

Il s'agit parfois aussi de l'utilisation des abréviations et des habitudes utilisées par des francophones dans leurs pays d'origine (ici, la France) et portées aussi vers le pays habité (ici, la Jordanie). Par exemple un Français en Jordanie va exprimer sa requête de recherche d'une voiture avec la spontanéité d'un Français pas celle d'un arabisant s'exprimant dans un français qui préserve le « style » jordanien.

7.2.3 Évaluation de cette hypothèse

Nous montrons que cette hypothèse est vérifiée sur le corpus de CATS, et généralisons ce résultat.

7.2.3.1 Méthode d'évaluation

Nous avons utilisé un petit ensemble de 200 SMS en français révisés et jugés fonctionnels. Nous avons utilisé les notions classiques **de rappel** et de **précision** définies ici comme suit :

Rappel=Nombre de traductions correctes/Nombre de traductions de référence
Précision=Nombre de traductions correctes/Nombre de traductions proposées

Afin de trouver les valeurs des mesures de rappel et précision, nous avons utilisé le calcul de distance d'édition entre deux chaînes de caractères en appliquant l'algorithme de Wagner & Fischer (Wagner and Fischer 1974). Cet algorithme peut déterminer une plus longue sous-chaîne (non connexe) commune à deux chaînes données (voir algorithme dans Figure 48).

Cette distance d'édition est une bonne métrique pour évaluer un alignement car elle représente l'effort qu'il faut fournir pour passer d'une chaîne à une autre. Les algorithmes d'alignement de 2

chaînes ont pour but de trouver un alignement pas forcément unique, qui minimise cette distance (Pineau 2004).

Les Figure 72 et Figure 73 (similaires aux Figure 50 et Figure 51) donnent deux représentations différentes pour un même exemple de calcul de distance, dont la valeur est égale à 19.

Traduction brute : Recherche de Mercedes 81 avec contrôle technique et toutes les options.

Traduction fonctionnelle : Recherche Mercedes 81 avec ctl tech et ttes options.

Représentation en suivi des modifications

Représentation : Recherche ~~de~~ Mercedes 81 avec ~~contrôle~~ technique et ~~toutes~~ les options

Préférence

	Coût
insertion	1
Suppression :	1
Changement :	1
Egalité :	0

Coût uniforme

distance de : 19

Figure 72 : exemple de calcul de distance avec une représentation en suivi des modifications

Comme le montre la Figure 74, une traduction brute produite par un non-Français est généralement très différente d'une traduction naturelle produite par un Français. La distance d'édition moyenne trouvée est de 21,88 (Hajlaoui 2006) pour 50 SMS pris aléatoirement du « CorpusEvalAr200SMS ».

Représentation sur 3 lignes

Recherche de Mercedes 81 avec contr

Recherche Mercedes 81 avec c t

Préférence

↓ : insertion 1 Coût unifor... Préciser les coûts

↓ : suppression 1

◇ : échange 1

|| : égalite 0

distance de : 19

Calculer Tracer Reset Fermer

Figure 73 : même exemple de calcul de distance avec une représentation sur 3 lignes

Traduction brute de NH	h_debut	h_fin	duree_incr	Traduction littérale de CB	h_debut_f	Traduction fonctionnelle de CB	Distance (brute, fonctionnelle)
						Distance moyenne	21,88
Recherche d'achat de voiture Toyota 4*4 une seule cabine	49	49	0	Recherche à l'achat une voiture Toyota 4*4 à une seule cabine	46	Recherche à l'achat une Toyota 4*4	29
Recherche Kia 95 première achat et le reste sous forme de tranches	49	49	0	Recherche Kia 95 première main	46	Recherche Kia 95 première main	38
A vendre Renault Megane m 2000	49	49	0	À vendre Renault Megane modèle 2000	46	À vendre Renault Mégane modèle 2000	7
Toyota corolla modèle 2004 couleur grise assurance complete	49	49	0	Toyota corolla modèle 2004 couleur grise, assurance complète	46	Toyota corolla modèle 2004 couleur grise, assurance complète	4

Figure 74 : exemples de traductions brutes, littérales et fonctionnelles

7.2.3.2 Résultats et conclusion

Avec un coût de suppression, d'insertion ou d'échange fixé à 1, nous avons les valeurs suivantes :

Distance minimale = 0
 Distance maximale = 88
 Rappel = 0,04

Distance moyenne = 21,88
 Distance médiane = 17,5

La valeur de rappel trouvée signifie qu'il y a très peu de correspondances entre les traductions brutes et les traductions fonctionnelles de référence.

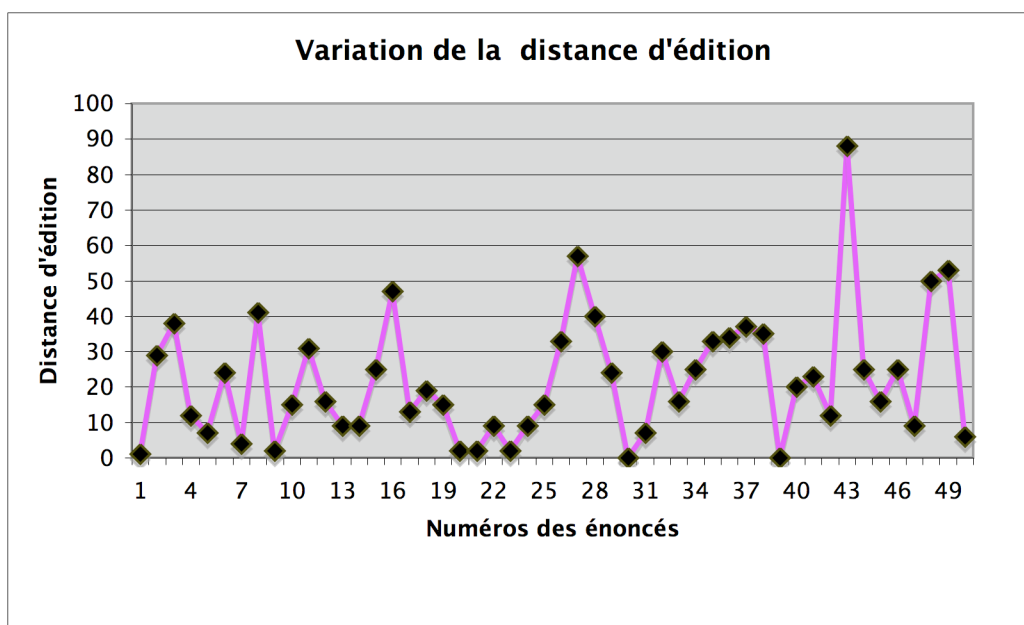


Figure 75 : mesures de distance d'édition

La Figure 75 montre la distance d'édition pour une cinquantaine de SMS. Notons que la plupart des énoncés pour lesquels la distance d'édition est très proche de zéro sont des SMS de taille très courte et de structure très simple, qui ne demandent pas d'effort de rédaction pour imiter la façon d'écrire d'un Français.

Par exemple, la traduction brute du SMS numéro 8 est « *A vendre Kia rio 2003* », une traduction fonctionnelle de ce SMS peut être « *À vendre Kia Rio 2003* ». Il s'agit d'un court SMS de structure très simple, la valeur de la distance correspondante est alors très faible (égale à 2).

Nous mesurons la similarité des traductions brutes et fonctionnelles comme suit :

$$S(i, j) = 1 - \frac{\text{edit}(i, j)}{\text{longueur}(i) + \text{longueur}(j)}$$

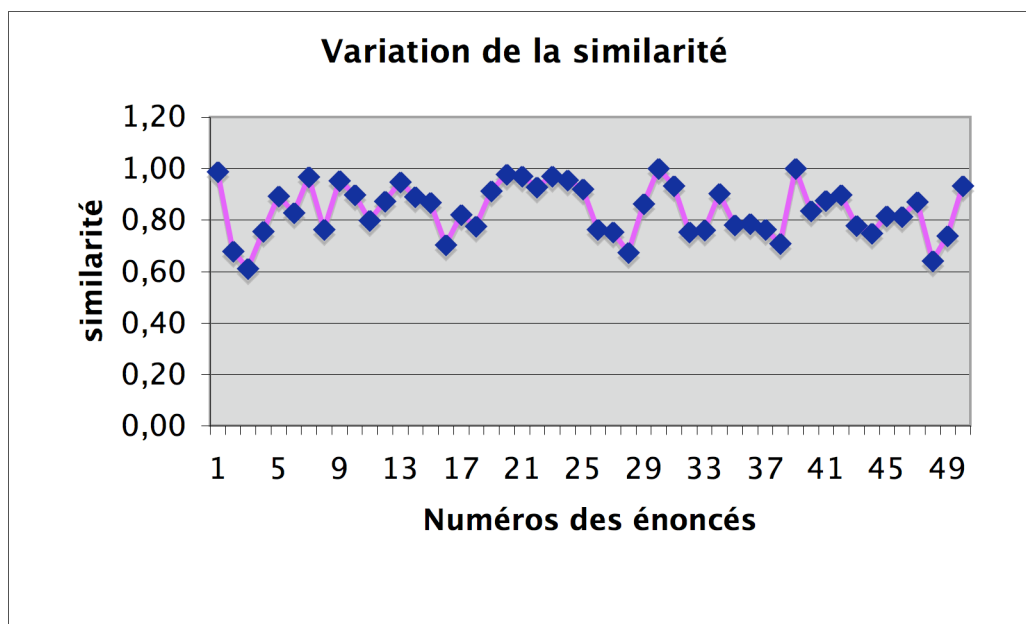


Figure 76 : mesures de similarité

Si on fixe l'acceptabilité à 0,90, on voit qu'en moyenne on est inacceptable (similarité moyenne égale à 0,84).

La valeur de similarité est très faible, cela signifie que notre hypothèse est vérifiée pour ce sous-langage.

On ne peut évidemment pas prouver cette hypothèse en passant en revue tous les sous-langages d'énoncés spontanés possibles. Mais le cas étudié montre pourquoi on peut penser que cette hypothèse est vraie dans le cas général :

- la formation des termes spécifiques en SL1 et SL2 à partir de L1 et L2 n'a pas de raison d'être parallèle (ex : first buy, première main),
- de même les transformations de structures syntaxiques standard vers des structures spécifiques n'ont pas de raison d'être parallèles (ex : recherche d'achat → recherche à l'achat).

7.2.3.3 Production de corpus réel

Tenant compte de cette hypothèse, nous avons essayé de produire un petit corpus français fonctionnellement équivalent au corpus arabe initial. Afin de développer ce corpus, nous avons adopté la technique suivante : à partir d'un petit ensemble de traductions révisées et jugées fonctionnelles, nous avons construit un ensemble plus grand en formant des combinaisons différentes des arguments utilisés (type, modèle, année, couleur, prix...). Par exemple, on remplace une année par une autre (*je cherche une voiture modèle 98*) → (*je cherche une voiture modèle 99* ...) ou une marque par une autre (*A vendre BMW rouge*) → (*A vendre PEUGEOT rouge*), etc.

Nous avons amélioré encore le corpus obtenu pour pouvoir tester l'extracteur de contenu du système CATS porté vers le français en supprimant beaucoup de mots inutiles (articles, prépositions longues, etc.), en introduisant des abréviations usuelles dans les petites annonces, et en ajoutant des abréviations de type phonétique ("ke" pour "que", "g" pour "j'ai", etc.).

Nous avons traduit manuellement le corpus d'évaluation « CorpusEvalAr200SMS » utilisé pour l'évaluation de la version arabe (originale) du système. Rappelons que c'est un corpus constitué de 200 SMS réels (100 SMS d'achat + 100 SMS de vente) envoyés par des utilisateurs réels en Jordanie.

Nous avons mis au total 289 mn pour traduire les 200 SMS arabes (2082 mots, à raison de 10 mots/SMS, soit environ 8 pages standard³¹) de l'arabe vers une traduction jugée brute, soit 35 mn par page. Nous avons mis 10 mn par page standard pour passer d'une traduction brute à une traduction fonctionnelle. Nous avons obtenu 200 SMS français jugés fonctionnels (1361 mots, soit 6,8 mots/SMS, environ 5 pages standard).

La différence en nombre de pages s'explique par le fait qu'on a éliminé tout le bruit, les erreurs et les expressions incompréhensibles pour passer à une traduction fonctionnelle.

7.3 Dictionnaire

Nous sommes partis de 1914 entrées du dictionnaire arabe concernant le domaine de l'automobile. Nous avons détecté 638 formes de base. Nous avons fabriqué 638 formes de base analogues en français.

Ensuite, nous avons cherché à trouver les différentes données dérivées qui peuvent être connectées à chacune des formes de base françaises. Nous avons produit de nouvelles entrées en introduisant des erreurs, des variations en genre (masculin/féminin), en nombre (singulier/pluriel), en casse (minuscule/majuscule), et en proposant aussi des transcriptions multiples des noms étrangers.

```
[ALFA ROMEO] {} "ALFA ROMEO (country>Italy, country>europe) " (make, car) <F, 3, 3>;
[Alfa Romeo] {} "ALFA ROMEO (country>Italy, country>europe) " (make, car) <F, 3, 3>;
[Alfa] {} "ALFA ROMEO (country>Italy, country>europe) " (make, car) <F, 3, 3>;
[ALphA ROMEO] {} "ALFA ROMEO (country>Italy, country>europe) " (make, car)
<F, 3, 3>;
[Alpha Romeo] {} "ALFA ROMEO (country>Italy, country>europe) " (make, car)
<F, 3, 3>;
[Alpha] {} "ALFA ROMEO (country>Italy, country>europe) " (make, car) <F, 3, 3>;
*****

[moteur] {} "motor" (motor, car) <F, 1, 1>;
[Moteur] {} "motor" (motor, car) <F, 1, 1>;
[engine] {} "motor" (motor, car) <F, 1, 1>;
[MOTEUR] {} "motor" (motor, car) <F, 1, 1>;
[Engine] {} "motor" (motor, car) <F, 1, 1>;
*****

[automatique] {} "automatic gear" (feature, gear, car) <F, 3, 3>;
[Automatique] {} "automatic gear" (feature, gear, car) <F, 3, 3>;
[automatic] {} "automatic gear" (feature, gear, car) <F, 3, 3>;
[boite vitesse automatique] {} "automatic gear" (feature, gear, car) <F, 3, 3>;
[B V automatique] {} "automatic gear" (feature, gear, car) <F, 3, 3>;
[BV automatique] {} "automatic gear" (feature, gear, car) <F, 3, 3>;
*****

[climatisée] {} "air condition" (feature, air, car) <F, 3, 3>;
[avec climatisation] {} "air condition" (feature, air, car) <F, 3, 3>;
[air conditionné] {} "air condition" (feature, air, car) <F, 3, 3>;
[Air conditionné] {} "air condition" (feature, air, car) <F, 3, 3>;
[avec conditionnement d'air] {} "air condition" (feature, air, car) <F, 3, 3>;
```

Figure 77 : quelques exemples d'entrées dictionnaires secondaires

³¹ Une page standard contient 250 mots.

Nous sommes ainsi arrivé à 1761 entrées en français, soit un coefficient d'expansion lexicale de 2,7. Ce coefficient est presque égal à celui de l'arabe (égal à 3). La Figure 77 présente quelques exemples de ce qu'on a pu imaginer comme entrées secondaires liées aux différentes raisons possibles.

Nous avons conservé la même structure du dictionnaire arabe. L'exemple de la Figure 78 montre qu'il faut tenir compte dans le dictionnaire du sous-langage français : un Français peut dire « *je cherche une A3...* » au lieu de dire « *je cherche une voiture AUDI A3...* ». Ainsi, on doit ajouter les entrées suivantes qui doivent pointer sur le même concept CRL-CATS.

```
[AUDI] {} "AUDI (country>germany, country>europe) " (make, car) <A, 3, 3>;
[A3] {} "AUDI (country>germany, country>europe) " (model, car) <A, 3, 3>;
[A4] {} "AUDI (country>germany, country>europe) " (model, car) <A, 3, 3>;
[A6] {} "AUDI (country>germany, country>europe) " (model, car) <A, 3, 3>;
```

Figure 78 : convergence d'entrées dictionnaires vers un même CW

7.3.1 Comparaison dictionnaire

Dans le travail de préparation du dictionnaire nécessaire à la tâche, le nombre d'entrées est dans certains cas beaucoup plus réduit que celui de l'arabe, car une seule entrée peut être écrite en arabe de plusieurs façons avec ou sans ECHAKEL³², avec ou sans ELHAMZA³³ et avec ou sans voyelles diacritiques, ce qui n'est pas le cas pour le français. Dans d'autres cas, le nombre d'entrées doit augmenter, car il faut tenir compte de la casse et des abréviations utilisées dans le sous-langage des SMS en français. Par exemple, « *cse départ* » à la place de « *cause départ* ». Pour l'arabe, le problème de la casse ne s'est pas posé étant donné qu'il n'y a pas cette distinction.

Une comparaison entre les coefficients d'expansion obtenus (2,7 pour le français et 3 pour l'arabe) montre une légère différence. Cela s'explique par le fait qu'on a peut-être oublié d'autres variantes d'entrées secondaires, qu'on pourra découvrir avec l'exploitation de la version française du système sur le terrain.

7.3.2 Recouvrement du domaine

Les expériences d'extraction de contenu que nous avons menées (voir Annexe III) montrent que les entrées dictionnaires obtenues couvrent bien le domaine de l'automobile. Au départ, leur nombre a augmenté en fonction de la taille du corpus puis s'est stabilisé après 650 à 700 SMS. Comme le montre la figure suivante représentant l'évolution de la taille des données en fonction du nombre des entrées dictionnaires, il semble se stabiliser à 1761 entrées.

³² Avec ou sans les voyelles.

³³ L'équivalent en arabe de la lettre « a ».

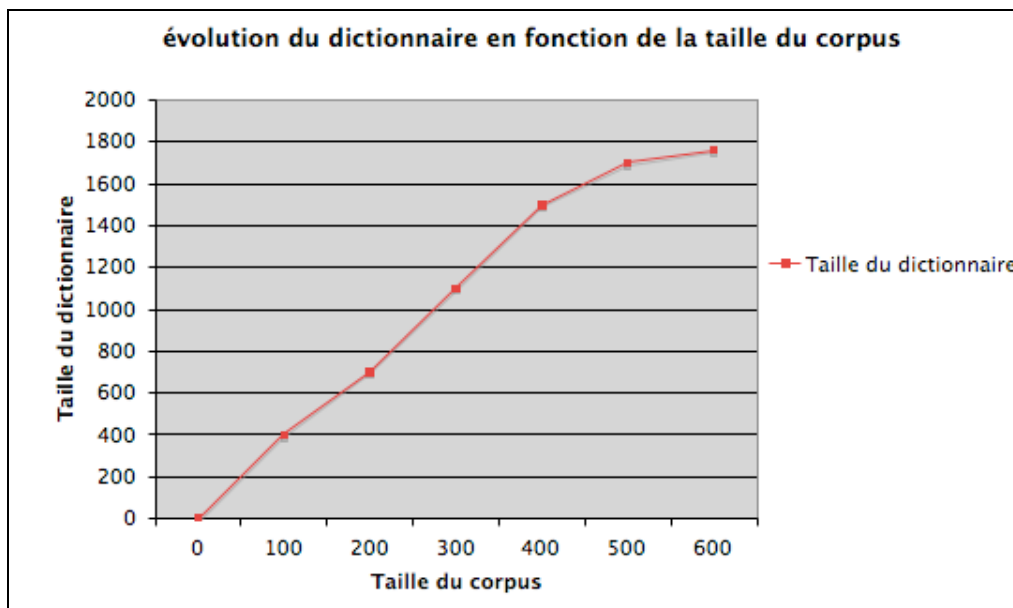


Figure 79 : évolution de la taille du dictionnaire en fonction du corpus

7.3.3 Conservation de l'utilisation de l'ontologie

La conservation de la même structure que celle du dictionnaire de CATS implique une conservation de l'utilisation de l'ontologie dans ce système. En effet, il y a bien une ontologie dans CATS même si elle n'est pas représentée séparément : on n'a pas seulement une hiérarchie d'objets et de classes comme *Megane(country>france, country>europe)* ou *Megane(country>france)* mais on sait qu'une voiture a une cylindrée, une couleur, un âge, etc.

Des processus externes aux objets, dans lesquels une propriété n'est pas inhérente à une voiture, existent aussi, comme celle du contrôle technique :

"contrôle technique de moins d'un an"

L'ontologie dans CATS est très partiellement explicitée dans la dénotation des concepts (CW) par des restrictions (*mak>Renault, country>france*) qui sont incomplètes, mais cela est compensé par le fait que CATS sait que France est dans Europe.

L'ontologie dans CATS est utilisée en 3 points, de façon procédurale.

- Création manuelle des concepts : choix manuel des restrictions liées à l'ontologie.
- Traitement d'informations incomplètes (élisions) par l'utilisation de ce qui est déjà reconnu : typage simple, par exemple par intervalles (année, prix...).

s'il reste "2000" et si on a l'année et le prix → c'est la cylindrée.

s'il reste "1850" et si on a le prix → c'est la cylindrée, pas l'année !

- Inférence d'informations implicites : inférences à la création de requêtes SQL. Comme le montre l'exemple de la Figure 80, CATS ajoute des informations complémentaires lors de la création de la requête. Par exemple, dans l'énoncé de la figure suivante, on ne trouve pas d'information indiquant le pays "Japon", mais CATS sait que "Mitsubishi" est une voiture japonaise. Cela permettra de répondre positivement à une requête de recherche de voiture dont les critères sont limités au choix du pays, par exemple de type "*Recherche voiture japonaise*".

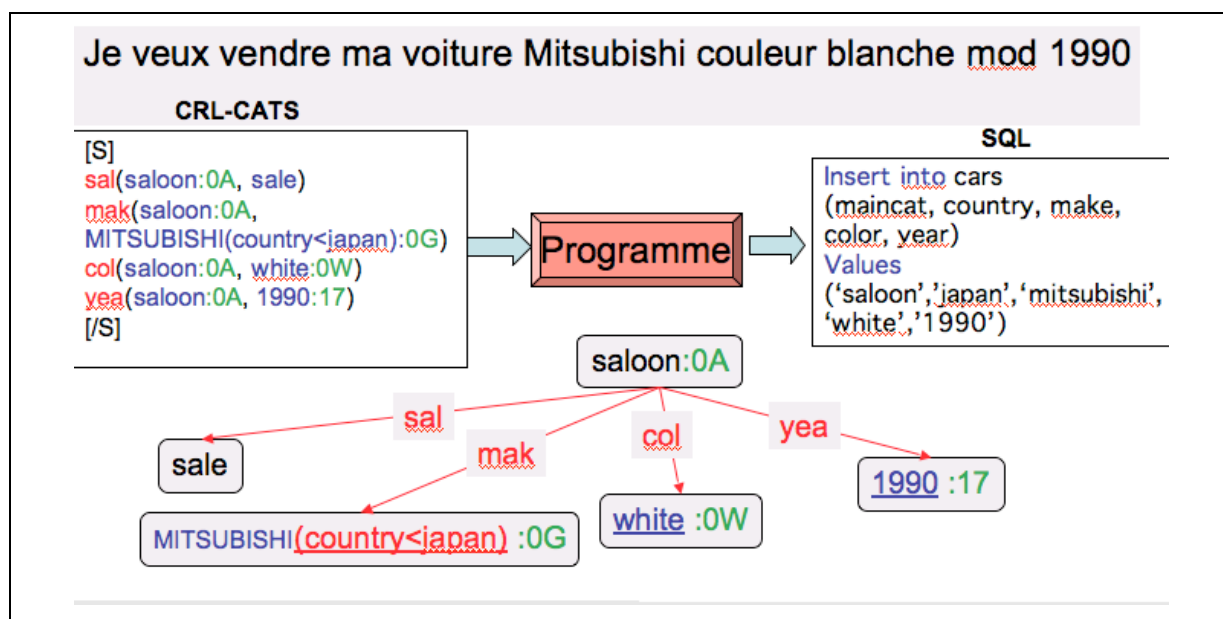


Figure 80 : passage du sens à la signification en utilisant l'ontologie

Chapitre 8 Portage « interne »

Introduction

Nous mettons ici en œuvre la première méthode de portage étudiée dans la section 3.1.2.1 en adaptant l'extracteur de contenu de CATS au français.

Dans la première partie, nous décrivons l'architecture et le fonctionnement de l'outil EnCo utilisé pour extraire le contenu des SMS arabes. Dans la deuxième partie, nous présentons le travail d'adaptation que nous avons effectué pour obtenir un extracteur de contenu pour les SMS français. Nous consacrons la troisième partie à l'évaluation des résultats en comparant les résultats obtenus avec la version originale et avec ceux d'une version de référence.

8.1 Extracteur de contenu pour les SMS en arabe

L'outil EnCo (Uchida and Zhu 1999) est un LSPL développé dans le cadre du projet UNL (Uchida, Zhu et al. 2005-2006) pour écrire des « enconvertisseurs » vers le langage pivot UNL.

8.1.1 Architecture générale de l'outil EnCo

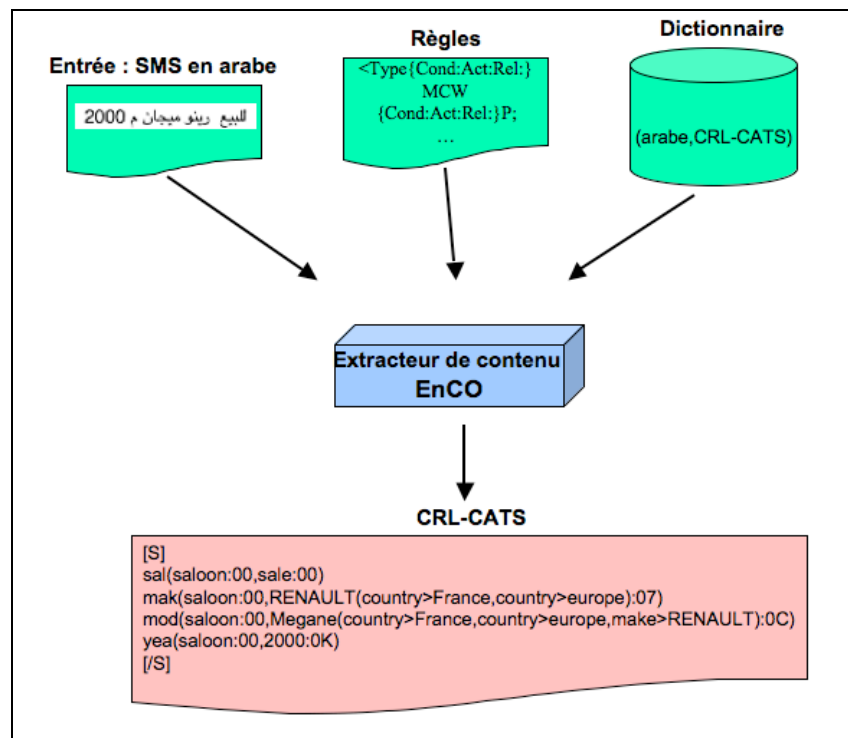


Figure 81 : architecture de l'extracteur de contenu pour l'arabe

EnCo a été utilisé dans CATS pour produire une représentation syntaxiquement semblable à UNL (Uchida and Zhu 2003), mais qui ne correspond pas du tout à la représentation UNL standard. En

effet, CRL-CATS (voir section 5.1.2.2) n'est pas un graphe qui représente l'analyse sémantique d'un énoncé, mais une structure représentant le contenu, sans lien avec un énoncé de surface en anglais.

La Figure 81 résume l'ensemble des éléments mis à notre disposition au départ du travail de portage interne. Cet ensemble est constitué essentiellement de l'outil EnCo, d'un dictionnaire et de règles.

8.1.2 Fonctionnement d'EnCo

EnCo attend en entrée :

- Un dictionnaire et une grammaire (linguiciel).
- Un texte découpé en phrases.
- Éventuellement une « base de connaissances » (les probabilités des triplets relations, UNL, UWL) non utilisée par CATS.

Le linguiciel est compilé, puis chaque phrase est traitée successivement. Les structures de données manipulées par EnCo sont :

- Une liste de nœuds avec deux têtes de lecture/écriture placées sur deux nœuds successifs (LAW « Left Analysis Windows », RAW « Right Analysis Windows ») et deux têtes de lecture (LCW « Left Conditions Windows », RCW « Right Conditions Windows ») pour les contextes gauche et droit.
- Un graphe de nœuds, initialement vide, pouvant contenir des nœuds de la liste, et dont les arcs portent des « relations » identifiées par des symboles à trois caractères alphabétiques.

Au départ, la liste comporte trois nœuds : la limite gauche, le nœud courant et la limite droite. Le nœud courant contient comme chaîne la phrase à traiter.

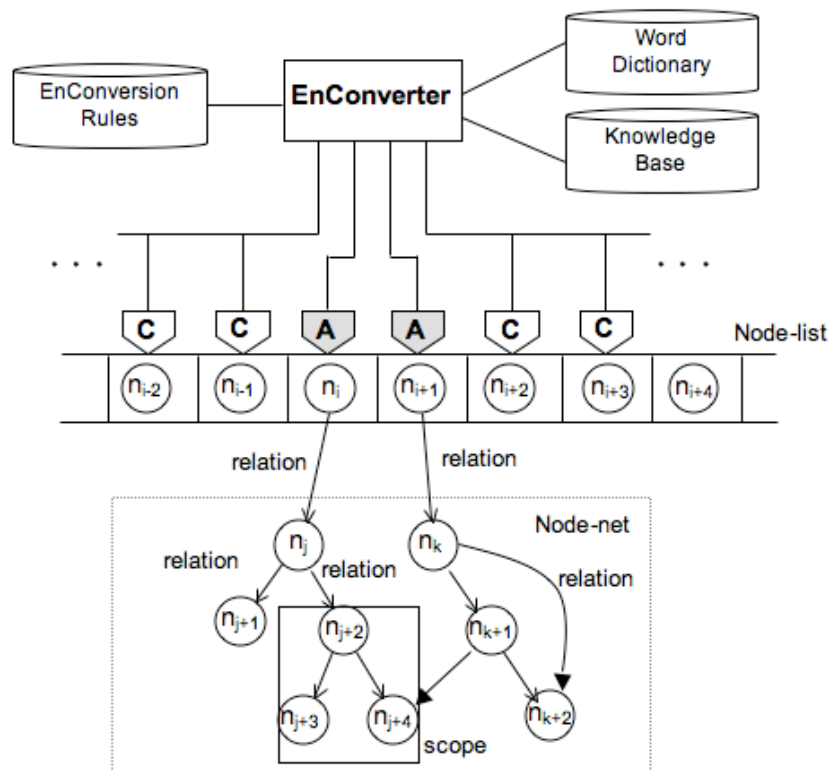


Figure 82 : structure d'EnCo, "A" indique la fenêtre d'analyse, "C" étiquette une fenêtre de Condition (Uchida and Zhu 1999).

De façon générale, un nœud peut contenir quatre éléments : une chaîne, un ensemble d'attributs « de chaîne » (initialisés lors des appels au dictionnaire), une UW (référence lexicale, venant du dictionnaire ou créée par une règle), et un ensemble d'attributs « de graphe » (préfixés par « @ »). Les attributs sont booléens, et ne sont pas déclarés. Seul « @entry » a un rôle spécial.

La Figure 83 montre la structure d'EnCo. EnCo utilise les fenêtres de condition pour tester si les nœuds voisins des deux côtés des fenêtres d'analyse remplissent les conditions pour l'application d'une règle d'analyse. Les fenêtres d'analyse sont utilisées pour vérifier l'existence de deux nœuds adjacents afin d'appliquer une règle d'analyse. S'il existe une règle applicable aux nœuds courants, EnCo ajoute ou supprime pour ces derniers nœuds les propriétés grammaticales indiquées par la règle appliquée et/ou insère un nouveau nœud dans le graphe, ainsi qu'une relation sémantique, selon le type de la règle.

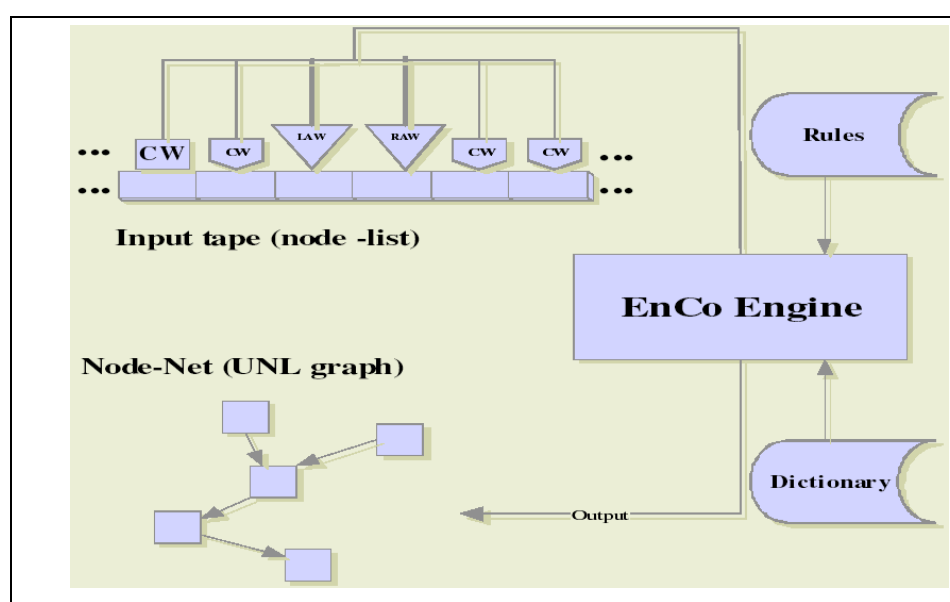


Figure 83 : fonctionnement d'EnCo

8.1.3 Règles

La syntaxe des règles d'EnCo est la suivante (Uchida, Zhu et al. 2005-2006) :

```
<TYPE>... (<PRE2>) (<PRE1>) {<LNODE>} {<RNODE>} (<SUF1>) (<SUF2>)... P<PRI>;
avec
<LNODE>:="{ " [<COND1>] ":" [<ACTION1>] ":" [<RELATION1>] ":" [<ROLE1>] "}"
<RNODE>:="{ " [<COND2>] ":" [<ACTION2>] ":" [<RELATION2>] ":" [<ROLE2>] "}"
```

Figure 84 : syntaxe d'une règle en EnCo

Notons que, pour une règle d'analyse donnée, il est possible d'insérer (respectivement de supprimer) un seul nœud dans (respectivement à partir de) la liste des nœuds. Ici, LNODE fait référence au nœud sous la fenêtre d'analyse gauche (LAW) et RNODE fait référence au nœud sous la fenêtre d'analyse droite (RAW).

L'interprétation générale de la syntaxe d'une règle en EnCo est alors la suivante.

Une règle peut s'appliquer si sous la fenêtre d'analyse gauche (LAW) se trouve un nœud qui satisfait la condition <COND1> et sous la fenêtre d'analyse droite (RAW) se trouve un nœud qui satisfait la condition <COND2>. Quand il y a des nœuds qui remplissent les conditions trouvées dans <PRE1> et <SUF1> du côté gauche (respectivement dans <PRE2> et <SUF2> du côté droit) des

fenêtres d'analyse, les propriétés grammaticales dans les fenêtres d'analyse sont réécrites selon les actions <ACTION1> (respectivement <ACTION2>). Nous illustrons cette syntaxe par l'exemple de la section 8.2.1.

Le champ <TYPE> porte le type de la règle à appliquer, qui indique l'opération qui doit être effectuée dans la liste des nœuds (par exemple une opération d'insertion, de suppression, etc.).

<COND1> et <COND2> sont des conjonctions de conditions élémentaires testant la présence (ATTR) ou l'absence (^ATTR) de certains attributs grammaticaux.

<ACTION1> et <ACTION2> contiennent des attributs grammaticaux à insérer ou à supprimer dans des nœuds sous les fenêtres d'analyse.

Les champs <RELATION1> et <RELATION2> sont utilisés pour créer des relations UNL entre les nœuds sous les fenêtres d'analyse.

<ROLE1> et <ROLE2> sont des attributs de la base de connaissances qui sont optionnels et qui ne sont pas utilisés dans CATS.

<PRI> indique la valeur de priorité de la règle, qui doit être comprise entre 0 et 255. Une règle dont la valeur de priorité n'est pas indiquée est considérée comme une règle de priorité 0.

Les 710 règles utilisées dans le système CATS réalisent l'extraction des informations utiles, et non pas l'analyse linguistique au sens classique. Elles affectent des valeurs à des objets préfinis dans le dictionnaire pour construire des relations semblables au type *Propriété{objet, valeur}* en les représentant dans le graphe construit par *objet --propriété→ valeur*. L'ensemble de ces relations forme la représentation CRL-CATS.

8.1.4 Dictionnaire

EnCo recherche dans le dictionnaire tous les lexèmes candidats à partir du premier caractère, il leur affecte ensuite des priorités basées sur la fréquence d'apparition et la longueur du lexème comme suit :

D'abord, il accorde la priorité la plus élevée à l'entrée qui possède la fréquence d'apparition la plus élevée. Ensuite, en cas de plusieurs candidats (qui ont la même fréquence la plus élevée), il choisit le lexème le plus long.

En cas d'échec ultérieur, ou d'activation directe d'un retour arrière, le système effectue un retour arrière (backtrack) et le lexème suivant dans la liste est alors découpé, puis le traitement reprend sur cette nouvelle « branche » du calcul. L'extracteur de contenu de CATS est construit pour utiliser cette possibilité le plus rarement possible.

Les entrées du dictionnaire ont la syntaxe suivante :

```
[HW]"UW" (ATTR1, ATTR2, ...) <FLG, FRE, PRI>;
HW : (Head Word) est un lexème qui peut être un candidat
UW : (Universal Word) est une référence lexicale
ATTR1, ATTR2 sont des attributs
FLG : indique la langue
FRE : indique la fréquence
PRI : indique la priorité
```

Figure 85 : syntaxe d'une entrée dictionnaire

Pour un blanc, et dans le code ASCII, EnCo crée automatiquement l'entrée suivante dont l'UW est la chaîne vide.

```
[ ]{} "" (BLK) <., 0, 0>;
```

Le dictionnaire utilisé dans la version arabe et pour l'ensemble des domaines (automobile, immobilier, divers...) a environ 10.000 CW et 30.000 lexèmes, dont 20.000 ont été générés automatiquement.

8.2 Adaptation de l'extracteur de contenu

Pour porter l'extracteur de contenu vers le français, nous avons effectué essentiellement un travail dictionnaire que nous avons décrit dans le Chapitre 7, et un travail de modification de règles que nous illustrons par un exemple. L'architecture générale reste la même.

8.2.1 Exemple de traitement

Afin de mieux comprendre le fonctionnement de l'outil EnCo, nous en montrons le détail sur un petit exemple de SMS en français « *recherche voiture* ».

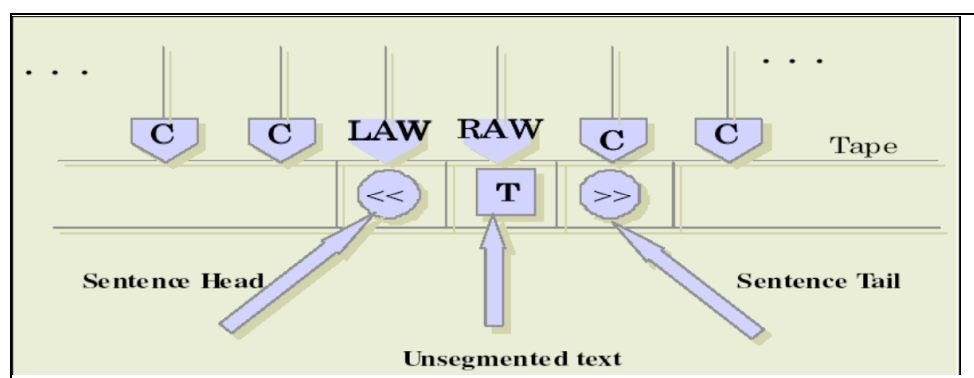


Figure 86 : configuration initiale d'EnCo (Daoud 2006)

Comme le montre la Figure 86, initialement la fenêtre d'analyse gauche (LAW) contient le symbole << appelé SHEAD et la fenêtre d'analyse droite (RAW) contient le texte non segmenté, c'est-à-dire « *recherche voiture* ».

La Figure 87 montre la configuration initiale. LAW pointe sur le symbole « << » et le nœud sous RAW contient la chaîne « *recherche voiture* ». EnCo consulte alors le dictionnaire pour chercher les préfixes de cette chaîne.

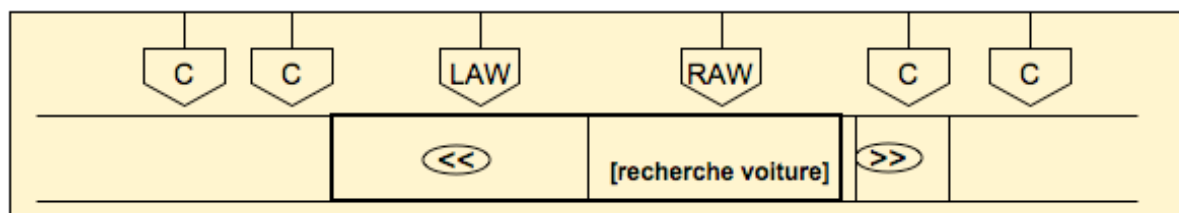


Figure 87 : configuration initiale de « *recherche voiture* »

Un seul article de dictionnaire est trouvé :

```
[recherche] {} "wanted" (want) <F,1,1>;
```

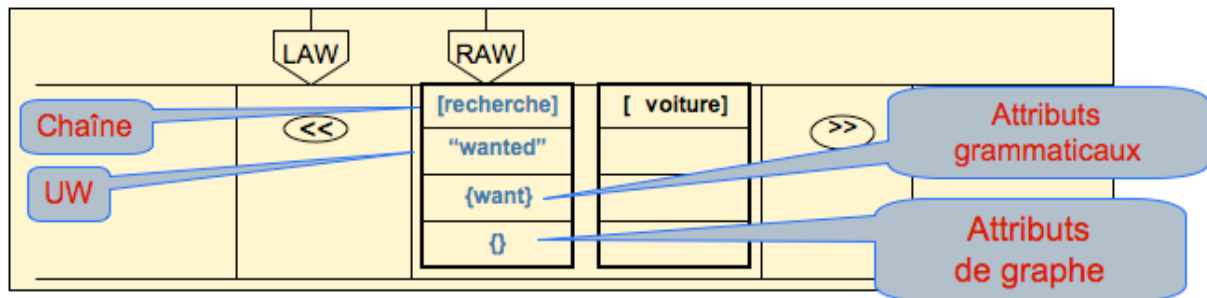


Figure 88 : segmentation de « recherche voiture »

La Figure 88 montre le résultat de segmentation de « recherche voiture ». RAW pointe sur un nœud contenant la chaîne « recherche », l'UW « wanted », et l'attribut dictionnaire « want », et aucun attribut de graphe. Par suite de la segmentation lexicale, le nœud initial a été divisé en deux nœuds, le deuxième contenant seulement la chaîne restant à analyser [voiture]. A ce niveau, EnCo cherche à trouver une règle qui satisfait la configuration actuelle : LAW pointe sur le symbole « << » et RAW pointe sur le nœud [recherche] {} "wanted" (want).

La première règle qui peut être appliquée est :

```
R{SHEAD:::}{want:::}P20;
```

Cette règle fait un « *shift right* » désigné par « R » (TYPE = R). P20 indique la priorité affectée à cette règle. Elle est appliquée sans aucune condition (car COND1 et COND2 sont vides) et sans aucun changement de propriétés grammaticales (car ACTION1 et ACTION2 sont vides). Comme le montre la Figure 89, après l'application de cette règle, LAW pointe sur « recherche » et RAW pointe sur un blanc « BLK ».

En effet, le « *shift* » provoque aussi la segmentation d'un préfixe, ici le blanc.

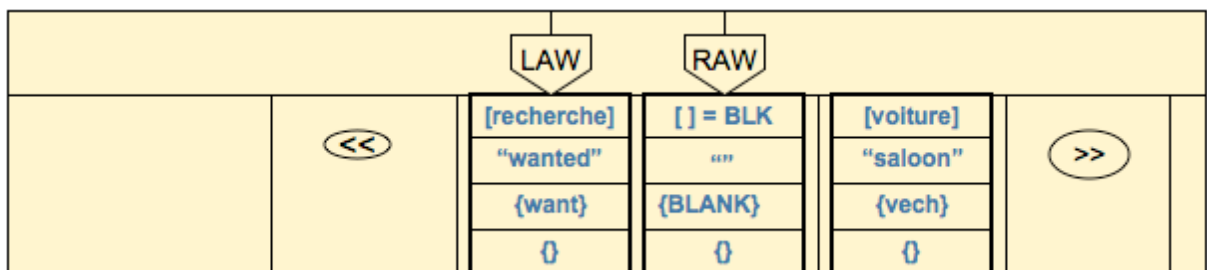


Figure 89 : configuration d'EnCo après un *shift right*

A ce moment, la première règle applicable est:

```
+{:+BLANK:::}{BLK:::}P255;
```

- Type de l'opération = + qui signifie la combinaison du nœud droit avec le nœud gauche (autrement dit, ajouter l'attribut de chaîne BLANK au nœud à gauche).
- Cond1 = rien
- Cond2 = BLK (présence de l'attribut BLK)
- Action1 = +BLANK (ajout du symbole BLANK à la liste des attributs du nœud à gauche).
- Action2 = rien
- P255 = Priorité 255 (élevée)

Suite à l'exécution de cette règle, LAW pointe sur le symbole SHEAD et RAW pointe sur le nœud « recherche ». EnCo fait ensuite un *shift right* en appliquant la règle suivante :

```
R{:::}{:::}()P1;
```

La Figure 90 montre l'état d'EnCo après l'exécution des deux précédentes règles. LAW pointe sur le nœud contenant la chaîne [recherche] et RAW pointe sur le dernier nœud, contenant (après segmentation) les informations du dictionnaire sur la chaîne [voiture]. EnCo n'a pas créé de nouveau nœud car la chaîne restante est vide.

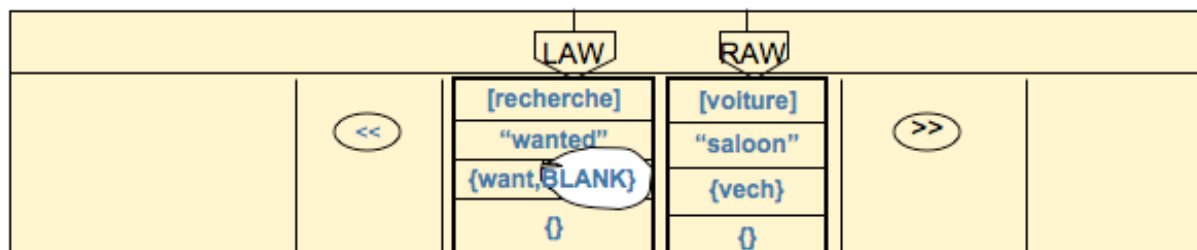


Figure 90 : suppression du BLANK et shift right

A ce niveau, nous avons besoin d'une règle qui crée, dans le graphe des nœuds, une relation [wan] qui part du nœud contenant [voiture] et arrive au nœud contenant [recherche], puisque LAW pointe sur un nœud contenant {want, BLANK} et RAW pointe sur un vech. Cette règle est la suivante :

```
>{want, BLANK:-want:want:}{vech, ^want_add:want_add:}()P70;
```

- Type d'opération = modification à droite désignée par > (le nœud gauche est supprimé de la liste des nœuds, le nœud à droite devient l'origine de l'arc inséré, portant le symbole wan, et allant vers l'ancien nœud gauche).
- Cond1 = want, BLANK (le nœud sous LAW doit contenir l'attribut want et un attribut BLANK).
- Cond2 = vech, ^want_add (le nœud sous RAW doit contenir l'attribut vech et ne doit pas contenir l'attribut want_add).
- Action1 = -want (supprimer l'attribut want au nœud gauche (placé sous LAW)).
- Action2 = want_add (ajouter l'attribut want_add au nœud droit (placé sous RAW)).
- RELATION1 = wan (la relation wan va du nœud droit au nœud gauche).

La Figure 91 montre le résultat de l'application de cette dernière règle.

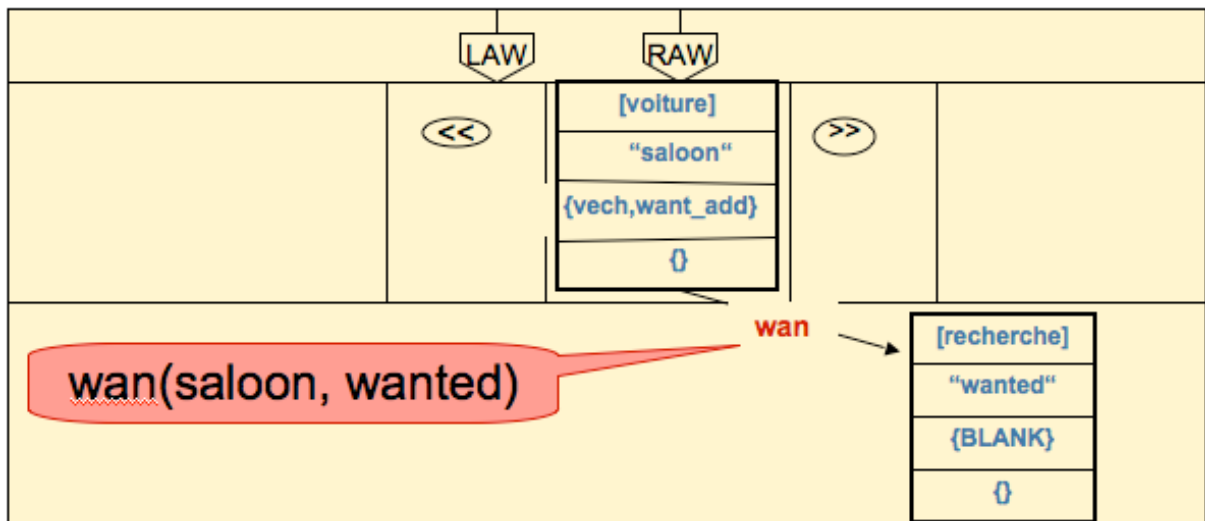


Figure 91 : création d'une relation dans le graphe des nœuds

Le résultat final est :

```

;===== UNL =====
;recherche voiture.
[S]
wan(saloon:0A,   wanted:00)
[/S]
;=====

```

8.2.2 Légère modification des règles

La bonne surprise de ce travail est que nous n'avons dû modifier que légèrement les règles fabriquées initialement pour la version arabe, et que l'extracteur de contenu obtenu fonctionne bien (voir résultat de l'évaluation dans la section 8.3) pour le sous-langage correspondant du français, celui des SMS spontanés pour l'achat et la vente des voitures d'occasion.

Comme le montre la Figure 92, cela confirme les théories linguistiques (Kittredge and Lehrberger 1982a) selon lesquelles deux sous-langages équivalents dans deux langues différentes sont proches (très proches ici) l'un de l'autre, même si leurs deux langues mères sont éloignées.

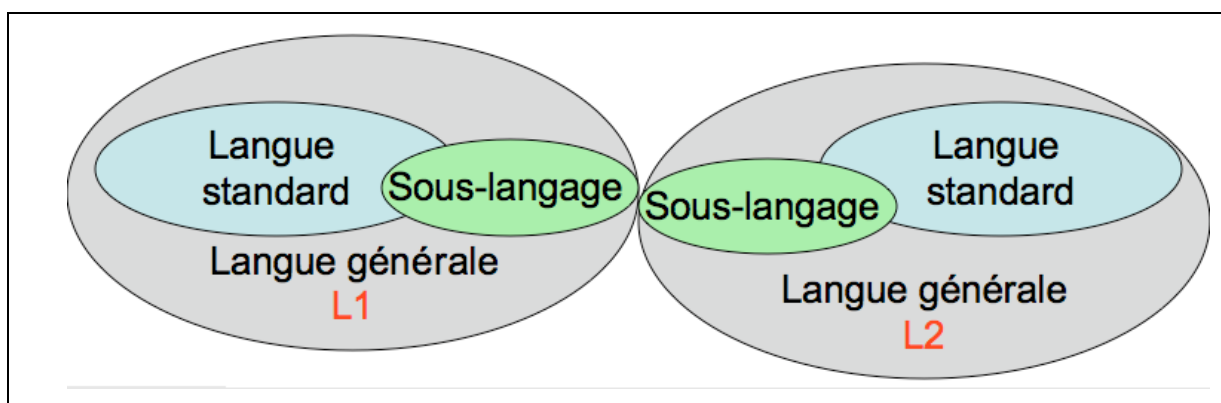


Figure 92 : les sous-langages sont proches

8.2.3 Exemple d'ambiguïté

Les résultats obtenus avec la version adaptée sont encourageants, mais il reste à résoudre quelques problèmes d'ambiguïté portant sur la sémantique des mots, par exemple le mot JEEP.

Pour le SMS suivant : « *Recherche voiture JEEP* », la version originale et la version française de CATS donnent le même résultat :

```

;Recherche voiture JEEP
[S]
wan(saloon:0A,    wanted:00)
mak(saloon:0A,    JEEP(country>America):0L)
[/S]

```

Figure 93 : exemple d'ambiguïté (première interprétation)

Une ambiguïté purement sémantique se présente ici et deux interprétations différentes sont possibles.

- Une première interprétation pour le mot « JEEP » peut être une marque de voiture américaine comme c'est le cas dans le résultat précédent (Figure 93).
- Une deuxième interprétation pour le mot « JEEP » peut être un modèle de voiture de type Mercedes. Dans ce cas, la propriété (mak) relie l'objet saloon et la valeur Mercedes et une deuxième propriété (mod) doit exister et relie l'objet saloon et la valeur JEEP (Figure 94).

```

;Recherche voiture JEEP
[S]
wan(saloon:0A,    wanted:00)
mak(saloon:00,    MERCEDES(country>germany,country>europe):09)
mod(saloon:00,    JEEP(country>germany,country>europe,make>MERCEDES):0D)
[/S]

```

Figure 94 : exemple d'ambiguïté (deuxième interprétation)

Aucune solution n'est proposée pour ce type de problème d'ambiguïté sémantique, ni dans la version originale, ni dans la version portée.

8.3 Résultats et évaluations

Comme nous l'avons expliqué dans la définition de notre démarche, nous évaluons ici le coût du portage en termes du temps mis pour effectuer la tâche, et les résultats obtenus en termes de la qualité d'extraction du contenu.

8.3.1 Coût du portage interne

Le Tableau 30 détaille les différents coûts en termes du temps passé à effectuer les différentes tâches dans le cadre du portage interne du système CATS.

Le travail du portage interne a demandé au total environ 1,9 homme/mois de linguiste-informaticien dont une bonne partie a été consacrée à la compréhension de l'environnement CATS, de la spécification CRL-CATS, et du fonctionnement de l'outil d'extraction de contenu utilisé (EnCo).

Le travail dictionnaire a demandé un total de 100 heures pour la construction du dictionnaire (français CRL-CATS) et son enrichissement pour couvrir le domaine de l'automobile.

Tâches	Durée (h)	Version	Remarques
Documentation sur CATS	18	version 2005	Support papier
Documentation sur la spécification CRL-CATS	14	version 2005	Support papier
Documentation sur la spécification EnCo	35	version 5/12/2002	Version papier + électronique
Compréhension du fonctionnement d'EnCo	15		Fonctionnement général (orienté vers UNL)
Petits tests	10		
Compréhension des règles dans CATS	30		
Modification des règles	15		
Travail dictionnaire	80		Construction du dictionnaire français
Premiers résultats	15		Résultats encourageants
Enrichissement du dictionnaire	20		Ajout de nouvelles entrées
Tests et évaluation	15		Résultats satisfaisants
Total	267		

Tableau 30 : évaluation de la durée des tâches pour le portage interne de CATS

Comme nous l'avons détaillé dans la section 6.2.2, nous utilisons trois mesures pour évaluer les résultats d'extraction de contenu obtenus : une « mesure informationnelle » basée sur le corpus nommé « CorpusEvalFr200SMS » (un corpus d'évaluation limité à 200 SMS en français), une « mesure simple », et une « mesure fine » ; les deux dernières appliquées à tout le corpus nommé « CorpusEvalFr1100SMS », constitué du reste de l'ensemble des énoncés français dont nous disposons et obtenus par déconversion automatique et révision humaine (voir Chapitre 11).

8.3.2 Mesure informationnelle

Nous rappelons que cette mesure est basée sur le corpus d'évaluation « CorpusEvalFr200SMS » obtenu par traduction du corpus « CorpusEvalAr200SMS » utilisé pour l'évaluation de la version originale du système (voir section 6.2.2.1).

8.3.2.1 Évaluation par rapport à la version originale

Les résultats d'extraction de contenu de la version française et de la version originale (arabe) pour les propriétés les plus importantes (comme achat/vente, année, prix, marque, modèle) sont regroupés dans le Tableau 31. Les pourcentages de portage par rapport à la version originale varient entre 95% et 100%, avec une moyenne de 98 % (Hajlaoui and Boitet 2007).

Propriété	EnCoAR			EnCoFR			% du portage
	Précision	Rappel	F-mesure (EnCoAR)	Précision	Rappel	F-mesure (EnCoFR)	
Achat/vente	0,956	0,970	0,963	0,994	0,855	0,919	95
Année	0,817	0,960	0,883	0,879	0,819	0,848	96
Prix	0,800	0,822	0,811	0,789	0,822	0,805	99
Marque	0,978	0,963	0,970	0,961	0,961	0,961	99
Modèle	0,901	0,837	0,868	0,842	0,903	0,872	100
Moyenne	0,890	0,910	0,899	0,893	0,872	0,881	98

Tableau 31 : comparaison entre les résultats d'extraction de contenu pour 200 SMS d'achat et de vente par rapport à la version originale

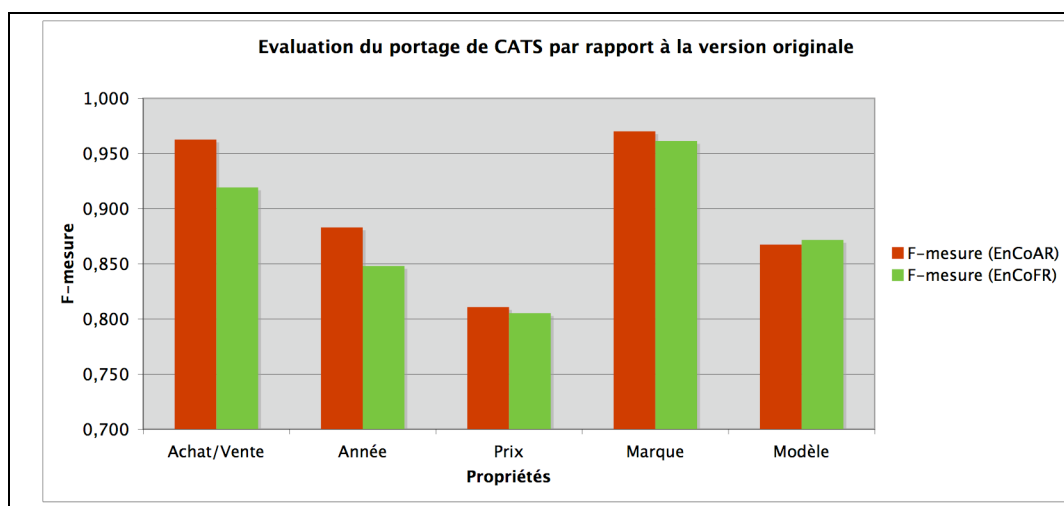


Figure 95 : comparaison entre F-mesure (par rapport à la version originale)

La Figure 95 permet de mieux visualiser la comparaison entre les valeurs de F-mesure trouvées pour le système original et le système porté.

8.3.2.2 Évaluation par rapport à une version de référence

Afin de comparer les résultats obtenus par rapport à une version de référence, nous avons corrigé manuellement les 200 CRL-CATS correspondant au corpus d'évaluation. Les types d'erreurs rencontrés sont essentiellement dans les propriétés dont les valeurs sont des chiffres comme le « prix » et « année ». Le Tableau 32 donne un exemple de correction de CRL-CATS.

SMS originale	مطلوب سياره ياباني أو الماني .0795777953
Traduction fonctionnelle	Recherche voiture japonaise ou allemande 0795777953
CRL-CATS	[S] wan(saloon:06, wanted:00) cou(saloon:06, japan:0I) yea(saloon:06, 0795777953:0Z) [/S]
CRL-CATS-REF	[S] wan(saloon:06, wanted:00) cou(saloon:06, japan:0I) cou(saloon:06, german:0I) [/S]

Tableau 32 : exemple de correction de la représentation CRL-CATS

Propriété	EnCoAR-Ref			EnCoFR			% du portage
	Précision	Rappel	F-mesure (EnCoAR-Ref)	Précision	Rappel	F-mesure (EnCoFR)	
Achat/Vente	1,000	0,970	0,985	0,994	0,855	0,919	93
Année	1,000	0,983	0,991	0,879	0,819	0,848	86
Prix	0,973	0,973	0,973	0,789	0,822	0,805	83
Marque	0,984	0,963	0,973	0,961	0,961	0,961	99
Modèle	0,969	0,901	0,934	0,842	0,903	0,872	93
Moyenne	0,985	0,958	0,971	0,893	0,872	0,881	91

Tableau 33 : comparaison entre les résultats d'extraction de contenu pour 200 SMS d'achat et de vente par rapport à une version de référence

Les résultats d'extraction de contenu de la version française et de la version de référence pour les mêmes propriétés les plus importantes sont regroupés dans le Tableau 33. Les pourcentages de portage par rapport à la version originale varient entre 83% et 99%, avec une moyenne de 91 %.

La Figure 96 permet une meilleure visualisation de la comparaison entre les valeurs de F-mesure trouvées pour la version française et la version de référence.

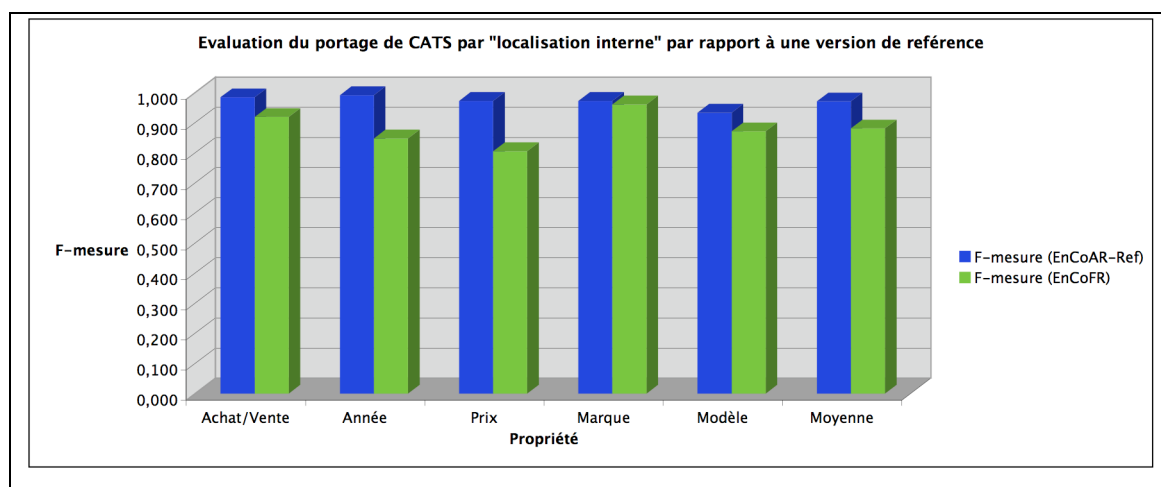


Figure 96 : comparaison entre F-mesures (par rapport à une version de référence)

8.3.3 Évaluation générale

8.3.3.1 Mesure simple

Nous avons appliqué la « mesure simple » (voir section 6.2.2.2) à cette méthode de portage. Les scores des attributs les plus importants ($m = 6$) sur le total du corpus ($n = 1100$) sont donnés dans le Tableau 34.

Propriétés	Version originale	Version candidate
	EnCoAr	EnCoFr
Score (wan) CorpusEvalFr1100SMS	49,33	60,67
Score (sal) CorpusEvalFr1100SMS	39,33	40,56
Score (yea) CorpusEvalFr1100SMS	98,22	108,44
Score (pri) CorpusEvalFr1100SMS	38,44	42,11
Score (mak) CorpusEvalFr1100SMS	124,67	119,67
Score (mod) CorpusEvalFr1100SMS	55,22	54,22
Score (wan, sal, yea, pri, mak, mod) CorpusEvalFr1100SMS	405,22	425,67

Tableau 34 : scores du portage interne sur la totalité du corpus de 1100 SMS

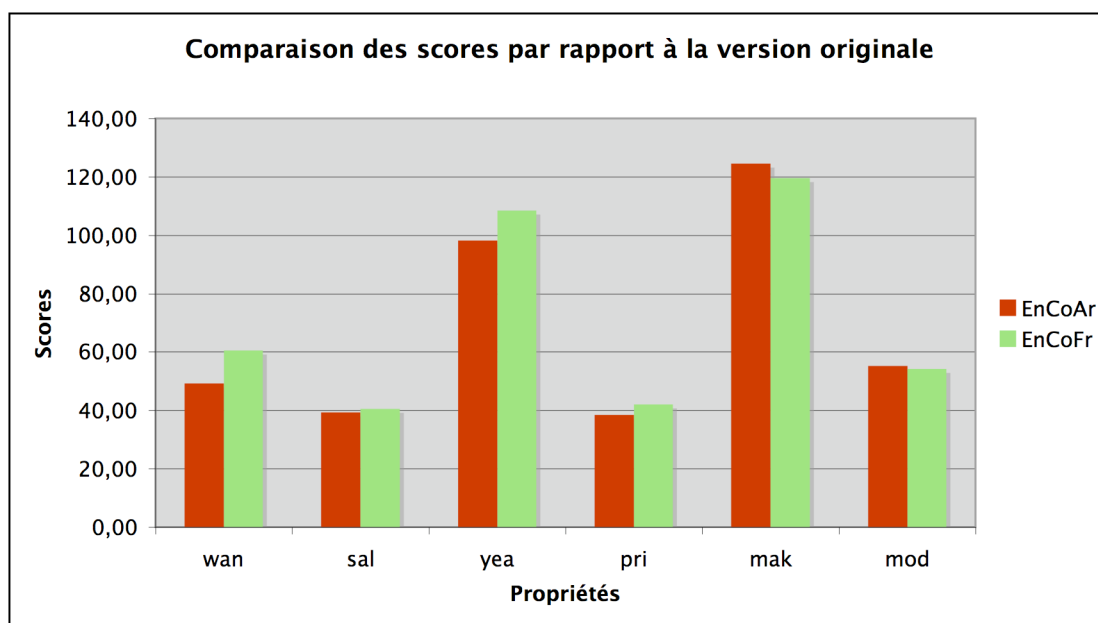


Figure 97 : comparaison des scores des attributs les plus importants par rapport à la version originale sur le corpus « CorpusEvalFr1100SMS »

Nous calculons ainsi le pourcentage de la différence par rapport à la version originale des scores trouvés, et trouvons les pourcentages du Tableau 35.

Propriétés	% de différence de EnCoFr par rapport à la version originale (EnCoAr)
wan	22,97
sal	3,11
yea	10,41
pri	9,54
mak	4,01
mod	1,81
Moyenne	8,64

Tableau 35 : comparaison des pourcentages de différences des scores d'extraction de contenu sur le corpus « CorpusEvalFr1100SMS »

En utilisant cette mesure, nous trouvons des résultats semblables à ceux trouvés sur le corpus d'évaluation limité à 200 SMS « CorpusEvalFr200SMS » et jugés par un humain. En effet, en passant à une extraction de contenu à partir d'un corpus plus grand, les différences de scores ne sont pas considérables par rapport à la version originale.

8.3.3.2 Mesure fine

Nous avons aussi appliqué la mesure fine décrite dans la section 6.2.2.3. Le Tableau 36 regroupe les valeurs moyennes de rappel, de la précision et de la F-mesure trouvées sur le corpus « CorpusEvalFr1100SMS » par rapport aux résultats de la version originale du système. La valeur de

F-mesure moyenne est de 0,718. Elle légèrement inférieure à celle trouvée par la mesure informationnelle (0,881).

	Rappel	Précision	F-mesure
Moyenne	0,690	0,658	0,718

Tableau 36 : résultats de la mesure fine sur le corpus « CorpusEvalFr1100SMS »

Conclusion

Nous avons présenté une première méthode de portage interne du système CATS qui consiste à localiser son extracteur de contenu prévu initialement pour des SMS en arabe. La bonne surprise dans ce travail est que cette méthode marche bien à cause de la proximité des deux sous-langages de l'arabe et du français et cela malgré la grande distance entre l'arabe et le français.

Nous avons effectué une comparaison des résultats obtenus (voir Annexe III) par rapport à la version originale et par rapport à une version de référence. En utilisant un jugement humain, nous avons obtenu un pourcentage de portage de 98% dans le premier cas et de 91% dans le deuxième cas.

Chapitre 9 Portage « externe »

Introduction

Avec la deuxième méthode de portage envisagée, nous nous mettons dans le cas où notre accès est limité à la représentation interne de l'application. Le portage est alors possible par adaptation au domaine de l'application d'un extracteur de contenu existant pour un autre domaine (tourisme) et pour la langue cible.

Dans la première partie, nous décrivons l'existant et le fonctionnement de cet extracteur de contenu dans le domaine du tourisme. Nous présentons ensuite le travail d'adaptation que nous avons effectué pour obtenir un extracteur de contenu pour les SMS concernant le domaine de l'occasion automobile. Nous détaillons dans la troisième partie la construction d'un traducteur qui permet de passer du format produit par cet extracteur de contenu (IF) à celui de l'extracteur de contenu natif (CRL-CATS).

Nous évaluons finalement les résultats obtenus en les comparant à ceux produits par la version originale de l'extracteur de contenu de CATS et ceux produits par une version de référence. Nous montrons alors la généralité de notre méthode par adaptation de ce même extracteur, construit initialement pour le domaine du tourisme, au domaine de la musique, dans le cadre du portage vers le français et l'anglais du système IMRS présenté dans le Chapitre 2.

9.1 Description de l'existant : EC pour le domaine du tourisme

Nous avons utilisé un extracteur de contenu pour le français construit par H.Blanchon dans le cadre des projets de traduction de parole C-STAR³⁴ « *Consortium for Speech Translation Advanced Research* » et Nespole! « *NEgotiating through SPOken Language in E-commerce* » (NESPOLE! 2000-2003).

L'objet du projet Nespole! est de permettre à un agent touristique italoophone situé à Trente en Italie de dialoguer avec un client parlant l'anglais, le français ou l'allemand, chacun disposant d'un simple PC.

Le projet Nespole! a donné lieu à la mise en place de deux démonstrateurs. Nous avons utilisé le code du second démonstrateur pour l'adaptation du système CATS au français. La traduction se fait via un pivot sémantique appelé IF "Interchange Format".

9.1.1 Spécification de l'IF

L'IF est fondé sur des actes de dialogue (DA, Dialogue Act) auxquels sont adjoints des arguments. Un acte de dialogue est constitué d'un acte de parole (SA, Speech Act) complété par des concepts. Les actes de dialogue décrivent les intentions, les besoins de celui qui parle (give-information, introduce-self,...). Les concepts précisent à propos de quoi l'acte de dialogue est exprimé (price, room, activity, ...). Les arguments permettent d'instancier les valeurs des variables du discours (room-spec, time, price, ...).

³⁴ <http://www.c-star.org/> : site web du projet CSTAR (1986-2003), visité en 2006.

Dans le projet Nespole!, la spécification IF est définie par :

- 75 actes de parole,
- 123 concepts,
- 189 arguments,
- 276 ensembles de valeurs.

Elle est répartie en trois fichiers :

- spécification des actes de parole (2 300 lignes) et des concepts (2 500 lignes),
- spécification des constructions des arguments (13 000 lignes),
- spécification des listes de valeurs (5 600 lignes).

Comme le montre la Figure 98, un client peut effectuer une *réservation* de chambres d'hôtel ainsi qu'une *réservation de voiture*. Nous nous sommes basé sur cette ouverture vers le domaine de l'automobile pour adapter cette spécification au domaine de l'achat et de la vente de voitures.

```

DA = speech act + (attitudes) + (main-predication) + (pred-participants)
>>>speech act = { give-information request-information offer ... } required

>>>attitudes = { +disposition +obligation +feasibility ... } one or more, optional
>>>main-predication (one of the following -- may be required, see note below):
relations = { +contain +feature +inclusion +exclusion +inclusion-in +exclusion-from}
actions = { +rent +proceed +reservation ... }
           or +concept, special concept used for fragments
;; note: main-predication is required unless following a "self-contained" speech act
;; like greeting, introduce-self, thank, apologize, etc.
>>>pred-participants (one or more of the following, optional):
objects = { +accommodation +transportation +activity ... }
;; sequences of concepts are ordered from general => specific
;; so that the concepts on the right specify the concepts to the left
DA = speech act +.....+vehicle....

```

Figure 98 : extrait de la spécification de l'IF

La Figure 99 montre un exemple d'extraction vers l'IF :

«la semaine du 12, nous avons des chambres simples et doubles disponibles».

Il s'agit d'une réponse de l'agent touristique qui correspond à l'IF suivante :

```

a:give-information+availability+room (room-type=(single & double),
time=(week, md12))

```

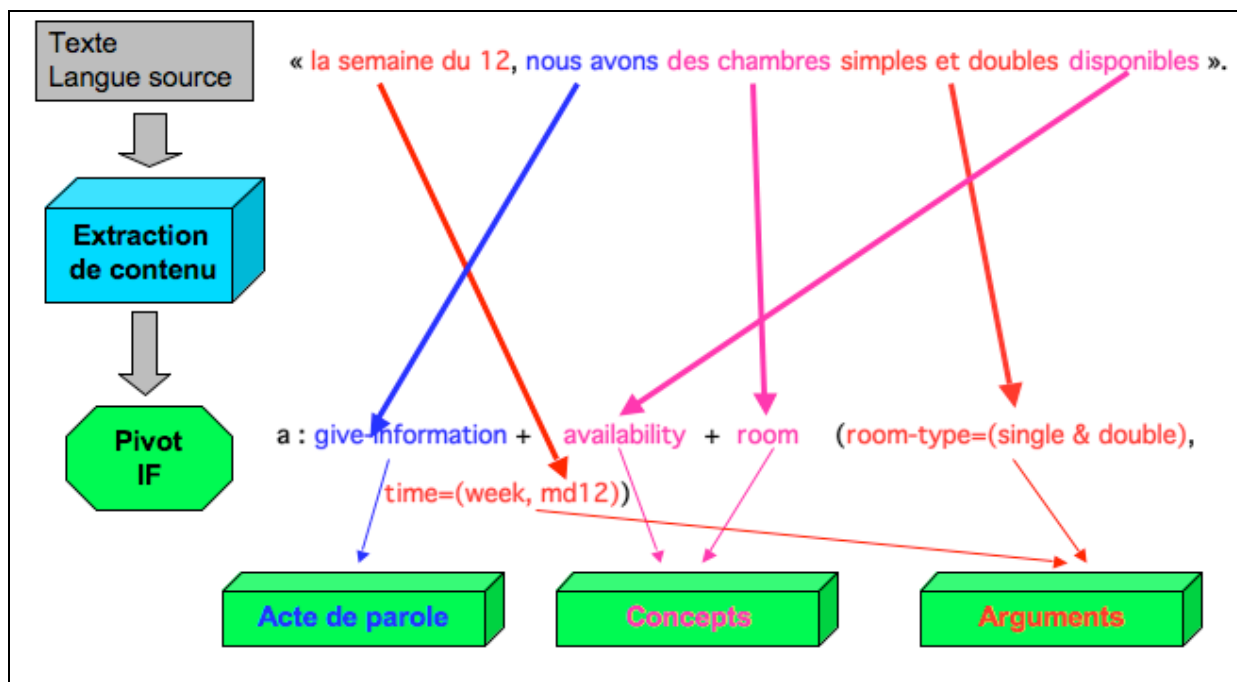


Figure 99 : exemple d'extraction vers l'IF

9.1.2 Ce dont on dispose et ce qu'on souhaite

Au début de ce travail, nous disposions :

- du code du second démonstrateur de Nespole! (31 918 lignes de code en Tcl/Tk).
- de la version papier et électronique de la spécification IF (version du 18-08-2002).

Nous avons adapté la spécification de l'IF au domaine de l'achat et de la vente de voitures d'occasion pour que l'extracteur (analyseur) puisse traiter l'exemple suivant :

je veux vendre une voiture Renault Clio bleue TBE
en produisant l'IF suivant :

```
a:give-information+disposition+vehicule(disposition=(desire, who=i),
action=e_sell, vehicule-spec=(car, vehicule-make=Renault, vehicule-model=Clio,
vehicule-color=blue, vehicule-condition=good))
```

qu'on appellera IF-CATS. Pour cela, regardons la méthode d'analyse utilisée par (Blanchon 2004) dans son second module.

9.1.3 Méthode d'analyse dans Nespole!

Comme le montre la Figure 100, la méthode utilisée pour l'analyse du français vers l'IF est composée des étapes suivantes :

- Segmentation des SDU (Unités Sémantiques de Dialogue).
- Détection du *domaine*.
- Construction d'un préfixe de l'acte de dialogue et instanciation des arguments liés.
- Instanciation des arguments liés au *domaine* et gestion des subordinations.
- Complémentation de l'acte de dialogue.

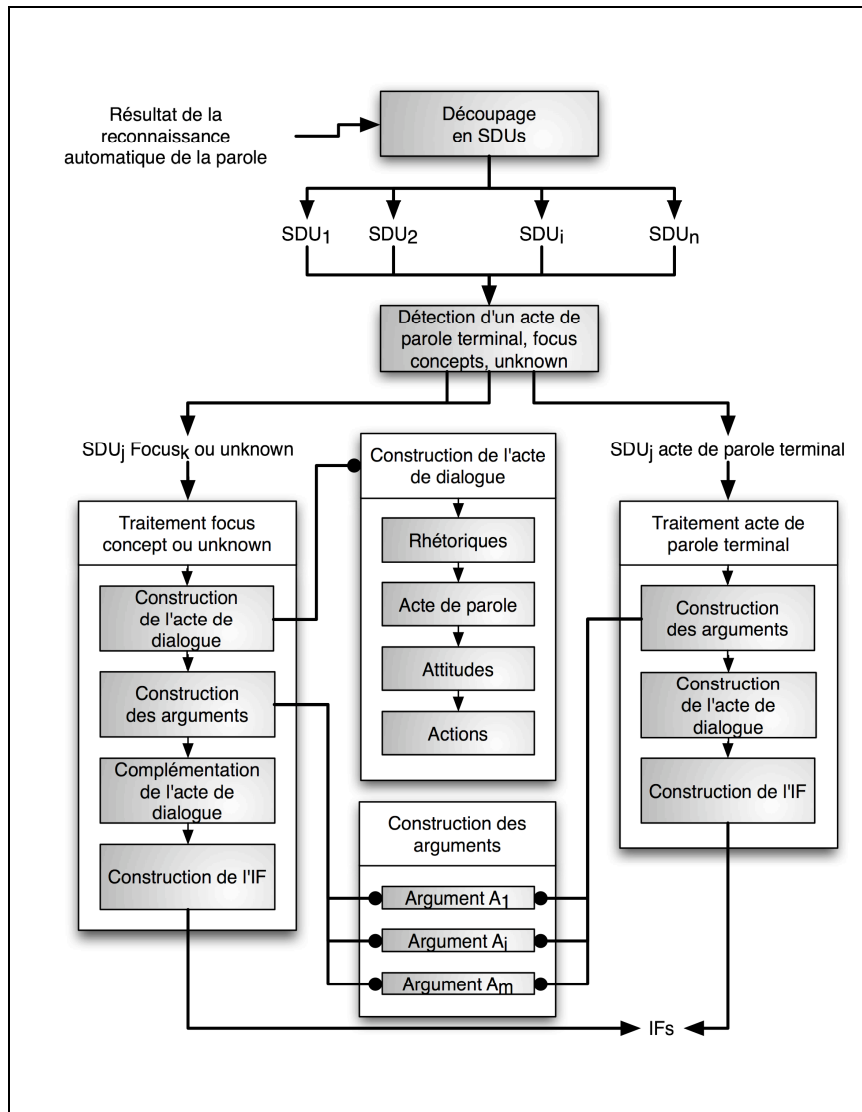


Figure 100 : architecture du module d'analyse français → IF du second démonstrateur Nespole! (Blanchon 2004).

❖ **Segmentation en SDU**

Au moment de l'analyse, un tour de parole est d'abord découpé en « unités sémantiques de dialogue » (SDU, Semantic Dialogue Unit). Une SDU est l'unité maximale du texte à analyser qui peut être représentée par un IF. Un énoncé peut correspondre à une ou plusieurs SDU.

Voici un exemple de découpage en SDU :

<p>Bonjour je voudrais réserver un hôtel du 4 juillet au 7 juillet</p> <p>SDU1 = Bonjour</p> <p>SDU2 = je voudrais réserver un hôtel du 4 juillet au 7 juillet</p>
--

Tableau 37 : exemple de découpage en SDU

Pour SDU1, l'IF obtenue est la suivante :

IF : c :greeting

Pour SDU2, l'IF obtenue est la suivante :

IF : c :request-action+reservation+temporal+hotel (time=(start-time=md4, end-time=(md7, july)))

Dans cet exemple, l'étiquette c indique que c'est le client qui parle. request-action est l'acte de parole, puis il y a trois concepts : reservation+temporal+hotel.

Pour faire la segmentation en SDU, l'analyseur utilise des groupes simples et des articulations. Ces groupes permettent de construire les actes de parole. Parmi les groupes utilisés, on trouve par exemple

variable acknowledge "je suis d' accord|nous sommes d' accord|c' est d' accord|d' accord|c' est ok|c' est okay|c' est bien|ok"³⁵

variable affirm "oui c' est ça|oui c' est bien ça|c' est bien ça|bien sûr"

Les articulations sont des arguments rhétoriques (**conjonctions simples** "et donc|donc|alors|et ensuite|ensuite|et notamment|notamment|et si|et puis si possible|et puis|et alors|et bien|et ben|mais si|mais s'|parce que|parce qu'|ou si", **syntagmes de conjonction...**) éventuellement suivis d'un pronom ("donc je|et puis je|donc on").

❖ Détection du domaine

La liste des domaines qui ont été utilisés est composée de 23 « focus concept » et de 22 actes de dialogue terminaux.

Dans un premier temps, Hervé Blanchon essaie de trouver un « focus concept » c'est à dire un concept terminal, sans successeur, parmi l'ensemble des 123 concepts possibles dans l'IF. Voici quelques exemples :

variable domvehicle "cars?|voitures?|attelages?|bacs?|ferrys?|planeurs?"

variable domaccommodation "auberges?|chalets?|pensions? de familles?|pensions?|chambres?"

variable domroom "places? de camping|emplacements? de camping|chambres? double|chambres? pour deux personnes|chambres? pour 2 personnes|chambres? familiale|chambres? simple|chambres? pour une personne|chambres? en tatami|chambres? avec deux lits|chambres? avec 2 lits|chambres? avec 1"

variable domservice "baby-sitting|garderies? d'enfant|garderie|gardes? d'enfant|classe|cours"

Si le domaine détecté après cette recherche est égal à la valeur unknown, on essaie de trouver un acte de dialogue terminal parmi les 75 actes de l'IF, associé à une liste de mots ou/et d'expressions. Par exemple :

variable domSAacknowledge "d'accord|ok|okay|c'est parfait|je veux bien|c'est ça|oui tout à fait|tout à fait|c'est très bien|bien sûr|ça va|c'est bien|c'est bon|c'est exactement ça|ça y est|c'est entendu|volontiers|effectivement|voilà"

variable domSAaffirm "oui|ouais|mouais"

domSAapologize "excusez?(?: |-)moi|pardonnez?(?: |-)moi|je m'excuse|pardon"

domSAexclamation "c'est excellent|c'est parfait|c'est

formidable|dommage|zut|ah non|oh non|merveilleu(?:x|ses?)|magnifiques?|super|ha|ah|ho|oh"

domSAgreeting "bonjour|salut"

domSAintroduceself "blanchon|besacier|guilbaud|caelen|boitet|serignat"

³⁵ Nous donnons l'écriture exacte en telkt, en mettant seulement en gras le début (variable <nom de variable>). La valeur d'une variable ainsi déclarée est une expression régulière.

Lorsque plusieurs domaines sont instanciés, celui qui est choisi est celui qui apparaît le plus rapidement après le début de la SDU, il sera traité par une fonction `domaine2if` qui construira l'IF associé à la SDU ; si aucun domaine n'est détecté, la valeur associée est `unknown`.

Pour SDU2 (du Tableau 37), le domaine détecté est `room`.

❖ **Construction d'un préfixe de l'acte de dialogue et instanciation des arguments liés**

Cette étape est réalisée par six sous-étapes :

- Construction des arguments rhétoriques.
- Construction de l'acte de parole.
- Construction des attitudes.
- Construction des actions (comme elles sont définies dans la spécification).
- Construction des concepts (`+price`, `+reception...`).
- Construction du concept `+action` qui permet d'instancier des actions qui ne sont pas des concepts.

Cette étape donne un préfixe de l'acte de dialogue enrichi avec une partie des arguments.

Dans l'exemple du Tableau 37, les concepts `reservation`, `temporal`, et `hotel` complètent l'acte de dialogue.

❖ **Instanciation des arguments liés au domaine et gestion des subordinations**

Dans cette étape, les arguments liés au « focus concept » en cours de traitement sont instanciés par exemple dans le domaine `room`, on instancie `room-spec`, `location...` et des fonctions `Argument2if` construisent leurs valeurs IF. D'autres types d'arguments sont traités dans ce second démonstrateur, ce sont les arguments qui prennent en compte la subordination.

Dans l'exemple du Tableau 37, `time` est l'argument supérieur qui comprend deux arguments inférieurs : `start-time` et `end-time`, dont `md4` et `md7`, `july` sont les valeurs.

❖ **Complémentation de l'acte de dialogue**

Dans cette étape, l'acte de dialogue est éventuellement complété par des concepts auxiliaires.

Dans son deuxième démonstrateur et par rapport au premier, Hervé Blanchon a intégré l'étape d'instanciation des arguments dans l'étape de construction de l'acte de dialogue en la décomposant en deux sous-parties (construction d'un préfixe et construction d'un suffixe). Il a changé le traitement des instanciation des arguments en détaillant leur construction par liaison à l'acte de dialogue (qui l'a précédée en construisant son préfixe) et par liaison au domaine détecté, avec une gestion des subordinations.

9.2 Adaptation à CATS

Nous montrons ici comment nous avons adapté l'extracteur d'IF construit pour le domaine du tourisme au domaine de l'achat et de la vente de voitures d'occasion.

9.2.1 Spécification de l'IF-CATS

La spécification IF telle quelle ne couvre pas le domaine de l'automobile. Nous avons donc dû enrichir cette spécification en ajoutant d'autres informations liées à ce domaine comme celle qui précise l'action d'achat ou de vente ou d'autres informations liées à l'état du véhicule, le moteur...

Nous notons IF-CATS la spécification IF adaptée au domaine de l'automobile. La transformation se compose de deux opérations essentielles :

- Adaptation et enrichissement de l'IF : c'est l'opération du passage du domaine du tourisme à celui de la vente et de l'achat de voitures d'occasion, autrement dit, on ne parle plus de chambres simples et doubles... mais de voitures, de modèles, de marques, de kilométrage, etc. Cela consiste à ajouter de nouveaux arguments tels que `vehicule-motor-type`, `vehicule-hand...` et de nouvelles actions, essentiellement l'action d'achat `e-buy` et l'action de vente `e-sell`. Afin de répondre à ce besoin, nous avons construit manuellement les représentations IF-CATS d'un corpus de 100 exemples (tirés du corpus CATS) pour ce nouveau domaine.
- Nettoyage de l'IF : il s'agit de supprimer les domaines inutiles parmi la liste des domaines programmés tels que `domrestaurant`, `domequipment...`

Nous ajoutons à l'ensemble de ces opérations une transformation qui tient compte de la nature de l'entrée. En effet, dans le projet Nespole!, l'étape d'analyse vers l'IF est précédée par une étape de reconnaissance de la parole qui gère le dialogue de communication et les phatiques comme `Allô`, `tu m'écoutes`, `oui j'entends`, `Oh`, `Ah`, `d'accord`, `ok`, `ouais`, `okay`, `excellent...`

Dans notre cas, la nature de l'entrée du système CATS est un texte envoyé par SMS. Nous devons donc tenir compte de la gestion des abréviations : `TBE` pour `très bon état`, `CT` pour `contrôle technique`, `ttes options` pour `toutes options...`

9.2.2 Adaptation de l'extracteur

Nous avons gardé les mêmes étapes d'analyse, en introduisant des changements par endroits, surtout au niveau de la sortie et de l'entrée de l'extracteur de contenu.

❖ Segmentation en SDU

Pour le découpage en SDU, nous gardons la même technique que celle utilisée précédemment, sauf qu'on enlève quelques groupes qui posent des problèmes de segmentation, comme `Ok` ou `excellent`. Ces deux termes peuvent être, dans le domaine du tourisme, une réponse d'un client ou d'un agent.

Comme le montre l'exemple suivant, ces deux termes peuvent constituer des SDUs et nous font sortir du domaine du `vehicule` vers un premier domaine `acknowledge` et un deuxième `unknown`. Ils posent des problèmes de découpage dans le domaine de l'automobile respectivement dans `...CT ok...` et `...excellent prix...`

Exemple :

```
je veux une grande voiture BM 4 portes 325 diesel première main bleue TBE
CT ok excellent prix

SDU : je veux une grande voiture BM 4 portes 325 diesel première main
bleue TBE CT
==> DOMAINE : vehicule
SDU : ok ==> DOMAINE : acknowledge
SDU : excellent ==> DOMAINE : unknown
SDU : prix ==> DOMAINE : unknown
```

Figure 101 : exemple de mauvaise segmentation en SDU

Nous réduisons alors la liste des groupes qui permettent la segmentation. Par exemple la liste **Acknowledge** devient :

```
variable acknowledge "je suis d'accord|nous sommes d'accord|c'est d'accord|d'accord|c'est ok|c'est okay|c'est bien"
```

❖ Détection et limitation du domaine

Nous nous limitons aux domaines qui ont des relations directes ou indirectes avec le domaine de l'automobile.

Nous avons gardé la même méthode en essayant d'améliorer la qualité du résultat et d'éviter les problèmes causés par l'étape précédente. Comme le montre l'exemple de la Figure 101, un mauvais découpage en SDU laisse le domaine inconnu pour le reste des segments.

Il s'agit aussi de supprimer tous les domaines (focus concept et acte de parole terminal et tout ce qui en dérive) qui n'ont pas de relation directe ou indirecte avec le domaine du commerce de l'automobile.

On limite alors la liste des concepts centraux alors à l'ensemble suivant :

```
variable domvehicule "cars?|voitures?|attelages?|bacs?|ferrys?|planeurs?|... "
variable dominformationobject "e-mail|adresse|brochure|num,..."
variable domroute "itinéraire|route ..."
variable domspelling "[a-z\]+ ..."
variable domtransportation "bateaux?|cars?|bus| ..."
```

❖ Construction d'un préfixe de l'acte de dialogue et instanciation des arguments liés

Dans cette partie, nous avons gardé les mêmes sous-étapes (construction des arguments rhétoriques, construction de l'acte de parole, construction des attitudes, construction des actions, construction des concepts, construction du concept +action qui permet d'instancier des actions qui ne sont pas des concepts).

De la même façon, cette étape donne un préfixe de l'acte de dialogue enrichi avec une partie des arguments.

Afin de réduire le temps de calcul, nous avons supprimé les arguments et les tests inutiles en nettoyant le code.

❖ Instanciation des arguments liés au domaine et gestion des subordinations

Le plus grand travail d'adaptation au domaine de l'automobile a été fait dans cette étape où nous traitons les arguments liés au domaine du véhicule. Dans ce domaine, on instancie essentiellement la spécification du véhicule `vehicle-spec`, ainsi que d'autres arguments moins intéressants tels que : `theDistance`, `theLocation`, `theDuration`, `theOrigin`, `theDestination`, `theTime`, `thePrice`, `theWithWhom`, `theForWhom`.

La fonction `VehicleSpec2If` permet la recherche et la construction des arguments liés au focus concept `vehicle` : le seul argument qui existe et qui était programmé dans le code du second démonstrateur est `frenchvehicle` qui peut avoir comme valeur {voiture, ski, camion, bus, train, avion... }.

D'autres arguments existent dans la spécification originale de l'IF, mais n'étaient pas programmés, tels que : `makevehicle`, `modelvehicle`, `sizevehicle`, `shapevehicle`, `frenchcolor`, `nationalityvehicle`, `agevehicle`, `pricevehicle`.

Afin d'adapter la spécification IF au domaine de l'automobile, nous avons enrichi cette spécification en ajoutant d'autres arguments liés à ce domaine tels que `motortypevehicle`, `handvehicle`, `conditionvehicle`, `assurancevehicle`, `controlevehicle`, `airconditionvehicle`, `mileagevehicle`.

De la même façon, des fonctions `Argument2if` construisent les valeurs IF associées.

Traitement des arguments

Deux méthodes sont possibles pour le traitement des arguments :

La première méthode consiste à calculer puis rechercher toutes les combinaisons possibles des arguments. La figure ci-dessous montre un exemple de combinaison possible qui correspond à la phrase suivante : « *Je veux une voiture Renault Clio* ».

L'expression régulière décrit un patron correspondant à la succession d'un véhicule (voiture), d'un type (Renault) et d'un modèle (Clio).

- Elle est applicable à la chaîne en entrée `$inString` (le SMS).
- La recherche se fait de gauche à droite.
- En cas de succès, le résultat est mis dans les variables `lMatch`³⁶, `lTransType`, `lMake` et `lModel`.
- On concatène à la droite du résultat courant `the_result` les trois arguments correspondant avec leurs valeurs, en ajoutant le « sucre syntaxique » de l'IF. Ici, si la variable `the_result` contient la valeur "c:request", elle contient après "c:request+vehicule(type=car, make=Renault, model=Clio)".

```
If {[regexp "($fifservdico::frenchvehicule)
($fifservdico::typevehicule) ($fifservdico::modelvehicule) "$inString
lMatch lTransType lMake lModel]!=0} {
    append the_result
" vehicle-spec=( " [fifservdico::Vehicle2If $lTransType]
" vehicle-make=" [fifservdico::Vehicle2If $lMake]
" vehicle-model=" [fifservdico::Vehicle2If $lModel] )"
```

Figure 102 : exemple de combinaison possible des arguments

Cette méthode n'est pas la plus efficace car le nombre de combinaisons à programmer est considérable, puisque l'ordre des arguments est très variable. De plus, on a ajouté de nouveaux arguments.

La deuxième méthode consiste à rechercher les arguments un par un en consommant³⁷ la chaîne en entrée, comme le montre le code de la Figure 103.

Par la suite, nous produisons une sortie standard pour tous les arguments dans un seul ordre choisi. Si l'argument est trouvé, sa valeur est affichée, sinon elle est remplacée par la chaîne vide.

³⁶ `lMath` : nom de la variable.

`$lMath` : valeur de cette variable.

³⁷ C'est la fonction `regsub` qui permet le remplacement d'une partie d'une chaîne.

```

if {[regexp " ($fifservdico::frenchvehicle) " $inString lMatch
lTransType]!=0} {
    set TransType $lTransType
    regsub $lMatch $partialIF(CleanedString) " " partialIF(CleanedString)
} else {set TransType ""}
if {[regexp " ($fifservdico::makevehicle) " $inString lMatch
lMakeVehicle]!=0} {
    set MakeVehicle $lMakeVehicle
    regsub $lMatch $partialIF(CleanedString) " " partialIF(CleanedString)
} else {set MakeVehicle ""}

```

Figure 103 : recherche et de traitement des arguments

Exemple d'ambiguïté

Nous observons la même ambiguïté que dans la première méthode (voir 8.2.3). Le mot JEEP est-il un type de voiture américain ou un modèle de voiture de type Mercedes? C'est une question qu'on ne sait pas non plus résoudre dans cette méthode.

9.2.3 Traduction IF-CATS vers CRL-CATS

Afin d'arriver à la représentation interne CRL-CATS, nous avons écrit un traducteur qui permet de passer de la sortie IF-CATS à CRL-CATS.

9.2.3.1 Passage de IF-CATS à CRL-CATS

Nous rappelons d'abord la structure des deux représentations CRL-CATS et IF-CATS.

❖ CRL-CATS

Nous avons décrit la structure de la représentation de contenu CRL-CATS dans la section 5.1.2.2. Nous la rappelons ici par un simple exemple de la Figure 104.

```

;Recherche RENAULT Clio mod 1998
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, RENAULT(country>France,country>europe):0A)
mod(saloon:00, Clio(country>France,country>europe,make>RENAULT):0I)
yea(saloon:00, 1998:14)
[/S]

```

Figure 104 : exemple de sortie CRL-CATS

❖ IF-CATS

Pour le même exemple, nous obtenons avec la deuxième méthode le résultat de la Figure 105.

Notons que les arguments constituant la sortie IF-CATS sont plus nombreux que ceux de CRL-CATS, car ils sont mentionnés, bien qu'avec une valeur vide, si aucune information n'est trouvée. Dans l'exemple ci dessous, trois arguments sont représentés : `vehicle-make=RENAULT`, `vehicle-model=Clio` et `vehicle-age=1998` ; les valeurs des autres arguments sont vides.

```

;Recherche RENAULT Clio mod 1998

{a:give-information+concept(action=e_buy, vehicle-spec=(car, vehicle-
make=RENAULT, vehicle-model=Clio, vehicle-age=1998, vehicle-price=, vehicle-
color=, vehicle-condition=, vehicle-assurance=, vehicle-controle=, vehicle-
air-condition=, vehicle-size=, vehicle-shape=, vehicle-motor-type=, vehicle-
hand=, vehicle-nationality=, vehicle-mileage=))}

```

Figure 105 : exemple de sortie IF-CATS

La sortie IF-CATS est composée du texte de la phrase en entrée placé en commentaire suivi du symbole { indiquant le début l'analyse de la phrase et fini par } indiquant sa fin. Dans cette sortie, le symbole a veut dire « acheteur » et v veut dire « vendeur ».

Nous avons construit un programme Tcl/Tk qui permet de prendre comme entrée un fichier texte, d'appeler la fonction qui réalise l'enconversion du français vers l'IF, et d'enregistrer le résultat dans un fichier texte.

9.2.3.2 Traducteur IF-CATS → CRL-CATS

Nous avons programmé en Tcl/Tk un traducteur qui permet d'analyser la sortie IF-CATS et qui la transforme dans le format CRL-CATS. Comme l'indique la Figure 106, l'entrée et la sortie sont des fichiers texte.

Ce traducteur nous permet ainsi d'effectuer une comparaison d'un côté entre les résultats obtenus et ceux obtenus par la version originale de l'extracteur de contenu de CATS, et d'un autre côté entre ces mêmes résultats obtenus et les résultats de référence.

La méthode que nous avons utilisée pour passer du pivot IF-CATS vers la représentation CRL-CATS est basée sur la recherche de segments d'IF-CATS et leur remplacement par leurs équivalents en CRL-CATS en utilisant un dictionnaire (IF-CATS,CRL-CATS) qui permet la substitution des arguments.

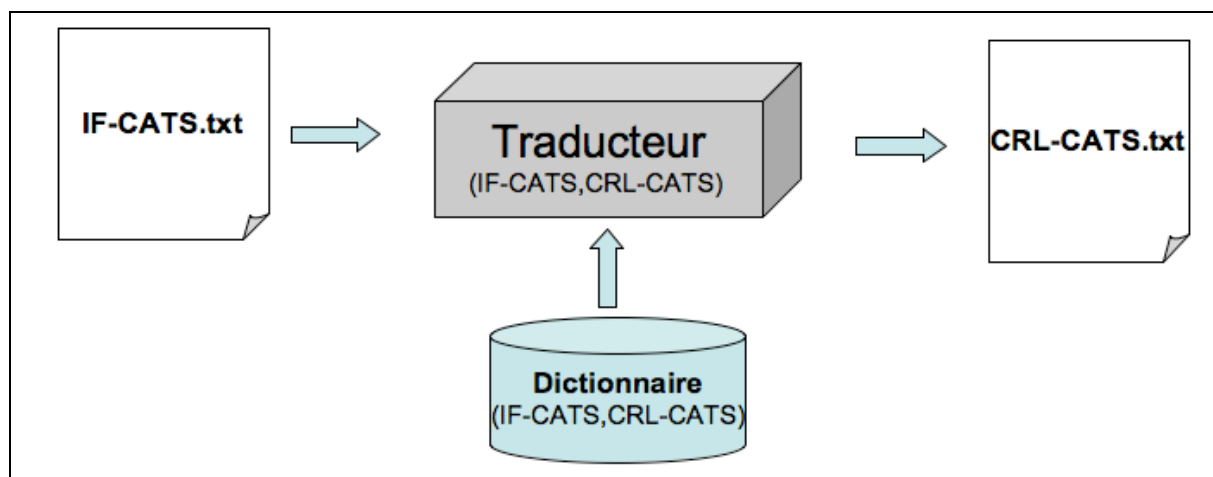


Figure 106 : transformation de IF-CATS vers CRL-CATS

La procédure de la Figure 107 permet la recherche et la substitution de chaque segment pertinent à l'aide de la fonction regsub, sinon elle rend le texte placé en entrée.


```

proc IF2CRL {inString} {
    variable if2crl #variable contenant la paire (texte IF,texte CRL)
    set lRes $inString
    foreach lLin $if2crl(textIF,textCRL) {
        if {[regsub -all [lindex $lLin 0] $lRes [lindex $lLin 1] lRes] != 0} {
            return $lRes
        }
    }
    return $lRes
}

```

Figure 107 : procédure permettant le passage de IF-CATS à CRL-CATS

La Figure 108 représente un exemple du dictionnaire (IF-CATS,CRL-CATS), c'est un dictionnaire qui se présente sous la forme d'une table de paires ({" entrée IF-CATS" équivalent CRL-CATS}).

```

{"Renault"  RENAULT(country>France,country>europe)}
{"Megane"   Megane(country>France,country>europe,make>RENAULT)}
{"Clio"     Clio(country>France,country>europe,make>RENAULT)}
{"Hyundai"  HYUNDAI(country>korea)}
{"AUDI"     AUDI(country>germany,country>europe)}

```

Figure 108 : exemple du dictionnaire (IF-CATS, CRL-CATS)

Pour établir la correspondance entre les deux représentations, nous avons été obligé de réduire le nombre d'arguments utilisés dans IF-CATS en faisant converger plusieurs arguments vers une seule propriété dans CRL-CATS, dont l'objet est le même, mais dont les valeurs varient suivant l'information. (voir Figure 109).

```

;=====CRL-CATS=====
;je veux une grande voiture francaise Renault Clio bleue TBE CT assure
prix 15000 euros an 1999 assuree avec climatisation ;

;a:give-information+disposition+vehicule(disposition=(desire,who=i),
action=e_sell, vehicule-spec=(car, vehicule-make=RENAULT, vehicule-model=Clio,
vehicule-age=1999, vehicule-price=15000, vehicule-color=blue, vehicule-
condition=good,
vehicule-assurance=insured, vehicule-controle=total_check,
vehicule-air-condition=air_condition, vehicule-size=, vehicule-shape=big,
vehicule-motor-type=, vehicule-hand=,
vehicule-nationality=french, vehicule-mileage=))}

[S]
sal(saloon, sale)
mak(saloon, RENAULT(country>France,country>europe))
mod(saloon, Clio(country>France,country>europe,make>RENAULT))
col(saloon, blue)
fea(saloon, good_condition)
fea(saloon, insured)
fea(saloon, air_condition)
fea(saloon, total_check)
cou(saloon, french)
yea(saloon, 1999)
pri(saloon, 15000)
[/S]
;=====

```

Figure 109 : exemple de convergence d'arguments IF-CATS vers une seule propriété CRL-CATS

Par exemple, dans la Figure 109 : l'ensemble des arguments IF-CATS `vehicle-condition`, `vehicle-assurance`, `vehicle-contrôle`, `vehicle-air-condition` convergent tous vers une seule propriété CRL-CATS dite `fea` pour `feature` en anglais.

9.2.3.3 Exemple et correspondance entre les arguments

L'exemple de la Figure 110 montre qu'en passant par cette transformation, on arrive à la même sortie que celle donnée par l'outil d'extraction d'information EnCo, à l'exception des symboles 00, 0J, 10, 0R ajoutés par ledit outil (EnCo).

```
;Recherche RENAULT Clio mod 1998

; {a:give-information+concept(action=e_buy, vehicle-spec=(car, vehicle-
make=RENAULT, vehicle-model=Clio, vehicle-age=1998, vehicle-price=, vehicle-
color=, vehicle-condition=, vehicle-assurance=, vehicle-controle=, vehicle-
air-condition=, vehicle-size=, vehicle-shape=, vehicle-motor-type=, vehicle-
hand=, vehicle-nationality=, vehicle-mileage=))}

[S]
wan(saloon, wanted)
mak(saloon, RENAULT(country>France, country>europe))
mod(saloon, Clio(country>France, country>europe, make>RENAULT))
yea(saloon, 1998)
[/S]
```

Figure 110 : exemple de sortie du traducteur IF-CATS

9.2.4 Évaluation et résultats

Comme dans l'évaluation du portage interne, nous évaluons d'abord le coût de cette méthode en termes de temps, puis de qualité des résultats d'extraction de contenu.

9.2.4.1 Coût du portage externe

Le Tableau 38 montre le détail des différents coûts en termes de temps, pour les différentes tâches, dans le cadre du portage externe du système CATS.

Tâches	Durée (h)	Version	Remarques
Apprentissage TCLTK	20	8.4.10	TCL surtout (livres + Web)
Documentation sur l'IF	15		
Compréhension focalisée de la spécification IF	40	18/08/02	Version papier
Compréhension focalisée du code	40		Orienté vers le domaine des véhicules
Petits tests	10		
documentation de la spécification	10		
Modification de la spécification IF + Adaptation	30	06-04-06 version IF-adaptée, validée par HB	
Création manuelle de corpus IF	12	100 SMS	But : détection du besoin
Enrichissement de la spécification IF	15		Spécification du nouveau besoin
Définition et programmation de la MAJ de la spécification IF	100		Expressions régulières
Programmation du dico fr-if	50		
Tests et modifications	10		
Premiers résultats	10		Résultats encourageants
Tests et évaluation	15		
Enrichissement du dictionnaire	30		Pour accélérer le temps de réponse
Construction d'un traducteur	20		transformation IF-CATS vers CRL-CATS
Total	427		

Tableau 38 : évaluation de la durée des tâches pour le portage externe

Le travail du portage externe a demandé 3 homme/mois d'informaticien, dont une bonne partie a été consacrée à la compréhension du code du second démonstrateur qui est très gros (31 918 lignes de code en Tcl/Tk) et à la spécification de l'IF.

Le reste du temps a été consacré au travail de programmation des modifications apportées au code, et aux tests.

9.2.4.2 Mesure informationnelle

Nous avons utilisé la même mesure informationnelle (voir section 6.2.2.1) que sur le corpus d'évaluation « CorpusEvalFr200SMS » utilisé pour évaluer la version portée au français par la première méthode. C'est l'équivalent du corpus « CorpusEvalAr200SMS » utilisé pour l'évaluation de la version arabe (originale) du système.

Cette mesure est basée sur le calcul du rappel R, de la précision P et de la F-mesure F pour les mêmes propriétés les plus importantes (action de vente ou d'achat, marque, modèle, année, prix).

9.2.4.2.1 Évaluation par rapport à la version en arabe

Les résultats d'extraction de contenu de la version française obtenue par portage externe et ceux de la version originale (arabe) pour les propriétés les plus importantes sont regroupés dans le Tableau 39. Les pourcentages de portage par rapport à la version originale varient entre 46% et 99%, avec une moyenne de 77 % (Hajlaoui and Boitet 2007).

Propriété	RegExpFR			EnCoAR			% portage
	Précision	Rappel	F-mesure (ExpRegFR)	Précision	Rappel	F-mesure (EnCoAR)	
Achat vente	1,000	0,835	0,910	0,956	0,970	0,963	95
Année	0,828	0,271	0,409	0,817	0,960	0,883	46
Prix	0,955	0,288	0,442	0,800	0,822	0,811	55
Marque	0,994	0,928	0,960	0,978	0,963	0,970	99
Modèle	0,965	0,661	0,785	0,901	0,837	0,868	90
Moyenne	0,948	0,597	0,701	0,890	0,910	0,899	77

Tableau 39 : comparaison entre les résultats d'extraction de contenu pour 200 SMS obtenus par portage externe par rapport à la version originale

Notons que ce sont les propriétés traitant les chiffres comme prix et années qui rendent faible la valeur du pourcentage du portage par adaptation externe, mais son avantage est quelle nécessite un simple accès à la représentation interne de l'application.

La Figure 111 permet de mieux visualiser la comparaison entre les valeurs de F-mesure trouvées pour chacune des versions du système.

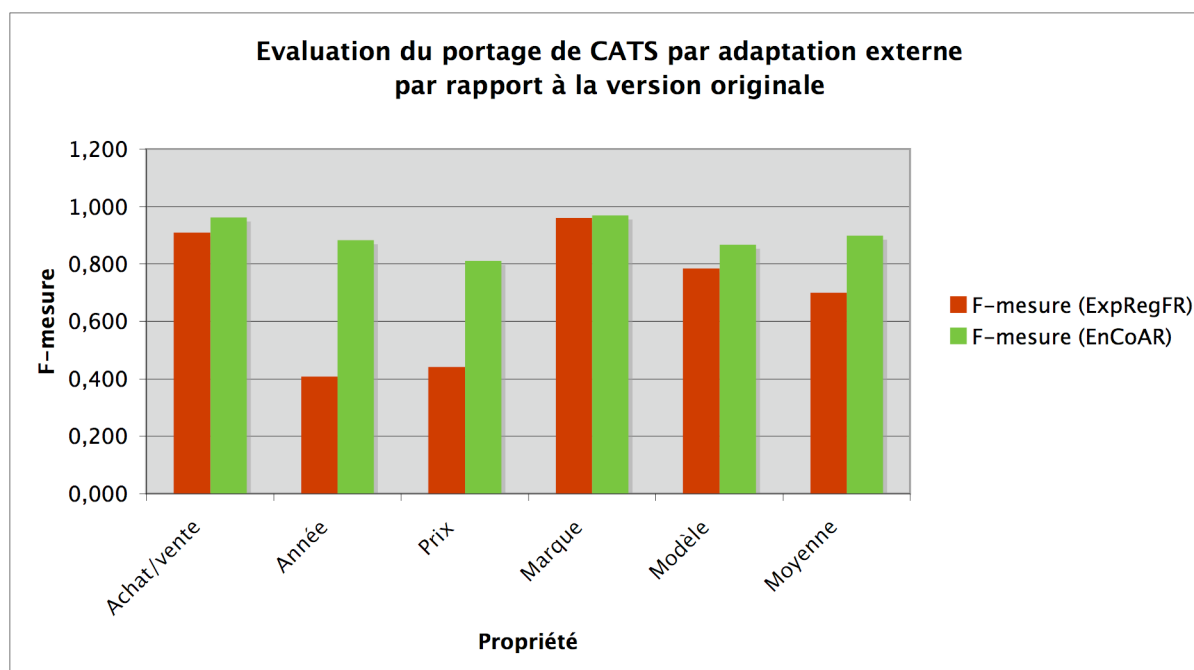


Figure 111 : comparaison entre F-mesures (par rapport à la version originale)

9.2.4.2.2 Évaluation par rapport à une version de référence

Nous avons aussi évalué la version française obtenue par portage externe en comparant ses résultats par rapport à ceux de la version de référence (obtenue par correction manuelle). Le Tableau 40 résume les résultats de la comparaison.

Propriété	EnCoAR-Ref			ExpRegFR			ExpRegFR % EnCoAR- Ref
	Précision	Rappel	F-mesure (EnCoAR- Ref)	Précision	Rappel	F-mesure (ExpRegFR)	% du portage
Achat/Vente	1,000	0,970	0,985	1,000	0,835	0,910	92
Année	1,000	0,983	0,991	0,828	0,271	0,409	41
Prix	0,973	0,973	0,973	0,955	0,288	0,442	45
Marque	0,984	0,963	0,973	0,994	0,928	0,960	99
Modèle	0,969	0,901	0,934	0,965	0,661	0,785	84
Moyenne	0,985	0,958	0,971	0,948	0,597	0,701	72

Tableau 40 : comparaison entre les résultats d'extraction de contenu obtenus par portage externe pour 200 SMS par rapport à une version de référence

Les pourcentages de portage par rapport à la version originale varient entre 41% et 99%, avec une moyenne de 72 %.

La Figure 112 permet une meilleure visualisation de la comparaison entre les valeurs de F-mesure trouvées pour la version française et la version de référence.

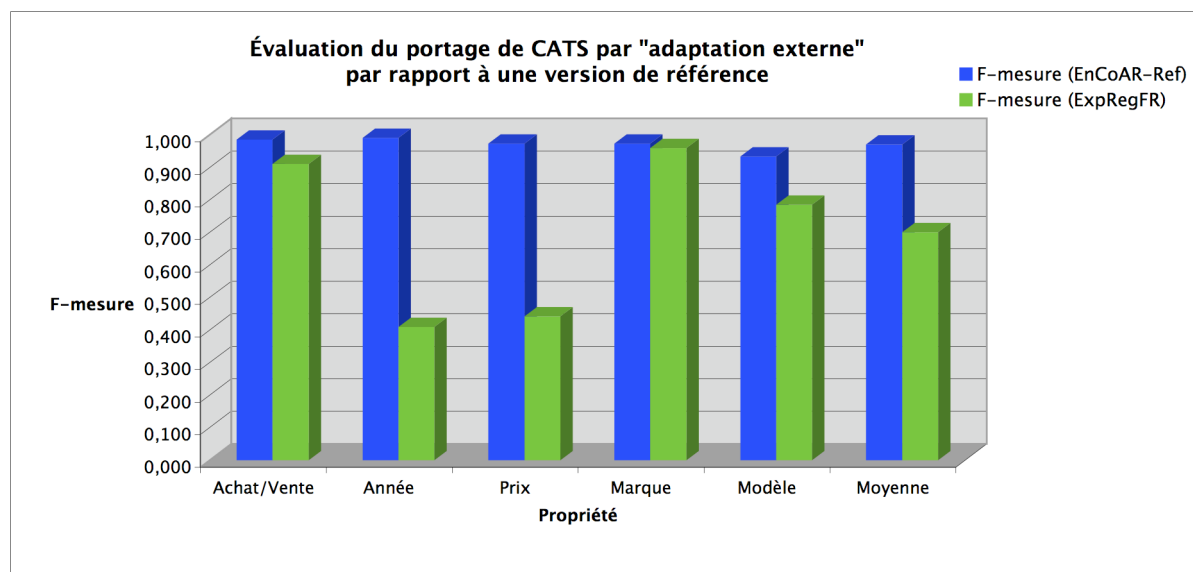


Figure 112 : comparaison entre F-mesures (par rapport à une version de référence)

9.2.4.3 Évaluation générale

Dans la suite, nous présentons les résultats des deux mesures décrites précédemment : la mesure simple et la mesure fine.

❖ Mesure simple

Par la même méthode que celle utilisée dans le portage interne et décrite dans la section 6.2.2.3, nous présentons les scores obtenus pour les attributs les plus importants ($m = 6$) sur la totalité du corpus « CorpusEvalFr1100SMS » ($n = 1100$) dans le Tableau 41.

Propriétés	Version originale	Version candidate
	EnCoAr	ExpRegFr
Score (wan) corpus	49,33	49,56
Score (sal) corpus	39,33	33,00
Score (yea) corpus	98,22	64,89
Score (pri) corpus	38,44	24,22
Score (mak) corpus	124,67	96,44
Score (mod) corpus	55,22	39,56
Score (wan, sal, yea, pri, mak, mod) corpus	405,22	307,67

Tableau 41 : scores trouvés pour le portage externe sur tout le corpus de 1100 SMS

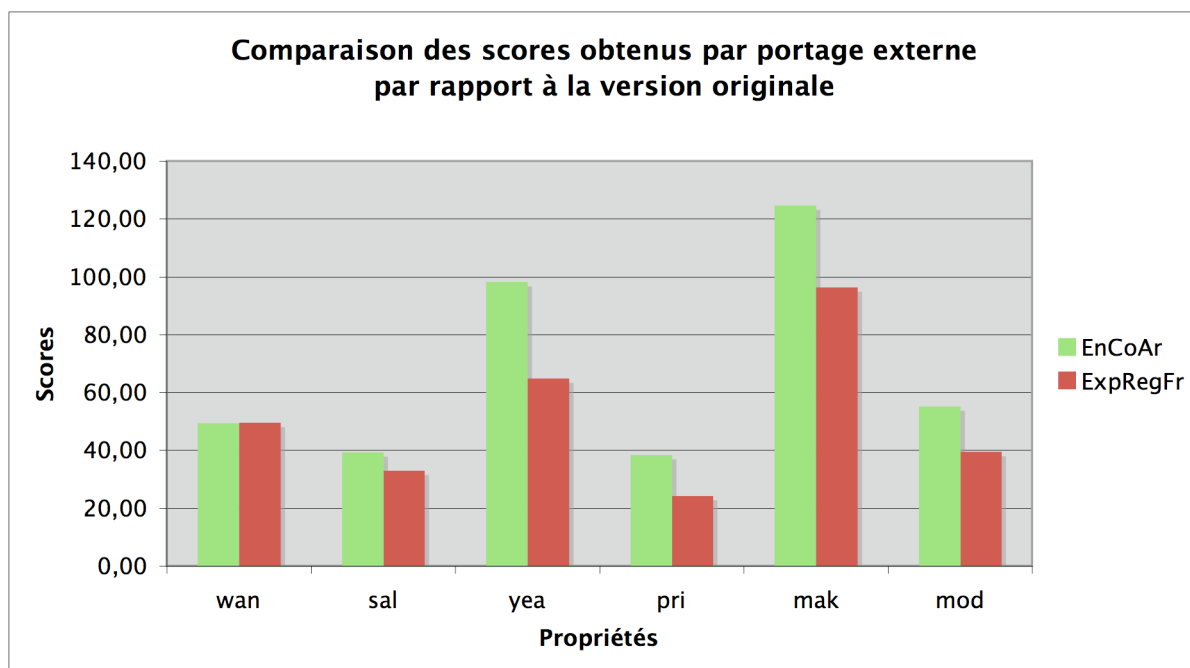


Figure 113 : comparaison des scores obtenus par portage externe par rapport à la version originale

De la même façon que dans la première méthode de portage, nous calculons le pourcentage de la différence par rapport à la version originale des scores trouvés, ce qui donne les pourcentages du Tableau 42.

Propriétés	% de différence de ExpRegFr par rapport à la version originale (EnCoAr)
wan	0,45
sal	16,10
yea	33,94
pri	36,99
mak	22,64
mod	28,37
Moyenne	23,08

Tableau 42 : comparaison des pourcentages de différences des scores d'extraction de contenu sur le total du corpus

Nous observons toujours que les résultats obtenus sur tout le corpus restent semblables à ceux obtenus sur le corpus d'évaluation limité à 200 SMS.

❖ **Mesure fine**

Nous avons enfin appliqué la mesure fine décrite dans la section 6.2.2.3. La distance entre l'ensemble des CRL-CATS sur le total du corpus constitué de 1100 SMS par rapport à l'ensemble des CRL-CATS de la version originale du système est de 2587 (somme des distances de chaque couple, donc 2,29 par couple en moyenne). Nous pouvons constater alors que cette mesure confirme les

résultats trouvés par les précédentes mesures (informationnelle, simple). Le Tableau 43 regroupe les valeurs moyennes de rappel, de la précision et de la F-mesure trouvées sur le corpus « CorpusEvalFr1100SMS ». La valeur de F-mesure moyenne est de 0,540. Elle se dégrade sur le total du corpus, par rapport à celle trouvée par la mesure informationnelle (0,701 sur « CorpusEvalFr200SMS ») à peu près de la même façon que la dégradation observée dans les F-mesures du portage interne.

	Rappel	Précision	F-mesure
Moyenne	0,410	0,670	0,540

Tableau 43 : résultats du portage externe par la mesure fine sur « CorpusEvalFr1100SMS »

9.3 Adaptation à IMRS

Nous avons aussi adapté le code de l'extracteur/analyseur de Nespole! au domaine de la musique pour les deux langues française et anglaise (portage vers l'arabe en cours). Voulant expérimenter le portage linguistique sur le système de recherche de morceaux de musique IMRS (Kumamoto 2007), présenté dans la section 2.2.4, et étant donné que nous ne pouvons pas adapter l'extracteur de contenu de ce dernier système, car nous n'y avons pas accès, et nous ne connaissons pas le japonais, nous avons adapté alors la spécification IF pour arriver au même résultat que celui produit par l'extracteur de IMRS. Après adaptation, nous obtenons un vecteur requête de dimension 10 à partir d'un énoncé en français ou en anglais avec presque toujours les mêmes valeurs.

9.3.1 Étude du sous-langage de IMRS

IMRS (Kumamoto 2007) traite des données d'un petit domaine. La taille du vocabulaire utilisé est d'environ 80 entrées principales qui peuvent s'étendre jusqu'à 300 entrées secondaires.

Il s'agit d'un vocabulaire facile et plus ou moins translittéré (*je veux un morceau joyeux*). Nous le voyons grâce aux exemples publiés, qui sont accompagnés de leurs traductions en anglais.

On a des phrases très simples et très courtes (zone sombre dans la figure suivante) avec une convergence grammaticale très rapide, mais une convergence lexicale moins rapide, à cause des nouveaux motifs comme les nouveaux genres de musique qui peuvent apparaître après un certain temps. On peut aussi avoir des phrases simples et courtes (zone claire dans la figure suivante) avec une convergence grammaticale et une convergence lexicale très rapides (*je veux un morceau assez doux, je préfère un morceau solennel...*).

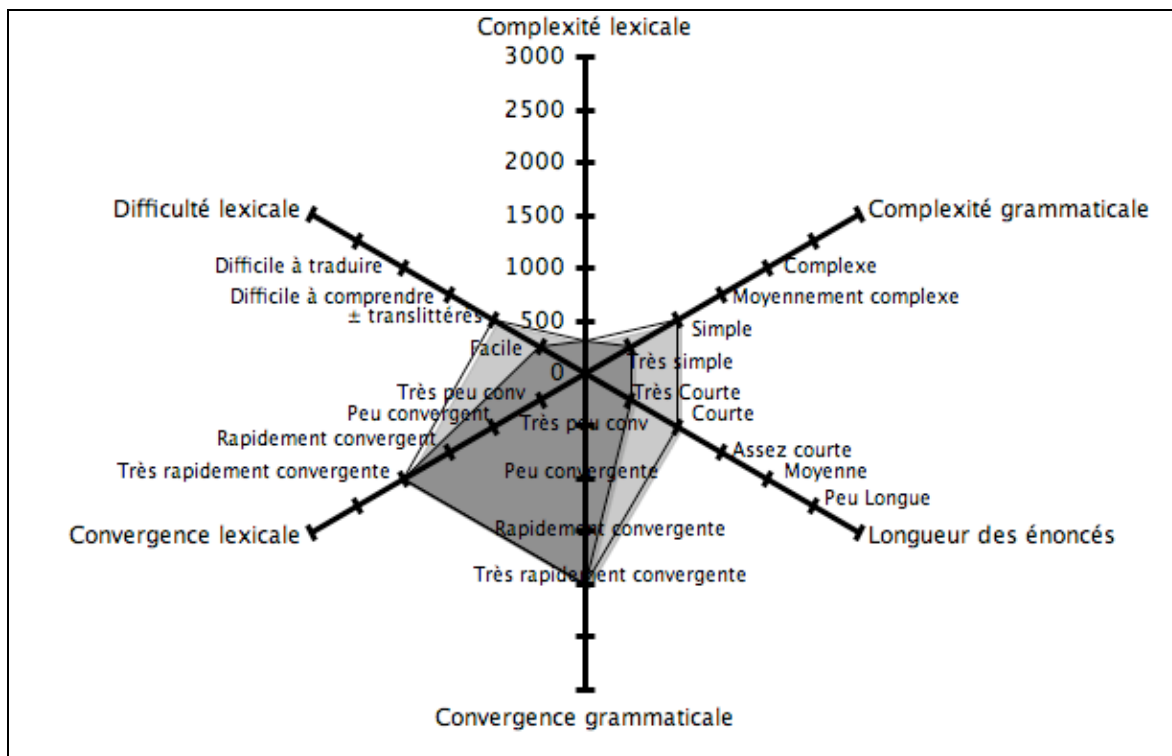


Figure 114 : analyse du sous-langage de IMRS

9.3.2 Méthode d'adaptation

Nous avons adopté les mêmes étapes d'analyse que pour le second démonstrateur (voir section 9.1.3) et la même démarche d'adaptation que celle utilisée pour CATS (voir section 9.2.2). Comme pour CATS, le travail d'adaptation le plus important a été effectué au niveau de l'étape d'instanciation des arguments liés au domaine. Ainsi, de la même façon, nous avons procédé à la recherche des arguments un par un en consommant la chaîne en entrée comme le montre le code de la Figure 115.

```
if {[regexp "($fifservdico::quietnoisy)" $inString lMatch lQuietNoisy]!=0}
{
    set QuietNoisy $lQuietNoisy
    regsub $lMatch $partialIF(CleanedString) " " partialIF(CleanedString)
} else {set QuietNoisy "nil"}
if {[regexp "($fifservdico::calmagitated)" $partialIF(CleanedString)
lMatch lCalmAgitated]!=0} {
    set CalmAgitated $lCalmAgitated
    regsub $lMatch $partialIF(CleanedString) " " partialIF(CleanedString)
} else {set CalmAgitated "nil"}
```

Figure 115 : recherche et de traitement des arguments liés au domaine de la musique

Par la suite, nous avons utilisé une sortie standard pour tous les arguments qui contient directement un vecteur requête de dimension 10. Si l'argument est trouvé, sa valeur est affichée sinon elle sera remplacée par la chaîne nil qui veut dire « *don't care* ».

Chaque vecteur a dix composantes. Chaque composante correspond à l'un des dix axes présentés dans le Tableau 44 : la valeur d'une composante est un nombre réel entre 1 et 7 qui correspond à sept degrés de l'échelle associée à l'axe en question (voir Figure 20 :). Par exemple, le premier axe « Quiet – Noisy » est caractérisé par sept valeurs intermédiaires : *very quiet*, *quiet*, *a little quiet*, *average*, *a little noisy*, *noisy*, et *very noisy* qui correspondent respectivement aux valeurs 7.0, 6.0, 5.0, 4.0, 3.0, 2.0, et 1.0.

Score N°	Paire de mots d'impressions anglais	Paire de mots d'impressions français
1	Quiet – Noisy	Doux – Fort
2	Calm – Agitated	Calme – Agité
3	Refreshing – Depressing	Joyeux – Triste
4	Bright – Dark	Clair – Sombre
5	Solemn – Flippant	Solennel – Léger
6	Leisurely – Restricted	À l'aise – À tempo strict
7	Pretty – Unattractive	Agréable – Sévère
8	Happy – Sad	Joyeux – Triste
9	Relax – Arouse	Apaisant – Excitant
10	The min is restored – The mind is vulnerable	Rassérénant – Inquiétant

Tableau 44 : dix axes (définis par des paires) pour représenter les impressions musicales en français et en anglais

9.3.3 Résultats et coût

Nous avons fait l'expérimentation sur une centaine d'énoncés (créés manuellement) et voici quelques exemples de résultats pour le français :

```

Exemple_Fr 1 : je veux un morceau de musique calme et très solennel
IFMusique_Fr 1:{c:give-information+disposition+service(disposition=(desire,
who=i), service=music, musique-spec=(nil 6,0 nil nil 7,0 nil nil nil nil
nil))}

Exemple_Fr 2 : je veux un morceau de musique assez fort et clair
IFMusique_Fr 2:{c:give-information+disposition+service(disposition=(desire,
who=i), service= music, musique-spec=(3,0 nil nil 6,0 nil nil nil nil nil
nil))}

Exemple_Fr 3 : je veux un morceau de musique très agréable
IFMusique_Fr 3:{c:give-information+disposition+service(disposition=(desire,
who=i), service=music, musique-spec=(nil nil nil nil nil nil 7,0 nil nil
nil))}

```

Figure 116 : exemple de résultats pour le français

Les mêmes exemples sont traduits en anglais et donnent ainsi les résultats suivants qui correspondent bien aux mêmes vecteurs :

```

Exemple_En 1 : I want a calm and very solemn music
IFMusique_En 1 : {c:give-information+disposition+service(service=music,
music-spec=(nil 6,0 nil nil 7,0 nil nil nil nil nil))}

Exemple_En 2 : I want a little noisy and bright music
IFMusique_En 2 : {c:give-information+disposition+service(service=music,
music-spec=(3,0 nil nil 6,0 nil nil nil nil nil))}

Exemple_En 3 : I want a very enticing music
IFMusique_En 3 : {c:give-information+disposition+service(service=music,
music-spec=(nil nil nil nil nil nil 7,0 nil nil nil))}

```

Figure 117 : exemple de résultats pour l'anglais

Une adaptation et une expérimentation sur des données de musique en arabe sont en cours.

Ce travail a nécessité beaucoup moins de temps que celui passé pour la première adaptation de l'IF du domaine du tourisme à celui de l'automobile. Nous avons passé environ une semaine pour chaque langue. Cela revient essentiellement à la maîtrise préalable du code de l'extracteur de contenu de Nespole!, et aussi à la taille du vocabulaire du domaine de la musique qui est très petit devant celui de l'automobile.

Malheureusement, nous n'avons pas accès au système complet pour pouvoir évaluer les résultats d'extraction trouvés.

Ce travail montre aussi que rien n'empêche de partir d'un même extracteur d'une langue donnée (le français ici) pour l'adapter à la fois à un nouveau domaine, et à de nouveaux sous-langages, du français (« natif ») et de l'anglais (et bientôt de l'arabe).

Conclusion

Nous avons présenté dans ce chapitre une deuxième méthode de portage externe pour le système CATS, basée sur l'adaptation d'un extracteur de contenu pour la même langue mais prévu pour le domaine du tourisme.

Cet extracteur de contenu utilise son propre format interne IF. IF est une représentation assez riche, complète et complexe. Son adaptation demande beaucoup d'effort de compréhension et de patience. La spécification de l'IF est très intéressante, elle est bien structurée mais très complexe.

Le code qu'on a adapté est très sensible aux problèmes de codage et de format des fichiers manipulés (expressions régulières). Par comparaison avec la représentation CRL-CATS, la spécification IF-CATS est beaucoup plus complète. Elle a été prévue et conçue dans le cadre d'un projet international.

Pour passer alors de IF-CATS à CRL-CATS, et dans une restriction au sous-langage du français en Jordanie, nous sommes obligés de réduire certains arguments de IF-CATS tels que `vehicle-size`, `vehicle-shape`, `vehicle-mileage` et de faire converger certains autres arguments vers un seul argument en CRL-CATS.

Nous avons présenté le passage du format IF-CATS à CRL-CATS basé sur la construction d'un traducteur de format. Ce passage permet d'arriver à la représentation utilisée par CATS.

Cette deuxième méthode donne aussi de bons résultats. Ils sont cependant un peu moins bons (voir Annexe IV) que la première méthode. Son avantage principal est qu'elle demande seulement d'avoir accès à la représentation interne de l'application. Son deuxième avantage est qu'elle est plus

générique, ce que nous avons montré par son adaptation à un troisième domaine, celui de la musique, pour la même langue (français) et aussi pour une troisième langue (l'anglais).

Conclusion de la partie C

Nous avons pu assurer le besoin essentiel et commun pour toutes méthodes de portage, à savoir un corpus de développement et un dictionnaire. Nous avons pu obtenir en deux étapes ces données nécessaires et préparatoires pour l'opération du portage.

Nous avons appliqué la démarche prévue pour une première méthode de portage interne du système CATS. Cette méthode nécessite un accès total aux ressources et au code. Pour cela, nous avons porté vers le français l'extracteur de contenu prévu initialement pour des SMS en arabe. Cette méthode marche bien à cause de la proximité de ces deux sous-langages de l'arabe et du français. Nous avons effectué une comparaison des résultats obtenus par rapport à la version originale et par rapport à une version de référence, et les résultats sont excellents.

Nous avons mis en œuvre la démarche prévue pour une deuxième méthode de portage externe du système CATS qui consiste à changer de domaine et à rester dans la même langue. Cette méthode nécessite un simple accès à la représentation interne. Nous avons localisé un extracteur de contenu de français prévu initialement pour le domaine du tourisme. Cette méthode marche bien et nécessite un traducteur de format. Nous avons effectué une comparaison des résultats obtenus par rapport à la version originale et par rapport à une version de référence.

Dans les deux méthodes, nous évaluons le portage effectué en termes de résultats d'extraction de contenu obtenus. En effet, Daoud a démarré son système à partir d'énoncés Web corrigés et édités qui sont plus loin de la réalité. Dans les deux cas, nous avons assuré le portage de CATS, et nous avons obtenu une évaluation objective positive, qui pourrait dans le futur être confirmée par une évaluation subjective en déployant la nouvelle version sur le terrain.

Partie D

Exploration d'une méthode de portage par TA

Si l'on veut porter une application d'extraction de contenu à partir d'énoncés spontanés sans aucun accès aux ressources internes de l'application, la seule approche envisageable est d'utiliser la traduction automatique. L'idée consiste à traduire les énoncés du sous-langage approprié de la nouvelle langue L2 vers le sous-langage original de L1. L2 est alors « cible » du portage, mais « source », pour la traduction.

Quelle que soit l'approche linguistique choisie pour cette TA, il faut créer un système spécialisé, et donc disposer d'un corpus parallèle L2//L1. On a vu précédemment comment obtenir un tel corpus « fonctionnel » en L1. La question qui se pose est la taille du corpus nécessaire.

Si l'on utilise une approche computationnelle de la TA « fondée sur des corpus » (TA statistique, TA par analogie), on sait qu'il faut d'énormes corpus s'il s'agit de langue générale (entre 50 et 200 millions de mots d'après K.Knight et Ph.Koehn), bien plus grands que ceux disponibles après deux ou trois ans de fonctionnement d'un e-service. Il est toujours possible que dans le cas de sous-langages restreints, des corpus beaucoup plus petits suffisent, mais ce n'est qu'une hypothèse, et nous n'avons trouvé aucune étude sur le sujet.

Si l'on utilise une approche computationnelle « par règles », le corpus de développement n'a pas besoin d'être très grand, il suffit qu'il soit représentatif, mais il faut disposer de linguistes computationnels pour chaque couple (L2,L1), ce qui est rare.

En résumé, le portage par réalisation d'un système de TA $L2 \rightarrow L1$ est possible en théorie, mais nous ne sommes pas encore en mesure de déterminer si on peut le faire, sans linguistes qualifiés, par des méthodes d'apprentissage automatique. Nous voulons répondre à cette question dans cette partie.

Chapitre 10 Architectures possibles pour la TA

Introduction

Il faut d'abord noter que, si on fait un portage par TA (SL2→SL1), il ne s'agit pas de produire des résultats de traduction parfaits, mais seulement des résultats tels que l'extracteur de contenu original puisse en extraire l'information pertinente.

Nous étudions le choix du type de système de TA à mettre en œuvre dans la première section en classant des architectures candidates potentielles. Avoir des données, éventuellement un corpus parallèle (L1,L2) est un besoin commun pour toutes ces architectures. Nous proposons alors dans la deuxième section une méthode de génération de données dans L2. Dans la troisième section, nous décrivons la démarche à suivre et les outils que nous pouvons utiliser.

10.1 Architectures calculatoires envisageables

Dans ce qui suit, nous classons les architectures de système de TA envisageables dans un ordre qui demande de moins en moins de compétences linguistiques. Nous nous basons sur des définitions tirées de (Boitet 2007) :

Les sources de connaissances utilisables pour construire un système de TA sont d'abord symboliques : linguistiques (lexique, grammaire, étude précise d'une typologie), sémantiques « internes » (propriétés véhiculées par la langue comme les relations sémantiques via les prépositions, etc.), sémantiques « externes » (ontologie), et pragmatiques (situation et locuteur dans un dialogue). De plus, on utilise souvent des connaissances « stratégiques » qui permettent d'améliorer la résolution automatique des ambiguïtés. Elles se présentent soit comme des « préférences » (symboliques), soit comme des statistiques (fréquences de bigrammes, trigrammes...) ou des poids (coefficients de confiance). Une direction assez récente consiste à marier les connaissances symboliques et numériques, ces dernières étant obtenues par apprentissage sur de gros corpus.

Les approches envisageables sont décrites dans les sections suivantes.

❖ TA « experte »

Actuellement, la plupart des systèmes de TA opérationnels sont à base de procédures et/ou règles. L'avantage de ce type de systèmes est que, contrairement aux systèmes statistiques, on peut les améliorer de façon ciblée. L'inconvénient majeur est qu'ils demandent des compétences linguistiques (personnel qualifié). Si on souhaite par exemple étendre CATS à 10 langues européennes, il nous faut $20 = 2 * 10$ spécialistes pour chaque langue (un linguiste et un lexicographe). Pour cette raison (coût humain important), nous ne pouvons pas essayer cette méthode et par conséquent nous n'allons pas garder ce premier candidat.

❖ TA empirique « par l'exemple »

Ce type de système entre dans la catégorie des systèmes de traduction fondés sur les données. L'idée est d'abandonner la construction à la main de règles complexes et de dictionnaires très détaillés, en utilisant directement des exemples de traductions lors du processus de traduction (TA par l'exemple) ou des tables de probabilités (TA statistique).

La TA par l'exemple est un paradigme le plus souvent hybride. À ATR, IBM-Japon et MS (MicroSoft Research), par exemple, on procède à une analyse assez classique produisant une structure de dépendance, puis on utilise une base d'exemples de correspondances entre morceaux d'arbres, on construit une meilleure couverture de l'arbre source, et on en déduit un arbre cible qu'on linéarise classiquement.

À l'USM (Penang), l'UTMK a développé un système de TA anglais-malais de bonne qualité (voir <http://utmk.cs.usm.my:8080/ebmt/index.jsp>) basé sur un corpus bilingue annoté par des arbres syntaxiques, leurs correspondances avec les chaînes (SSTC), et les correspondances entre ces correspondances (S-SSTC = Synchronized Structured String-Tree Correspondences).

Dans sa thèse (Planas 1998), E. Planas a établi un lien avec les mémoires de traduction, en utilisant des "mémoires à étages", et en restreignant la recherche à des exemples "proches" au sens d'une distance d'édition adéquate, et contenant le segment à traduire à un certain étage, par exemple à celui des catégories. On peut alors établir les correspondances aux niveaux inférieurs et utiliser le dictionnaire bilingue pour proposer une traduction complète (il manque encore la génération flexionnelle).

Pour la même raison (nécessité importante de compétences linguistiques pour créer les annotations), nous ne pouvons pas tester cette méthode de traduction.

❖ TA empirique « par analogie »

Le terme de TA par analogie a été introduit en 1984 par le Pr. Nagao, mais il s'agissait en fait de TA par similarité, consistant, pour une chaîne à traduire T_s , à trouver dans un corpus de bitextes analysés une plus proche chaîne C_s et son analyse A_s , ainsi que son image cible (C_c , A_c), et à modifier A_s , puis A_c et C_c en conséquence. Mais on n'a pas de méthode efficace pour calculer ces modifications.

Y. Lepage à ATR est ensuite reparti de la définition fondamentale de l'analogie, qui fait intervenir 4 objets homogènes, et permet d'en calculer un à partir des trois autres ($a:b :: c:d$, ou "a est à b ce que c est à d"). Le paradigme en cours d'expérimentation consiste alors à faire l'analyse, puis le transfert, puis la génération, par analogie : trouver $Cs1, Cs2, Cs3$ t.q. $Ts:Cs1 :: Cs2:Cs3$, résoudre $x:As1 :: As2:As3$ et $y:Ac1 :: Ac2:Ac3$ donnant As et Ac , résoudre enfin $z:Cc1 :: Cc2:Cc3$ (ou projeter Ac en z , ce point n'est pas encore clair).

L'avantage de la réalisation d'un système de TA par analogie est qu'il ne demande pas de compétences linguistiques. Par contre, il faut disposer des outils nécessaires pour effectuer une telle tâche. Ces outils ne sont pas disponibles sur le Web et ne sont pas en utilisation libre (Lepage 2006) (Denoual 2006).

Y. Lepage utilise le corpus BTEC (Basic Travel Expression Corpus) (Boitet 2004) constitué de 163.000 phrases de tourisme alignés en japonais-anglais de taille 6,5 mots. Ses expériences montrent qu'il faut avoir une taille de corpus entre 40.000 et 163.000 phrases, pour obtenir un début de résultat positif par traduction analogique.

❖ TA empirique « statistique »

(E.Brown, Pietra et al. 1993) ont formalisé au début des années 90s la traduction « statistique » qu'on devrait plutôt appeler « probabiliste ». Traduire un document source se résume à trouver le document cible ayant la plus grande probabilité d'être sa traduction (Besacier 2007).

Dans le cas de la traduction statistique, on suppose que n'importe quel énoncé f est une traduction possible de l'énoncé e . À chaque paire (e,f) , on assigne une valeur $Pr(f|e)$, qui représente la probabilité que la machine produise l'énoncé f comme traduction de l'énoncé e . En utilisant la formule de Bayes :

$$Pr(e|f) = Pr(f|e)Pr(e)/Pr(f)$$

Puisque le dénominateur est indépendant de e , on peut utiliser seulement le numérateur pour arriver à l'équation fondamentale en traduction automatique statistique.

$$\hat{e} = \operatorname{argmax}_e \Pr(e) \Pr(f|e)$$

Cette équation résume les trois défis informatiques présentés par la pratique de la traduction statistique : le « modèle de langage » $\Pr(e)$ (pour la langue cible), le modèle de traduction $\Pr(f|e)$, et l'algorithme de recherche pour trouver la traduction optimale (argmax), dit « décodage ».

Un modèle statistique de langue donne la probabilité d'observer un mot connaissant ceux qui le précèdent. Avec cette information, nous pouvons déterminer si un énoncé est plus ou moins probable dans la langue cible. Ces modèles sont obtenus à partir de données d'entraînement sur des corpus de la langue cible.

Les systèmes à base de modèles statistiques de langue permettent d'estimer la probabilité a priori de la séquence de mots $S = m_1, m_2, \dots, m_i$ selon l'équation suivante :

$$P(S) = P(m_1) \times P(m_2|m_1) \times \dots \times P(m_i|m_1, m_2, \dots, m_{i-1})$$

Il y a plusieurs méthodes pour calculer les probabilités d'un modèle de langue. Celles-ci sont généralement estimées sur des corpus d'apprentissage censés représenter au mieux le langage à modéliser. Il existe de nombreuses variantes, qui utilisent différentes informations, depuis le simple graphème (pour certaines langues n -grammes) jusqu'à des classes (n -classes) ou des séquences de mots.

Pour construire des systèmes de TA statistique, des outils gratuits sont disponibles sur le Web³⁸ tels que les décodeurs Pharaoh et Moses, Giza++ pour l'alignement, et d'autres outils pour le passage d'une format à l'autre. Des documentations et des articles concernant ce travail sont aussi disponibles sur le Web.

Cette méthode semble être intéressante et faisable. Elle demande moins de compétences linguistiques que les précédentes, mais demande un travail de prétraitement (balisage, segmentation). Le gros problème, a priori, est la taille du corpus parallèle nécessaire.

10.2 TA statistique à petits corpus pour de petits sous-langages

On sait que la TA experte (par règles, ou procédurale) peut donner des résultats quasi-parfaits (comme dans METEO, ALTFLASH) avec des algorithmes simples, des ressources de petite taille, et de faibles temps de calcul pour des sous-langages très restreints, alors que le « passage à l'échelle » se passe très mal, sauf si on utilise de très grosses ressources, une architecture et une programmation beaucoup plus élaborée (comme le transfert multiniveau avec des LSPL modulaires très puissants), et des temps de calcul bien plus longs.

Nous avons donc voulu voir s'il en va de même avec la TA statistique, c'est-à-dire si l'on pourrait obtenir de bons résultats sur un sous-langage très restreint avec un très petit corpus d'apprentissage (quelques milliers d'énoncés plutôt que plusieurs millions ou dizaines de millions) et un dictionnaire assez complet.

❖ Les systèmes de TA statistique s'améliorent

La citation de (Boitet 2007) dans la section 3.3.1.1 montre que les systèmes de TA statistique s'améliorent et donnent des résultats de plus en plus satisfaisants, surtout dans des cas comme les textes journalistiques (cas où le corpus d'entraînement est de même type), alors que les résultats sont moins satisfaisants dans le cas des textes techniques.

³⁸ <http://www.statmt.org/wmt06/shared-task/baseline.html>

De plus, les boîtes à outils de TA statistique comme Pharaoh, Moses sont maintenant disponibles et leur mise en place est assez facile : si on dispose d'un corpus parallèle, on peut fabriquer, en trois à quatre jours, un système de TA statistique.

❖ Un petit corpus représentatif pourrait suffire

Si les sous-langages sont très proches, comme dans les cas que nous visons, cela peut aider la TA statistique et nécessiter une taille de corpus beaucoup moins importante.

Avec les outils d'alignement disponibles sur le Web comme GIZA++ (version améliorée et complète de GIZA) et avec l'éventuel ajout d'un dictionnaire, nous pensons que le corpus nécessaire pour le cas d'un sous-langage très petit comme celui de l'occasion auto dans CATS (Cars) pourra ne pas être très volumineux. L'espoir est que les corpus dont nous disposons suffisent (entre 1000 et 2000 SMS), mais cela reste à confirmer ou infirmer.

10.3 Démarche

Nous décomposons le travail de traduction par méthode statistique en quatre étapes :

- Préparer les données (corpus bilingue et dictionnaire bilingue L2-L1).
- Aligner le corpus parallèle et générer le système de TA.
- Expérimenter avec le décodeur.
- Évaluer les résultats trouvés.

Dans ce qui suit, nous décrivons pour chaque étape, la démarche pratique que nous avons suivie.

10.3.1 Génération de corpus

Nous avons proposé une méthode pour construire un corpus bilingue parallèle par déconversion d'un corpus représenté sous forme CRL-CATS vers la langue L2. Les conditions identifiées dans le paragraphe 3.1.4.3 sont vérifiées. En effet, nous avons bien :

- un corpus en L1 (arabe) même s'il est petit,
- la représentation interne (CRL-CATS) correspondante du corpus L1,
- un environnement (langage) de déconversion,
- des compétences en langage DeCo (langage de déconversion),
- des compétences langagières, en effet nous maîtrisons les deux langues en question (arabe et français).

N'oublions pas que l'outil DeCo demande un dictionnaire (français CRL-CATS). Nous utilisons alors le même dictionnaire que celui utilisé dans la première méthode de portage basée sur l'outil EnCo. Il reste à construire les règles qui permettent la génération, à partir de l'ensemble des CRL-CATS, de SMS français en tenant compte de la spontanéité du résultat voulu.

Dans le cas où le corpus parallèle obtenu ne suffit pas, la même méthode de déconversion ne permette de générer un corpus parallèle vers les deux langues (originale et cible).

Dans tout les cas, nous avons prévu une étape de révision et de nettoyage manuelle ou semi-automatique dans laquelle nous vérifions que le résultat de déconversion correspond à une traduction fonctionnelle.

10.3.2 Traduction Automatique Statistique : TAS

10.3.2.1 Alignement

L'alignement est, de manière générale, le fait de mettre en correspondance les parties des deux textes d'un corpus bilingue qui sont traductions mutuelles l'une de l'autre. Un alignement peut être

plus ou moins fin. On peut trouver plusieurs niveaux d'alignement : au niveau des phrases, des groupes élémentaires, ou des « mots » (linguistiques, pas typographiques).

Les traductions d'une langue dans une autre sont rarement littérales. Si on peut en général effectuer un alignement systématique au niveau des phrases dans un corpus bilingue, il n'est pas toujours possible d'établir une correspondance entre tous les mots d'une phrase et de sa traduction.

De nombreuses recherches ont déjà été menées sur les problèmes d'alignement sous-phrastique. (Och and Ney 2003) ont effectué une comparaison entre les différentes méthodes d'alignement au niveau du mot. Ils ont présenté plusieurs méthodes pour combiner ces différentes méthodes pour une meilleure performance d'un modèle statistique d'alignement. Ce travail a abouti à la production d'un outil d'alignement qui devient de plus en plus utilisé, appelé GIZA++.

Nous utiliserons GIZA++ comme outil d'alignement au niveau des mots. GIZA++ est une extension de l'outil GIZA qui a été développé en 1999 par le groupe de traduction statistique du centre de recherche CLSP (Center for Language and Speech Processing) à l'université Johns-Hopkins.

10.3.2.2 Traduction

Plusieurs décodeurs pour la traduction statistique sont disponibles sur le Web. Les plus connus sont Pharaoh (Koehn 2004)³⁹, et Moses⁴⁰, mais d'autres boîtes à outil sont aussi disponibles, comme Fickle, et Tackle. Ils sont gratuits et en utilisation libre.

Nous cherchons à utiliser un décodeur qui peut être entraîné avec les outils GIZA et GIZA++. Nous choisissons alors Pharaoh, gratuit pour une utilisation non commerciale.

Pharaoh est un décodeur de traduction automatique réservé à la communauté de recherche pour faciliter la recherche en traduction automatique statistique. Il a été développé par Philipp Koehn comme partie de sa thèse de doctorat à l'université de la Californie du Sud (USC) Institut des Sciences de l'Information (ISI).

Le travail en question nécessite un modèle de langage pour la langue cible. Des outils gratuits sont disponibles sur le Web, comme SRILM (Stolcke 2002), mais il faut avoir une quantité minimale suffisante de texte pour construire un modèle de langage.

10.3.3 Évaluation

Plusieurs méthodes numériques ont été proposées pour l'évaluation des systèmes de traduction automatique, comme NIST, BLEU, ORANGE, METEOR... L'idée est de comparer la distribution des n-grammes à celle observée dans les traductions de référence produites par des humains pour mesurer la « qualité ». (Boitet 2007) dit que :

Il s'est avéré depuis que cette affirmation est fausse. Ce qui est mesuré, ce n'est pas la qualité des traductions soumises au test, mais leur similarité avec les traductions de référence. Or, la "densité" des traductions correctes dans l'espace des traductions possibles est faible. Autrement dit, la distance d'une excellente traduction à une traduction "de référence" peut être très grande. Ainsi, plusieurs expériences ont montré que des traductions humaines, révisées, jugées parfaites, étaient reléguées au fond du classement, derrière des traductions souvent gauches, inexactes, voire incompréhensibles, produites par des systèmes de TA (Boitet 2007).

³⁹ <http://www.isi.edu/licensed-sw/pharaoh/>

⁴⁰ <http://www.statmt.org/moses/>

D'autres critères objectifs sont utilisés dans l'évaluation d'une traduction. Les principaux sont essentiellement la grammaticalité, l'exactitude terminologique, le coût de la TA (prix, temps, espace), le temps de post-édition, et le temps d'ajustement des dictionnaires.

Nous pouvons utiliser dans notre évaluation les méthodes NIST et BLEU. Notre objectif n'est pas d'avoir une traduction parfaite (car le résultat est destiné à un programme, l'EC, et pas à un humain), mais d'extraire l'information pertinente dans un énoncé. Il s'agit par exemple dans le cas de CATS d'extraire les propriétés importantes (type, modèle, prix, année...) d'une voiture à partir d'un SMS arabe qui le résultat d'une traduction statistique d'un SMS français. De la même façon que pour les deux méthodes de portage utilisées précédemment, nous mesurerons alors le rappel et la précision pour la tâche d'extraction de contenu.

Chapitre 11 Production de corpus parallèles par déconversion

Introduction

Nous avons déjà construit un petit corpus « fonctionnel » pour CATS en français pour les deux expériences de portage « interne » et « externe », et disposons donc d'un corpus dans le sous-langage SL2 (ici, celui de Cars en français). Comme nous ne savions pas si ce petit corpus suffirait, ou si nous aurions besoin d'un corpus beaucoup plus grand, nous avons cherché une méthode automatique qui permette de produire un corpus parallèle automatiquement. Si l'on a accès au code, comme pour CATS, le plus simple semble être de « déconvertir » un ensemble suffisamment grand de représentations de contenu (en CRL-CATS) vers SL1 et SL2.

Disposant déjà de 1100 SMS en arabe associés à leurs représentations de contenu, nous nous sommes limités à expérimenter la construction d'un déconvertisseur vers le français, et cela a suffi. S'il avait fallu obtenir un corpus beaucoup plus grand, nous aurions généré aléatoirement un grand nombre de représentations de contenu, et nous aurions aussi écrit un déconvertisseur vers l'arabe (SL1).

Dans la première partie, nous décrivons les ressources et les outils que nous allons utiliser dans ce travail de déconversion. Dans la deuxième partie, nous présentons la méthode utilisée pour la déconversion vers le sous-langage du français en question. Dans la troisième partie, nous présentons les résultats et le coût de ce travail.

11.1 Ressources et outils mis en oeuvre

Puisque les représentations CRL-CATS sont obtenues à partir de l'arabe en utilisant le langage EnCo, il nous a paru naturel de faire l'opération inverse (déconversion) en utilisant le langage DeCo (conçu aussi pour le projet UNL et très semblable à EnCo).

11.1.1 Ressources

Actuellement nous disposons :

- d'un petit ensemble de 1100 SMS en arabe, filtré et lié au domaine de l'automobile,
- des représentations CRL-CATS correspondant à ces SMS,
- d'un dictionnaire (français CRL-CATS) de 1761 entrées,

Ci-dessous, nous donnons de nouveau un exemple de SMS arabe et de sa représentation CRL-CATS.

لبيع سيارة رينو كليو موديل 2002 بحاله ممتازه بسعر 5500 دينار

```

; A vendre RENAULT CLIO modèle 2002 bon état 5500
[S]
sal(saloon:06,    sale:00)
mak(saloon:06,    RENAULT(country<France, country<europe):0C)
mod(saloon:06,    Clio(country<France, country<europe, make<RENAULT):0H)
yea(saloon:06,    2002:0S)
fea(saloon:06,    good condition:0X)
pri(saloon:06,    5500:1F)
[/S]

```

Figure 118 : exemple d'un SMS arabe et de sa représentation CRL-CATS

11.1.2 Outils

Nous disposons de l'outil DeCo et de sa spécification version 3.0 (24-01-2003) (Uchida and Zhu 2003). DeCo permet d'écrire des règles en interaction avec un dictionnaire. Plusieurs décodeurs ont été écrits dans ce langage et dans d'autres environnements.

Dans la suite, nous décrivons l'outil de décodage DeCo en précisant son architecture et son fonctionnement.

11.1.2.1 Architecture et fonctionnement de DeCo

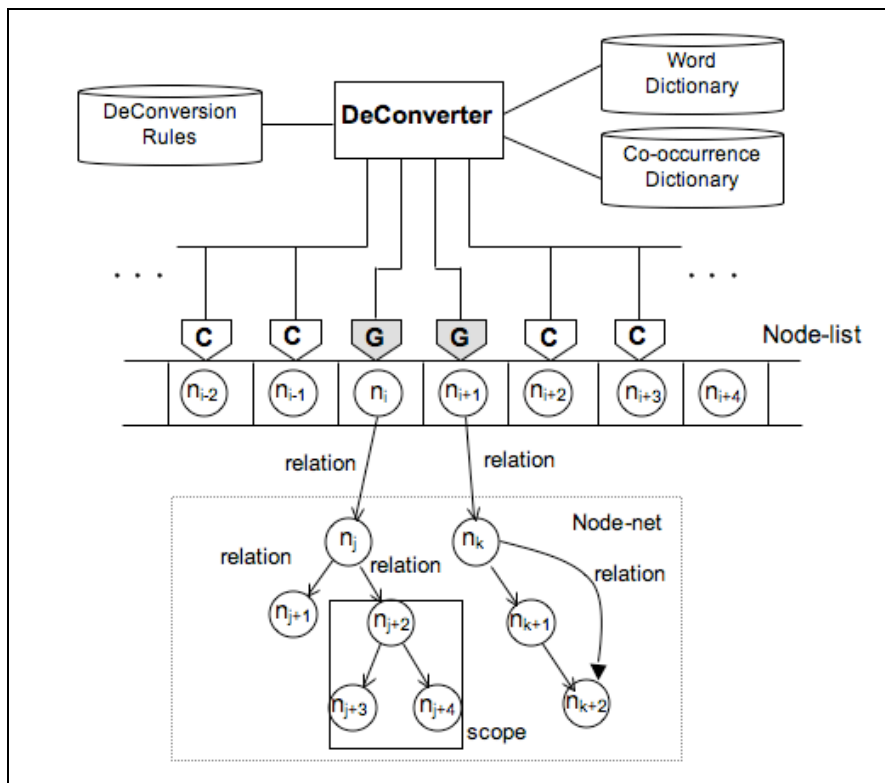


Figure 119 : modèle de calcul du langage DeCo

« G » indique fenêtre de Génération, « C » indique la fenêtre de Condition et « n » indique un nœud de génération (Uchida 2002).

La Figure 119 représente l'architecture du fonctionnement de DeCo. DeCo manipule les nœuds dans une liste de nœuds "Node-list" à travers ses fenêtres. Il insère des nœuds à partir d'un graphe de

nœuds “*Node-net*” dans la liste de nœuds. Il y a deux types de fenêtres, des fenêtres de génération “*Generation Window*” (GW) et des fenêtres de condition “*Condition Window*” (CW).

DeCo génère une phrase en utilisant un dictionnaire “*Word Dictionary*”, des règles “*Deconversion Rules*”, et un dictionnaire de co-occurrences “*Co-occurrence Dictionary*”. Il retrouve les entrées dictionnaires dans le “*Word Dictionary*”, manipule ou insère les nœuds en appliquant les règles de déconversion, et sélectionne les items lexicaux à partir des UW (dans notre cas, les CW de CATS) et en faisant éventuellement référence au dictionnaire de co-occurrences. L’utilisation du dictionnaire de co-occurrences est optionnelle (dans notre cas, nous ne l’utiliserons pas).

L’outil DeCo utilise deux fenêtres de condition (CW) pour tester les nœuds voisins (à gauche ou à droite) des deux nœuds « courants », placés sous deux fenêtres de génération (GW), contiguës, selon la ou les conditions correspondantes éventuellement présentes dans la règle de déconversion.

Un nœud du graphe ou de la liste des nœuds a la même structure qu’une entrée dictionnaire composée d’une entête “*Headword*”, d’un mot universel “*Universal Word*”, et des attributs grammaticaux et de graphe.

La fenêtre de génération (GW) est utilisée pour chercher deux nœuds adjacents pour pouvoir leur appliquer une règle de déconversion. S’il y a une règle applicable, DeCo modifie les attributs grammaticaux des deux nœuds, et/ou insère un nœud dans la liste des nœuds à partir du graphe de nœuds. Ce processus continue jusqu’à ce que tous les éléments du graphe de nœuds soient insérés dans la liste. Les chaînes contenues dans les nœuds de la liste des nœuds composent alors la phrase générée.

DeCo prend comme entrée une représentation UNL constituant le graphe de nœuds. Il retrouve pour chaque nœud les entrées candidates dans le dictionnaire (à partir de leurs UW). Chaque nœud du graphe aura alors une liste d’entrées candidates. Une seule entrée doit être choisie. Le nœud sera alors modifié :

- sa chaîne et son UW deviennent celles de l’entrée,
- ses attributs grammaticaux sont modifiés par la règle à appliquer, en fonction de ceux du nœud et de ceux de l’entrée. Le processus s’arrête avec succès quand le graphe de nœuds est vide. En cas de blocage (ou d’instruction explicite d’une règle) un retour arrière est effectué.
- ses attributs de graphe restent inchangés.

11.1.2.2 Dictionnaire

Chaque entrée du dictionnaire est composée de trois type d’éléments.

- Le *Headword* ou « mot-vedette » est une chaîne, le plus souvent un lemme, une forme fléchiée ou un affixe qui peut être utilisée pour générer du texte.
- L’UW représente un des sens du mot-vedette.
- Les attributs grammaticaux sont des informations comme la classe morphosyntaxique (nom, verbe, adjectif) ou les valences syntaxiques.

La syntaxe générale du dictionnaire est la même que celle d’EnCo (voir 8.1.4) :

Exemple :

```
[Passat] {"Passat (country>germany, country>europe, make>VOLKSWAGEN) " (model, car)
<F, 3, 3>;
[Passat 1.6] {"Passat (country>germany, country>europe, make>VOLKSWAGEN) " (model, car)
<F, 3, 3>;
[Passat 1.8] {"Passat (country>germany, country>europe, make>VOLKSWAGEN) " (model, car)
<F, 3, 3>;
```



```
[Passat 1.8-5V-
Turbo]{} "Passat(country>germany,country>europe,make>VOLKSWAGEN)" (model,car)
<F,3,3>;
[Passat]{} "Passat(country>germany,country>europe,make>VOLKSWAGEN)" (model,car)
<F,3,3>;
[Passat V6 Cmeffrout
line]{} "Passat(country>germany,country>europe,make>VOLKSWAGEN)" (model,car) <F,3,3>;
[Passat VR5]{} "Passat(country>germany,country>europe,make>VOLKSWAGEN)" (model,car)
<F,3,3>;
[Polo]{} "Polo(country>germany,country>europe,make>VOLKSWAGEN)" (model,car) <F,3,3>;
[Polo 1.0]{} "Polo(country>germany,country>europe,make>VOLKSWAGEN)" (model,car)
<F,3,3>;
[Polo Fox]{} "Polo(country>germany,country>europe,make>VOLKSWAGEN)" (model,car)
<F,3,3>;
[Polo Variant
1.4]{} "Polo(country>germany,country>europe,make>VOLKSWAGEN)" (model,car) <F,3,3>;
```

Figure 120 : exemple d'entrées dictionnaires en DeCo

De façon parallèle à EnCo, si une UW a pour en-tête un nombre en chiffres arabes, DeCo crée automatiquement une entrée de dictionnaire temporaire et lui affecte les attributs NUM, ANUM (NUM pour « NUMerals », ANUM pour « Arabic NUMerals »).

Comme le montre l'exemple suivant, l'entrée pour un ANUM est alors de la forme suivante et sa chaîne est le même que son UW :

```
[106]{} "106" (NUM, ANUM) <., 0, 0>;
```

La Figure 121 montre trois exemples d'entrées numériques permanentes, insérées à la main dans le dictionnaire.

```
[206]{} "206" (model,car) <F,3,3>;
[106]{} "106" (model,car) <F,3,3>;
[306]{} "306" (model,car) <F,3,3>;
```

Figure 121 : exemple d'entrées dictionnaires numériques

Pour chaque UW, DeCo trie toutes les entrées trouvées dans le dictionnaire. Il les place d'abord suivant la valeur de priorité et ensuite suivant la longueur du mot. Premièrement, il donne la priorité à l'entrée qui a la valeur de priorité la plus élevée. Deuxièmement et dans le cas où deux entrées ou plus ont la même valeur de priorité, il donne la priorité à l'entrée la plus longue.

11.1.2.3 Base de règles

Une règle de déconversion est composée de :

- Conditions sur les nœuds placés dans la fenêtre de génération (GW) et la fenêtre de condition (CW).
- Actions et/ou opérations pour les nœuds placés dans la fenêtre de génération (GW).
- Relations entre les nœuds.

Chaque règle décrit les actions et/ou opérations (insertion, suppression, ajout, décalage...) à exécuter et sous quelle condition. DeCo trouve la règle qu'il faut appliquer à chaque fois, insère un nouveau mot ou un suffixe ou un préfixe pour compléter un mot et ainsi de suite... jusqu'à la construction de la phrase.

La syntaxe générale pour écrire une règle de déconversion est la suivante :

```

<TYPE>
["("<PRE>")" ["*"]]....
"{ " | " " " " [ <COND1> ] ":" [ <ACTION1> ] ":" [ <RELATION1> ] ":" [ <ROLE1> ]
}" | " " " "
["("<MID>")" ["*"]]....
"{ " | " " " " [ <COND2> ] ":" [ <ACTION2> ] ":" [ <RELATION2> ] ":" [ <ROLE2> ]
}" | " " " "
["("<SUF>")" ["*"]]....
"P(" <PRIORITY> ");"

```

Tableau 45 : syntaxe des règles de DeCo (1)

Les caractères entre doubles quotes sont prédéfinis comme des délimiteurs de la règle.

{ }	indique la rangée,
{ }... ou []...	indique la répétition de la même description pour plus de 0 fois,
	indique la sélection,
[]	Indique les parties admissibles
* après ()	Indique la répétition des nœuds qui ont la même condition pour plus de 0 fois.

Tableau 46 : syntaxe des règles de DeCo (2)

Les expressions entre { } et " " sont utilisées pour décrire les règles pour le nœud désigné par la fenêtre de génération. { } sont utilisées pour désigner le nœud qui existe dans la liste des nœuds, et les guillemets " " sont utilisés pour désigner le nœud qui doit être inséré.

{ }	Indique un nœud qui existe dans la liste des nœuds
" "	Fait référence à un nœud qui doit être inséré dans la liste des nœuds

Tableau 47 : syntaxe des règles de DeCo (3)

Notons que, pour n'importe quelle règle de déconversion, il est possible d'insérer un seul nœud à la fois depuis la liste des nœuds.

En général, on peut lire une règle de la façon suivante. Si les conditions suivantes sont satisfaites :

- le nœud de la fenêtre de génération de gauche satisfait les attributs de <COND1>.
- le nœud de la fenêtre de génération de droite satisfait les attributs de <COND2>.
- les deux mots des deux nœuds en question remplissent la Co-occurrence Relation de <ROLE1> ou <ROLE2>.
- les nœuds des côtés gauche, milieu et droite de la fenêtre de génération remplissent les conditions décrites dans <PRE>, <MID> et <SUF> de la fenêtre de condition « *Condition Windows* ».

alors les attributs grammaticaux du nœud dans la fenêtre de génération sont réécrits en accord avec, respectivement, à <ACTION1> et <ACTION2>. Si un autre nœud dans la fenêtre de génération est indiqué comme "Insertion", le nœud lié à un autre nœud par <RELATION1> ou <RELATION2> dans « *Node-net* » est inséré dans la liste des nœuds. Et les opérations sont exécutées vers l'extrémité de la liste des nœuds en tenant compte du type de règle montré dans le champ <TYPE>.

<PRIORITY> : décrit l'ordre d'interprétation des règles.

La valeur de la priorité est comprise entre 0 et 255. Les règles sans valeur de priorité sont considérées comme ayant une priorité nulle.

11.1.3 Ce que nous souhaitons

Nous souhaitons déconvertir en français toutes les représentations CRL-CATS dont nous disposons. Pour l'exemple de la Figure 118, nous voudrions obtenir un résultat de déconversion qui ressemble à :

A vendre RENAULT Clio modèle 2002 très bon état 5500 dinars

Avec notre déconvertisseur, nous avons en fait obtenu après déconversion à partir de CRL-CATS le texte suivant :

```
[S]
Vente voiture RENAULT Clio 2002 TB état 5500 dinars
;;Time 0.0 Sec
[S]
```

Le résultat produit correspond bien à la représentation CRL-CATS de départ. Sémantiquement, il s'agit de la même phrase. Le choix du terme « A vendre » ou « vente » revient à l'attribution des priorités dans le dictionnaire. Voyons dans ce qui suit le détail du travail de la déconversion.

11.2 Déconversion à partir de CRL-CATS

La Figure 122 résume l'architecture de notre cas d'utilisation de DeCo. Comme entrées, on donne une représentation CRL-CATS, des règles et un dictionnaire, et comme sortie, on obtient un texte en français.

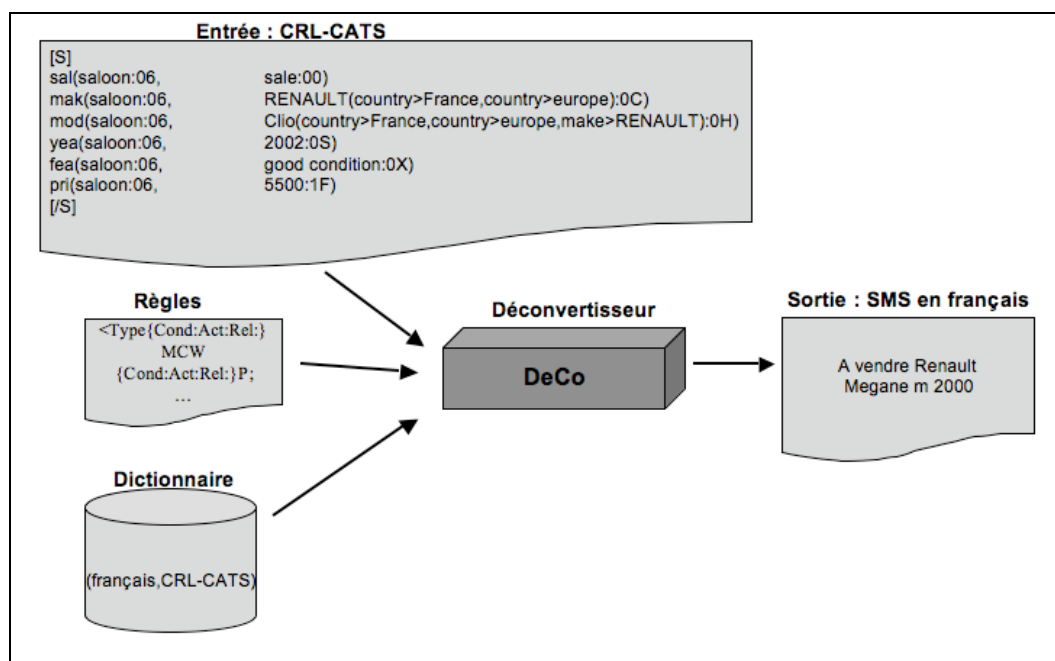


Figure 122 : architecture de notre cas d'utilisation de DeCo

11.2.1 Entrée

Comme le montre l'exemple de la Figure 122, l'entrée est une représentation CRL-CATS. Dans CRL-CATS, il n'y a aucune indication sur le nœud d'entrée (@entry). Nous avons donc édité toutes les représentations CRL-CATS pour préciser le nœud d'entrée en ajoutant l'attribut @entry. C'est généralement saloon ou bus ou pick up ou van. Dans le cas de l'exemple de la Figure 122, nous ajoutons @entry dans les six relations, ainsi la première relation devient sal(saloon.@entry:06,sale :00).

Les représentations CRL-CATS à déconvertir sont mises dans un fichier texte, où elles sont délimitées par un début [S] et une fin [/S]. Les lignes précédées par “;” sont traitées comme des commentaires.

11.2.2 Sortie

DeCo donne comme sortie l'énoncé obtenu après déconversion ainsi que des informations de trace suivant le niveau de trace demandé en paramètre. La trace montre l'ordre des règles appliquées, les règles appliquées, l'état des nœuds dans la fenêtre de génération avant et après l'application de chaque règle, la liste des nœuds, et l'arbre syntaxique.

Cinq niveaux de trace sont définis par DeCo, comme indiqué dans le Tableau 48 :

Niveaux	0	1	2	3	4
Sortie de la phrase cible	+	+	+	+	+
Entrée UNL	-	+	+	+	+
Liste des mots	-	+	+	+	+
Liste des nœuds	-	-	+	+	+
Nœuds dans la fenêtre de génération (avant et après l'application de la règle)	-	-	+	+	+
Règles appliquées	-	-	+	+	+
Règle de retour arrière	-	-	-	+	+
Nœuds de retour arrière	-	-	-	+	+
Règles candidates qui satisfaisaient la condition de la fenêtre de génération	-	-	-	-	+

Tableau 48 : niveaux de la sortie de l'outil DeCo

+ : affichage de l'option

- : non affichage de l'option

11.2.3 Dictionnaire

Nous avons utilisé le dictionnaire de 1761 entrées fabriqué dans la section 7.3 pour le travail d'enconversion du même sous-langage du français vers CRL-CATS. Nous donnons ici un exemple de traitement des entrées candidates.

Pour l'UW `B.M.W(country>germany, country>europe)`, la liste des entrées candidates est la suivante :

```
[B.M.W.] {"B.M.W(country>germany, country>europe)" (make, nmodel, car) <F, 3, 3>;
[B.M.W] {"B.M.W(country>germany, country>europe)" (make, nmodel, car) <F, 3, 3>;
 [BMW] {"B.M.W(country>germany, country>europe)" (make, nmodel, car) <F, 3, 3>;
 [Bmw] {"B.M.W(country>germany, country>europe)" (make, nmodel, car) <F, 3, 3>;
 [bmw] {"B.M.W(country>germany, country>europe)" (make, nmodel, car) <F, 3, 3>;
 [BM] {"B.M.W(country>germany, country>europe)" (make, nmodel, car) <F, 3, 3>;
 [W] {"B.M.W(country>germany, country>europe)" (make, nmodel, car) <F, 3, 3>;
```

Figure 123 : liste d'entrées candidates triées pour un exemple d'UW

Les six entrées candidates ont la même priorité (3), donc DeCo choisira l'entrée la plus longue, ici la première et généra donc « B.M.W. ».

11.2.4 Construction du déconvertisseur

Nous avons d'abord fabriqué une première version de règles qui déconvertissent les dix relations de CRL-CATS (*wan, sal, cou, yea, col, fea, pri, mak, mod, mot*), puis avons exécuté le déconvertisseur et étudié les résultats obtenus.

Ensuite, nous avons développé une deuxième version des règles qui tient compte des erreurs et des mauvais résultats trouvés dans la première version.

Voici un exemple : dans la figure suivante, nous attendions une sortie correcte comme *recherche une voiture TOYOTA noire*, mais l'article « *une* » a été ajouté à la chaîne en cours de création avant « *recherche* » et pas après.

```

===== UNL =====
saloon(,@entry,>wan,>mak,>col)
  -wan>wanted(<wan)
  -mak>TOYOTA(country>japan):00(<mak)
  -col>black:0V(<col)
=====
une recherche voiture TOYOTA noire
;;Time 0.0 Sec

```

Figure 124 : exemple de mauvaise sortie

L'erreur provient des deux règles suivantes. Pour la corriger, nous devons modifier leurs valeurs de priorité. La première est :

```
: "[une],ART::" {car,@entry:-@entry::} P19;
```

Cette règle peut être lue de la façon suivante : il s'agit d'une insertion (TYPE = :) d'un nœud contenant comme chaîne l'article [une] et comme attribut grammatical ART (sans aucune action (ACTION1 = rien) et sans aucune relation (RELATION1 = rien)), à gauche d'un terme défini comme [car] et comme une entrée @entry avec l'action de retirer l'attribut @entry figurant en condition -@entry (car ACTION2 = -@entry) pour que cette règle soit exécutée une seule fois et sans aucune relation). La valeur de priorité de cette règle est 19. L'exécution de cette règle crée un nœud de chaîne [une] à gauche du nœud d'UW « saloon ».

[une]	[voiture]
" "	"saloon"
{ART}	{car}
{}	{}

La deuxième règle est :

```
: "want::wan:" {car,@entry:-@entry::} P17;
```

Cette règle peut être lue de la façon suivante : il s'agit d'une insertion (TYPE = :) d'un terme défini dans le dictionnaire comme *want* (sans aucune action car ACTION1 = rien mais il y a une relation : ce terme doit avoir la relation *wan* avec le nœud *car*) à gauche d'un terme défini comme *car* avec la condition qu'il soit une entrée (@entry) (avec l'action de retirer la condition -@entry (car ACTION2 = -@entry) pour que cette règle soit exécutée une seule fois et sans aucune relation). La valeur de priorité de cette règle est 17. L'exécution de cette règle donne les trois nœuds suivants.

[une]	[recherche]	[voiture]
" "	"wanted"	"saloon"
{ART}	{}	{car}
{}	{}	{}

Or, nous voulons la sortie « *rechercheunevoiture* ». Pour cela, nous commençons par inverser les valeurs de priorité de ces deux règles, et insérer deux blancs :

```
: "want::wan:" {car,@entry:-@entry::} P19;
: "[une],ART:::" {car,@entry:-@entry::} P17;
```

(NB : l'ordre d'écriture des règles dans le fichier n'a pas d'importance.)

La règle d'insertion de l'article « *une* » est exécutée alors après l'insertion du terme « *want* » et nous obtenons réellement l'état suivant (correspondant à la chaîne « *rechercheunevoiture* ») :

[recherche]	[une]	[voiture]
"wanted"	" "	"saloon"
{want}	{ART}	{car}
{}	{}	{}

Nous insérons un blanc après « *recherche* » en appliquant la règle suivante (avec la condition qu'il n'existe pas déjà, exprimée par `^blk`) :

```
: {want,^blk:blk::} "[ ],BLK:::" P4;
```

et nous obtenons des nœuds qui correspondent à « *recherche unevoiture* ».

Nous insérons un blanc après le terme « *une* » en appliquant la règle suivante :

```
: {ART,^blk:blk::} "[ ],BLK:::" P4;
```

et nous obtenons les nœuds suivants (« *recherche une voiture* ») :

[recherche]	BLK	[une]	BLK	[voiture]
"wanted"		" "		"saloon"
{want,blk}		{ART,blk}		{car}
{}		{}		{}

Ainsi de suite, nous obtenons le résultat final correct représenté par la Figure 125.

```
===== UNL =====
saloon(,@entry,>wan,>mak,>col)
  -wan>wanted(<wan)
  -mak>TOYOTA(country>japan):00(<mak)
  -col>black:0V(<col)
=====
une recherche voiture TOYOTA noire
;;Time 0.0 Sec
```

Figure 125 : exemple de sortie correcte

Voyons ici un autre type de mauvaise sortie que nous avons obtenu suite à l'exécution d'une première version des règles :

```
A vendre une voiture clioRENAULT en bon état
;;Time 0.0 Sec
```

Figure 126 : deuxième exemple de mauvaise sortie

L'exemple de la Figure 126, montre un cas de mauvaise sortie. Il nécessite une intervention au niveau des priorités des règles pour pouvoir afficher le type de voiture « *RENAULT* » avant le modèle « *clio* » ainsi qu'une insertion d'un blanc entre les deux noms.

Ce sont les mêmes règles qu'il faut modifier pour corriger l'exemple de la Figure 127, sauf que dans cet exemple, on doit aussi modifier la priorité d'application des règles pour que la relation couleur `col` s'applique après la relation modèle `mod`, car c'est plus pratique de dire « *A vendre une voiture BMW X3 blanche toutes options* ».

```
A vendre une voiture X3BMW toutes options blanche
;;Time 0.0 Sec
```

Figure 127 : troisième exemple de mauvaise sortie

Ce travail d'amélioration et de correction de cette première version a abouti à une deuxième version des règles qui tient compte de toutes ces erreurs et donne de bons résultats, comme le montre la Figure 128.

```
recherche voiture TOYOTA 2000
recherche voiture MITSUBISHI toutes options
recherche voiture KIA 95 première main
A vendre voiture RENAULT CLIO 91 assurance complète
A vendre voiture RENAULT CLIO contrôle technique assuré
A vendre voiture PEUGEOT 306 contrôle technique assuré
```

Figure 128 : quelques exemples de bonnes sorties de déconversion

11.3 Résultats et coûts

Nous avons déconverti en français tout le corpus de CRL-CATS que nous avons, et évalué le coût de cette opération.

11.3.1 Résultats

Nous disposons d'un total de 1100 SMS arabes pour le domaine de l'automobile. À partir des 1100 représentations CRL-CATS correspondant à ces SMS, nous avons généré par déconversion une première version de textes français. Nous avons mis 704 mn (11,7 h) pour réviser le résultat obtenu qui donne un corpus français de 10800 mots (l'équivalent à 43 pages standard).

Notons qu'ici, notre but n'est pas d'obtenir une traduction parfaite, mais plutôt une traduction fonctionnelle qui donne un corpus bilingue parallèle pour la traduction statistique. Rappelons aussi que nous utilisons la traduction pour l'extraction d'information ; nous évaluerons donc cette tâche dans le prochain chapitre.

Le Tableau 49 montre quelques exemples de résultats de déconversion. Dans certains cas, la qualité du résultat n'est pas parfaite, ce qui justifie le temps mis pour réviser tout le corpus de 1100 SMS français obtenu.

Recherche Voiture KIA 95
Recherche Voiture TOYOTA 4 4
A vendre Voiture RENAULT Megane 2000
Voiture TOYOTA Corolla gris
Voiture LAND ROVER
Voiture MITSUBISHI LANCER gris
Recherche Voiture PEUGEOT 307
Voiture GOLF 1600 99 TB état 10000
Recherche Voiture MERCEDES 81
Recherche Voiture HYUNDAI 3800
Recherche Voiture MITSUBISHI LANCER 4 99 97
A vendre Voiture 5600 86
Recherche Voiture MITSUBISHI Galant 1991 1988 TB état boîte normale
Recherche Voiture MITSUBISHI TB état boîte normale
Recherche Voiture honda 1300 2500
Pour vente Voiture MITSUBISHI LANCER 1300 2002
Recherche Voiture GOLF 91 86
Vends bus TOYOTA 1986
Recherche Voiture GOLF 90 85
Recherche Voiture GOLF
Pour vente Voiture OPEL 87 TB état
Recherche Voiture PEUGEOT 207
Recherche Voiture PEUGEOT
Recherche Voiture PEUGEOT 206
Voiture MAZDA 626
Recherche Voiture NISSAN jeep 95777953 95
Voiture MERCEDES 4 gris TB état air conditionné
Recherche Voiture TOYOTA 88
Recherche transport commun NISSAN
Recherche Voiture PAJERO 2003
Recherche bus HYUNDAI 2000 Neuve
Recherche Voiture MERCEDES 200

Tableau 49 : exemples de résultat de déconversion

11.3.2 Coût

Lors de l'atelier UNL organisé à Genève en avril 2006, nous avons eu un début d'expérience de déconversion vers le français à partir d'UNL. Le présent travail a pris environ 72 heures, venant après environ 180 heures passées sur EnCo et DeCo à l'occasion de cet atelier.

Notons que nous avons utilisé le dictionnaire fabriqué dans la première méthode de portage, dont la préparation avait coûté 80 heures, ce qui donne un total de 152 heures pour ce travail de déconversion.

En conclusion, ce travail de déconversion a coûté presque 1 homme/mois d'informaticien linguiste.

Tâches	Durée (h)	Version ou nombre d'items	Remarques
DeCo : spécification + expérimentation	20	2.7	Spécification obtenue par l'UNDL en tant que membre
Étude de la spécification de CRL-CATS	6	Version originale	Révision profonde
Dictionnaire	6	Version 2	Version utilisée dans un précédent travail d'enconversion
Programmation des règles : version 1	10	Version 1	Résultats encourageants
Tests	2	10 CRL-CATS	Simple tests de déconversion
Programmation des règles : version 2	6	Version 2	Meilleurs résultats
Analyse des niveaux de sorties de DeCo	4	10 CRL-CATS	But : amélioration du résultat
Tests et modifications	4	100 CRL-CATS	Modification des règles
Analyse des résultats	2	100 CRL-CATS	Vérification du classement et priorité des arguments
Déconversion du corpus	2	1100 CRL-CATS	Déconversion du corpus de représentations CRL-CATS
Révision du résultat obtenu	11,7	1100 CRL-CATS	Révision du corpus français obtenu (1100 SMS)
Total	71,7		

Tableau 50 : durée des tâches de déconversion

11.3.3 Remarques sur la gestion des arguments

Dans ce travail, il est possible de produire des versions différentes de messages en jouant sur l'ordre de déconversion des arguments (*type, modèle, année, couleur...*) d'une voiture pour produire des corpus réels différents utilisant un même dictionnaire mais une version différentes des règles. En effet, à chaque ordre de classement des arguments correspond un classement de priorité d'application des règles.

On peut produire d'autres versions de SMS en jouant aussi sur la priorité du choix des entrées dans le dictionnaire, par exemple donner la priorité à l'entrée « *en très bon état* » au lieu de « *TB état* » ou « *Je suis à la recherche* » au lieu de « *Recherche* ». Pour avoir un corpus bilingue parallèle (arabe, français) avec un alignement plus correct, nous avons appliqué ici l'ordre suivant qui est le plus proche de celui des SMS en arabe :

Voiture + nationalité + type + modèle + moteur + année + couleur + traits (fea) + prix

Chapitre 12 Portage par traduction statistique

Introduction

Après avoir mis en place et validé une méthode permettant de produire, si nécessaire, un gros corpus parallèle Fr-Ar pour construire un système de TA statistique des SMS français vers les SMS arabes, nous passons à la construction d'un tel système.

L'intérêt, ici, est d'essayer de répondre à la question suivante, pour laquelle nous avons pas trouvé de précédente étude ni de chercheurs qui travaillent sur ce point : est-il possible, dans le cas de sous-langages restreints, de construire un système de TA statistique (donnant des résultats exploitables pour un portage linguistique) avec un dictionnaire complet (ou presque complet), et un corpus très petit ou assez petit ? Et, si oui, comment peut-on déterminer la taille minimale suffisante du corpus parallèle nécessaire ?

Dans la première partie, nous présentons le travail de préparation des données. Dans la deuxième partie, nous décrivons notre utilisation du décodeur Pharaoh en mode d'apprentissage et en mode de développement. Nous évaluons les résultats produits pour la traduction et pour l'extraction de contenu dans la troisième partie.

12.1 Travail préparatoire pour la traduction

La mise en place d'un système de traduction automatique statistique nécessite une étape préparatoire des données, essentiellement une étape d'alignement.

12.1.1 Ressources utilisées

Nous avons utilisé deux types de données dans cette expérience : des données d'entraînement et des données de développement.

Les données d'entraînement sont constituées d'un corpus parallèle, composé d'un corpus arabe (transtac_train.ar) et d'un corpus français obtenu par déconversion suivi d'une révision humaine (transtac_train.fr).

Le Tableau 51 résume les différentes informations concernant les données d'entraînement.

Le résultat de la comparaison entre la taille en mots du corpus français et arabe est variable. Ici, la longueur moyenne d'un SMS français est plus élevée (9,93) que celle de l'arabe (8,75). Dans d'autres blocs de données, ça peut être l'inverse. Cela peut varier en fonction de la nature des données source utilisées et de l'utilisation des variantes lexicales.

Nous expliquons cela par le fait que les données source sont des données réelles rédigées par de vrais utilisateurs réels en Jordanie. Aucune règle ne les empêche d'écrire « VOLKSWAGEN » sous forme d'un seul mot ou de deux mots « VOLKS WAGEN » ou encore « LANDROVER » et « LAND ROVER » etc.

Notons que le même type de variation peut exister en français. Puisque deux mots en arabe ou en français sont séparés par un blanc, dans certains cas « VOLKSWAGEN » peut être compté comme un seul mot et dans d'autres comme deux mots.

Taille du corpus d'entraînement	Taille du corpus arabe en mots	Nombre de mots par SMS arabe	Taille du corpus français en mots	Nombre de mots par SMS français
100	860	8,60	1010	10,10
200	1788	8,94	2051	10,26
300	2696	8,99	3052	10,17
400	3575	8,94	4006	10,02
500	4424	8,85	4954	9,91
600	5299	8,83	5907	9,85
700	6090	8,70	6809	9,73
800	6867	8,58	7742	9,68
900	7729	8,59	8811	9,79
1000	8685	8,69	9961	9,96
1100	9446	8,59	10803	9,82
Moyenne		8,75		9,93

Tableau 51 : taille des données d'entraînement

12.1.2 Alignement GIZA++

Nous avons utilisé GIZA++ comme outil d'alignement. GIZA++ prend comme format d'entrée un fichier dont l'extension est « .snt ». C'est un format segmenté et codé en chiffres.

Comme le montre la Figure 129, ce format peut être obtenu par l'exécution du script *plain2snt* qui permet la conversion et l'obtention dudit format, ainsi que le vocabulaire de chaque langue à partir de deux fichiers textes.

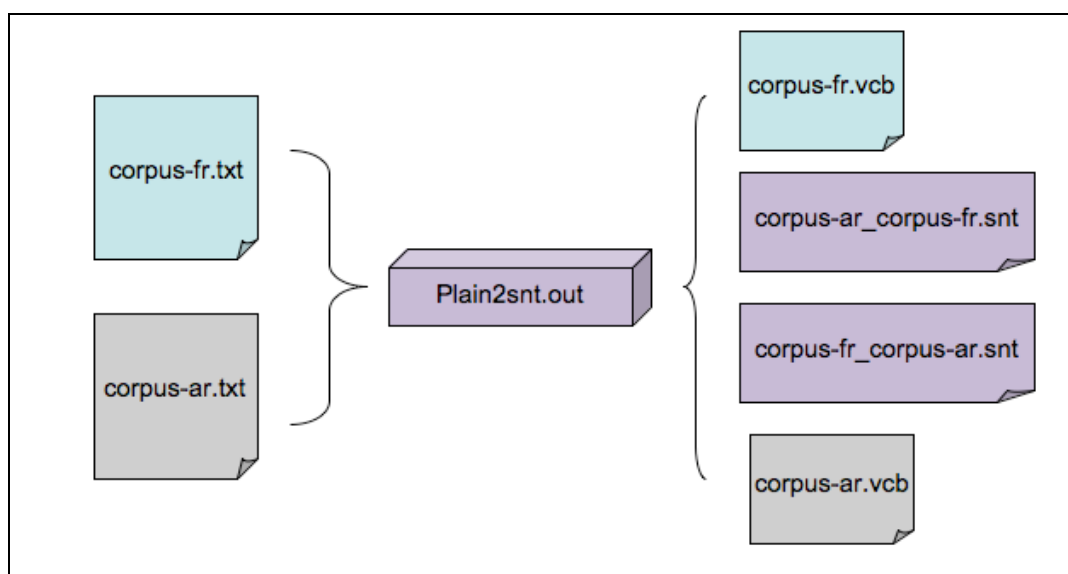


Figure 129 : conversion des données du format texte vers snt

Prenons comme exemple les segments de la figure suivante :

corpus-ar.txt	corpus-fr.txt
.مطلوب سياره من400إلي 6500 دديزل أوبنزين/0795601711.	Recherche voiture de 400 à 6500 dinars diesel ou essence /0795601711
.مطلوب سياره مقسطة ياباني أو الماني 0795777953.	Recherche voiture avec facilités de paiement japonaise ou allemande 0795777953
.مطلوب سيارة اوبل فكترا.	recherche voiture OPEL VECTRA

Figure 130 : exemple de corpus-ar et de corpus-fr

Comme le montre l'exemple du Tableau 52, les fichiers vocabulaires contiennent les mots du corpus, des identificateurs entiers et le nombre d'occurrences de ces mots.

corpus-ar.vcb	corpus-fr.vcb
1 مطلوب 2	2 Recherche 1
3 سياره 58	3 voiture 67
4 من400إلي 1	4 de 24
5 6500 دديزل 1	5 400 1
6 أوبنزين/0795601711. 1	6 à 47
7 ;مطلوب 82	7 6500 3
8 مقسطة 1	8 dinars 68
9 ياباني 1	9 diesel 3
10 أو 4	10 ou 22
11 الماني 1	11 essence 2
12 0795777953. 1	12 /0795601711
13 سيارة 42	13 Recherche 4
14 اوبل 5	14 avec 22
15 فكترا. 1	15 facilités 3
	16 paiement 4
	17 japonaise 1
	18 allemande 2
	19 0795777953 1
	20 recherche 80
	21 OPEL 11
	22 VECTRA 4

Tableau 52 : fragments du vocabulaires générés par le script plain2snt

Chaque paire de phrases est représentée par trois lignes comme le montre le tableau suivant. La première ligne indique la fréquence de la phrase dans le corpus (habituellement égale à 1 dans un processus d'entraînement). Les deux autres lignes contiennent les identificateurs des mots dans les phrases arabe et française.

>head -9 corpus/corpus-ar_corpus-fr.snt	>head -9 corpus/Corpus-fr_corpus-ar.snt
1	1
2 3 4 5 6	2 3 4 5 6 7 8 9 10 11 12
2 3 4 5 6 7 8 9 10 11 12	2 3 4 5 6
1	1
7 3 8 9 10 11 12	13 3 14 15 4 16 17 10 18 19
13 3 14 15 4 16 17 10 18 19	7 3 8 9 10 11 12
1	1
7 13 14 15	20 3 21 22
20 3 21 22	7 13 14 15

Tableau 53 : fragments du format numérique généré par le script plain2snt

Une fois que le format snt est obtenu, un fichier de configuration doit être édité pour préciser les paramètres d'entrée pour GIZA++, ainsi que le répertoire où ranger les résultats, comme le montre la Figure 131 :

```
SOURCE VOCABULARY FILE : ./corpus/corpus-fr.vcb
TARGET VOCABULARY FILE : ./corpus/corpus-ar.vcb
CORPUS FILE : ./corpus/corpus-fr_corpus-ar.snt

TEST CORPUS FILE : ./corpus/corpus-fr_corpus-ar.snt
...
OUTPUT PATH : /Users/hajlaoui/Documents/These/STA/giza++N/corpus/
```

Figure 131 : fichier de configuration (config.default) pour GIZA++

12.1.3 Analyse des résultats d'alignement

GIZA++ est lancé par l'exécution d'une simple commande. La Figure 132 est un exemple de résultat obtenu.

```
# Sentence pair (1) source length 11 target length 5 alignment score :
5.59757e-07
.0795601711/أوتوبيل/6500دينار إلى400من مازول سياره من
NULL ({} ) Recherche ({} 1 {}) voiture ({} 2 {}) de ({} ) 400 ({} 3 4 {}) à
({} ) 6500 ({} ) dinars ({} ) diesel ({} ) ou ({} ) essence ({} ) /
0795601711 ({} 5 {})

# Sentence pair (2) source length 10 target length 7 alignment score :
1.23282e-05
.0795777953 مازول سياره منسظمه باباني أو الماني
NULL ({} ) Recherche ({} 1 {}) voiture ({} 2 {}) avec ({} ) facilités ({} )
de ({} ) paiement ({} 5 6 {}) japonaise ({} 7 {}) ou ({} ) allemande
({} 3 {}) 0795777953 ({} 4 {})

# Sentence pair (3) source length 4 target length 4 alignment score :
0.00138728
. مازول سياره اوبل فكترا.
NULL ({} ) recherche ({} 1 {}) voiture ({} 2 {}) OPEL ({} 3 {}) VECTRA
({} 4 {})
```

Figure 132 : résultat d'alignement pour le corpus exemple

Pour la première paire de phrases, le premier mot « *Recherche* » est aligné avec le premier mot de la phrase arabe, le mot « *voiture* » est aligné avec le deuxième, le mot « *de* » n'est pas aligné, et ainsi de suite.

Comme le montrent les exemples d'alignements de la Figure 133, la sortie de GIZA++ est un fichier qui contient le résultat d'alignement, les phrases cibles ainsi que d'autres informations concernant la longueur de la phrase source et cible.

```
# Sentence pair (18) source length 3 target length 3 alignment score : 0.00385455
#مطلوب سيارة ميتسوبيشي.
NULL ({} ) recherche ({} 1) voiture ({} 2) MITSUBISHI ({} 3)
```

Figure 133 : exemple d'alignement correct

Le mot « *recherche* » est aligné avec le premier mot de la phrase arabe, le mot « *voiture* » est aligné avec le deuxième mot de la phrase arabe et le mot « *MITSUBISHI* » est aligné avec le troisième mot de la phrase arabe. Cela constitue un alignement correct.

```
# Sentence pair (16) source length 8 target length 9 alignment score :
1.36541e-08
#مطلوب زينو كليو غير اونوماتيك موديل 95 الهى 2000.
NULL ({} ) recherche ({} 1) RENAULT ({} 2) Clio ({} 3 9) manuelle ({} 4
5) modèle ({} 6) 95 ({} 7) à ({} 8) 2000 ({} )
```

Figure 134 : exemple d'alignement d'un mot composé

Dans l'exemple précédent, « *manuelle* » est aligné avec les mots arabes numéros 4 et 5. C'est correct, et, sémantiquement, ces deux mots constituent bien un mot composé en arabe (non automatique) mais le mot « *Clio* » est mal aligné.

Malgré la différence de la longueur moyenne des SMS utilisés, les mots sont dans la plupart des cas alignés correctement.

Dans certains cas, l'alignement est mauvais car les SMS arabes contiennent parfois des mots étrangers, comme «E200»⁴¹ dans l'exemple suivant, ou «CC», ce qui provoque un changement de sens d'écriture.

```
# Sentence pair (11) source length 5 target length 7 alignment score :
1.02623e-08
#مربيدس 200Eكلاسك جميع الاضافات 22000 دينار.
NULL ({} ) MERCEDES ({} 1) classique ({} 2) E200 ({} 3 6 7) toutes ({} 4)
options ({} 5)
```

Figure 135 : exemple de mauvais alignement

12.2 Mise en place du système

La Figure 136 montre l'architecture générale d'un système de traduction statistique.

⁴¹ On peut trouver des erreurs comme E200 au lieu de 200E.

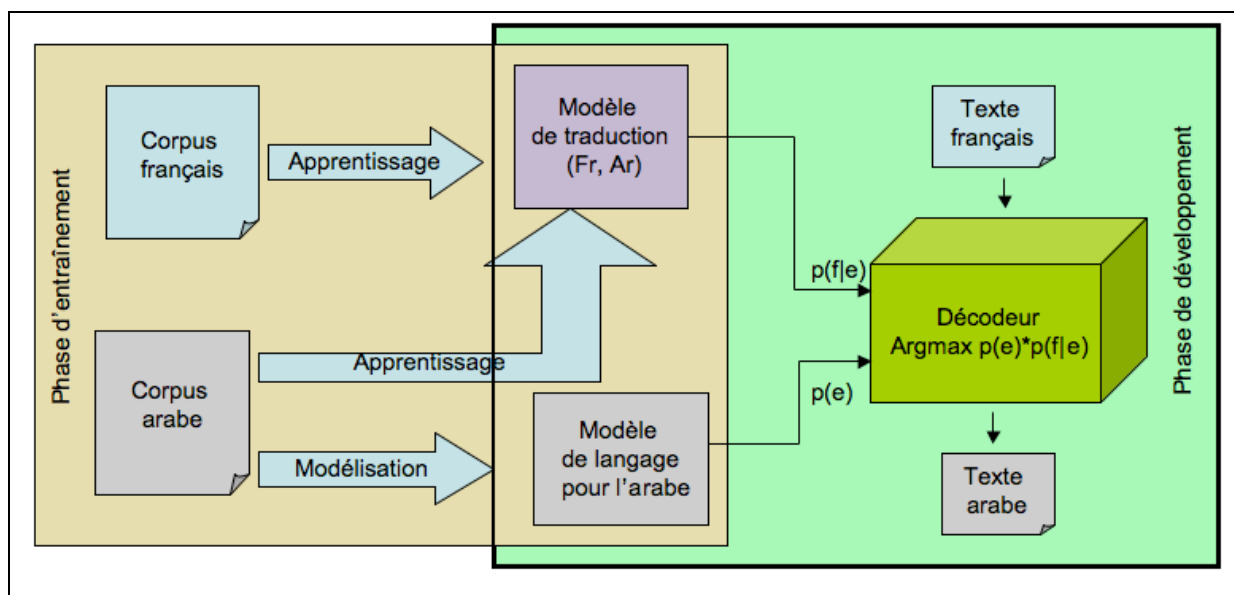


Figure 136 : architecture générale d'un système de traduction statistique

12.2.1 Modélisation du langage cible

Un modèle de langage pour la langue cible est nécessaire. Dans notre cas, l'arabe est la langue source de l'application et la langue cible pour la traduction. Nous devons donc construire un modèle de langage pour l'arabe. C'est une langue pour laquelle très peu de ressources sont disponibles et gratuites. Nous avons seulement trouvé un modèle de langage pour Coran, ce qui n'est pas du tout adapté au cas du sous-langage traité. Nous disposons aussi au laboratoire d'un modèle de langage, pour l'arabe mésopotamien (iraki), mais il s'agit d'un dialecte et d'une typologie très différentes de ceux des SMS concernant l'occasion auto à Amman.

Nous avons donc dû construire un modèle de langage pour le sous-langage du domaine de l'automobile en nous basant sur les données disponibles.

Nous avons utilisé le constructeur de modèle de langage SRILM disponible gratuitement sur le web⁴² (Stolcke 2002). Nous avons entraîné le modèle de langage sur le même corpus d'entraînement que celui utilisé pour l'entraînement du décodeur de traduction. De la même façon, nous sommes parti d'une taille de corpus minimale et nous avons augmenté la taille au fur et à mesure jusqu'à avoir des résultats satisfaisants.

Nous montrons dans la troisième partie de ce chapitre l'évolution des résultats en fonction de la taille du corpus d'entraînement.

12.2.2 Pharaoh en mode apprentissage

L'exécution du script perl « *train-phrase-model.perl* » prend comme paramètre un corpus parallèle d'entraînement « *transtac_train* » avec l'indication des extensions « *fr* » et « *ar* » et génère une table de correspondance entre les segments « *phrase-table* » et un fichier de configuration « *pharaoh.ini* ». Ces fichiers seront utilisés dans le mode développement.

Pharaoh cherche à maximiser une somme pondérée. Il y a huit poids à fixer dans le fichier de configuration du décodeur Pharaoh.ini présenté dans la Figure 137 : le poids de chacun des cinq coûts de la table de traduction, le poids du modèle de langage, le poids de la pénalité sur le réordonnement des syntagmes/fragments (phrase, chuntes), et le poids de la pénalité sur le nombre de mots dans la phrase cible.

⁴² <http://www.speech.sri.com/projects/srilm/>

```
#####  
### PHARAOH CONFIG FILE ###  
#####  
  
# phrase table f, n, p(n|f)  
[ttable-file]  
/home/hajlaoui/Desktop/SMTdownloads/training-release-1.3-V3/model/phrase-  
table  
  
# language model  
[lmodel-file]  
/home/hajlaoui/Desktop/SMTdownloads/training-release-1.3-V3/LM/ar5g.BO  
  
# limit on how many phrase translations e for each phrase f are loaded  
# 0 = all elements loaded  
[ttable-limit]  
20  
  
# distortion (reordering) weight  
[weight-d]  
0.5  
  
# language model weight  
[weight-l]  
0.5  
  
# translation model weight (phrase translation, lexical weighting)  
[weight-t]  
0.2  
0.2  
0.2  
0.2  
0.2  
  
# word penalty  
[weight-w]  
-1
```

Figure 137 : fichier de configuration pour Pharaoh « pharaohh.ini »

Le choix de ces huit poids est critique pour que le décodeur produise des traductions de qualité. Nous les avons donc ajustés à l'aide d'un critère de minimisation du taux d'erreur.

La Figure 138 montre quelques exemples de correspondances basées sur des statistiques et extraites du fichier *phrase-table*.


```

1995 toutes options sauf la boite ||| 19951 0.0457954 1 ||| كاملة الاضافات ما عدا الجير
2.718 0.000408244
1995 toutes options sauf la ||| 19951 0.0672316 1 ||| كاملة الاضافات ما عدا
2.718 0.00100758
1995 toutes options ||| /2.718 0.00117939 0.5 0.234694 1 ||| 1995 فل أبشن
1995 toutes options ||| 19952.718 0.00896158 0.5 0.241379 1 ||| كاملة الاضافات
1995 toutes ||| 19952.718 0.0379267 1 0.25 1 ||| كاملة
...
A vendre voiture taxi ||| 2.718 0.0535066 1 0.127334 1 ||| سيارة للبيع تكسي مكتب
A vendre voiture type OPEL Vectra 2200 ||| 1 ||| 2200 بيع سيارة نوع اوپل فكترا
2.718 0.00717725 1 0.00863003
A vendre voiture type OPEL Vectra ||| 1 0.0107875 1 ||| بيع سيارة نوع اوپل فكترا
2.718 0.00897156
A vendre voiture type OPEL ||| 2.718 0.0108603 1 0.0124908 1 ||| بيع سيارة نوع اوپل
A vendre voiture type ||| 2.718 0.017105 1 0.0128031 1 ||| بيع سيارة نوع
A vendre voiture ||| 2.718 0.0371907 0.00689655 0.0257904 0.0555556 ||| بيع
A vendre voiture ||| 2.718 0.0228067 0.0275862 0.0896217 0.363636 ||| بيع سيارة
A vendre voiture ||| 2.718 0.267533 0.0344828 0.193325 1 ||| سيارة للبيع
A vendre voiture ||| 2.718 0.51889 0.0758621 0.199979 0.846154 ||| سيارة للبيع
A vendre voiture ||| 2.718 0.00258173 0.00689655 0.037037 1 ||| سيارة-لبيع
A vendre voiture ||| 2.718 0.00577165 0.0137931 0.106513 1 ||| لبيع سيارة
A vendre voiture ||| 2.718 0.00154894 0.00689655 0.171359 1 ||| لبيع .. سيارة
A vendre voiture ||| 2.718 0.846151 0.0206897 0.057548 0.0108696 ||| لبيع
A vendre voiture ||| 2.718 0.267533 0.337931 0.193325 0.942308 ||| لبيع سيارة
A vendre voiture ||| 2.718 0.51889 0.462069 0.199979 0.943662 ||| لبيع سيارة
A vendre voiture ||| 2.718 0.0019239 0.00689655 0.239655 1 ||| لبيع سيارة
A vendre ||| +962796617671,2.718 0.0031373 0.00380228 0.25 1 ||| لبيع

```

Figure 138 : extrait du fichier phrase-table

La table “phrase-table” est le résultat de l’alignement par GIZA++. Dans la Figure 138, on voit bien que plusieurs formes d’écriture différentes en arabe sont correctement alignées au niveau sémantique avec une même entrée française.

12.2.3 Pharaoh en mode développement

L’exécution du décodeur en mode développement se fait par simple exécution du script perl *pharaoh.perl* qui prend essentiellement comme paramètres le fichier de configuration “*pharaoh.ini*” qui contient une indication du chemin du modèle de langage “*lmodel-file*”, le fichier “*phrase-table*” et le corpus de données à traduire, ici “*transtac_dev.fr*”. Il produit en sortie un fichier qui contient les phrases traduites (*translated*).

```

./run-filtered-pharaoh.perl ./evaluation/filtered ./pharaoh
./model/pharaoh.ini ./corpus/transtac_dev.fr > ./out/translated

```

Figure 139 : exécution du script *pharaoh.perl*

Nous avons essayé de nouvelles données qui n’ont pas été utilisées dans le mode apprentissage. Le résultat obtenu dépend de la taille du corpus passé en entraînement (voir section suivante).

La Figure 140 donne les résultats obtenus pour une taille de corpus d’entraînement limitée à 400 SMS.

Texte de référence (arabe original)	Texte source (français)	Texte cible obtenu par TA statistique (Fr → Ar)
مطلوب نيسان صني عادي موديل 93 الى 97	recherche NISSAN Sunny manuelle modèle 93 à 97	مطلوب نيسان صني عادي موديل 93 الى 97
مطلوب رينو كليو غير اوتوماتيك موديل 95 الى 2000	recherche RENAULT Clio manuelle modèle 95 à 2000	مطلوب رينو عادي موديل 95 الى 2000
مطلوب سيارة هونداي	recherche voiture HYUNDAI	مطلوب سيارة HYUNDAI
مطلوب سيارة ميتسوبيشي	recherche voiture MITSUBISHI	مطلوب سيارة ميتسوبيشي
أبحث عن سيارة بي إم دبليو موديل 92	je cherche une voiture BMW modèle 92	اريد a cherche سيارة بي إم دبليو موديل 92
مطلوب دايو لمنز	recherche DAEWOO	مطلوب دايو
مطلوب نيسان صني 93 الى 95 عادي	recherche NISSAN Sunny modèle 93 à 95 manuelle	مطلوب نيسان صني موديل 93 الى 95 عادي
مطلوب هوندا سيفك موديل 94 كل اوبشن غير اوتوماتيك	recherche HONDA Civic modèle 94 toutes options ma	مطلوب هوندا سيفك موديل 94 كل اوبشن عادي
مطلوب كيا سيفيا	recherche KIA Sephia	مطلوب كيا سيفيا
مطلوب دايو لانوس 95 الى 97 عادي	recherche DAEWOO Lanos 95 à 97 manuelle	مطلوب دايو لانوس 95 الى 97 عادي

Figure 140 : exemple de résultat obtenu par traduction statistique (Pharaoh)

Tenant compte de la complexité et de la richesse de la langue arabe, le résultat obtenu pour cette taille est encourageant. En effet, très peu de mots sont inconnus, comme HYUNDAI, et les SMS en arabe semblent assez bons (cela sera précisé plus loin).

12.3 Résultats et évaluations

Dans ce qui suit, nous évaluons d'abord, la traduction par les deux méthodes NIST et BLEU. Dans la deuxième partie, nous évaluons l'extraction d'information par une mesure informationnelle (voir section 6.2.2.1). Nous effectuons enfin une évaluation générale (sur la totalité du corpus) par la mesure simple (voir section 6.2.2.2) et la mesure fine.

12.3.1 Évaluation de la traduction

L'évaluation par NIST et BLEU suppose d'avoir au moins une version de référence, une version candidate et la version source. Dans notre cas, la version de référence est le SMS arabe original, la version candidate est le résultat produit par le système Pharaoh, et la version source est celle du corpus d'évaluation en français.

Le Tableau 54 présente les différents mesures de NIST et BLEU obtenues pour le corpus d'évaluation en fonction de la taille du corpus d'entraînement.

On observe que ces scores n'augmentent presque plus à partir de 500 SMS. Ils sont aux alentours de 25% pour BLEU, loin derrière les meilleurs systèmes statistiques arabe-anglais (qui atteignent un peu plus de 30 %), mais assez élevés tout de même pour espérer que les résultats soient utilisables pour en extraire l'essentiel du contenu.

Taille du corpus d'entraînement	NIST	BLEU
100	3,52	0,14
200	4,23	0,20
300	4,42	0,21
400	4,64	0,23
500	4,95	0,25
600	5,00	0,25
700	5,04	0,25
800	5,05	0,26
900	5,01	0,25
1000	5,07	0,26
1100	5,05	0,26

Tableau 54 : mesure de NIST et BLEU en fonction de la taille du corpus d'entraînement

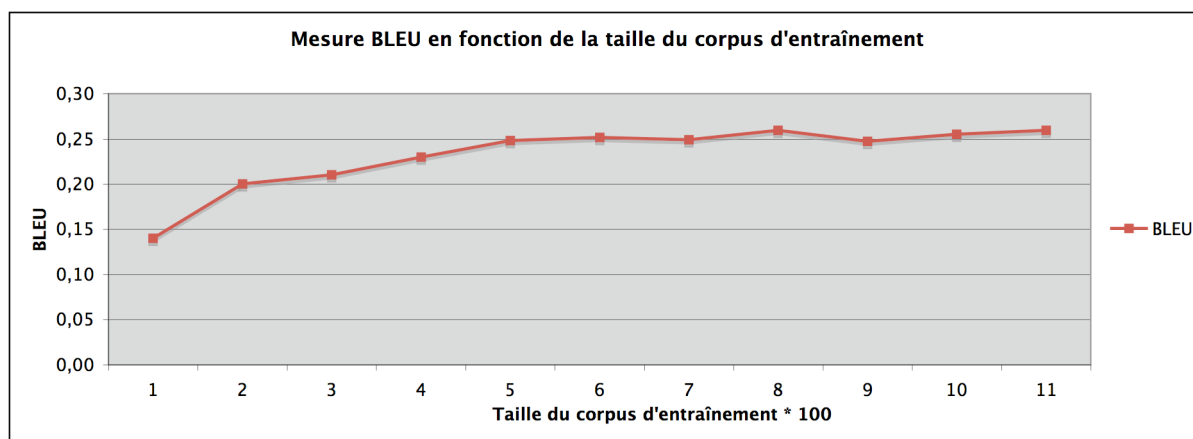


Figure 141 : BLEU en fonction de la taille du corpus d'entraînement

La courbe de la Figure 141 montre une croissance très faible de la mesure BLEU à partir de la valeur 0,26 qui correspond à une taille du corpus d'entraînement égale à 800 SMS. A partir de cette même taille de corpus, la courbe de la Figure 142, représentant la mesure NIST, croît (et décroît parfois) aussi très faiblement à partir de la valeur 5,05.

Cela veut dire qu'une augmentation de la taille du corpus d'entraînement ne modifie presque pas la valeur des mesures BLEU et NIST, entre 500 et 1100 SMS. Rien ne peut être conclu sur l'évolution de ces mesures si la taille croissait beaucoup (par exemple 50-110 K SMS), mais cela n'est pas important, puisque :

- la qualité obtenue avec 600 à 800 SMS est comme on va le voir, suffisante pour le portage visé ;
- pour obtenir un corpus 100 fois plus grand, il faudrait générer ce corpus (génération aléatoire de CRL-CATS et déconversion), et surtout le post-éditer dans les deux langues.

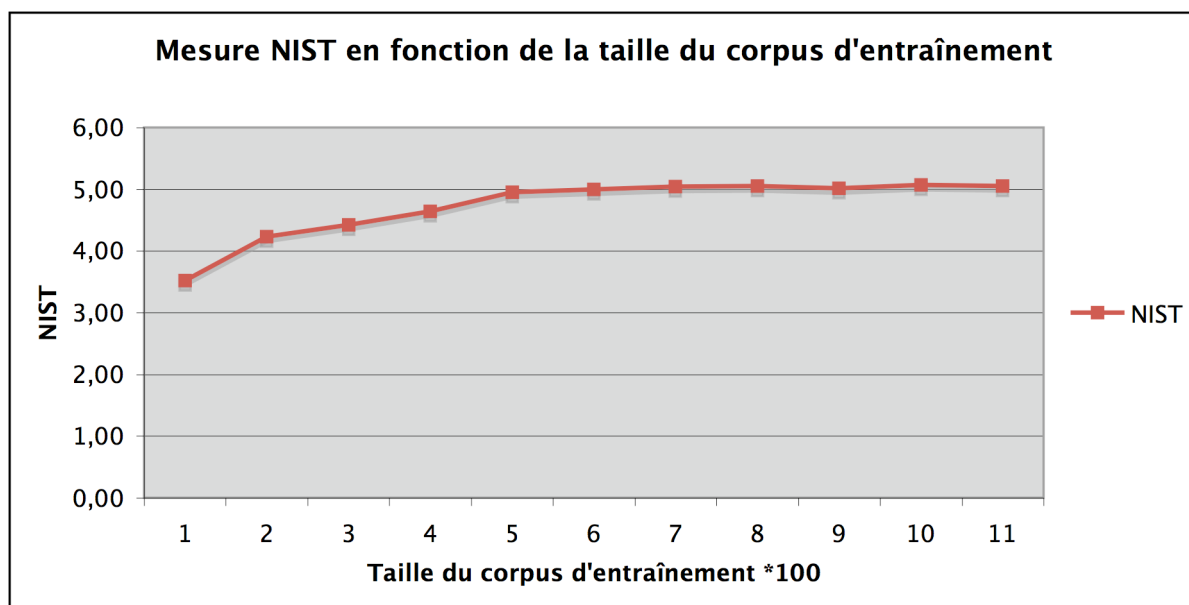


Figure 142 : NIST en fonction de la taille du corpus d'entraînement

Nous ne garantissons donc pas que les mesures ne peuvent pas s'améliorer après une grande augmentation de la taille du corpus d'entraînement ou l'ajout d'autres ressources et outils comme un analyseur morphologique pour le français. Mais, rappelons-le, notre objectif est de proposer des solutions de multilinguisation simples et applicables sur le terrain avec le coût le plus faible possible. Or, un examen rapide des résultats nous indique qu'il semble n'y avoir que peu ou pas de perte d'information. Nous allons vérifier ce point en appliquant l'extracteur de contenu à ces résultats, et en comparant les CRL-CATS obtenus avec ceux obtenus à partir des SMS originaux.

Notons qu'on arrive approximativement à la même mesure BLEU que d'autres expériences sur le couple de langues (anglais vers arabe). En effet, d'autres chercheurs de notre équipe sont arrivés à la valeur 0,25 pour BLEU, mais cela leur a demandé une taille de corpus d'environ 42000 phrases car il s'agissait d'une variante de l'arabe plus complexe et plus générale que la nôtre (dialecte irakien, dialogues informationnels).

12.3.2 Évaluation de l'extraction d'information

De la même façon que pour les deux méthodes de portage utilisées précédemment, nous appliquons la mesure informationnelle en calculant le rappel R, la précision P et la F-mesure F pour les mêmes propriétés les plus importantes (*action de vente ou d'achat, marque, modèle, année, prix*).

12.3.2.1 Mesure informationnelle

Rappelons que cette mesure informationnelle (voir section 6.2.2.1) est basée sur le corpus d'évaluation « CorpusEvalFr200SMS ».

❖ Évaluation par rapport à la version en arabe

Voici d'abord quelques exemples de résultats.

FR	FR→AR→CRL	AR→CRL (version originale)
Recherche voiture OPEL VECTRA	؛مطلوب سيارة اوپل فكترا [S] wan(saloon:06, wanted:00) mak(saloon:06, OPEL(country>germany, country>europe):0C) mod(saloon:06, Vectra(country>germany, country>europe,make>OPEL):0H) [/S]	؛مطلوب سيارة اوپل الفكترا. [S] wan(saloon:06, wanted:00) mak(saloon:06, OPEL(country>germany, country>europe):0C) mod(saloon:06, Vectra(country>germany, country>europe,make>OPEL):0H) [/S]
Recherche achat voiture modèle 99-2000 prix raisonnable	؛مطلوب شراء سيارة موديل 99-2000 وسعر مقبول [S] wan(saloon:0B, wanted:00) yea(saloon:0B, 99:0N) yea(saloon:0B, 2000:0Q) [/S]	؛مطلوب شراء سيارة موديل 99_2000 بسعر مناسب.. [S] wan(saloon:0B, wanted:00) yea(saloon:0B, 99:0N) yea(saloon:0B, 2000:0Q) [/S]
Recherche voiture modèle à partir de 85 avec tranches mensuelles de 100 dinars sans acompte	؛مطلوب سيارة موديل 90050L partir من 85 مع 100 دينار شهري بدون دفعة [S] wan(saloon:06, wanted:00) yea(saloon:06, 90050:0J) yea(saloon:06, 85:0Z) pri(saloon:06, 100:15) [/S]	؛مطلوب سيارة موديل 85 وقرى بالتقسط بدون دفعة اولي كل شهر 100 دينار . . [S] wan(saloon:06, wanted:00) yea(saloon:06, 85:0I.@more) pri(saloon:06, 100:1L) [/S]
recherche voiture	؛مطلوب سيارة [S] wan(saloon:06, wanted:00) [/S]	؛مطلوب سيارة. [S] wan(saloon:06, wanted:00) [/S]
recherche bus KIA transport commun paiement fractionné	؛مطلوب باص كيا ترانسپورت COMMUN paiement fractionné [S] wan(bus:06, wanted:00) mak(bus:06, KIA(country>korea):0A) fea(bus:06, power steering:0J) [/S]	؛مطلوب باص كيا نقل مشترك اقساط. [S] wan(bus:06, wanted:00) mak(bus:06, KIA(country>korea):0A) [/S]
recherche bus transport commun KIA ou HONDA ou SINGE YOUNG avec acompte trois mille et le reste par tranches	؛مطلوب باص ترانسپورت COMMUN كيا او هونداي او سانج يونج مع بدفعة trois mille و الباقى اقساط [S] wan(bus:06, wanted:00) fea(bus:06, power steering:0F) mak(bus:06, KIA(country>korea):0S) mak(bus:06, HYUNDAI(country>korea):0Z) [/S]	؛مطلوب باص نقل مشترك كيا او هونداي او سانج يونج بدفعة ثلاث الالف والباقي اقساط. [S] wan(bus:06, wanted:00) mak(bus:06, KIA(country>korea):0K) mak(bus:06, HYUNDAI(country>korea):0R) mak(bus:06, SangYong(country>korea):10) [/S]
recherche JEEP Cherokee 2000 1900 dinars	؛مطلوب جبب شيروكي 2000 1900 دينار [S] wan(saloon:00, wanted:00) mak(saloon:00, JEEP(country>Ameria):06) mod(saloon:00, Cherokee(country>America,make>JEEP	؛جبب شيروكي 2000 مطلوب . 19000 دينار [S] mak(saloon:00, JEEP(country>america):00) mod(saloon:00, Cherokee(country>America,make>JEEP

	Cherokee(country>America,make<JEEP):0A) yea(saloon:00, 2000:0H) pri(saloon:00, 1900:0M) [/S]):04) yea(saloon:00, 2000:0A) wan(saloon:00, wanted:0F) pri(saloon:00, 19000:0M) [/S]
MERCEDES classique E200 toutes options	ميرسيدس كلاسيك بظه فلل اوتشن [S] mak(saloon:00,MERCEDES(country>germany,country>europe):00) [/S]	ميرسيدس 200E كلاسيك جميع الاضافات 22000 دينار. [S] mak(saloon:00, MERCEDES(country>germany,country>europe):00) mod(saloon:00, 200:07) fea(saloon:00, full options:0J) pri(saloon:00, 22000:0X) [/S]
recherche NISSAN Sunny modèle 93 à 95 manuelle	مطرب نيسان صني عادي موديل 93 ل 90050 [S] wan(saloon:00, wanted:00) mak(saloon:00, NISSAN(country>japan):06) mod(saloon:00, Sunny(country>japan,make>NISSAN):0C) mak(saloon:00, AUDI(country>germany,country>europe):0H) yea(saloon:00, 93:0R) yea(saloon:00, 90050:0V) yea(saloon:00, 97:11) [/S]	مطرب نيسان صني 93 الي 95 غير اوتوماتيك. [S] wan(saloon:00, wanted:00) mak(saloon:00, NISSAN(country>japan):06) mod(saloon:00, Sunny(country>japan,make>NISSAN):0C) yea(saloon:00, 93:0G) yea(saloon:00, 95:0N) fea(saloon:00, automatic gear:0W) [/S]

Figure 143 : quelques exemples de CRL-CATS obtenus par traduction (FR-AR) et par la version originale

Les résultats d'extraction de contenu de la version française (voir Annexe V) obtenus par TA statistique (STA) et ceux de la version originale (arabe) pour les propriétés les plus importantes sont regroupés dans le Tableau 55. Les pourcentages de portage par rapport à la version originale varient entre 85% et 98%, avec une moyenne de 93 %.

Rappelons que l'avantage de cette méthode est qu'elle ne nécessite aucun accès aux ressources de l'application à porter. La Figure 144 permet de mieux visualiser la comparaison entre les valeurs de F-mesure trouvées pour chacune des versions du système.

Propriété	STA FR			EnCoAR			% portage
	Précision	Rappel	F-mesure (STA FR)	Précision	Rappel	F-mesure (EnCoAR)	
Achat/vente	1,000	0,800	0,889	0,956	0,970	0,963	92
Année	0,753	0,740	0,746	0,817	0,960	0,883	85
Prix	0,883	0,726	0,797	0,800	0,822	0,811	98
Marque	0,964	0,901	0,931	0,978	0,963	0,970	96
Modèle	0,957	0,718	0,820	0,901	0,837	0,868	95
Moyenne	0,912	0,777	0,837	0,890	0,910	0,899	93

Tableau 55 : comparaison entre les résultats d'extraction de contenu pour 200 SMS obtenus par traduction statistique par rapport à la version originale

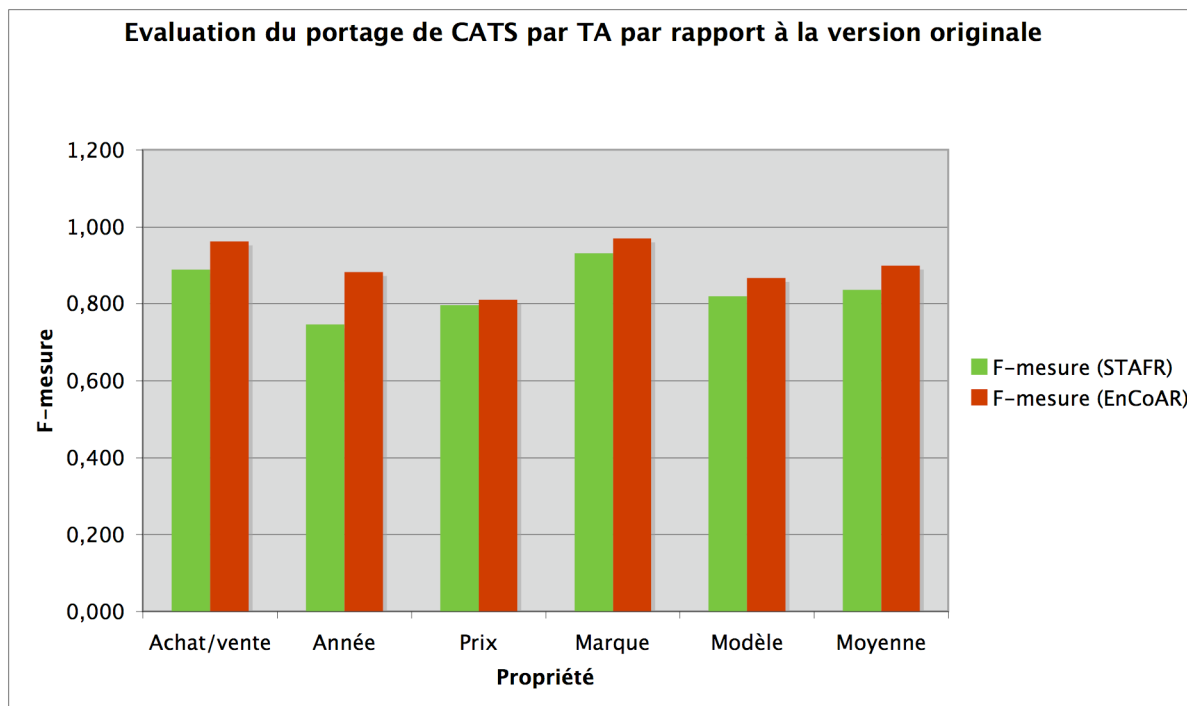


Figure 144 : comparaison entre F-mesure (par rapport à la version originale)

❖ **Évaluation par rapport à une version de référence**

De la même manière, nous avons utilisé une deuxième façon d'évaluer l'extraction de contenu pour les SMS en français. Il s'agit de comparer les résultats trouvés par rapport à une version de référence, obtenue par correction manuelle de la version originale. Voici quelques exemples concrets.

FR	FR→AR→CRL	AR→CRL (version de référence)
<p>A vendre MERCEDES m 1989 contrôle complet toutes options sauf l'ouverture 0795517487</p>	<p>؛ للبيع مرسيدس م 1989 فحص كامل 0795517487 فل اوشن ما عدا</p> <p>[S] sal(saloon:00, sale:00) mak(saloon:00, MERCEDES(country>germany,country>europe):06) yea(saloon:00, 1989:0F) fea(saloon:00, total check:0K) yea(saloon:00, 0795517487:19) [/S]</p>	<p>؛ للبيع /مرسيدس م 1989 فحص كامل كافة الاضافات عدا الفتحه .0795517487</p> <p>[S] sal(saloon:00, sale:00) mak(saloon:00, MERCEDES(country>germany,country>europe):06) mod(saloon:00, 200(country>germany, country>europe,make>MERCEDES):0D) yea(saloon:00, 1989:0J) fea(saloon:00, total check:00) fea(saloon, full options) [/S]</p>
<p>A vendre LAND ROVER DISCOVER mod 2001 toutes options 8 cylindres état original tél 0796635036</p>	<p>؛ للبيع لاندروفر ديسكفري مديل 2001 كاملة الاضافات 0796635036 8 سقندر original للفون</p> <p>[S] sal(saloon:00, sale:00) mod(saloon:00, Land Rover(country>England, country>europe,make>Rover):06) mod(saloon:00, Discovery(country>England, country>europe,make>Rover):0F) yea(saloon:00, 2001:0S) fea(saloon:00, full options:0X) yea(saloon:00, 0796635036:1Z) [/S]</p>	<p>؛ للبيع لاندروفر ديسكفري 2001 كاملة الاضافات 8 سقندر بحال الوكالة للفون .0796635036</p> <p>[S] sal(saloon:00, sale:00) mod(saloon:00, Land Rover(country>England, country>europe,make>Rover):06) mod(saloon:00, Discovery(country>England, country>europe,make>Rover):0F) yea(saloon:00, 2001:0N) fea(saloon:00, full options:0T) [/S]</p>
<p>A vendre KIA Sephia mod 96 toutes options avec boite automatique couleur verte contrôle complet 5800 dinars</p>	<p>؛ للبيع كيا سلفيا مديل 96 فل مع الجير اوتوماتيك لون اخضر فحص كامل 5800 دينار</p> <p>[S] sal(saloon:00, sale:00) mak(saloon:00, KIA(country>korea):06) mod(saloon:00, Sephia(country>korea,make>KIA):0A) yea(saloon:00, 96:0L) col(saloon:00, green:1C) fea(saloon:00, total check:1I) pri(saloon:00, 5800:1R) [/S]</p>	<p>؛ للبيع كيا سلفيا 96 فل اوشن مع جير اوتوماتيك لون اخضر زيني فحص كامل السعر 5800 دينار.</p> <p>[S] sal(saloon:00, sale:00) mak(saloon:00, KIA(country>korea):06) mod(saloon:00, Sephia(country>korea,make>KIA):0A) fea(saloon:00, automatic gear:11) col(saloon:00, green:1D) col(saloon:00, darkgreen:1I) fea(saloon:00, total check:1N) pri(saloon:00, 5800:22) [/S]</p>
<p>A vendre FORD GT neuve mod 2005 moteur 4600 couleur rouge toutes options</p>	<p>؛ للبيع فورد GT جديد كوشوك جديد مديل 2005 ، 4600 لون احمر فل اوشن</p> <p>[S] sal(saloon:00, sale:00) mak(saloon:00, FORD(country>Ameria):06) fea(saloon:00, new:0E) fea(saloon:00, new:0P) yea(saloon:00, 2005:0Z) col(saloon:00, red:1F) [/S]</p>	<p>؛ للبيع فورد موستنغ مديل 2005 جديدة موتور 4600 لون احمر فل اوشن GT</p> <p>[S] sal(saloon:00, sale:00) mak(saloon:00, FORD(country>america):06) yea(saloon:00, 2005:0N) fea(saloon:00, new:0S) mot(saloon:00, 4600:19) col(saloon:00, red:1I) [/S]</p>

A vendre BMW 520 mod 1992 grise toutes options en bon état, 14000 dinars	; للبيع بي ام دبليو 520 مديل 1992 سطر فل ايشن بحالة 14 , état دينار [S] sal(saloon:00, sale:00) mak(saloon:00, B.M.W(country>germany, country>europe):06) mod(saloon:00, 520:0I) yea(saloon:00, 1992:0R) col(saloon:00, silver:0W) pri(saloon:00, 14:1L) [/S]	; للبيع سياره 520 مديل 1996. سطر فل اكرزىكتف بحاله ممتازه السعر BMW14000 . [S] sal(saloon:06, sale:00) mak(saloon:06, B.M.W (country>germany, country>europe):10) mod(saloon:06, 520:0B) yea(saloon:06, 1996:0L) col(saloon:06, silver:0P) fea(saloon:06, good condition:16) pri(saloon:06, 14000:1R) [/S]
A vendre HYUNDAI mod 2001 toutes options sauf la boite, 8900 dinars	; للبيعهونداي مديل 2001 كامله الاضافات ما عدا 8900 , boite دينار [S] sal(saloon:00, sale:00) mak(saloon:00, HYUNDAI(country>korea):05) yea(saloon:00, 2001:0H) fea(saloon:00, full options:0M) pri(saloon:00, 8900:1F) [/S]	; للبيع سياره هونداي أفانتي 2001 كامله الاضافات عدا الجير بسعر 8900 دينار. [S] sal(saloon:06, sale:00) mak(saloon:06, HYUNDAI(country>korea):0C) yea(saloon:06, 2001:0Q) pri(saloon:06, 8900:1P) [/S]

Figure 145 : quelques exemples de CRL-CATS obtenus par traduction statistique (FR-AR) et par la version de référence

Le Tableau 56 résume les résultats de la comparaison.

Propriété	STAFR			EnCoAR-Ref			% du portage
	Précision	Rappel	F-mesure (EnCoAR-Ref)	Précision	Rappel	F-mesure (STAFR)	
Achat/Vente	1,000	0,800	0,889	1,000	0,970	0,985	90
Année	0,753	0,740	0,746	1,000	0,983	0,991	75
Prix	0,883	0,726	0,797	0,973	0,973	0,973	82
Marque	0,964	0,901	0,931	0,984	0,963	0,973	96
Modèle	0,957	0,718	0,820	0,969	0,901	0,934	88
Moyenne	0,912	0,777	0,837	0,985	0,958	0,971	86

Tableau 56 : comparaison entre les résultats d'extraction de contenu obtenus par TA statistique pour 200 SMS par rapport à une version de référence

Les pourcentages de portage par rapport à la version originale varient entre 75% et 96%, avec une moyenne de 86 %.

La Figure 146 permet une meilleure visualisation de la comparaison entre les valeurs de F-mesure trouvées.

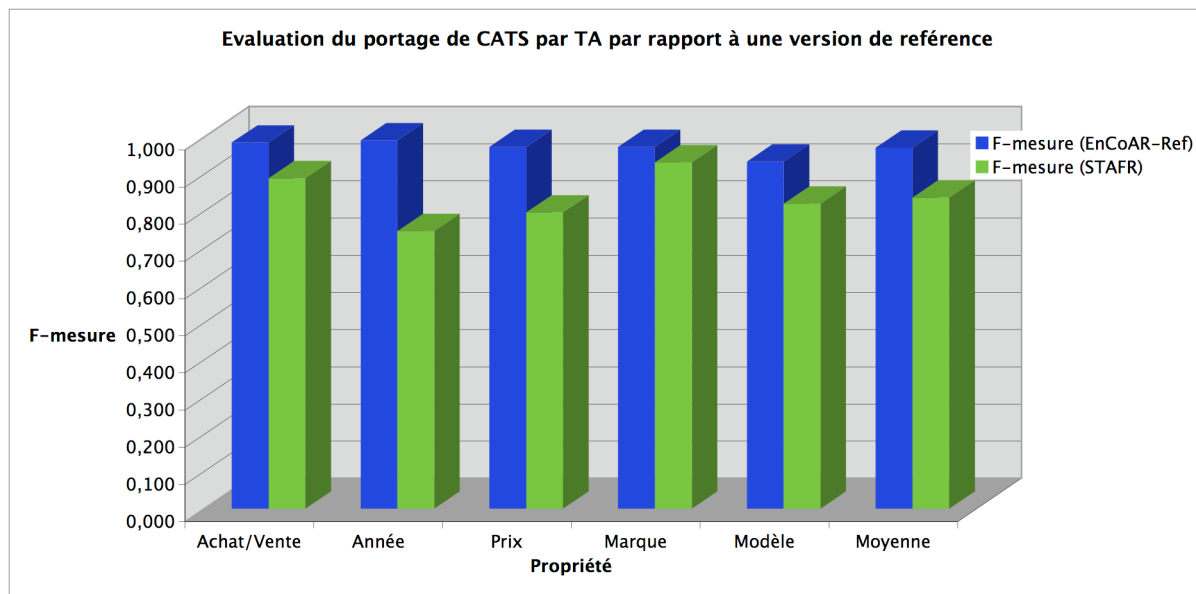


Figure 146 : comparaison entre F-mesures (par rapport à une version de référence)

12.3.3 Évaluation générale

Dans la suite, nous évaluons les résultats d'extraction de contenu sur tout le corpus appelé « CorpusEvalFr1100SMS » en utilisant les deux mesures suivantes : mesure simple et mesure fine.

12.3.3.1 Mesure simple

Les résultats des scores des attributs les plus importants ($m = 6$) sur le total du corpus ($n = 1100$) sont présentés dans le Tableau 57 et dans la Figure 147. On remarque que certains scores de la version candidates sont parfois supérieurs à ceux de la version originale, c'est en fait le principe de fonctionnement de cette mesure présentée dans la section 6.2.2.2. Elle évalue l'extracteur de contenu en terme de quantité d'attributs détectés, sans tenir compte de la qualité de l'attribut détecté.

Propriétés	Version candidate	Version originale
	STAFrAr	EnCoAr
Score (wan) _{corpus}	56,33	49,33
Score (sal) _{corpus}	41,89	39,33
Score (yea) _{corpus}	98,00	98,22
Score (pri) _{corpus}	41,44	38,44
Score (mak) _{corpus}	130,33	124,67
Score (mod) _{corpus}	57,33	55,22
Score (wan, sal, yea, pri, mak, mod) _{corpus}	425,33	405,22

Tableau 57 : scores trouvés pour la méthode par traduction sur le total du corpus de 1100 SMS

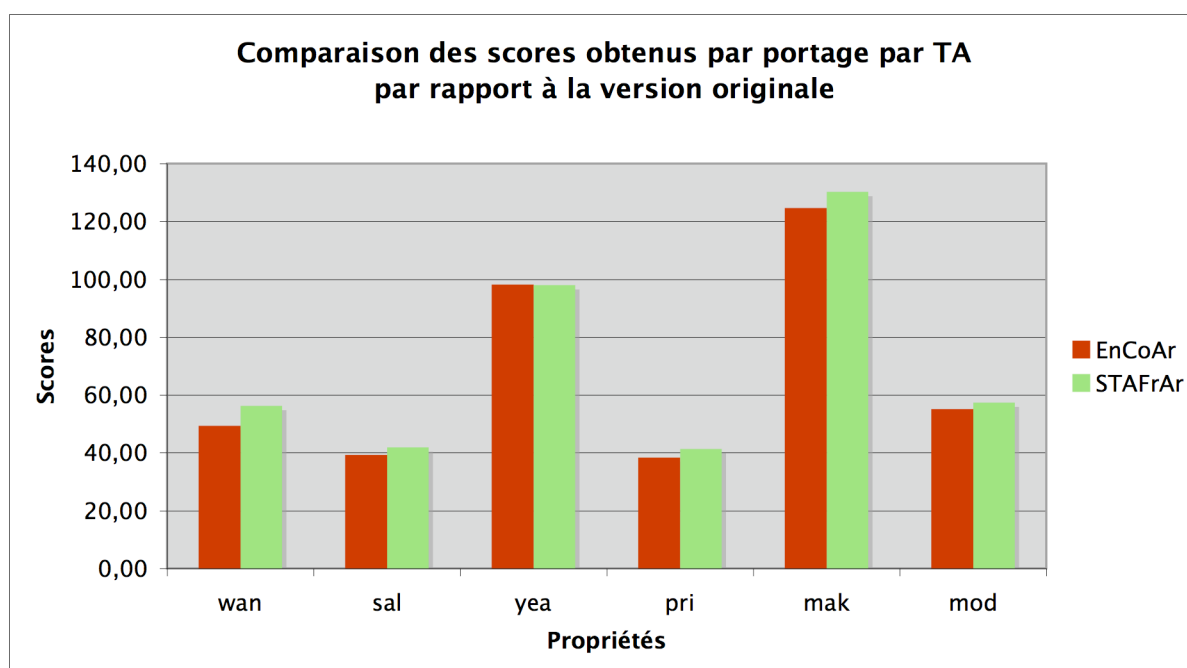


Figure 147 : comparaison des scores des attributs les plus importants obtenus par la méthode par traduction par rapport à la version originale

Les pourcentages de la différence par rapport à la version originale des scores trouvés sont résumés dans le Tableau 58.

Propriétés	% de différence de STAFrAr par rapport à la version originale (EnCoAr)
wan	14,19
sal	6,50
yea	0,23
pri	7,80
mak	4,55
mod	3,82
Moyenne	6,18

Tableau 58 : comparaison des pourcentages de différences des scores obtenus par la méthode de traduction sur le total du corpus

Les résultats obtenus sur le total du corpus confirment bien les résultats obtenus par un jugement humain sur un corpus limité à 200 SMS.

12.3.3.2 Mesure fine

Nous avons appliqué la mesure fine décrite dans la section 6.2.2.3. La distance de l'ensemble des CRL-CATS (1100 CRL-CATS générées par par le corpus « CorpusEvalFr1100SMS ») par rapport à l'ensemble des CRL-CATS de la version originale du système est de 566. Cette mesure donne une F-mesure moyenne de 0,901.

Le Tableau 59 regroupe les valeurs moyennes de rappel, de la précision et de la F-mesure trouvées sur le corpus « CorpusEvalFr1100SMS ». Nous constatons ici que la valeur de la F-mesure moyenne obtenue est la meilleure par rapport aux deux autres obtenues par les deux précédentes méthodes de portage. Pour cette mesure, les résultats d'extraction obtenus par cette méthode sur le total du corpus sont les meilleurs. Cela est dû peut être à l'utilisation de l'extracteur de contenu de l'application originale.

	Rappel	Précision	F-mesure
Moyenne	0,789	0,783	0,901

Tableau 59 : résultats du portage par TA par la mesure fine sur « CorpusEvalFr1100SMS »

12.3.4 Coût de la méthode

Le Tableau 60 montre les détails des différents coûts en termes du temps passé aux différentes tâches dans le cadre du portage par traduction statistique, y compris l'étude préalable sur les systèmes de TA en général et sur le fonctionnement des systèmes de TA statistique en particulier.

Le travail de TA statistique a demandé un total de 1,1 homme/mois d'informaticien, dont une bonne partie a été consacrée au travail préparatoire des données et à l'expérimentation avec les outils libres (GIZA++, SRLIM, Pharaoh).

À ce coût s'ajoute le coût du travail de déconversion qui a coûté presque 1 homme/mois d'informaticien linguiste.

Tâches	Durée (h)	Références	Remarques
Documentation sur les systèmes de TA	20	(Boitet 2007) Site de MT-list : (http://www.mail-archive.com/mt-list%40eamt.org/)	Rapport interne sur l'état de l'art de la traduction de l'écrit + Web
Documentation sur la TA statistique	15	Workshop sur la TA statistique (20 pages imprimées A4) http://www.statmt.org/wmt06/shared-task/baseline.html	Web
Documentation sur le système Pharaoh	15	(Koehn 2004) http://www.isi.edu/licensed-sw/pharaoh/ (Besacier 2007)	Web + articles de P.Koehn + HDR de L.Besacier
Documentation sur l'outil d'alignement GIZA et GIZA++	12	(Och and Ney 2000) http://www.fjoch.com/GIZA++.html	Web + articles d'Och &Ney
Documentation sur les méthodes d'évaluation de TA	10	(Blanchon 2004)	HDR de H.Blanchon
Expérimentation de GIZA++	10	Travail local (plate-forme Mac)	Premiers tests
Expérimentation de Pharaoh	15	Utilisation du serveur isis (plate-forme Linux)	Premiers tests
Préparation des données	30	Travail local (plates-formes Mac et Windows)	Construction du corpus parallèle
Expérimentation (GIZA++, SRLIM, Pharaoh)	35	Utilisation du serveur nemesis (plate-forme Windows)	Alignement + traduction des données
Evaluation NIST et BLEU	5	Utilisation du serveur nemesis (plate-forme Windows)	Évaluation de la traduction
Tests et évaluation	15	Travail local (plate-forme Mac)	Évaluation de l'extraction d'information
Total	182		

Tableau 60 : évaluation de la durée des tâches pour la méthode de portage externe.

Conclusion de la partie D

Nous avons présenté dans cette partie une troisième méthode de portage, par traduction automatique, qui ne demande pas d'accès au code de l'application, ni à la représentation interne.

Nous avons choisi d'expérimenter d'abord la traduction statistique, car les systèmes statistiques et les outils liés à ce travail sont disponibles et se sont améliorés ces dernières années. De plus, c'est la méthode qui demande le moins d'expertise linguistique et qui est donc a priori la plus largement applicable.

Nous avons présenté et expérimenté une méthode possible pour obtenir un corpus parallèle, besoin obligatoire pour la TA statistique. Dans le cas de CATS, cette méthode repose sur l'utilisation de l'outil de déconversion DeCo, conçu initialement pour UNL.

Nous avons décrit le travail de préparation des données et d'alignement qui précède la tâche de traduction statistique basée sur l'utilisation du décodeur Pharaoh. La mise en œuvre du décodeur Pharaoh se déroule en deux étapes : une première étape pour l'entraînement et une deuxième de développement.

Malgré la très petite taille du corpus d'entraînement utilisé (1100 SMS), nous avons obtenu de bons résultats d'extraction d'informations, ce qui donne un pourcentage de portage égal à 93 % par rapport à l'application originale et de 86 % par rapport à une version de référence. C'est une autre confirmation de l'hypothèse de Kittredge sur les sous-langages.

Conclusion et perspectives

La multilinguïisation des applications informatiques grand public, au sens de « localisation », est une activité en plein essor, qui pose encore de nombreux problèmes.

Nous avons étudié un sous-problème intéressant et réel, celui du portage linguistique d'applications supposant le traitement automatique du contenu d'énoncés spontanés en langue naturelle, comme le traitement de petites annonces (CATS), ou de requêtes en recherche de morceaux de musique (IMRS), ou de description d'incidents (pour du SAV) ou d'accidents, etc. Nous étions dans un contexte favorable, puisque nous disposions au départ d'un accès complet à l'application CATS, créée par D.Daoud (qui a passé sa thèse au GETA, CLIPS sur ce système en 2006).

Pour ce genre d'applications, la multilinguïisation, même réduite au portage, reste assez délicate, et rencontre plusieurs difficultés, vu la spécificité du type de textes à traiter. Il s'agit en effet de textes entrés à partir d'une interface libre et non contrôlée.

D'autre part, il y a a priori plusieurs méthodes possibles pour « porter » une application, et il faut tenir compte de la situation traductionnelle, qui est définie par plusieurs facteurs, essentiellement le niveau d'accès à l'application à porter, les ressources disponibles, et les compétences langagières et linguistiques des intervenants.

Le point le plus important est qu'on a toujours affaire à un « sous-langage naturel » (par opposition à la langue générale, à la langue standard, ou à un sous-langage « contrôlé »). Nous avons étudié les différentes approches liées à la notion de « sous-langage » et avons dégagé de cette étude six critères pour définir et identifier le type de langage traité, et une taxonomie basée essentiellement sur une mesure de la complexité lexicale et de la complexité syntaxique.

Nous avons illustré notre méthode par le portage de l'application d'extraction de contenu CATS, qui traite des énoncés très spontanés en arabe. L'étude de la situation traductionnelle nous a permis d'apporter trois solutions de portage pour cette application, pour un couple de langues assez distantes (arabe-français).

La première méthode, dite portage interne, nécessite un accès complet à l'application et consiste à adapter au français l'extracteur de contenu original. Cette méthode a donné de bons résultats et n'a demandé qu'une légère modification au niveau des règles, à cause de la proximité des sous-langages : en effet, deux sous-langages de deux langues distantes comme l'arabe et le français sont grammaticalement très proches.

La deuxième méthode, dite portage externe, nécessite seulement l'accès à la représentation interne de l'application. Elle consiste à adapter un extracteur de contenu en restant dans la même langue mais en changeant de domaine. Elle marche bien et donne de bons résultats, un peu moins bons cependant que ceux du portage interne. Cela est dû à un mauvais traitement des données contenant des chiffres comme le prix et l'année et qui peut être amélioré.

La troisième solution ne demande d'accès ni au code ni à la représentation interne de l'application originale. Elle consiste à traduire automatiquement les énoncés du nouveau sous-langage (SL2) vers le sous-langage « natif » (SL1) et à soumettre les résultats à l'EC natif de l'application. Cette méthode marche bien et donne d'aussi bons résultats que le portage interne. Ce résultat est d'autant plus intéressant que le système de TA statistique utilisé peut être développé (dans ce cas d'un sous-langage très restrictif) avec un corpus très petit par rapport à ce dont on a besoin pour la langue générale.

Le Tableau 61 récapitule les pourcentages de portage de ces trois dernières méthodes, obtenus par évaluation informationnelle sur le corpus limité « CorpusEvalFr200SMS ».

Portage par	Par rapport à la version originale			Par rapport à la version de référence		
	Minimum	Moyenne	Maximum	Minimum	Moyenne	Maximum
Adaptation interne	95%	98%	100%	83%	91%	99%
Adaptation externe	46%	77%	99%	41%	72%	99%
TA	85%	93%	98%	75%	86%	96%

Tableau 61 : récapitulatif des résultats d'évaluation informationnelle sur le corpus « CorpusEvalFr200SMS »

Le Tableau 62 résume les résultats de l'évaluation générale sur tout le corpus « CorpusEvalFr1100SMS » obtenus par calcul de distance entre les résultats d'extraction de contenu obtenus par chaque méthode de portage par rapport à ceux obtenus dans la version originale.

	Portage interne	Portage externe	Portage par SMT (TA stat)
F-mesures	0,718	0,540	0,901

Tableau 62 : récapitulatif des résultats d'évaluation générale sur « CorpusEvalFr1100SMS »

Nous sommes satisfaits des trois solutions apportées du point de vue des coûts estimés (voir Tableau 63). Les trois solutions sont applicables sur le terrain, mais le choix de celle à appliquer dépend de la situation traductionnelle.

	Portage interne	Portage externe	Portage par SMT (TA stat)
Homme/mois d'informaticien		3	1,1
Homme/mois d'informaticien linguiste	1,9		1

Tableau 63 : récapitulatif des coûts des trois méthodes de portage

Nous avons aussi expérimenté notre approche sur le portage linguistique d'une application de recherche de musique qui traite des énoncés spontanés en japonais. Bien que nous n'ayons pas d'accès au système IMRS, nous connaissons la représentation interne du système, ce qui nous a permis d'adapter le même extracteur de contenu du français que celui adapté pour le portage externe de CATS.

Nous concluons notre méthode de travail de portage linguistique applicable pour des applications qui traitent des énoncés spontanés en langue naturelle par le guide suivant :

1. Identification de la tâche.
2. Identification des langues cibles.
3. Identification du ou des sous-langages et évaluation suivant les axes définis.
4. Étude de la situation traductionnelle :
 - Types et niveaux d'accès aux ressources.
 - Compétences langagières et linguistiques.
5. Évaluation de la taille et du coût de construction du ou des dictionnaires.
6. Évaluation de la taille et du coût de construction du ou des corpus parallèles.
7. Identification des méthodes de portage possibles :
 - Coût
 - Délais
8. Décision (réaliser ou non le portage) :
 - Oui
 - Quelles langues
 - Quelle méthode
 - Quel prix/coût
 - Réalisation du portage
 - Évaluation de performance
 - Non

Perspectives

À court terme, nous désirons :

- Compléter l'évaluation de performance, actuellement basée sur la comparaison de représentations de contenu, par une évaluation « de bout en bout » grâce à une simulation et à des « cobayes ». Ce travail a déjà commencé dans le cadre d'un TER (Zair 2008). De même, pour le système IMRS.
- Essayer la méthode par TA « experte », par exemple, en passant par une représentation de type CRL-CATS et en utilisant EnCo et DeCo (par exemple pour les petites annonces auto de e-Bay).
- Évaluer le guide de portage en portant CATS et/ou IMRS vers une ou deux autres langues. Par exemple, il serait en principe facile de porter CATS vers l'anglais, mais un peu plus difficile

que vers le français, car grammaticalement l'anglais est plus loin de l'arabe que le français. Nous essayerons de faire quelque chose d'analogue pour le JRC « Joint Research Centre » où un postdoctorat semble possible.

À long terme, nous souhaitons :

- Essayer d'appliquer la méthode TBL basée sur l'apprentissage pour construire des extracteurs de contenu pour CATS et pour IMRS.
- Participer à des travaux sur la construction d'un extracteur de contenu « générique » dans une classe restreinte de situations (petits sous-langages, petits domaines ou tâches, tolérance aux erreurs), qui serait paramétrable et pourrait prendre comme entrée un dictionnaire, une grammaire, un texte, et un type de représentation de contenu, et qui donnerait en sortie le même texte représenté sous une forme donnée comme paramètre.

Bibliographie

- Atserias, J., S. Climent, et al. (1997). *Combining multiple methods for the automatic construction of multilingual WordNets*. Proc. RANLP'97. Tzigov Chark, Bulgaria. pp. 143-149.
- Bélanger, L. (2003) Le traitement automatisé des courriels pour les services aux investisseurs : une approche par la question-réponse. Rap. Université de Montréal. Montreal, Canada. 48 p.
- Besacier, L. (2007) *Transcription enrichie de documents dans un monde multilingue et multimodal*. Habilitation à diriger les recherches. Université Joseph Fourier. Grenoble, France. 11 janvier 2007. 64 p.
- Besacier, L., H. Blanchon, et al. (2001). *Speech Translation for French in the NESPOLE! European Project*. Proc. Eurospeech. Aalborg, Denmark. 3-7 september 2001. Vol. 2/4: pp. 1291-1294.
- Bhavsar, V. C., H. Boley, et al. (2004). *A weighted-tree similarity algorithm for multi-agent systems in e-business environments*. in Computational Intelligence. Vol. 4: pp. 584-602.
- Biber, D. (1993). *Using register-diversified corpora for general language studies*. in Computational Intelligence. Vol. 19(2): pp. 219-241.
- Blanchon, H. (2004). Comment définir, mesurer et améliorer la qualité, l'utilisabilité et l'utilité des systèmes de TAO de l'écrit et de l'oral - Une bataille contre le bruit, l'ambiguïté, et le manque de contexte. Habilitation à diriger les recherches. Université Joseph Fourier. Grenoble, France. Décembre 2004. 356 p.
- Blanchon, H. and L. Besacier (2004). Traduction de dialogues : mise en perspective des résultats du projet NESPOLE! et pistes pour le domaine. Proc. TALN. Fès, Maroc. 19-21 avril 2004. Vol. 1/1 : pp. 55-60.
- Blanchon, H. and C. Boitet (2000). *Speech Translation for French within the C-STAR II Consortium and Future Perspectives*. Proc. ICSLP 2000. Beijing, China. Vol. 4/4: pp. 412-417.
- Boitet, C. (1990a) *La TAO à Grenoble en 1990*. Rap. école d'été de Lannion sur le TALN. Lannion, France. juillet 1990. 65 p.
- Boitet, C. (1990b). *Towards Personel MT: general design, dialogue structure, potential role of speech*. Proc. COLING-90. Helsinki. August 20-25, 1990. Vol. 3/3: pp. 30-35.
- Boitet, C. (1994). *Promesses et problèmes de la "TAO pour tous" après LIDIA-1, une première maquette*. in Langages, "le traducteur et l'ordinateur". Vol. 116 : pp. 20-47.
- Boitet, C. (2002). *Advantages of the UNL language and format for web-oriented crosslingual applications*. Proc. Seminar on linguistic meaning representation and their applications over the World Wide Web. Penang, Malaisie. August 20-22, 2002. Vol. 4/4: 4p.
- Boitet, C. (2003). *Automated Translation*. in Revue française de linguistique appliquée. Décembre 2003. Vol. 3(2): pp. 99-121.
- Boitet, C. (2004). Progress report on building the French BTEC and participating in the MT evaluation campaign (CSTAR project). Rap. GETA, CLIPS. August 5, 2004. 10 p.
- Boitet, C. (2007). *Etat de l'art en traduction de l'écrit*. Rap. de Contrat TRANSAT pour France Telecom GETALP, LIG. Grenoble, France. Janvier 2007. 68 p.
- Boufaden, N. (2004). Extraction d'information à partir de transcriptions de conversations téléphoniques spécialisées. Thèse. Université de Montréal. Montreal, Canada. 203 p.
- Brill, E. (1995). Transformation-based error-driven learning and natural language processing: A case study in part of speech tagging. in Computational Linguistics. Vol. 4(21): pp. 543-565.

- Bross, I. D. J., P. A. Shapiro, et al. (1972). *How information is carried in scientific sub-languages*. In *science*. Vol. 176(4041): pp. 1303-1307. juin 1972.
- Brown, P., J. Cocke, et al. (1990). *A statistical approach to machine translation*. in *Computational Linguistics*. Vol. 16(2): pp. 79-85.
- Brown, P., V. J. D. Pietra, et al. (1993). *The Mathematics of Statistical Machine Translation: Paramter Estimation*. in *Computational Linguistics*. Vol. 19(2): pp. 263-311.
- Buitelaar, P., K. Netter, et al. (1998). *Integrating Different Strategies for Cross-Language Information Retrieval in the MIETTA Project*. Proc. 14th Twente Workshop on Language Technology (TWLT14). Enschede, Netherlands. December 7-8, 1998. pp. 9-17.
- Califf, M. E. (1998). *Relational learning techniques for natural language information extraction*. Thèse. The university University of Texas at Austin. Texas, U.S.A. 132 p.
- Candito, M.-H. (1999). *Structuration d'une grammaire LTAG : application au français et à l'italien*. Thèse. University Paris 7. Paris, France. 200 p.
- Cardie, C. (1997). *Empirical methods in information extraction*. in *AI Journal*. Vol 18(4) : pp. 65-79.
- Carpenter, B. and J. Chu-Carroll (1998). *Natural Language Call Routing: A Robust, Self-organizing Approach*. Proc. ICSLP'98. Sydney, Australia. pp. 2059-2062.
- Chai, J. Y., M. Budzikowska, et al. (2001). *Natural Language Sales Assistant - A Web-based Dialog System for Online Sales*. Proc. Thirteenth Conference on Innovative Applications of Artificial Intelligence Conference. pp. 19-26.
- Chai, J. Y., J. Lin, et al. (2000). *Comparative Evaluation of a Natural Language Dialog Based System and a Menu Driven System for Information Access: a Case Study*. Proc. RIAO 2000. Paris, France. pp. 1590-1600.
- Chandioux, J. (1988). *10 ans de METEO (MD)*. Traduction Assistée par Ordinateur. Actes du séminaire international sur la TAO et dossiers complémentaires. Paris Observatoire Francophone des Industries de la Langue (OFIL). A. Abbou. pp. 169-173.
- Chenon, C. (2005). *Vers une meilleure utilisabilité des mémoires de traduction, fondée sur un alignement sous-phrastique*. Thèse. Université Joseph Fourier. Grenoble, France. 28 octobre 2005. 200 p.
- Ciravegna, F. (2001a). *Adaptive information extraction from text by rule induction and generalisation*. Proc. IJCAI-01. Seattle. 10 p.
- Ciravegna, F. (2001b). *(LP)2, an adaptive algorithm for information extraction from web-related texts*. Proc. Workshop on Adaptive Text Extraction and Mining held in conjunction with IJCAI-01. Seattle. 9 p.
- Coupet, P., B. Buschbeck, et al. (2005). *Le Text Mining multilingue : application au monde de l'intelligence économique*. Proc. SFBA (Société Française de Bibliométrie Appliquée). Ile de France, France. 13-17 juin 2005. 13 p.
- Daoud, D. M. (2005). *Building SMS-based System using Information Extraction Technology*. Proc. ACIDCA-ICMI. Tozeur, Tunisia. November 5-7, 2005. 7 p.
- Daoud, D. M. (2006). *It is necessary and possible to build (multilingual) NL-based restricted e-commerce systems with mixed sublanguage and contend-oriented methods*. Thèse. Université Joseph Fourier. Grenoble, France. September 23, 2006. 296 p.
- Denos, N. (1994) *Typage interactif de fragments de textes en "Styles d'Enoncé" et "Genres de Textes"*. Rap. DEA, GETA, Institut National Polytechnique de Grenoble & Université Joseph Fourier. Grenoble, France. 22 juin 1993. 79 p.

- Denoual, E. (2006) *Méthodes en caractères pour le traitement automatique des langues*. Thèse. Université Joseph Fourier. Grenoble, France. 21 septembre 2006. 184 p.
- Deville, G. (1989). Modelization of task-oriented Utterances in a Man-Machine Dialogue System. Thèse. University of Antwerpen. Belgique. 200 p.
- Fairon, C., J. R. Klein, et al. (2006). Étude d'un corpus informatisé à partir de l'enquête 'Faites don de vos SMS à la science'. Press universitaires de Louvain. Vol. 3(1).
- Freitag, D. (1998). Toward general-purpose learning for information extraction. Proc. COLING-ACL-98. pp. 404-408.
- Grishman, R. and R. Kittredge.(1986). *Analyzing language in restricted domains*. Hillsdale NJ. Lawrence Erlbaum Associates. 248 p.
- Hajlaoui, N. (2006). *Recherche et production de corpus de messages pour la multilinguisation de sites de e-commerce en SMS, initialement en arabe*. Proc. 6th international IBIMA (International Business Information Association) Conference. Bonn, Allemagne. 5-6 juin 2006. 11 p.
- Hajlaoui, N. (2007). *Multilinguisation de services de gestion de contenu*. Proc. IC "Ingénierie des Connaissances", plate-forme AFIA "Association Française pour l'Intelligence Artificielle". Grenoble, France 2 p.
- Hajlaoui, N. (2007). *Localizing Content Management Application for Spontaneous Textual Utterances in Natural Language*. Proc. NLDB-2007 International Conference on Applications of Natural Language to Information System. CNAM, Paris, France. 27-29 juin 2007. 12 p.
- Hajlaoui, N. and C. Boitet (2004). *PolyphraZ: a tool for the quantitative and subjective evaluation of parallel corpora*. Proc. IWSLT 2004: International Workshop on Spoken Language Translation. Kyoto, Japan. September 30 – October 01, 2004.
- Hajlaoui, N. and C. Boitet (2007). *Portage linguistique d'applications de gestion de contenu*. Proc. TOTh 2007 Conférence sur la Terminologie & Ontologie : Théories et Applications. Annecy, France. 1 juin 2007. 13 p.
- Hajlaoui, N. and C. Boitet (2008). *TA statistique à petits corpus pour des petits sous-langages*. Proc. TOTh 2008 Conférence sur la Terminologie & Ontologie : Théories et Applications. Annecy, France. 15 p.
- Hajlaoui, N., D. M. Daoud, et al. (2008). *Methods for porting NL-based restricted e-commerce systems into other languages*. Proc. LREC 2008. Marrakech, Maroc. June 26-31, 2008. 7 p.
- Harris, Z. (1968). *Mathematical structures of language*. in The Mathematical Gazette. Vol. 54(388): pp. 173-174. May, 1970.
- Huberman, B., P. Pirolli, et al. (1998). *Strong Regularities in World Wide Web Surfing*. Science. in Science. Vol. 280. pp. 95-97.
- Humphreys, K., R. Gaizauskas, et al. (1998). *Description of the university of sheffield lasie-ii system as used for MUC-7*. Proc. MUC-7 (The 7th Message Understanding Conference).
- Ikezo, T., Y. Kajikawa, et al. (2001). *Music Database Retrieval System with Sensitivity Words Using Music Sensitivity Space*. Proc. IPS of Japan. Japon. Vol. 12/42: pp. 3201-3212.
- Isabelle, P. (1984). *Machine translation at the TAUM group*. Rap. King, M. (ed.) Machine Translation Today: The State of the Art. Cambridge University, Cambridge, Grande-Bretagne.
- Kamat, V. N. (1996) *Inductive learning with the evolving tree transformation system*. Thèse. University of New Brunswick. Canada. 200 p.

- Kim, J.-T. and D. I. Moldovan (1993). *Acquisition of semantic patterns for information extraction from corpora*. Proc. The Ninth IEEE Conference on Artificial Intelligence for Applications. pp. 171-176.
- Kittredge, R. (1978). *Textual cohesion within sublanguages: implications for automatic analysis and synthesis*. Proc. Coling-78. Bergen, Norvège. August 14-18, 1978. Vol. 1/1.
- Kittredge, R. (1982b). *Variation and Homogeneity of Sublanguages*. in *Sublanguage - Studies of Language in Restricted Semantic Domains*. Walter de Gruyter. Berlin / New York. 20 p.
- Kittredge, R. (1993). *Sublanguage Analysis for Natural Language Processing*. Proc. First Symposium on Natural Language Processing. Thailand, Bangkok pp. 69-83.
- Kittredge, R. and J. Lehrberger.(1982a). *Sublanguage - Studies of language in restricted semantic domain*. Walter de Gruyter. Berlin / New York.
- Koehn, P. (2004). *Pharaoh: a Beam Search Decoder for Phrase-Based SMT*. Proc. 6th AMTA. Washington, U.S.A. pp. 115-124.
- Kraif, O. (2001) *Constitution et exploitation de bi-textes pour l'aide à la traduction*. Thèse. Université de Nice-Sofia Antipolis. Nice, France. 679 p.
- Kumamoto, T. (2004). *Design and Implementation of Natural Language Interface for Impression-based Music-Retrieval Systems*. Knowledge-Based Intelligent Information and Engineering Systems. Springer Berlin / Heidelberg. October 14, 2004. Vol. 3214/2004: pp. 139-147.
- Kumamoto, T. (2007). *A Natural Language Dialogue System for Impression-based Music-Retrieval*. Proc. CICLING-07 (Computational Linguistics and Intelligent Text Processing). Mexico. February 12-24, 2007. 12 p.
- Kumamoto, T. and K. Ohta (2003). *Design and Development of Natural Language Interface for an Impression-based Music Retrieval System*. in *Joho Shori Gakkai Kenkyu Hokoku*. Vol. 4(NL-153): pp. 97-104.
- Kurohashi, S. and M. Nagao (1999) *Manual for Japanese Morphological Analysis System JUMAN*. Rap. Language Media Lab. School of Informatics, Kyoto University. Kyoto, Japan. November 1999.
- Lehman, A. (1996). *Construction d'un système de résumé automatique de textes de type scientifique et technique*. Proc. RECITAL (Rencontre des Etudiants Chercheurs en Informatique pour le Traitement Automatique des Langues). Paris, France. 25-27 septembre 1996. pp. 65-69.
- Lehrberger, J.(1982). *Automatic Translation and the Concept of Sublanguage*. in *Sublanguage - Studies of language in restricted semantic domain*. Berlin / New York. pp. 81-106.
- Lepage, Y. (2005). *Translation of sentences by analogy principale*. Proc. Language & Technology Conference. Poznań, Poland. April 21-23, 2005.
- Lepage, Y. (2006). *Traduction par analogie*. Proc. MT-Summit. Pukhet, IAMT, ed.
- Liu, T. and D. Geiger (1999). *Approximate tree matching and shape similarity*. Proc. Seventh International Conference on Computer Vision. Kerkyra. pp. 456-462.
- Lopez, P., C. Fay-Varnier, et al. (1999). *Sous-langages d'application et LTAG : le système EGAL*. Proc. Actes de TALN-99 (Traitement Automatique des Langues Naturelles). Cargèse. 12-17 juillet 1999.
- Lu, S. (1979). *A tree-to-tree distance and its application to cluster analysis*. Proc. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 1/2: pp. 219-224.
- Marsh, S., A. Ghorbani, et al. (2003). *The ACORN multi-agent system*. in *Web Intelligence and Agent Systems*. Vol. 1: pp. 1-21.

- Melamed, I. D. (1996) *A geometric approach to mapping bitext correspondence*. Rap. University of Pennsylvania. Pennsylvania, U.S.A. 12 p.
- Miller, S., M. Crystal, et al. (1998). *Description of the sift system as used for MUC7*. Proc. The 7th Message Understanding Conference.
- MUC7 (1998). Proceedings of the 7th Message Understanding Conference (MUC7).
- Nelson, F. W. (1982). *Problems of assembling and computerizing large corpora*. Proc. Computer Corpora in English Language Research. Bergen: Norwegian Computing Centre for the Humanities. pp. 7-42.
- Nelson, F. W. (1992). *Language Corpora B.C.* Proc. Directions in Corpus Linguistics: Proceedings of Nobel Symposium. Stockholm. pp. 17-32.
- Ngai, G. and R. Florian (2001). *Transformation-based learning in the fast lane*. Proc. The second Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL). Pittsburgh, U.S.A. June 2-7, 2001. pp. 40-47.
- Och, F. J. and H. Ney (2000). *Improved statistical alignment models*. Proc. The 38th Annual Meeting of the Association for Computational Linguistics. pp. 440-447.
- Och, F. J. and H. Ney (2003). *A systematic Comparison of Various Statistical Alignments Models*. Proc. Computational Linguistics. March 2003. Vol. 1/29: pp. 19-51.
- Papineni, K., S. Roukos, et al. (2002). *BLEU: a Method for Automatic Evaluation of Machine Translation*. Proc. the 40th Annual Meeting of the Association for Computational Linguistics (ACL). Philadelphia. July 2002. pp. 311-318.
- Pearson, J.(1998). *Terms in Context*. Amsterdam/Philadelphia. John Benjamins.
- Pineau, M. (2004) *Comparaison de résultats de traduction (automatique ou non)*. Rap. stage TER Université Joseph Fourier. Grenoble, France. 10 mai 2004. p. 24.
- Plamondon, L. (2002) *Le système de question-réponse QUANTUM*. Rap. Département d'informatique et de recherche opérationnelle Faculté des arts et des sciences, Université de Montreal. Montreal, Canada. Mars 2002. 49 p.
- Planas, E. (1998) *TELA Structures et algorithmes pour la Traduction Fondée sur la Mémoire*. Thèse. Université Joseph Fourier. Grenoble, France. 7 juillet 1998. 200 p.
- Rigamonti, M., K. Hadjar, et al. (2004). *Xed : un outil pour l'extraction et l'analyse de documents PDF*. Proc. Huitième Colloque International Francophone sur l'Ecrit et le Document (CIFED'04). La Rochelle, France. 21-25 juin 2004 . pp. 85-90.
- Riloff, E. (1993). *Automatically Constructing a dictionary for information extraction tasks*. Proc. The eleventh National Conference on Artificial Intelligence (AAAI-93). The AAAI Press/MIT Press. pp. 811-816.
- Riloff, E. (1996). *Automatically generating extraction patterns from untagged text*. Proc. The Thirteenth National Conference on Artificial Intelligence (AAAI-96). The AAAI Press/MIT Press. pp. 1044-1049.
- Riloff, E. and R. Jones (1999). *Learning dictionaries for information extraction by multi-level bootstrapping*. Proc. The sixteenth National Conference on Artificial Intelligence (AAAI-99). The AAAI Press/MIT Press. pp. 474-479.
- Riloff, E., C. Schafer, et al. (2002). *Inducing Information Extraction Systems for New Languages via Cross-Languages Projection*. Proc. COLING 2002. August 24-25, 2002. Taipei, Taiwan 7 p.
- Ritchie, G. D. and I. Androutsopoulos (1995). *Natural Language Interfaces to Databases Ð an Introduction*. Natural Language Engineering. in Cambridge University Press: pp. 29-81.

- Salton, G.(1971). *The SMART Retrieval System-Experiments in Automatic Documents Processing*. Englewood Cliffs (Prentice-Hall). New Jersey, U.S.A.
- Sato, A., J. Ogawa, et al. (2000). *An Impression-based Retrieval System of Music Collection*. Proc. 4th Int. Conf. on Knowledge-Based Intelligent System and Allied Technologies. Brighton, U.K. pp. 856-859.
- Sekine, S. (1994). *A new direction for sublanguage NLP*. Proc. International Conference on New Methods in Language Processing. Manchester, England. 8 p.
- Selkow, S. M. (1977). *The tree-to-tree editing problem*. in Information Processing Letters. Vol. 6: pp. 184-186. December, 1977.
- Sérasset, G. and C. Boitet (2000). On UNL as the future "html of the linguistic content" & the reuse of existing NLP components in UNL-related applications with the example of a UNL-French deconverter. Proc. COLING-2000. Saarbrücken. July 31-August 3, 2000. Vol. 2/2: pp. 768-774.
- Shasha, K., J. Wang, et al. (2001). "Treediff: approximate tree matcher for ordered trees." Retrieved October 20, 2003 from <http://www.cs.nyu.edu/cs/faculty/shasha/papers/tree.html>.
- Slocum, J. (1986). *How one might automatically identify and adapt to a sublanguage*. in Analyzing language in restricted domains. pp. 195-210.
- Soderland, S., D. Fisher, et al. (1995). *CRYSTAL: Inducing a conceptual dictionary*. Proc. The fourteenth International Joint Conference on Artificial Intelligence. pp. 1314-1319.
- Sowa, J. F. (1984). *Conceptual Structures: Information Processing in Mind and Machine*. Reading, Mass. Don Mills, Ont. Addison, Wesley.
- Stolcke, A. (2002). *SRILM - an Extensible Language Modeling Toolkit*. Proc. ICSLP. Vol. 2: pp. 901-904. Denver. USA.
- Tai, K.-C. (1979). *The tree-to-tree correction problem*. in ACM. Vol. 26: pp. 422-433.
- Tsai, W.-J. (2004) *La coédition langue – UNL pour partager la révision entre langues d'un document multilingue*. Thèse. Université Joseph Fourier, GETA, CLIPS. Grenoble, France. 9 juillet 04. 311 p.
- Tsuji, Y., M. Hoshi, et al. (1997). *Local Pattern of a Melody and Its Applications to Retrieval by Sensitivity Words*. in IEICE technical report. Speech. Vol. 96/565: pp. 17-24.
- Uchida, H. and M. Zhu (1999). *Enconverter Specifications*. Rap. UNL Center, UNU/IAS, Tokyo. 33 p.
- Uchida, H. and M. Zhu (2003). *Deconverter Specifications*. Rap. UNL Center, UNU/IAS, Tokyo. January 24, 2003. 30 p.
- Uchida, H. and M. Zhu (2003). *The Universal Networking Language specification*. Rap. UNU/IAS, Tokyo. 20p.
- Uchida, H., M. Zhu, et al. (2005-2006). *Universal Networking Language* 10 2-8399-0128-5. 218 p.
- Vauquois, B., G. Veillon, et al. (1969). Une notation des textes hors des contraintes morphologiques et syntaxiques de l'expression. Proc. COLING-69. Stockholm, Suède. 27 p.
- Véronis, J. and É. G. De. Neef, (2006). *Le traitement des nouvelles formes de communication écrite*. Compréhension automatique des langues et interaction. Paris Hermès Science. G. Sabah. pp. 227-248.
- Wagner, R. and M. Fischer (1974). *The String-to-String Correction Problem*. in ACM Journal Vol. 21: pp. 309-321.

- Walker, M. A., J. C. Fromer, et al. (1998). *Learning Optimal Dialogue Strategies: A Case Study of a Spoken Dialogue Agent for Email*. Proc. Association for Computation Linguistics. Montreal, Canada.
- Wang, J., B. A. Shapiro, et al. (1998). *An algorithm for finding the largest approximately common substructures of two trees*. Proc. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 20: pp. 889-895.
- Yangarber, R., R. Grishman, et al. (2000). *Automatic acquisition of domain knowledge for information extraction*. Proc. COLING 2000. pp. 940-946.
- Yarowsky, D., G. Ngai, et al. (2001). *Inducing multilingual text analysis tools via robust projection across aligned corpora*. Proc. First International Conference on Human Language Technologies. San Diego, California, U.S.A. March 2001. pp. 109-116.
- Zadrozny, W., C. Wolf, et al. (1998). *Conversation Machines for Transaction Processing*. Proc. The fifteenth National Conference on Artificial Intelligence and Tenth Conference on Innovative Applications of Artificial Intelligence Conference (AAAI). Madison Wisconsin U.S.A.
- Zair, H. (2008) *Évaluation interne et externe de système de traitement de contenu*. Rap. Université Joseph Fourier. GETALP, LIG. Grenoble, France. 26 juin 2008. 17 p.

Annexes

Annexe I. Grille estimative d'un portage linguistique et/ou localisation de CATS

Les chiffres qui représentent les pourcentages sont des estimations du taux de modifications nécessaires pour la tâche du portage linguistique et/ou de la localisation. Ça varie entre des légères modifications de critère de recherche ou des légères modifications liées à l'arabe dialectal utilisée dans d'autres pays voisins et des grandes modifications liées à un changement de langue et de pays.

Pour le cas d'une tâche de multilinguïsation, qui consiste à avoir l'application CATS qui fonctionne dans plusieurs langues, ces grandeurs dépendront du nombre n des langues souhaitées et du pays.

Service.	Sous-langage	Grammaires	Variation géographique						Variation temporelle
			Pays origine	Pays arabe voisin	Autre pays arabe	France	Europe	Monde	
Portage linguistique et/ou localisation	Automobile	Dictionnaire	2000	2000 ± 5%	2000 ± 10%	2000 ± 20%	2000 ± 30%	2000 ± 40%	2000 + 5% tous les 10 ans
		Règles	710	710 ± 2%	710 ± 4%	710 ± 10%	710 ± 20%	710 ± 40%	710 + 5% tous les 10 ans
	Immobilier	Dictionnaire	7000	7000 ± 40%	7000 ± 50%	7000 ± 60%	7000 ± 70%	7000 ± 80%	7000 ± 20% tous les 10 ans
		Règles	Y	Y ± 2%	Y ± 4%	Y ± 10%	Y ± 20%	Y ± 40%	Y ± 5% tous les 10 ans

Annexe II. DTD CPXM « Common Parallel eXample Markup ».

CPXM est une dtd qui permet de représenter des documents multilingues et multi-versions. Chaque document contient plusieurs polyphrases. Chaque polyphrase contient plusieurs composants monolingue. Chaque composant monolingue contient plusieurs propositions.

```
>!-- CPXM.dtd (Common Parallel eXample Markup ) is a DTD which
describes
multilingual documents (m languages), multiversions(n versions) (n>m),
it allows the description of the notion of the polyphrase represented by a
single coding.
  $Author: Najeh Hajlaoui najeh.hajlaoui@imag.fr
  $Date: 2004/07/06 15:28:30 $
-->
<!ELEMENT document (information, polyphrase*) >
<!ELEMENT information (#PCDATA) >
<!ATTLIST information    document-name          CDATA #REQUIRED>
<!ATTLIST information    creation-date          CDATA #IMPLIED>
<!ATTLIST information    modification-date      CDATA #IMPLIED>
<!ATTLIST information    coding-set            CDATA #IMPLIED>
<!ATTLIST information    number-of-languages   CDATA #IMPLIED>
<!ATTLIST information    number-of-polyphrases CDATA #IMPLIED>

<!ATTLIST polyphrase     polyphrase-id         CDATA #REQUIRED>
<!ELEMENT polyphrase (monolingual-component*) >

<!ATTLIST monolingual-component xml:lang      CDATA #REQUIRED>
<!ELEMENT monolingual-component (proposal*) >
<!ATTLIST proposal       proposal-id         CDATA #REQUIRED>
<!ELEMENT proposal (#PCDATA) >

<!-- number-of-languages is the total number of languages appearing in the
document;
if the document is monolingual, number-of-languages =1 -->
<!-- number-of-polyphrases is the total number of polyphrases constituting
the document -->
<!-- A polyphrase is a set of monolingual components, each containing 1 or
more possible proposals.
Every polyphrase is identified by a number called polyphrase-id -->
<!-- Each monolingual component is a set of one or more possible renderings
of the proposal in question;
it is identified by an attribute which indicates the language -->
<!-- Proposal represents the level of alignment, it is usually a sentence -
->
```

Annexe III. Résultats de la méthode de portage interne de CATS en français

```

;===== UNL
=====
;Recherche voiture de 400 à 6500 dinars
diesel ou essence /0795601711
[S]
[W]
wanted:03
[/W]
[/S]
;=====
=====
;;Time 0.3 Sec
;;Done!
;===== UNL
=====
;Recherche voiture avec facilités de
paiement japonaise ou allemande 0795777953
[S]
wan(saloon:0A, wanted:00)
cou(saloon:0A, japan:1A)
cou(saloon:0A, germany:1N)
yea(saloon:0A, 0795777953:1X)
[/S]
;=====
=====
;;Time 0.3 Sec
;;Done!
;===== UNL
=====
;recherche voiture OPEL VECTRA
[S]
wan(saloon:0A, wanted:00)
mak(saloon:0A,
      OPEL(country>germany,country>europe)
:0I)
mod(saloon:0A,
      Vectra(country>germany,country>europ
e,make>OPEL):0N)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
;Recherche achat voiture modèle 99 2000
prix raisonnable
[S]
wan(saloon:0G, wanted:00)
yea(saloon:0G, 99:0W)
pri(saloon:0G, 2000:0Z)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
;Recherche voiture modèle à partir de 85
avec tranches mensuelles de 100 dinars sans
acompte
[S]
wan(saloon:0A, wanted:00)
yea(saloon:0A, 85:13)
[/S]
;=====
=====
;;Time 0.2 Sec
;;Done!
;===== UNL

```

```

=====
;recherche voiture
[S]
wan(saloon:0A, wanted:00)
[/S]
;=====
=====
;;Time 0.0 Sec
;;Done!
;===== UNL
=====
;recherche voiture
[S]
wan(saloon:0A, wanted:00)
[/S]
;=====
=====
;;Time 0.0 Sec
;;Done!
;===== UNL
=====
;recherche bus KIA transport commun
paiement fractionné
[S]
wan(bus:0A, wanted:00)
mak(bus:0A, KIA(country>korea):0E)
[/S]
;=====
=====
;;Time 0.2 Sec
;;Done!
;===== UNL
=====
;recherche bus transport commun KIA ou
HONDA ou SINGE YOUNG avec acompte 3000 et
le reste par tranches
[S]
wan(bus:0A, wanted:00)
mak(bus:0A, KIA(country>korea):0V)
mak(bus:0A, HONDA(country>japan):12)
mak(bus:0A, SangYong(country>korea):1B)
pri(bus:0A, 3000:20.@downpayment)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
;recherche JEEP Cherokee 2000 1900 dinars
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, JEEP(country>America):0A)
mod(saloon:00,
      Cherokee(country>AmericF,make<JEEP):
0F)
yea(saloon:00, 2000:00)
pri(saloon:00, 1900:0T)
[/S]
;=====
=====
;;Time 0.0 Sec
;;Done!
;===== UNL
=====
;MERCEDES classique E200 toutes options
22000 dinars
[S]
mak(saloon:00,
      MERCEDES(country>germany,country>eur

```

```

ope):00)
mod(saloon:00, E 200:0J)
fea(saloon:00, full options:00)
pri(saloon:00, 22000:13)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
;je cherche une voiture moteur 1300 cc
modèle 85 2500 dinars
[S]
wan(saloon:0F, wanted:00)
mot(saloon:0F, 1300:0U)
mot(saloon:0F, 85:1A)
pri(saloon:0F, 2500:1D)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
;recherche voiture américaine à condition
que le prix ne dépasse pas 1000 dinars
[S]
wan(saloon:0A, wanted:00)
pri(saloon:0A, 1000:1Z)
[/S]
;=====
=====
;;Time 0.2 Sec
;;Done!
;===== UNL
=====
;recherche DAEWOO
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, DAEWOO(country>korea):0A)
[/S]
;=====
=====
;;Time 0.0 Sec
;;Done!
;===== UNL
=====
;recherche NISSAN Sunny manuelle modèle 93
à 97
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, NISSAN(country>japan):0A)
mod(saloon:00,
Sunny(country>japan,make>NISSAN):0H)
fea(saloon:00, manual gear:0N)
yea(saloon:00, 93:14)
yea(saloon:00, 97:1A)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
;recherche RENAULT Clio manuelle modèle 95
à 2000
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
RENAULT(country>France,country>europ
e):0A)
mod(saloon:00,
Clio(country>France,country>europe,m
ake>RENAULT):0I)
fea(saloon:00, manual gear:0N)
yea(saloon:00, 95:14)

```

```

yea(saloon:00, 2000:1A)
[/S]
;=====
=====
;;Time 0.2 Sec
;;Done!
;===== UNL
=====
;recherche voiture HYUNDAI
[S]
wan(saloon:0A, wanted:00)
mak(saloon:0A, HYUNDAI(country>korea):0I)
[/S]
;=====
=====
;;Time 0.0 Sec
;;Done!
;===== UNL
=====
;recherche voiture MITSUBISHI
[S]
wan(saloon:0A, wanted:00)
mak(saloon:0A,
MITSUBISHI(country>japan):0I)
[/S]
;=====
=====
;;Time 0.0 Sec
;;Done!
;===== UNL
=====
;je cherche une voiture BMW modèle 92
[S]
wan(saloon:0F, wanted:00)
mak(saloon:0F,
B.M.W(country>germany,country>europe
):0N)
mod(saloon:0F, 92:0Z)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
;recherche DAEWOO
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, DAEWOO(country>korea):0A)
[/S]
;=====
=====
;;Time 0.0 Sec
;;Done!
;===== UNL
=====
;recherche NISSAN Sunny modèle 93 à 95
manuelle
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, NISSAN(country>japan):0A)
mod(saloon:00,
Sunny(country>japan,make>NISSAN):0H)
yea(saloon:00, 93:0W)
yea(saloon:00, 95:12)
fea(saloon:00, manual gear:15)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
;recherche HONDA Civic modèle 94 toutes
options manuelle
[S]
wan(saloon:00, wanted:00)

```

```

mak(saloon:00, HONDA(country>japan):0A)
mod(saloon:00,
    Civic(country>japan,make>HONDA):0G)
yea(saloon:00, 94:0U)
fea(saloon:00, full options:0X)
fea(saloon:00, manual gear:1C)
[/S]
;=====
;Time 0.1    Sec
;;Done!
;===== UNL
;=====
;recherche KIA Sephia
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, KIA(country>korea):0A)
mod(saloon:00,
    Sephia(country>koreF,make>KIA):0E)
[/S]
;=====
;Time 0.0    Sec
;;Done!
;===== UNL
;=====
;recherche DAEWOO Lanos 95 à 97 manuelle
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, DAEWOO(country>korea):0A)
mod(saloon:00,
    Lanos(country>koreF,make>DAEWOO):0H)
yea(saloon:00, 95:0N)
yea(saloon:00, 97:0T)
fea(saloon:00, manual gear:0W)
[/S]
;=====
;Time 0.1    Sec
;;Done!
;===== UNL
;=====
;recherche Lada SAMARA
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, Lada:0A)
mod(saloon:00, SAMARA(make>Lada):0F)
[/S]
;=====
;Time 0.0    Sec
;;Done!
;===== UNL
;=====
;recherche voiture HYUNDAI
[S]
wan(saloon:0A, wanted:00)
mak(saloon:0A, HYUNDAI(country>korea):0I)
[/S]
;=====
;Time 0.1    Sec
;;Done!
;===== UNL
;=====
;recherche voiture Excel
[S]
wan(saloon:0A, wanted:00)
mod(saloon:0A,
    Excel(country>koreF,make>HYUNDAI):0I
)
[/S]
;=====
;Time 0.0    Sec
;;Done!
;===== UNL
;=====
;recherche SKODA ou LADA
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, SKODA(country>europe):0A)
mak(saloon:00, Lada:0J)
[/S]
;=====
;Time 0.1    Sec
;;Done!
;===== UNL
;=====
;cherche à acheter JEEP avec acompte de
3000 dinars et des tranches mensuelles de
200 contact 0795291030
[S]
mak(saloon:00, JEEP(country>Ameria):0J)
pri(saloon:00, 3000:14)
pri(saloon:00, 200:2A.@installment)
pri(saloon:00, 0795291030:2M.@downpayment)
[/S]
[S]
mak(saloon:00, JEEP(country>Ameria):0J)
pri(saloon:00, 3000:14)
pri(saloon:00, 200:2A.@installment)
pri(saloon:00, 0795291030:2M.@downpayment)
[/S]
[S]
mak(saloon:00, JEEP(country>Ameria):0J)
pri(saloon:00, 3000:14)
pri(saloon:00, 200:2A.@installment)
pri(saloon:00, 0795291030:2M.@downpayment)
[/S]
[S]
mak(saloon:00, JEEP(country>Ameria):0J)
pri(saloon:00, 3000:14)
pri(saloon:00, 200:2A.@installment)
pri(saloon:00, 0795291030:2M.@downpayment)
[/S]
;=====
;Time 0.2    Sec
;;Done!
;===== UNL
;=====
;recherche voiture américaine Caprice de
préférence
[S]
wan(saloon:0A, wanted:00)
mod(saloon:0A, Caprice
Classic(country>AmericF,make>CHEVROLET):0U)
[/S]
;=====
;Time 0.1    Sec
;;Done!
;===== UNL
;=====
;recherche voiture sport de préférence OPEL
Calibra
[S]
wan(saloon:0A, wanted:00)
col(saloon:0A, goldr:0K)
mak(saloon:0A,
    OPEL(country>germany,country>europe)
:14)
[/S]
;=====
;Time 0.2    Sec
;;Done!
;===== UNL
;=====
;recherche BMW 316

```

```

[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    B.M.W(country>germany, country>europe
):0A)
mod(saloon:00, 316:0E)
[/S]
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
;recherche à l'achat MERCEDES CLK modèle
2000 ou 99
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    MERCEDES(country>germany, country>eur
ope):0L)
mod(saloon:00, 2000:16)
yea(saloon:00, 99:1E)
[/S]
=====
;;Time 0.2 Sec
;;Done!
===== UNL
=====
;recherche BMW 316
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    B.M.W(country>germany, country>europe
):0A)
mod(saloon:00, 316:0E)
[/S]
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
;recherche à l'achat bus TOYOTA 1990 ou
après numero tel 0795133829
[S]
wan(bus:0L, wanted:00)
mak(bus:0L, TOYOTA(country>japan):0P)
yea(bus:0L, 1990:0W)
yea(bus:0L, 0795133829:1M)
[/S]
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
;recherche à l'achat NISSAN Sunny modèle 92
à 95
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, NISSAN(country>japan):0L)
mod(saloon:00,
    Sunny(country>japan, make>NISSAN):0S)
yea(saloon:00, 92:16)
yea(saloon:00, 95:1C)
[/S]
=====
;;Time 0.2 Sec
;;Done!
===== UNL
=====
;cherche à vendre BMW m5
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,

```

```

    B.M.W(country>germany, country>europe
):0I)
mod(saloon:00, 5:0N)
[/S]
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
;recherche HONDA Civic contrôle boîte
automatique modèle 98 ou 99 avec acompte et
tranches
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, HONDA(country>japan):0A)
mod(saloon:00,
    Civic(country>japan, make>HONDA):0G)
fea(saloon:00, automatic gear:12)
yea(saloon:00, 98:1M)
yea(saloon:00, 99:1S)
[/S]
=====
;;Time 0.3 Sec
;;Done!
===== UNL
=====
;recherche MERCEDES
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    MERCEDES(country>germany, country>eur
ope):0A)
[/S]
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
;je veux vendre voiture DAEWOO LIMINZ
modèle 92 contrôle complet toutes options
[S]
sal(saloon:0F, sale:00)
mak(saloon:0F, DAEWOO(country>korea):0N)
yea(saloon:0F, 92:19)
fea(saloon:0F, full options:1U)
[/S]
=====
;;Time 0.2 Sec
;;Done!
===== UNL
=====
;recherche voiture BM 520 m à partir de 90
[S]
wan(saloon:0A, wanted:00)
mak(saloon:0A,
    B.M.W(country>germany, country>europe
):0I)
mod(saloon:0A, 520:0L)
yea(saloon:0A, 90:14)
[/S]
=====
;;Time 0.2 Sec
;;Done!
===== UNL
=====
;recherche MITSUBISHI
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    MITSUBISHI(country>japan):0A)
[/S]
=====

```

```

=====
;;Time 0.0 Sec
;;Done!
;===== UNL
=====
;recherche MITSUBISHI Galant modèle 80 à 86
en bon état avec toutes les options
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    MITSUBISHI(country>japan):0A)
mod(saloon:00,
    Galant(country>japan,make>MITSUBISHI
):0L)
yea(saloon:00, 80:10)
yea(saloon:00, 86:16)
fea(saloon:00, full options:1M)
[/S]
;=====
=====
;;Time 0.3 Sec
;;Done!
;===== UNL
=====
;recherche voiture MERCEDES
[S]
wan(saloon:0A, wanted:00)
mak(saloon:0A,
    MERCEDES(country>germany,country>eur
ope):0I)
[/S]
;=====
=====
;;Time 0.0 Sec
;;Done!
;===== UNL
=====
;recherche achat voiture
[S]
wan(saloon:0G, wanted:00)
[/S]
;=====
=====
;;Time 0.0 Sec
;;Done!
;===== UNL
=====
;recherche voiture prix 3500 à 3850 dinars
[S]
wan(saloon:0A, wanted:00)
pri(saloon:0A, 3500:0N)
pri(saloon:0A, 3850:0V)
[/S]
;=====
=====
;;Time 0.0 Sec
;;Done!
;===== UNL
=====
;achat voiture
[S]
wan(saloon:06, wanted:00)
[/S]
;=====
=====
;;Time 0.0 Sec
;;Done!
;===== UNL
=====
;recherche MERCEDES
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    MERCEDES(country>germany,country>eur
ope):0A)
[/S]

```

```

=====
;;Time 0.0 Sec
;;Done!
;===== UNL
=====
;recherche KIA Sephia
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, KIA(country>korea):0A)
mod(saloon:00,
    Sephia(country>koreF,make>KIA):0E)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
;recherche HYUNDAI Accent ou Elantra ou KIA
Sephia à 3800 dinars
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, HYUNDAI(country>korea):0A)
mod(saloon:00,
    Accent(country>koreF,make>HYUNDAI):0
I)
mod(saloon:00,
    Elantra(country>koreF,make>HYUNDAI):
0S)
mak(saloon:00, KIA(country>korea):13)
mod(saloon:00,
    Sephia(country>koreF,make>KIA):17)
pri(saloon:00, 3800:1H)
[/S]
;=====
=====
;;Time 0.2 Sec
;;Done!
;===== UNL
=====
;recherche à l'achat MERCEDES 2005
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    MERCEDES(country>germany,country>eur
ope):0L)
mod(saloon:00, 2005:0U)
[/S]
;=====
=====
;;Time 0.0 Sec
;;Done!
;===== UNL
=====
;recherche voiture DAEWOO
[S]
wan(saloon:0A, wanted:00)
mak(saloon:0A, DAEWOO(country>korea):0I)
[/S]
;=====
=====
;;Time 0.0 Sec
;;Done!
;;Time Over!
;===== UNL
=====
;recherche MERCEDES 200 modèle 1995
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    MERCEDES(country>germany,country>eur
ope):0A)
mod(saloon:00, 200:0J)
yea(saloon:00, 1995:0V)
[/S]

```



```

=====
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
;Besoin d'une voiture ne dépassant pas 1500
dinars en bon état et économique première
main
[S]
wan(saloon:0D, wanted:00)
pri(saloon:0D, 1500:13)
[/S]
=====
=====
;;Time 0.3 Sec
;;Done!
===== UNL
=====
;Recherche voiture KIA
[S]
wan(saloon:0A, wanted:00)
mak(saloon:0A, KIA(country>korea):0I)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
;recherche à l'achat BMW modèle 86 ou KIA
modèle 90 à 96
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
      B.M.W(country>germany, country>europe
):0L)
mod(saloon:00, 86:0X)
mak(saloon:00, KIA(country>korea):13)
yea(saloon:00, 90:1F)
yea(saloon:00, 96:1L)
[/S]
=====
=====
;;Time 0.3 Sec
;;Done!
===== UNL
=====
;recherche CHEVROLET Caprice paiement en
espèces
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
      CHEVROLET(country>Ameria):0A)
mod(saloon:00, Caprice
Classic(country>AmericF, make>CHEVROLET):0K)
[/S]
=====
=====
;;Time 0.2 Sec
;;Done!
===== UNL
=====
;recherche pickup NISSAN
[S]
wan(pickup:0A, wanted:00)
mak(pickup:0A, NISSAN(country>japan):0H)
[/S]
=====
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
;je souhaite acheter MERCEDES modèle 82
toutes options
[S]

```

```

wan(saloon:00, wanted:00)
mak(saloon:00,
      MERCEDES(country>germany, country>eur
ope):0K)
mod(saloon:00, 82:11)
fea(saloon:00, full options:14)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
;recherche pickup NISSAN modèle 95-96-97
[S]
wan(pickup:0A, wanted:00)
mak(pickup:0A, NISSAN(country>japan):0H)
yea(pickup:0A, 95:0W)
yea(pickup:0A, 96:0Z)
yea(pickup:0A, 97:12)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
;recherche voiture acompte 1000 et le reste
par tranches
[S]
wan(saloon:0A, wanted:00)
pri(saloon:0A, 1000:0Q.@downpayment)
[/S]
=====
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
;Besoin d'une voiture ne dépassant pas 2000
dinars en bon état et économique en essence
[S]
wan(saloon:0D, wanted:00)
pri(saloon:0D, 2000:13)
[/S]
=====
=====
;;Time 0.3 Sec
;;Done!
===== UNL
=====
;recherche voiture DAEWOO acompte 1000 et
le reste par tranches
[S]
wan(saloon:0A, wanted:00)
mak(saloon:0A, DAEWOO(country>korea):0I)
pri(saloon:0A, 1000:0X.@downpayment)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
;recherche Lada station modèle 91-92-93-94-
95
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, Lada:0A)
yea(saloon:00, 91:0V)
yea(saloon:00, 92:0Y)
yea(saloon:00, 93:11)
yea(saloon:00, 94:14)
yea(saloon:00, 95:17)
[/S]
=====
=====

```

```

;;Time 0.2 Sec
;;Done!
;===== UNL
;recherche Golf m 89 90 ou KIA 94 95
contrôle complet
[S]
mak(saloon:00, KIA(country>korea):0Q)
[/S]
[S]
mak(saloon:00, KIA(country>korea):0Q)
[/S]
[S]
mak(saloon:00, KIA(country>korea):0Q)
[/S]
[S]
mak(saloon:00, KIA(country>korea):0Q)
[/S]
;=====
;;Time 0.2 Sec
;;Done!
;===== UNL
;je veux acheter une voiture a compte 500
dinars et le reste par tranches mensuelles
de 100 dinars
[S]
wan(saloon:0K, wanted:00)
pri(saloon:0K, 500:10)
[/S]
;=====
;;Time 0.2 Sec
;;Done!
;===== UNL
;recherche voiture Golf
[S]
wan(saloon:0A, wanted:00)
mod(saloon:0A,
    Golf(country>germany,country>europe,
    make>VOLKS WAGEN):0I)
[/S]
;=====
;;Time 0.0 Sec
;;Done!
;===== UNL
;recherche KIA Sephia
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, KIA(country>korea):0A)
mod(saloon:00,
    Sephia(country>koreF,make>KIA):0E)
[/S]
;=====
;;Time 0.0 Sec
;;Done!
;===== UNL
;recherche voiture MITSUBISHI boite normale
et en bon état
[S]
wan(saloon:0A, wanted:00)
mak(saloon:0A,
    MITSUBISHI(country>japan):0I)
fea(saloon:0A, manual gear:0T)
[/S]
;=====
;;Time 0.2 Sec
;;Done!
;===== UNL
;recherche voiture MITSUBISHI Galant modèle
1988 à 1991 boite manuelle et en bon état
[S]
wan(saloon:0A, wanted:00)
mak(saloon:0A,
    MITSUBISHI(country>japan):0I)
mod(saloon:0A,
    Galant(country>japan,make>MITSUBISHI
    ):0T)
yea(saloon:0A, 1988:18)
yea(saloon:0A, 1991:1G)
fea(saloon:0A, manual gear:1R)
[/S]
;=====
;;Time 0.2 Sec
;;Done!
;===== UNL
;recherche HYUNDAI Elantra à 3800 dinars
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, HYUNDAI(country>korea):0A)
mod(saloon:00,
    Elantra(country>koreF,make>HYUNDAI):
    0I)
pri(saloon:00, 3800:0T)
[/S]
;=====
;;Time 0.0 Sec
;;Done!
;===== UNL
;recherche HONDA Civic ou MITSUBISHI Lancer
modèle 97 à 99 a compte 3000 dinars et le
reste par tranches mensuelles
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, HONDA(country>japan):0A)
mod(saloon:00,
    Civic(country>japan,make>HONDA):0G)
mak(saloon:00,
    MITSUBISHI(country>japan):0P)
mod(saloon:00,
    Lancer(country>japan,make>MITSUBISHI
    ):10)
yea(saloon:00, 97:1F)
pri(saloon:00, 99:1L)
pri(saloon:00, 3000:1W)
[/S]
;=====
;;Time 0.2 Sec
;;Done!
;===== UNL
;recherche voiture MITSUBISHI Lancer modèle
de 97 à 99 avec un a compte de 4000 dinars
et le reste par tranches mensuelles
[S]
wan(saloon:0A, wanted:00)
mak(saloon:0A,
    MITSUBISHI(country>japan):0I)
mod(saloon:0A,
    Lancer(country>japan,make>MITSUBISHI
    ):0T)
yea(saloon:0A, 97:1B)
yea(saloon:0A, 99:1H)
pri(saloon:0A, 4000:23)
[/S]
;=====
;;Time 0.3 Sec
;;Done!

```

```

;===== UNL
;recherche à l'achat MERCEDES 81 contrôle
complet et toutes options
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    MERCEDES(country>germany,country>eur
ope):0L)
mod(saloon:00, 81:0U)
fea(saloon:00, full options:1I)
[/S]
;=====
;Time 0.1 Sec
;;Done!
;===== UNL
;recherche HYUNDAI à 3800 dinars
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, HYUNDAI(country>korea):0A)
pri(saloon:00, 3800:0M)
[/S]
;=====
;Time 0.1 Sec
;;Done!
;===== UNL
;recherche PEUGEOT 307 SALFA
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    PEUGEOT(country>France,country>europ
e):0A)
mod(saloon:00, 307:0I)
[/S]
;=====
;Time 0.1 Sec
;;Done!
;===== UNL
;recherche BMW toutes options modèle 98 à
2003
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    B.M.W(country>germany,country>europ
e):0A)
fea(saloon:00, full options:0E)
mod(saloon:00, 98:11)
yea(saloon:00, 2003:17)
[/S]
;=====
;Time 0.1 Sec
;;Done!
;===== UNL
;recherche BMW 740 toutes options modèle 98
à 2003
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    B.M.W(country>germany,country>europ
e):0A)
mod(saloon:00, 740:0E)
fea(saloon:00, full options:0I)
yea(saloon:00, 98:15)
yea(saloon:00, 2003:1B)
[/S]
;=====
;Time 0.1 Sec
;;Done!

```

```

;===== UNL
;recherche KIA Sephia 95 acompte et le
reste par tranches
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, KIA(country>korea):0A)
mod(saloon:00,
    Sephia(country>koreF,make>KIA):0E)
pri(saloon:00, 95:0L)
[/S]
;=====
;Time 0.1 Sec
;;Done!
;===== UNL
;J'ai une voiture Land Rover je veux la
vendre
[S]
sal(saloon:09, sale:00)
mod(saloon:09, Land
Rover(country>England,country>europe,make>R
over):0H)
sal(saloon:09, sale:0S)
[/S]
;=====
;Time 0.1 Sec
;;Done!
;===== UNL
;recherche voiture HYUNDAI ATOS avec
facilités de paiement
[S]
wan(saloon:0A, wanted:00)
mak(saloon:0A, HYUNDAI(country>korea):0I)
mod(saloon:0A,
    Atos(country>koreF,make>HYUNDAI):0Q)
[/S]
;=====
;Time 0.2 Sec
;;Done!
;===== UNL
;recherche BMW
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    B.M.W(country>germany,country>europ
e):0A)
[/S]
;=====
;Time 0.1 Sec
;;Done!
;===== UNL
;recherche JEEP ou BM ou NISSAN modèle à
partir de 95 contact 95777953
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, JEEP(country>Ameria):0A)
mak(saloon:00,
    B.M.W(country>germany,country>europ
e):0I)
mak(saloon:00, NISSAN(country>japan):0O)
yea(saloon:00, 95:1G)
yea(saloon:00, 95777953:1R)
[/S]
;=====
;Time 0.3 Sec
;;Done!
;===== UNL

```

```

;recherche voiture BM modèle 75 = 77
[S]
wan(saloon:0A, wanted:00)
mak(saloon:0A,
    B.M.W(country>germany,country>europe
):0I)
mod(saloon:0A, 75:0T)
yea(saloon:0A, 77:0Z)
[/S]
;=====
;Time 0.1 Sec
;Done!
;===== UNL
;=====
;je veux une voiture MERCEDES 200 modèle
[S]
wan(saloon:0C, wanted:00)
mak(saloon:0C,
    MERCEDES(country>germany,country>eur
ope):0K)
mod(saloon:0C, 200:0T)
[/S]
;=====
;Time 0.1 Sec
;Done!
;===== UNL
;=====
;recherche PEUGEOT 206
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    PEUGEOT(country>France,country>europ
e):0A)
mod(saloon:00, 206:0I)
[/S]
;=====
;Time 0.1 Sec
;Done!
;===== UNL
;=====
;recherche PEUGEOT
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    PEUGEOT(country>France,country>europ
e):0B)
[/S]
;=====
;Time 0.1 Sec
;Done!
;===== UNL
;=====
;recherche PEUGEOT 207
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    PEUGEOT(country>France,country>europ
e):0A)
mod(saloon:00, 207:0I)
[/S]
;=====
;Time 0.0 Sec
;Done!
;===== UNL
;=====
;recherche datsun modèle de 80 à 82 moteur
de 1200 à 1400 CC
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, DATSUN(country>japan):0A)
yea(saloon:00, 80:0S)
yea(saloon:00, 82:0Y)
mot(saloon:00, 1200:1B)
mot(saloon:00, 1400:1J)
[/S]
;=====
;Time 0.2 Sec
;Done!
;===== UNL
;=====
;recherche à l'achat voiture modèle à
partir de 90 à 3000 dinars avec facilités
de paiement
[S]
wan(saloon:0L, wanted:00)
yea(saloon:0L, 90:1E)
pri(saloon:0L, 3000:1K)
[/S]
;=====
;Time 0.2 Sec
;Done!
;===== UNL
;=====
;recherche pickup TOYOTA modèle environ 80
contrôle ok
[S]
wan(pickup:0A, wanted:00)
mak(pickup:0A, TOYOTA(country>japan):0H)
yea(pickup:0A, 80:14)
[/S]
;=====
;Time 0.1 Sec
;Done!
;===== UNL
;=====
;recherche OPEL VECTRA moteur 1600 modèle à
partir de 94
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    OPEL(country>germany,country>europe)
:0A)
mod(saloon:00,
    Vectra(country>germany,country>europ
e,make>OPEL):0F)
mot(saloon:00, 1600:0T)
yea(saloon:00, 94:1J)
[/S]
;=====
;Time 0.2 Sec
;Done!
;===== UNL
;=====
;recherche voiture HONDA Civic 2003 en bon
état couleur grise ou dorée
[S]
wan(saloon:0A, wanted:00)
mak(saloon:0A, HONDA(country>japan):0I)
mod(saloon:0A,
    Civic(country>japan,make>HONDA):0O)
yea(saloon:0A, 2003:0U)
col(saloon:0A, silver:1K)
col(saloon:0A, goldr:1U)
[/S]
;=====
;Time 0.3 Sec
;Done!
;===== UNL
;=====
;recherche à l'achat voiture modèle à
partir de 90 acompte 500 dinars et tranches
[S]

```

```

wan(saloon:0L, wanted:00)
pri(saloon:0L, 90:1E)
pri(saloon:0L, 500:1P)
[/S]
=====
;;Time 0.2 Sec
;;Done!
===== UNL
=====
;recherche HONDA Civic 96 97 98
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, HONDA(country>japan):0A)
mod(saloon:00,
    Civic(country>japan,make>HONDA):0G)
yea(saloon:00, 96:0M)
yea(saloon:00, 97:0P)
yea(saloon:00, 98:0S)
[/S]
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
;recherche HYUNDAI Excel 1994
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, HYUNDAI(country>korea):0A)
mod(saloon:00,
    Excel(country>koreF,make>HYUNDAI):0I
)
yea(saloon:00, 1994:00)
[/S]
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
;recherche HONDA Civic 96 97 98
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, HONDA(country>japan):0A)
mod(saloon:00,
    Civic(country>japan,make>HONDA):0G)
yea(saloon:00, 96:0M)
yea(saloon:00, 97:0P)
yea(saloon:00, 98:0S)
[/S]
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
;recherche HONDA Civic 2003 en bon état
couleur grise ou dorée
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, HONDA(country>japan):0A)
mod(saloon:00,
    Civic(country>japan,make>HONDA):0G)
yea(saloon:00, 2003:0M)
col(saloon:00, silver:1C)
col(saloon:00, goldr:1M)
[/S]
=====
;;Time 0.2 Sec
;;Done! ;===== UNL
=====
;recherche HONDA Civic 2003 en bon état
couleur grise ou dorée
[S]
wan(saloon:00, wanted:00)

```

```

mak(saloon:00, HONDA(country>japan):0A)
mod(saloon:00,
    Civic(country>japan,make>HONDA):0G)
yea(saloon:00, 2003:0M)
col(saloon:00, silver:1C)
col(saloon:00, goldr:1M)
[/S]
=====
;;Time 0.2 Sec
;;Done!
===== UNL
=====
;je veux vendre une MERCEDES 280C modèle
2006 jordanienne kilométrage 4000 KM
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
    MERCEDES(country>germany,country>eur
ope):0J)
mod(saloon:00, 280:0S)
yea(saloon:00, 2006:15)
col(saloon:00, goldr:1B)
yea(saloon:00, 4000:1Z)
[/S]
=====
;;Time 0.2 Sec
;;Done!
===== UNL
=====
;OPEL kadett pack modèle 90 ou GOLF modèle
90 pack ou CITROEN pack
[S]
mak(saloon:00,
    OPEL(country>germany,country>europe)
:00)
mod(saloon:00,
    kadet(country>germany,country>europe
,make>OPEL):05)
yea(saloon:00, 90:0P)
mod(saloon:00,
    Golf(country>germany,country>europe,
make>VOLKS WAGEN):0V)
yea(saloon:00, 90:18)
mak(saloon:00,
    CITROEN(country>France,country>europ
e):1J)
[/S]
=====
;;Time 0.3 Sec
;;Done!
===== UNL
=====
;A vendre OPEL ASTRA rouge (avec ouverture
centrale vitre et rétro électrique)
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
    OPEL(country>germany,country>europe)
:09)
mod(saloon:00,
    Astra(country>germany,country>europe
,make>OPEL):0E)
col(saloon:00, red:0K)
fea(saloon:00, Fatha:0R)
[/S]
=====
;;Time 0.2 Sec
;;Done!
===== UNL
=====
;pickup datsun NISSAN diesel modèle 83
assuré un an en bon état à vendre, 2000
dinars

```

```

[S]
mak(pickup:00, DATSUN(country>japan):07)
mak(pickup:00, NISSAN(country>japan):0E)
yea(pickup:00, 83:10)
pri(pickup:00, 2000:25)
[/S]
=====
=====
;;Time 0.3    Sec
;;Done!
===== UNL
=====
;BMW, 1988, 318 en bon état, 6250 dinars
contact 0795322 079
[S]
mak(saloon:00,
    B.M.W(country>germany, country>europe
):00)
mod(saloon:00, 1988:05)
[/S]
[S]
mak(saloon:00,
    B.M.W(country>germany, country>europe
):00)
mod(saloon:00, 1988:05)
[/S]
[S]
mak(saloon:00,
    B.M.W(country>germany, country>europe
):00)
mod(saloon:00, 1988:05)
[/S]
[S]
mak(saloon:00,
    B.M.W(country>germany, country>europe
):00)
mod(saloon:00, 1988:05)
[/S]
=====
;;Time 0.2    Sec
;;Done!
===== UNL
=====
;PEUGEOT 2002 T.O sauf la boîte contrôle
complet pour renseignement 0796702818
[S]
mak(saloon:00,
    PEUGEOT(country>France, country>europ
e):00)
mod(saloon:00, 2002:09)
fea(saloon:00, full options:0E)
fea(saloon:00, manual gear:0I)
[/S]
=====
=====
;;Time 0.3    Sec
;;Done!
===== UNL
=====
;Vente ou échange VOLVO modèle 8 contrôle
complet sport 2 portes toutes options
[S]
mak(saloon:00,
    VOLVO(country>Sweden, country>europe)
:0I)
mod(saloon:00, 8:0W)
col(saloon:00, goldr:1I)
[/S]
[S]
mak(saloon:00,
    VOLVO(country>Sweden, country>europe)
:0I)
mod(saloon:00, 8:0W)
col(saloon:00, goldr:1I)
[/S]

```

```

[S]
mak(saloon:00,
    VOLVO(country>Sweden, country>europe)
:0I)
mod(saloon:00, 8:0W)
col(saloon:00, goldr:1I)
[/S]
[S]
mak(saloon:00,
    VOLVO(country>Sweden, country>europe)
:0I)
mod(saloon:00, 8:0W)
col(saloon:00, goldr:1I)
[/S]
[S]
mak(saloon:00,
    VOLVO(country>Sweden, country>europe)
:0I)
mod(saloon:00, 8:0W)
col(saloon:00, goldr:1I)
[/S]
[S]
mak(saloon:00,
    VOLVO(country>Sweden, country>europe)
:0I)
mod(saloon:00, 8:0W)
col(saloon:00, goldr:1I)
[/S]
[S]
mak(saloon:00,
    VOLVO(country>Sweden, country>europe)
:0I)
mod(saloon:00, 8:0W)
col(saloon:00, goldr:1I)
[/S]
=====
=====
;;Time 0.3    Sec
;;Done!
===== UNL
=====
;GOLF GTI 89 en bon état originale première
main prix 6500 tel 0647799079
[S]
[W]
Golf(country>germany, country>europe, make>VO
LKS WAGEN):00
[/W]
[/S]
[S]
[W]
89:09
[/W]
[/S]
[S]
[W]
year:0M
[/W]
[/S]
[S]
[W]
""
[/W]
[/S]
[S]
[W]
"originale"
[/W]
[/S]
[S]
[W]
year:1A
[/W]
[/S]
[S]
[W]

```

```

""
[/W]
[/S]
[S]
[W]
Price:1E
[/W]
[/S]
[S]
[W]
6500:1J
[/W]
[/S]
;=====
=====
;;Time 1.2      Sec
;;Done!
;===== UNL
=====
;JEEP Cherokee toutes options couleur dorée
douane payée vitres électriques en état
original pas plus de20000 dinars
[S]
mak(saloon:00, JEEP(country>Ameria):00)
mod(saloon:00,
    Cherokee(country>AmericF,make<JEEP):
05)
fea(saloon:00, full options:0E)
col(saloon:00, goldr:12)
pri(saloon:00, 20000:2Z.@less)
[/S]
;=====
=====
;;Time 0.3      Sec
;;Done!
;===== UNL
=====
;DAEWOO Nubira modèle 1997 doré toutes
options automatique douane payée récemment
6500 dinars
[S]
mak(saloon:00, DAEWOO(country>korea):00)
mod(saloon:00,
    Nubira(country>koreF,make>DAEWOO):07
)
yea(saloon:00, 1997:0M)
col(saloon:00, goldr:0S)
fea(saloon:00, full options:0X)
fea(saloon:00, automatic gear:1C)
pri(saloon:00, 6500:2D)
[/S]
;=====
=====
;;Time 0.3      Sec
;;Done!
;===== UNL
=====
;range rover 82 renouvelée 95 en bon état
80000 KM 16000 dinars
[S]
[W]
range
rover(country>England, country>europe, make>R
over):00
[/W]
[/S]
[S]
[W]
82:0C
[/W]
[/S]
[S]
[W]
year:14
[/W]
[/S]
[S]

```

```

[W]
80000:17
[/W]
[/S]
;=====
=====
;;Time 0.2      Sec
;;Done!
;===== UNL
=====
;RENAULT 5 modèle 82 assurée jusqu'à
septembre en bon état 1000 dinars tél
0796232997
[S]
[W]
RENAULT(country>France, country>europe):00
[/W]
[/S]
[S]
[W]
5:08
[/W]
[/S]
[S]
[W]
year:0L
[/W]
[/S]
[S]
[W]
"sur"
[/W]
[/S]
[S]
[W]
ée:0Q
[/W]
[/S]
[S]
[W]
year:1N
[/W]
[/S]
[S]
[W]
1000:1Q
[/W]
[/S]
;=====
=====
;;Time 0.3      Sec
;;Done!
;===== UNL
=====
;voiture GOLF abs modèle 1998 nouvelle look
contrôle complet ouverture électrique
[S]
mod(saloon:00,
    Golf(country>germany, country>europe,
make>VOLKS WAGEN):08)
yea(saloon:00, 1998:0P)
fea(saloon:00, Fatha:1Q)
[/S]
;=====
=====
;;Time 0.2      Sec
;;Done!
;===== UNL
=====
;A vendre OPEL VECTRA modèle 1990 toutes
options contrôle complet en bon état
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
    OPEL(country>germany, country>europe)
:09)
mod(saloon:00,

```



```

[S]
mak(saloon:00,
    PEUGEOT(country>France,country>europ
e):08)
mod(saloon:00, 307:0G)
[/S]
[S]
mak(saloon:00,
    PEUGEOT(country>France,country>europ
e):08)
mod(saloon:00, 307:0G)
[/S]
;=====
;Time 0.3    Sec
;;Done!
;===== UNL
;=====
;TOYOTA Celica 1981 boîte automatique 1800
CC en bon état sauf le moteur
[S]
mak(saloon:00, TOYOTA(country>japan):00)
mod(saloon:00,
    Celica(country>japan,make>TOYOTA):07
)
yea(saloon:00, 1981:0E)
fea(saloon:00, automatic gear:0P)
mot(saloon:00, 1800:11)
[/S]
;=====
;Time 0.3    Sec
;;Done!
;===== UNL
;=====
;voiture DAEWOO 93 vitres électriques
contrôle parfait couleur verte
[S]
mak(saloon:00, DAEWOO(country>korea):08)
yea(saloon:00, 93:0F)
col(saloon:00, green:1S)
[/S]
;=====
;Time 0.2    Sec
;;Done!
;===== UNL
;=====
;A vendre voiture taxi NISSAN Sunny modèle
2005 taxe libre pour contact 0795597203
[S]
sal(saloon:09, sale:00)
mak(saloon:09, NISSAN(country>japan):0M)
mod(saloon:09,
    Sunny(country>japan,make>NISSAN):0T)
yea(saloon:09, 2005:17)
yea(saloon:09, 0795597203:20)
[/S]
;=====
;Time 0.3    Sec
;;Done!
;===== UNL
;=====
;A vendre voiture MITSUBISHI LANCER 93
douane payée récemment 5900 dinars à
débattre
[S]
sal(saloon:09, sale:00)
mak(saloon:09,
    MITSUBISHI(country>japan):0H)
mod(saloon:09,
    Lancer(country>japan,make>MITSUBISHI
):0S)
yea(saloon:09, 93:0Z)
pri(saloon:09, 5900:1R)
[/S]

```

```

;=====
;=====
;Time 0.3    Sec
;;Done!
;===== UNL
;=====
;MERCEDES 99 grise 23000 dinars voiture
HONDA CIVIC 2001 marron 10500 dinars
[S]
mak(saloon:00,
    MERCEDES(country>germany,country>eur
ope):00)
mod(saloon:00, 99:09)
col(saloon:00, silver:0C)
pri(saloon:00, 23000:0I)
mak(saloon:00, HONDA(country>japan):13)
mod(saloon:00,
    Civic(country>japan,make>HONDA):19)
yea(saloon:00, 2001:1F)
pri(saloon:00, 10500:1R)
[/S]
;=====
;Time 0.3    Sec
;;Done!
;===== UNL
;=====
;A vendre MERCEDES modèle 2001 couleur
grise toutes options contrôle complet
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
    MERCEDES(country>germany,country>eur
ope):09)
mod(saloon:00, 2001:0Q)
col(saloon:00, silver:13)
fea(saloon:00, full options:19)
[/S]
;=====
;Time 0.2    Sec
;;Done!
;===== UNL
;=====
;A vendre VOLVO modèle 1980 et OPEL modèle
1992 contactez 0795902118
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
    VOLVO(country>Sweden,country>europ
e):09)
mod(saloon:00, 1980:0N)
mak(saloon:00,
    OPEL(country>germany,country>europ
e):0V)
yea(saloon:00, 1992:18)
yea(saloon:00, 0795902118:1N)
[/S]
;=====
;Time 0.1    Sec
;;Done!
;===== UNL
;=====
;voiture bmw toutes options sauf la
climatisation avec contrôle technique
modèle 1989 couleur grise jantes 17
[S]
mak(saloon:00,
    B.M.W(country>germany,country>europ
e):08)
fea(saloon:00, full options:0C)
fea(saloon:00, no air condition:0R)
col(saloon:00, silver:2N)
yea(saloon:00, 17:30)
[/S]
;=====

```

```

=====
;;Time 0.3 Sec
;;Done!
;===== UNL
=====
;A vendre voiture américaine modèle 1981 en
très bon état
[S]
sal(saloon:09, sale:00)
yea(saloon:09, 1981:11)
[/S]
;=====
=====
;;Time 0.3 Sec
;;Done!
;===== UNL
=====
;OPEL kadett modèle 86 moteur 1200CC en bon
état
[S]
mak(saloon:00,
      OPEL(country>germany,country>europe)
:00)
mod(saloon:00,
      kadet(country>germany,country>europe
,make>OPEL):05)
yea(saloon:00, 86:0K)
mot(saloon:00, 1200:0U)
[/S]
;=====
=====
;;Time 0.2 Sec
;;Done!
;===== UNL
=====
;TOYOTA Corolla modèle 99 couleur verte
toutes options sauf la boîte 8800 dinars
contrôle complet
[S]
mak(saloon:00, TOYOTA(country>japan):00)
mod(saloon:00,
      Corolla(country>japan,make>TOYOTA):0
7)
yea(saloon:00, 99:0N)
col(saloon:00, green:0Y)
fea(saloon:00, full options:14)
fea(saloon:00, manual gear:1J)
pri(saloon:00, 8800:1X)
[/S]
;=====
=====
;;Time 0.2 Sec
;;Done!
;===== UNL
=====
;voiture DAEWOO Lanos modèle 99 couleur
blanche contrôle assuré toutes options sauf
la boîte
[S]
mak(saloon:00, DAEWOO(country>korea):08)
mod(saloon:00,
      Lanos(country>koreF,make>DAEWOO):0F)
yea(saloon:00, 99:0T)
col(saloon:00, white:14)
fea(saloon:00, full options:1U)
fea(saloon:00, manual gear:29)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
;voiture DAEWOO modèle 99 contrôle complet
économique couleur jaune vitres électriques
0796864442 Kaled Achour
[S]
mak(saloon:00, DAEWOO(country>korea):08)
yea(saloon:00, 99:0N)
yea(saloon:00, 0796864442:2I)
[/S]
;=====
=====
;;Time 0.3 Sec
;;Done!
;===== UNL
=====
;A vendre voiture DAEWOO NUBIRA PREMIUM 2
modèle 2000 8000 dinars propriétaire
0795080690
[S]
sal(saloon:09, sale:00)
mak(saloon:09, DAEWOO(country>korea):0H)
mod(saloon:09,
      Nubira(country>koreF,make>DAEWOO):00
)
yea(saloon:09, 2:13)
yea(saloon:09, 2000:1D)
pri(saloon:09, 8000:1I)
yea(saloon:09, 0795080690:28)
[/S]
;=====
=====
;;Time 0.2 Sec
;;Done!
;===== UNL
=====
;A vendre KIA Shuma modèle 98 toutes
options contrôle complet à 6800 dinars
[S]
sal(saloon:00, sale:00)
mak(saloon:00, KIA(country>korea):09)
mod(saloon:00,
      Shuma(country>koreF,make>KIA):0D)
yea(saloon:00, 98:0R)
fea(saloon:00, full options:0U)
pri(saloon:00, 6800:1U)
[/S]
;=====
=====
;;Time 0.2 Sec
;;Done!
;===== UNL
=====
;A vendre KIA Sephia modèle 1994 toutes
options contrôle non assuré 3800 dinars
[S]
sal(saloon:00, sale:00)
mak(saloon:00, KIA(country>korea):09)
mod(saloon:00,
      Sephia(country>koreF,make>KIA):0D)
yea(saloon:00, 1994:0S)
fea(saloon:00, full options:0X)
pri(saloon:00, 3800:1Y)
[/S]
;=====
=====
;;Time 0.3 Sec
;;Done!
;===== UNL
=====
;A vendre HONDA Rio modèle 89 blanche
toutes options sauf la boîte 4000 dinars
[S]
sal(saloon:00, sale:00)
mak(saloon:00, HONDA(country>japan):09)
mod(saloon:00,
      Rio(country>koreF,make>KIA):0F)
yea(saloon:00, 89:0R)
col(saloon:00, white:0U)
fea(saloon:00, full options:12)
fea(saloon:00, manual gear:1H)

```

```

pri(saloon:00, 4000:1V)
[/S]
;=====
=====
;;Time 0.3 Sec
;;Done!
;===== UNL
=====
;A vendre HONDA CIVIC LXI climatisée
automatique contrôle complet modèle 96
[/S]
sal(saloon:00, sale:00)
mak(saloon:00, HONDA(country>japan):09)
mod(saloon:00, Civic
Exi(country>japan,make>HONDA):0F)
fea(saloon:00, automatic gear:11)
yea(saloon:00, 96:23)
[/S]
;=====
=====
;;Time 0.3 Sec
;;Done!
;===== UNL
=====
;A vendre HONDA CIVIC LXI climatisée
automatique contrôle complet modèle 96
[/S]
sal(saloon:00, sale:00)
mak(saloon:00, HONDA(country>japan):09)
mod(saloon:00, Civic
Exi(country>japan,make>HONDA):0F)
fea(saloon:00, automatic gear:11)
yea(saloon:00, 96:23)
[/S]
;=====
=====
;;Time 0.3 Sec
;;Done!
;===== UNL
=====
;MERCEDES modèle 99 classique 200 toutes
options sauf l'ouverture à 22000 dinars ou
acompte et tranches bancaires
[/S]
mak(saloon:00,
MERCEDES(country>germany, country>eur
ope):00)
mod(saloon:00, 99:0H)
yea(saloon:00, 200:0U)
fea(saloon:00, full options:0Y)
fea(saloon:00, Fatha:1K)
pri(saloon:00, 22000:1X)
[/S]
;=====
=====
;;Time 0.3 Sec
;;Done!
;===== UNL
=====
;MERCEDES C180 modèle 1996 grise toutes
options contrôle complet 16000 dinars
[/S]
mak(saloon:00,
MERCEDES(country>germany, country>eur
ope):00)
mod(saloon:00, 180:0A)
[/S]
[/S]
mak(saloon:00,
MERCEDES(country>germany, country>eur
ope):00)

```

```

mod(saloon:00, 180:0A)
[/S]
[/S]
mak(saloon:00,
MERCEDES(country>germany, country>eur
ope):00)
mod(saloon:00, 180:0A)
[/S]
[/S]
mak(saloon:00,
MERCEDES(country>germany, country>eur
ope):00)
mod(saloon:00, 180:0A)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
;A vendre voiture taxi HONDA modèle 2001
taxe libre 23000 dinars
[/S]
sal(saloon:09, sale:00)
mak(saloon:09, HONDA(country>japan):0M)
yea(saloon:09, 2001:10)
pri(saloon:09, 23000:1G)
[/S]
;=====
=====
;;Time 0.2 Sec
;;Done!
;===== UNL
=====
;A vendre voiture HONDA CIVIC modèle 96
injection moteur 1005 cc assurée et
contrôle complet
[/S]
sal(saloon:09, sale:00)
mak(saloon:09, HONDA(country>japan):0H)
mod(saloon:09,
Civic(country>japan,make>HONDA):0N)
mot(saloon:09, 1005:1L)
[/S]
;=====
=====
;;Time 0.3 Sec
;;Done!
;===== UNL
=====
;A vendre HONDA CIVIC modèle 2000 toutes
options vitres manuelles couleur noire
0795066927
[/S]
sal(saloon:00, sale:00)
mak(saloon:00, HONDA(country>japan):09)
mod(saloon:00,
Civic(country>japan,make>HONDA):0F)
yea(saloon:00, 2000:0T)
fea(saloon:00, full options:0Y)
fea(saloon:00, manual gear:1K)
col(saloon:00, black:22)
yea(saloon:00, 0795066927:28)
[/S]
;=====
=====
;;Time 0.3 Sec
;;Done!
;===== UNL
=====
;A vendre HONDA SONATA modèle 97 toutes
options couleur jaune
[/S]
sal(saloon:00, sale:00)
mak(saloon:00, HONDA(country>japan):09)
mod(saloon:00,
Sonata(country>koreF,make>HYUNDAI):0

```

```

F)
yea(saloon:00, 97:0U)
fea(saloon:00, full options:0X)
[/S]
;=====
=====
;;Time 0.2      Sec
;;Done!
;===== UNL
=====
;A vendre HONDA modèle 1998 toutes options
sauf boîte contrôle complet, verte, 5500
dinars
[S]
sal(saloon:00, sale:00)
mak(saloon:00, HONDA(country>japan):09)
yea(saloon:00, 1998:0N)
fea(saloon:00, full options:0S)
col(saloon:00, green:21)
pri(saloon:00, 5500:28)
[/S]
;=====
=====
;;Time 0.3      Sec
;;Done!
;===== UNL
=====
;A vendre MERCEDES modèle 96 rouge CT OK
13000 dinars première main tél 0795535225
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
      MERCEDES(country>germany,country>eur
ope):09)
mod(saloon:00, 96:0Q)
col(saloon:00, red:0T)
fea(saloon:00, checked:0Z)
pri(saloon:00, 13000:15)
yea(saloon:00, 0795535225:22)
[/S]
;=====
=====
;;Time 0.3      Sec
;;Done!
;===== UNL
=====
;A vendre VOLKSWAGEN BEETLE modèle 1999
toutes options sauf la boîte rouge douane
payée récemment 10500 dinars
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
      VOLKSWAGEN(country>germany,country>e
urope):09)
mod(saloon:00,
      Beetle(country>germany,country>europ
e,make>VOLKS WAGEN):0K)
yea(saloon:00, 1999:10)
fea(saloon:00, full options:15)
fea(saloon:00, manual gear:1K)
col(saloon:00, red:1Y)
pri(saloon:00, 10500:2T)
[/S]
;=====
=====
;;Time 0.4      Sec
;;Done!
;===== UNL
=====
;A vendre KIA sport 4*4 m 2003 toutes
options sauf la boîte douane non payée 7000
dinars couleur marron
[S]
sal(saloon:00, sale:00)
mak(saloon:00, KIA(country>korea):09)
col(saloon:00, goldr:0F)

yea(saloon:00, 4:0J)
yea(saloon:00, 4:0L)
yea(saloon:00, 2003:0P)
fea(saloon:00, full options:0U)
fea(saloon:00, manual gear:19)
pri(saloon:00, 7000:25)
[/S]
;=====
=====
;;Time 0.3      Sec
;;Done!
;===== UNL
=====
;A vendre KIA Sportage JEEP 4*4 m 2004
toutes options sauf la boîte couleur noire
11800 dinars assurance complète
[S]
sal(saloon:00, sale:00)
mak(saloon:00, KIA(country>korea):09)
mod(saloon:00,
      Sportage(country>koreF,make>KIA):0D)
mak(saloon:00, JEEP(country>Ameria):0M)
yea(saloon:00, 4:0R)
yea(saloon:00, 4:0T)
yea(saloon:00, 2004:0X)
fea(saloon:00, full options:12)
fea(saloon:00, manual gear:1H)
col(saloon:00, black:23)
pri(saloon:00, 11800:29)
[/S]
;=====
=====
;;Time 0.3      Sec
;;Done!
;===== UNL
=====
;A vendre MITSUBISHI LANCER modèle 93
automatique climatisée en bon état le prix
4000 d
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
      MITSUBISHI(country>japan):09)
mod(saloon:00,
      Lancer(country>japan,make>MITSUBISHI
):0K)
yea(saloon:00, 93:0Z)
fea(saloon:00, automatic gear:12)
pri(saloon:00, 4000:2B)
[/S]
;=====
=====
;;Time 0.3      Sec
;;Done!
;===== UNL
=====
;A vendre voiture KIA CIVIC modèle 94
toutes options sauf la boîte contrôle
complet
[S]
sal(saloon:09, sale:00)
mak(saloon:09, KIA(country>korea):0H)
mod(saloon:09,
      Civic(country>japan,make>HONDA):0L)
yea(saloon:09, 94:0Z)
fea(saloon:09, full options:12)
fea(saloon:09, manual gear:1H)
[/S]
;=====
=====
;;Time 0.2      Sec
;;Done!
;===== UNL
=====
;A vendre bus KIA BASTA commun blanc modèle
95 5300 dinars à débattre

```

```

[S]
sal(bus:09, sale:00)
mak(bus:09, KIA(country>korea):0D)
col(bus:09, white:0U)
yea(bus:09, 95:18)
pri(bus:09, 5300:1B)
[/S]
;=====
;=====
;;Time 0.3 Sec
;;Done!
;===== UNL
;=====
;A vendre GrandCherokee modèle 1994 couleur
noire 6 cylindres à traction arrière toutes
options
[S]
sal(saloon:00, sale:00)
mod(saloon:00, Grand
Cherokee(country>AmericF,make<JEEP):09)
yea(saloon:00, 1994:0V)
col(saloon:00, black:18)
yea(saloon:00, 6:1E)
fea(saloon:00, full options:2B)
[/S]
;=====
;=====
;;Time 0.3 Sec
;;Done!
;===== UNL
;=====
;A vendre GOLF 1600CC distribution
électrique ABS noire toutes options
[S]
sal(saloon:00, sale:00)
mod(saloon:00,
Golf(country>germany,country>europe,
make>VOLKS WAGEN):09)
mot(saloon:00, 1600:0E)
[/S]
[S]
sal(saloon:00, sale:00)
mod(saloon:00,
Golf(country>germany,country>europe,
make>VOLKS WAGEN):09)
mot(saloon:00, 1600:0E)
[/S]
[S]
sal(saloon:00, sale:00)
mod(saloon:00,
Golf(country>germany,country>europe,
make>VOLKS WAGEN):09)
mot(saloon:00, 1600:0E)
[/S]
[S]
sal(saloon:00, sale:00)
mod(saloon:00,
Golf(country>germany,country>europe,
make>VOLKS WAGEN):09)
mot(saloon:00, 1600:0E)
[/S]
[S]
sal(saloon:00, sale:00)
mod(saloon:00,
Golf(country>germany,country>europe,
make>VOLKS WAGEN):09)
mot(saloon:00, 1600:0E)
[/S]
;=====
;=====

```

```

;;Time 0.3 Sec
;;Done!
;===== UNL
;=====
;A vendre GOLF GTI modèle 89 noire jantes
15 contrôle original avec ouverture système
5
[S]
sal(saloon:00, sale:00)
mod(saloon:00,
Golf(country>germany,country>europe,
make>VOLKS WAGEN):09)
yea(saloon:00, 89:0Q)
col(saloon:00, black:0T)
yea(saloon:00, 15:16)
fea(saloon:00, Fatha:1S)
yea(saloon:00, 5:2G)
[/S]
;=====
;=====
;;Time 0.3 Sec
;;Done!
;===== UNL
;=====
;A vendre bmw modèle 94 toutes options avec
cuir sauf la boîte à 11000 dinars
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
B.M.W(country>germany,country>europe
):09)
mod(saloon:00, 94:0L)
fea(saloon:00, full options:00)
fea(saloon:00, manual gear:1D)
pri(saloon:00, 11000:1U)
[/S]
;=====
;=====
;;Time 0.2 Sec
;;Done!
;===== UNL
;=====
;A vendre AUDI allemande modèle 78 moteur
et boîte neufs à traction avant 800 dinars
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
AUDI(country>germany,country>europe)
:09)
cou(saloon:00, germany:0E)
yea(saloon:00, 78:0W)
fea(saloon:00, new:1F)
pri(saloon:00, 800:23)
[/S]
;=====
;=====
;;Time 0.4 Sec
;;Done!
;===== UNL
;=====
;A vendre OPEL VECTRA modèle 97 toutes
options sauf la boîte et la climatisation
9750 dinars
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
OPEL(country>germany,country>europe)
:09)
mod(saloon:00,
Vectra(country>germany,country>europ
e,make>OPEL):0E)
yea(saloon:00, 97:0T)
fea(saloon:00, full options:0W)
fea(saloon:00, manual gear:1B)
pri(saloon:00, 9750:29)
[/S]
;=====
;=====

```

```

=====
;;Time 0.3    Sec
;;Done!
;===== UNL
=====
;A vendre BM 520 modèle 93 sans contrôle
toutes options sauf la boîte excellente
10500 dinars
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
    B.M.W(country>germany, country>europ
):09)
mod(saloon:00, 520:0C)
yea(saloon:00, 93:00)
fea(saloon:00, full options:16)
fea(saloon:00, manual gear:1L)
pri(saloon:00, 10500:2A)
[/S]
;=====
=====
;;Time 0.3    Sec
;;Done!
;===== UNL
=====
;A vendre BMW 520 modèle 2000 couleur dorée
avec cuir blanc 0795610755
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
    B.M.W(country>germany, country>europ
):09)
mod(saloon:00, 520:0D)
yea(saloon:00, 2000:0P)
col(saloon:00, goldr:13)
col(saloon:00, white:1J)
yea(saloon:00, 0795610755:1P)
[/S]
;=====
=====
;;Time 0.2    Sec
;;Done!
;===== UNL
=====
;A vendre BMW modèle 95 contrôle complet
options complètes avec des options
extraordinaires
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
    B.M.W(country>germany, country>europ
):09)
mod(saloon:00, 95:0L)
col(saloon:00, goldr:2B)
[/S]
;=====
=====
;;Time 0.3    Sec
;;Done!
;===== UNL
=====
;A vendre voiture PEUGEOT 307 modèle 2005
6000 KM compteur 1,4l toutes options sauf
la boîte 0795321200
[S]
sal(saloon:09, sale:00)
mak(saloon:09,
    PEUGEOT(country>France, country>europ
e):0H)
mod(saloon:09, 307:0P)
yea(saloon:09, 2005:11)
yea(saloon:09, 6000:16)
yea(saloon:09, 1:1N)
fea(saloon:09, full options:1S)
fea(saloon:09, manual gear:27)
yea(saloon:09, 0795321200:2L)

[/S]
;=====
=====
;;Time 0.3    Sec
;;Done!
;===== UNL
=====
;A vendre TOYOTA Corolla modèle 1982 boîte
automatique en bon état batterie neuve et
fauteuils neufs
[S]
sal(saloon:00, sale:00)
mak(saloon:00, TOYOTA(country>japan):09)
mod(saloon:00,
    Corolla(country>japan, make>TOYOTA):0
G)
yea(saloon:00, 1982:0W)
fea(saloon:00, automatic gear:17)
fea(saloon:00, new:25)
fea(saloon:00, new:20)
[/S]
;=====
=====
;;Time 0.4    Sec
;;Done!
;===== UNL
=====
;A vendre GOLF modèle 1998 comme neuve
contrôle complet ouverture électrique
[S]
sal(saloon:00, sale:00)
mod(saloon:00,
    Golf(country>germany, country>europ
e,
make>VOLKS WAGEN):09)
yea(saloon:00, 1998:0M)
fea(saloon:00, new:0X)
fea(saloon:00, Fatha:1L)
[/S]
;=====
=====
;;Time 0.2    Sec
;;Done!
;===== UNL
=====
;A vendre voiture GOLF modèle 92 noire
système d'alarme vitre électrique
[S]
sal(saloon:09, sale:00)
mod(saloon:09,
    Golf(country>germany, country>europ
e,
make>VOLKS WAGEN):0H)
yea(saloon:09, 92:0U)
col(saloon:09, black:0X)
[/S]
;=====
=====
;;Time 0.2    Sec
;;Done!
;===== UNL
=====
;A vendre voiture DAEWOO modèle 97 noire
très économique première main assurée
jusqu'en mai
[S]
sal(saloon:09, sale:00)
mak(saloon:09, DAEWOO(country>korea):0H)
yea(saloon:09, 97:0W)
col(saloon:09, black:0Z)
[/S]
;=====
=====
;;Time 0.4    Sec
;;Done!
;===== UNL
=====
;A vendre voiture DAEWOO espero modèle 92

```

```

toutes options sauf la boite en bon état
3800 dinars
[S]
sal(saloon:09, sale:00)
mak(saloon:09, DAEWOO(country>korea):0H)
mod(saloon:09,
    espero(country>koreF,make>DAEWOO):00
)
yea(saloon:09, 92:13)
fea(saloon:09, full options:16)
fea(saloon:09, manual gear:1L)
pri(saloon:09, 3800:2C)
[/S]
;=====
=====
;;Time 0.4 Sec
;;Done!
;===== UNL
=====
;A vendre SUBARU modèle 1983 toutes
options sauf la boite 0795131719
[S]
sal(saloon:00, sale:00)
mak(saloon:00, SUBARU(country>japan):0A)
yea(saloon:00, 1983:0P)
fea(saloon:00, full options:0V)
fea(saloon:00, manual gear:1A)
yea(saloon:00, 0795131719:10)
[/S]
;=====
=====
;;Time 0.2 Sec
;;Done!
;===== UNL
=====
;A vendre SUZUKI modèle 97 en bon état très
économique 2700 dinars à débattre
[S]
sal(saloon:00, sale:00)
mak(saloon:00, SUZUKI(country>japan):09)
yea(saloon:00, 97:00)
pri(saloon:00, 2700:1M)
[/S]
;=====
=====
;;Time 0.4 Sec
;;Done!
;===== UNL
=====
;A vendre VOLKSWAGEN GOLF modèle 94 toutes
options sauf la boite et la climatisation
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
    VOLKSWAGEN(country>germany,country>e
urope):09)
mod(saloon:00,
    Golf(country>germany,country>europe,
make>VOLKS WAGEN):0K)
yea(saloon:00, 94:0X)
fea(saloon:00, full options:10)
fea(saloon:00, manual gear:1F)
[/S]
;=====
=====
;;Time 0.3 Sec
;;Done!
;===== UNL
=====
;A vendre VOLKSWAGEN POLO m 2003 grise avec
climatisation contrôle complet prix
définitif 10000 dinars
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
    VOLKSWAGEN(country>germany,country>e
urope):09)

```

```

mod(saloon:00, 2003:0R)
col(saloon:00, silver:0W)
fea(saloon:00, air condition:12)
pri(saloon:00, 10000:2J)
[/S]
;=====
=====
;;Time 0.3 Sec
;;Done!
;===== UNL
=====
;A vendre MAZDA sport modèle 85 toutes
options sauf la boite 2400 dinars
[S]
sal(saloon:00, sale:00)
mak(saloon:00, MAZDA(country>japan):09)
col(saloon:00, goldr:0H)
mod(saloon:00, 85:0T)
fea(saloon:00, full options:0W)
fea(saloon:00, manual gear:1B)
pri(saloon:00, 2400:1P)
[/S]
;=====
=====
;;Time 0.3 Sec
;;Done!
;===== UNL
=====
;A vendre MITSUBISHI LANCER modèle 90
contrôle OK ouverture et rétro électriques
prix demandé 4500 dinars
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
    MITSUBISHI(country>japan):09)
mod(saloon:00,
    Lancer(country>japan,make>MITSUBISHI
):0K)
yea(saloon:00, 90:0Z)
fea(saloon:00, Fatha:1F)
pri(saloon:00, 4500:2Q)
[/S]
;=====
=====
;;Time 0.3 Sec
;;Done!
;===== UNL
=====
;A vendre MERCEDES C180 modèle 95 noire
automatique contrôle complet
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
    MERCEDES(country>germany,country>eur
ope):09)
mod(saloon:00, 180:0K)
yea(saloon:00, 95:0W)
col(saloon:00, black:0Z)
fea(saloon:00, automatic gear:15)
[/S]
;=====
=====
;;Time 0.3 Sec
;;Done!
;===== UNL
=====
;A vendre MITSUBISHI coupé modèle 99
ouverture automatique 1500 dinars
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
    MITSUBISHI(country>japan):09)
yea(saloon:00, 99:0Z)
fea(saloon:00, Fatha:12)
fea(saloon:00, automatic gear:1C)
pri(saloon:00, 1500:10)
[/S]

```

```

;=====
;Time 0.2      Sec
;;Done!
;===== UNL
;=====
;A vendre voiture DAEWOO mod 94 toutes
options sauf la boite, couleur marron
[S]
sal(saloon:09, sale:00)
mak(saloon:09, DAEWOO(country>korea):0H)
yea(saloon:09, 94:0S)
fea(saloon:09, full options:0V)
fea(saloon:09, manual gear:1A)
[/S]
;=====
;Time 0.2      Sec
;;Done!
;===== UNL
;=====
;A vendre voiture de sport SUBARU modèle
2001 double turbo couleur très spéciale
[S]
sal(saloon:09, sale:00)
col(saloon:09, goldr:0M)
mak(saloon:09, SUBARU(country>japan):0Q)
yea(saloon:09, 2001:16)
[/S]
;=====
;Time 0.2      Sec
;;Done!
;===== UNL
;=====
;A vendre voiture samsung modèle 99 du
concessionnaire en très bon état toutes
options
[S]
sal(saloon:09, sale:00)
mak(saloon:09, samsung(country>korea):0I)
yea(saloon:09, 99:0Y)
fea(saloon:09, full options:22)
[/S]
;=====
;Time 0.3      Sec
;;Done!
;===== UNL
;=====
;A vendre VOLVO 240 avec climatisation
boite automatique 3000 dinars assurance
complète contrôle m 1983
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
VOLVO(country>Sweden, country>europe)
:09)
mod(saloon:00, 240:0F)
fea(saloon:00, air condition:0J)
fea(saloon:00, automatic gear:18)
pri(saloon:00, 3000:1K)
yea(saloon:00, 1983:2S)
[/S]
;=====
;Time 0.4      Sec
;;Done!
;===== UNL
;=====
;A vendre voiture cresseda 1900 CC modèle
86 ou échange avec MERCEDES blanche m 82
[S]
sal(saloon:09, sale:00)
mak(saloon:09, cresseda(country>japan):0H)
mot(saloon:09, 1900:0Q)

yea(saloon:09, 86:16)
mak(saloon:09,
MERCEDES(country>germany, country>eur
ope):1Q)
col(saloon:09, white:1Z)
mod(saloon:09, 82:29)
[/S]
;=====
;Time 0.2      Sec
;;Done!
;===== UNL
;=====
;A vendre MITSUBISHI LANCER 89 bleu ciel
contrôle complet douane payée récemment
prix définitif 4800 dinars
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
MITSUBISHI(country>japan):09)
mod(saloon:00,
Lancer(country>japan, make>MITSUBISHI
):0K)
yea(saloon:00, 89:0R)
col(saloon:00, blue:0U)
pri(saloon:00, 4800:2R)
[/S]
;=====
;Time 0.4      Sec
;;Done!
;===== UNL
;=====
;A vendre voiture MITSUBISHI LANCER modèle
82 en bon état prix définitif 3000 dinars
[S]
sal(saloon:09, sale:00)
mak(saloon:09,
MITSUBISHI(country>japan):0H)
mod(saloon:09,
Lancer(country>japan, make>MITSUBISHI
):0S)
yea(saloon:09, 82:17)
pri(saloon:09, 3000:23)
[/S]
;=====
;Time 0.4      Sec
;;Done!
;===== UNL
;=====
;A vendre MERCEDES modèle 84 toutes options
sauf la boite avec climatisation et
ouverture centrale 5350 dinars
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
MERCEDES(country>germany, country>eur
ope):09)
mod(saloon:00, 84:0Q)
fea(saloon:00, full options:0T)
fea(saloon:00, manual gear:18)
fea(saloon:00, air condition:1M)
fea(saloon:00, Fatha:28)
pri(saloon:00, 5350:2R)
[/S]
;=====
;Time 0.3      Sec
;;Done!
;===== UNL
;=====
;A vendre voiture MERCEDES C180 modèle 98
toutes options sauf l'ouverture, couleur
grise 18500 dinars
[S]

```



```

sal(saloon:09, sale:00)
mak(saloon:09,
  MERCEDES(country>germany, country>eur
ope):0H)
mod(saloon:09, 180:0R)
yea(saloon:09, 98:13)
fea(saloon:09, full options:16)
fea(saloon:09, Fatha:1S)
col(saloon:09, silver:2B)
pri(saloon:09, 18500:2H)
[/S]
;=====
;Time 0.3 Sec
;Done!
;===== UNL
;A vendre KIA CIVIC toutes options boite
automatique
[S]
sal(saloon:00, sale:00)
mak(saloon:00, KIA(country>korea):09)
mod(saloon:00,
  Civic(country>japan, make>HONDA):0D)
fea(saloon:00, full options:0J)
fea(saloon:00, automatic gear:14)
[/S]
;=====
;Time 0.1 Sec
;Done!
;===== UNL
;A vendre NISSAN MAXIMA mod 97 toutes
options cuir ouverture 11200 dinars
[S]
sal(saloon:00, sale:00)
mak(saloon:00, NISSAN(country>japan):09)
mod(saloon:00,
  Maxima(country>japan, make>NISSAN):0G
)
yea(saloon:00, 97:0R)
fea(saloon:00, full options:0U)
fea(saloon:00, Fatha:1E)
pri(saloon:00, 11200:10)
[/S]
;=====
;Time 0.2 Sec
;Done!
;===== UNL
;A vendre voiture HONDA CIVIC mod 1997
boite automatique avec climatisation, 7750
dinars
[S]
sal(saloon:09, sale:00)
mak(saloon:09, HONDA(country>japan):0H)
mod(saloon:09,
  Civic(country>japan, make>HONDA):0N)
yea(saloon:09, 1997:0X)
fea(saloon:09, automatic gear:18)
fea(saloon:09, air condition:1K)
pri(saloon:09, 7750:24)
[/S]
;=====
;Time 0.1 Sec
;Done!
;===== UNL
;A vendre HYUNDAI Accent mod 1995 toutes
options sauf la boite 5000 dinars
[S]
sal(saloon:00, sale:00)
mak(saloon:00, HYUNDAI(country>korea):09)
mod(saloon:00,

```

```

Accent(country>korea, make>HYUNDAI):0
H)
yea(saloon:00, 1995:0S)
fea(saloon:00, full options:0X)
fea(saloon:00, manual gear:1C)
pri(saloon:00, 5000:1Q)
[/S]
;=====
;Time 0.2 Sec
;Done!
;===== UNL
;A vendre HYUNDAI mod 2001 toutes options
sauf la boite, 8900 dinars
[S]
sal(saloon:00, sale:00)
mak(saloon:00, HYUNDAI(country>korea):09)
yea(saloon:00, 2001:0L)
fea(saloon:00, full options:0Q)
fea(saloon:00, manual gear:15)
pri(saloon:00, 8900:1K)
[/S]
;=====
;Time 0.3 Sec
;Done!
;===== UNL
;A vendre BMW 520 mod 1992 grise toutes
options en bon état, 14000 dinars
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
  B.M.W(country>germany, country>europe
):09)
mod(saloon:00, 520:0D)
yea(saloon:00, 1992:0L)
col(saloon:00, silver:0Q)
fea(saloon:00, full options:0W)
pri(saloon:00, 14000:1P)
[/S]
;=====
;Time 0.3 Sec
;Done!
;===== UNL
;A vendre FORD GT neuve mod 2005 moteur
4600 couleur rouge toutes options
[S]
sal(saloon:00, sale:00)
mak(saloon:00, FORD(country>America):09)
fea(saloon:00, new:0H)
yea(saloon:00, 2005:0R)
mot(saloon:00, 4600:14)
col(saloon:00, red:1H)
fea(saloon:00, full options:1N)
[/S]
;=====
;Time 0.2 Sec
;Done!
;===== UNL
;A vendre Caprice mod 86 modernisée 90
jantes aluminum nouvelle peinture prix
définitif 4600 dinars
[S]
sal(saloon:00, sale:00)
mod(saloon:00, Caprice
Classic(country>America, make>CHEVROLET):09)
yea(saloon:00, 86:0L)
yea(saloon:00, 90:10)
pri(saloon:00, 4600:2H)
[/S]
;=====

```

```

=====
;;Time 0.4 Sec
;;Done!
;===== UNL
=====
;A vendre KIA Sephia mod 96 verte contrôle
complet toutes options prix définitif 5250
tél 95918898
[S]
sal(saloon:00, sale:00)
mak(saloon:00, KIA(country>korea):09)
mod(saloon:00,
    Sephia(country>koreF,make>KIA):0D)
yea(saloon:00, 96:00)
col(saloon:00, green:0R)
fea(saloon:00, full options:1F)
pri(saloon:00, 5250:2A)
pri(saloon:00, 95918898:2K)
[/S]
;=====
;;Time 0.3 Sec
;;Done!
;===== UNL
=====
;A vendre KIA Sephia mod 96 toutes options
avec boîte automatique couleur verte
contrôle complet 5800 dinars
[S]
sal(saloon:00, sale:00)
mak(saloon:00, KIA(country>korea):09)
mod(saloon:00,
    Sephia(country>koreF,make>KIA):0D)
yea(saloon:00, 96:00)
fea(saloon:00, full options:0R)
fea(saloon:00, automatic gear:1H)
col(saloon:00, green:2I)
pri(saloon:00, 5800:2P)
[/S]
;=====
;;Time 0.3 Sec
;;Done!
;===== UNL
=====
;A vendre LAND ROVER DISCOVERY mod 2001
toutes options 8 cylindres état original
tél 0796635036
[S]
sal(saloon:00, sale:00)
mod(saloon:00, Land
Rover(country>England, country>europe, make>R
over):09)
mod(saloon:00,
    Discovery(country>England, country>eu
rope, make>Rover):0K)
yea(saloon:00, 2001:0Y)
fea(saloon:00, full options:13)
yea(saloon:00, 8:1I)
[/S]
;=====
;;Time 0.3 Sec
;;Done!
;===== UNL
=====
;A vendre MERCEDES mod 76 modernisée 84
avec ouverture climatisation automatique
cuir couleur grise
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
    MERCEDES(country>germany, country>eur
ope):09)
mod(saloon:00, 76:0M)
yea(saloon:00, 84:1I)

fea(saloon:00, Fatha:14)
fea(saloon:00, automatic gear:1X)
col(saloon:00, silver:2N)
[/S]
;=====
;;Time 0.5 Sec
;;Done!
;===== UNL
=====
;A vendre HONDA Accord mod 91 couleur
blanche avec climatisation automatique
contrôle complet
[S]
sal(saloon:00, sale:00)
mak(saloon:00, HONDA(country>japan):09)
mod(saloon:00,
    Accord(country>japan, make>HONDA):0F)
yea(saloon:00, 91:0Q)
col(saloon:00, white:1I)
fea(saloon:00, air condition:19)
fea(saloon:00, automatic gear:1S)
[/S]
;=====
;;Time 0.4 Sec
;;Done!
;===== UNL
=====
;A vendre MERCEDES m 1989 contrôle complet
toutes options sauf l'ouverture 0795517487
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
    MERCEDES(country>germany, country>eur
ope):09)
mod(saloon:00, 1989:0K)
fea(saloon:00, full options:17)
fea(saloon:00, Fatha:1T)
yea(saloon:00, 0795517487:23)
[/S]
;=====
;;Time 0.3 Sec
;;Done!
;===== UNL
=====
;A vendre LADA SAMARA modèle 88 couleur
marron 900 dinars
[S]
sal(saloon:00, sale:00)
mak(saloon:00, Lada:09)
mod(saloon:00, SAMARA(make>Lada):0E)
yea(saloon:00, 88:0T)
pri(saloon:00, 900:1B)
[/S]
;=====
;;Time 0.2 Sec
;;Done!
;===== UNL
=====
;A vendre MAZDA noire contrôlée et assurée
prix définitif 2500 dinars son état est
très bonne
[S]
sal(saloon:00, sale:00)
mak(saloon:00, MAZDA(country>japan):09)
col(saloon:00, black:0F)
pri(saloon:00, 2500:1P)
[/S]
;=====
;;Time 0.5 Sec
;;Done!
;===== UNL

```

```
=====
;A vendre DAEWOO NUBIRA mod 2000 prix
intéressant +0795200986 MITSUBISHI LANCER
2003
[S]
sal(saloon:00, sale:00)
mak(saloon:00, DAEWOO(country>korea):09)
mod(saloon:00,
      Nubira(country>koreF,make>DAEWOO):0G
)
pri(saloon:00, 2000:0R)
yea(saloon:00, 0795200986:1F)
mak(saloon:00,
      MITSUBISHI(country>japan):1Q)
mod(saloon:00,
      Lancer(country>japan,make>MITSUBISHI
):21)
yea(saloon:00, 2003:28)
[/S]
=====
```

```
=====
;;Time 0.3 Sec
;;Done!
;===== UNL
=====
;A vendre pickup NISSAN diesel modèle 1985
assuré jusqu'au 8/12/2005 contrôle complet.
[S]
sal(pickup:09, sale:00)
mak(pickup:09, NISSAN(country>japan):0G)
yea(pickup:09, 1985:12)
yea(pickup:09, 8:10)
yea(pickup:09, 12:1Q)
yea(pickup:09, 2005:1T)
[/S]
;=====
=====
;;Time 0.3 Sec
;;Done!
```



```

vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(bus, wanted)
mak(bus, KIA(country>korea))
[/S]
;=====
;===== CRLcats
;=====
; recherche bus transport commun KIA ou
HONDA ou SINGE YOUNG avec acompte trois
mille et le reste par tranches
; {a:give-information+concept(action=e_buy,
vehicle-spec=(bus, vehicle-make=KIA,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(bus, wanted)
mak(bus, KIA(country>korea))
[/S]
;=====
;===== CRLcats
;=====
; recherche JEEP Cherokee 2000 1900 dinars
; {a:accept+concept(action=e_buy, vehicle-
spec=(, vehicle-make=JEEP, vehicle-model=,
vehicle-age=, vehicle-price=, vehicle-
color=, vehicle-condition=, vehicle-
assurance=, vehicle-controle=, vehicle-air-
condition=, vehicle-size=, vehicle-shape=,
vehicle-motor-type=, vehicle-hand=,
vehicle-nationality=, vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, JEEP(country>Ameria))
[/S]
;=====
;===== CRLcats
;=====
; MERCEDES classique E200 toutes options
; {a:give-information+concept(action=,
vehicle-spec=(, vehicle-make=MERCEDES,
vehicle-model=E2060, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
mak(saloon,
MERCEDES(country>germany,country>eur
ope))
[/S]
;=====
;===== CRLcats
;=====
; je cherche une voiture moteur 1300 cc
modèle 85 2500 dinars
; {a:give-information+action+vehicule(who=i,
action=e-find-3, action=, vehicle-
spec=(car, vehicle-make=, vehicle-model=,
vehicle-age=, vehicle-price=, vehicle-

```

```

color=, vehicle-condition=, vehicle-
assurance=, vehicle-controle=, vehicle-air-
condition=, vehicle-size=, vehicle-shape=,
vehicle-motor-type=1300, vehicle-hand=,
vehicle-nationality=, vehicle-mileage=),
time=(md=85))}
[S]
mot(saloon, 1300)
[/S]
;=====
;===== CRLcats
;=====
; recherche voiture américaine ^ condition
que le prix ne dépasse pas 1000 dinars
; {a:give-
information+price+concept(action=e_buy,
vehicle-spec=(car, vehicle-make=, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
[/S]
;=====
;===== CRLcats
;=====
; recherche DAEWOO
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=DAEWOO,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, DAEWOO(country>korea))
[/S]
;=====
;===== CRLcats
;=====
; recherche NISSAN Sunny manuelle modèle 93
^ 97
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=NISSAN,
vehicle-model=Sunny, vehicle-age=97,
vehicle-price=, vehicle-color=, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=93))}
[S]
wan(saloon, wanted)
mak(saloon, NISSAN(country>japan))
mod(saloon,
Sunny(country>japan,make>NISSAN))
yea(saloon, 97)
[/S]
;=====
;===== CRLcats
;=====
; recherche RENAULT Clio manuelle modèle 95
^ 2000
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=RENAULT,
vehicle-model=Clio, vehicle-age=2000,

```

```

vehicle-price=, vehicle-color=, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=95))}
[S]
wan(saloon, wanted)
mak(saloon,
    RENAULT(country>France, country>europ
e))
mod(saloon,
    Clio(country>France, country>europe, m
ake>RENAULT))
yea(saloon, 2000)
[/S]
;=====
;===== CRLcats
=====
; recherche voiture HYUNDAI
; {a:give-information+concept(action=e_buy,
vehicle-spec=(car, vehicle-make=HYUNDAI,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, HYUNDAI(country>korea))
[/S]
;=====
;===== CRLcats
=====
; recherche voiture MITSUBISHI
; {a:give-information+concept(action=e_buy,
vehicle-spec=(car, vehicle-make=ITSUBISHI,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, MITSUBISHI(country>japan))
[/S]
;=====
;===== CRLcats
=====
; je cherche une voiture BMW modèle 92
; {a:give-information+action+vehicle(who=i,
action=e-find-3, action=, vehicle-
spec=(car, vehicle-make=BMW, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=92))}
[S]
mak(saloon,
    B.M.W(country>germany, country>europe
))
[/S]
;=====
;===== CRLcats
=====
; recherche DAEWOO
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=DAEWOO,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, DAEWOO(country>korea))
[/S]
;=====
;===== CRLcats
=====
; recherche NISSAN Sunny modèle 93 ^ 95
manuelle
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=NISSAN,
vehicle-model=Sunny, vehicle-age=95,
vehicle-price=, vehicle-color=, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=93))}
[S]
wan(saloon, wanted)
mak(saloon, NISSAN(country>japan))
mod(saloon,
    Sunny(country>japan, make>NISSAN))
yea(saloon, 95)
[/S]
;=====
;===== CRLcats
=====
; recherche HONDA Civic modèle 94 toutes
options manuelle
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=HONDA,
vehicle-model=Civic, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=94))}
[S]
wan(saloon, wanted)
mak(saloon, HONDA(country>japan))
mod(saloon,
    Civic(country>japan, make>HONDA))
[/S]
;=====
;===== CRLcats
=====
; recherche KIA Sephia
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=KIA, vehicle-
model=Sephia, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, KIA(country>korea))
mod(saloon,
    Sephia(country>koreF, make>KIA))

```

```

[/S]
;=====
=====
;===== CRLcats
=====
; recherche DAEWOO Lanos 95 ^ 97 manuelle
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=DAEWOO,
vehicle-model=Lanos, vehicle-age=97,
vehicle-price=, vehicle-color=, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, DAEWOO(country>korea))
mod(saloon,
Lanos(country>koreF,make>DAEWOO))
yea(saloon, 97)
[/S]
;=====
=====
;===== CRLcats
=====
; recherche Lada SAMARA
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=Lada, vehicle-
model=SAMARA, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, Lada)
mod(saloon, SAMARA(make>Lada))
[/S]
;=====
=====
;===== CRLcats
=====
; recherche voiture HYUNDAI
; {a:give-information+concept(action=e_buy,
vehicle-spec=(car, vehicle-make=HYUNDAI,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, HYUNDAI(country>korea))
[/S]
;=====
=====
;===== CRLcats
=====
; recherche voiture Excel
; {a:give-information+concept(action=e_buy,
vehicle-spec=(car, vehicle-make=, vehicle-
model=Excel, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mod(saloon,
Excel(country>koreF,make>HYUNDAI))

```

```

[/S]
;=====
=====
;===== CRLcats
=====
; recherche SKODA ou LADA
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=SKODA,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, SKODA(country>europe))
[/S]
;=====
=====
;===== CRLcats
=====
; cherche ^ acheter JEEP avec acompte de
3000 dinars et des tranches mensuelles de
200 contact 0795291030
; {a:give-information+concept(action=acher,
vehicle-spec=(, vehicle-make=JEEP, vehicle-
model=2060, vehicle-age=200, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
mak(saloon, JEEP(country>Ameria))
yea(saloon, 200)
[/S]
;=====
=====
;===== CRLcats
=====
; recherche voiture américaine Caprice de
préférence
; {a:give-information+concept(action=e_buy,
vehicle-spec=(car, vehicle-make=, vehicle-
model=Caprice, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mod(saloon, Caprice)
[/S]
;=====
=====
;===== CRLcats
=====
; recherche voiture sport de préférence
OPEL Calibra
; {a:give-information+concept(action=e_buy,
vehicle-spec=(car, vehicle-make=OPEL,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon,
OPEL(country>germany,country>europe))

```

```

)
[/S]
;=====
=====
;===== CRLcats
=====
; recherche BMW 316
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=BMW, vehicle-
model=316, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon,
    B.M.W(country>germany,country>europe
))
mod(saloon, 316)
[/S]
;=====
=====
;===== CRLcats
=====
; recherche ^ l'achat MERCEDES CLK modèle
2000 ou 99
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=MERCEDES,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=2000))}
[S]
wan(saloon, wanted)
mak(saloon,
    MERCEDES(country>germany,country>eur
ope))
[/S]
;=====
=====
;===== CRLcats
=====
; recherche BMW 316
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=BMW, vehicle-
model=316, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon,
    B.M.W(country>germany,country>europe
))
mod(saloon, 316)
[/S]
;=====
=====
;===== CRLcats
=====
; recherche ^ l'achat bus TOYOTA 1990 ou
après numero tel 0795133829
; {a:give-information+concept(action=e_buy,
vehicle-spec=(bus, vehicle-make=TOYOTA,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon,
    TOYOTA(country>japan))
[/S]
;=====
=====
;===== CRLcats
=====
; recherche ^ l'achat NISSAN Sunny modèle
92 ^ 95
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=NISSAN,
vehicle-model=Sunny, vehicle-age=95,
vehicle-price=, vehicle-color=, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=92))}
[S]
wan(saloon, wanted)
mak(saloon,
    NISSAN(country>japan))
mod(saloon,
    Sunny(country>japan,make>NISSAN))
yea(saloon, 95)
[/S]
;=====
=====
;===== CRLcats
=====
; cherche ^ vendre BMW m5
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=BMW, vehicle-model=,
vehicle-age=, vehicle-price=, vehicle-
color=, vehicle-condition=, vehicle-
assurance=, vehicle-controle=, vehicle-air-
condition=, vehicle-size=, vehicle-shape=,
vehicle-motor-type=, vehicle-hand=,
vehicle-nationality=, vehicle-mileage=))}
[S]
sal(saloon, sale)
mak(saloon,
    B.M.W(country>germany,country>europe
))
[/S]
;=====
=====
;===== CRLcats
=====
; recherche HONDA Civic contr™le boîte
automatique modèle 98 ou 99 avec acompte et
tranches
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=HONDA,
vehicle-model=Civic, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=98))}
[S]
wan(saloon, wanted)
mak(saloon,
    HONDA(country>japan))
mod(saloon,
    Civic(country>japan,make>HONDA))
[/S]
;=====
=====
;===== CRLcats
=====

```



```

; recherche MERCEDES
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=MERCEDES,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon,
MERCEDES(country>germany,country>eur
ope))
[/S]
;=====
;===== CRLcats
;=====
; je veux vendre voiture DAEWOO LIMINZ
modèle 92 contrTMle complet toutes options
; {a:give-
information+disposition+vehicule(disposition
=(desire, who=i), action=e_sell, vehicle-
spec=(car, vehicle-make=DAEWOO, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=92))}
[S]
sal(saloon, sale)
mak(saloon, DAEWOO(country>korea))
[/S]
;=====
;===== CRLcats
;=====
; recherche voiture BM 520 m ^ partir de 90
; {a:give-information+concept(action=e_buy,
vehicle-spec=(car, vehicle-make=BM,
vehicle-model=, vehicle-age=90, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon,
B.M.W(country>germany,country>europe
))
yea(saloon, 90)
[/S]
;=====
;===== CRLcats
;=====
; recherche MITSUBISHI
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make= Mitsubishi,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, MITSUBISHI(country>japan))
[/S]

```

```

;=====
;===== CRLcats
;=====
; recherche MITSUBISHI Galant modèle 80 ^
86 en bon état avec toutes les options
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=MITSUBISHI,
vehicle-model=Galant, vehicle-age=86,
vehicle-price=, vehicle-color=, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=80))}
[S]
wan(saloon, wanted)
mak(saloon, MITSUBISHI(country>japan))
mod(saloon,
Galant(country>japan,make>MITSUBISHI
))
yea(saloon, 86)
[/S]
;=====
;===== CRLcats
;=====
; recherche voiture MERCEDES
; {a:give-information+concept(action=e_buy,
vehicle-spec=(car, vehicle-make=MERCEDES,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon,
MERCEDES(country>germany,country>eur
ope))
[/S]
;=====
;===== CRLcats
;=====
; recherche achat voiture
; {a:give-information+concept(action=e_buy,
vehicle-spec=(car, vehicle-make=, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
[/S]
;=====
;===== CRLcats
;=====
; recherche voiture prix 3500 ^ 3850 dinars
; {a:give-information+concept(action=e_buy,
vehicle-spec=(car, vehicle-make=, vehicle-
model=, vehicle-age=3850, vehicle-
price=3850, vehicle-color=, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
yea(saloon, 3850)

```

```

pri(saloon,      3850)
[/S]
;=====
;=====
;===== CRLcats
;=====
; achat voiture
; {a:give-information+concept(action=achat,
vehicle-spec=(car, vehicle-make=, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
[/S]
;=====
;=====
;===== CRLcats
;=====
; recherche MERCEDES
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=MERCEDES,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon,      wanted)
mak(saloon,
      MERCEDES(country>germany,country>eur
ope))
[/S]
;=====
;=====
;===== CRLcats
;=====
; recherche KIA Sephia
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=KIA, vehicle-
model=Sephia, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon,      wanted)
mak(saloon,      KIA(country>korea))
mod(saloon,
      Sephia(country>koreF,make>KIA))
[/S]
;=====
;=====
;===== CRLcats
;=====
; recherche HYUNDAI Accent ou Elantra ou
KIA Sephia ^ 3800 dinars
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=HYUNDAI,
vehicle-model=Elantra, vehicle-age=3800,
vehicle-price=3800, vehicle-color=,
vehicle-condition=, vehicle-assurance=,
vehicle-controle=, vehicle-air-condition=,
vehicle-size=, vehicle-shape=, vehicle-
motor-type=, vehicle-hand=, vehicle-
nationality=, vehicle-mileage=))}
[S]
wan(saloon,      wanted)
mak(saloon,      HYUNDAI(country>korea))

mod(saloon,
      Elantra(country>koreF,make>HYUNDAI))
yea(saloon,      3800)
pri(saloon,      3800)
[/S]
;=====
;=====
;===== CRLcats
;=====
; recherche ^ l'achat MERCEDES 2005
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=MERCEDES,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon,      wanted)
mak(saloon,
      MERCEDES(country>germany,country>eur
ope))
[/S]
;=====
;=====
;===== CRLcats
;=====
; recherche voiture DAEWOO
; {a:give-information+concept(action=e_buy,
vehicle-spec=(car, vehicle-make=DAEWOO,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon,      wanted)
mak(saloon,      DAEWOO(country>korea))
[/S]
;=====
;=====
;===== CRLcats
;=====
; vente de OPEL Vectra modèle 1990 et achat
de MERCEDES CD200 modèle 1995
; {a:give-information+concept(action=achat,
vehicle-spec=(, vehicle-make=OPEL, vehicle-
model=Vectra, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(operator=conjunct,
[(md=1990), (md=1995)]))}
[S]
mak(saloon,
      OPEL(country>germany,country>europe)
)
mod(saloon,
      Vectra(country>germany,country>europ
e,make>OPEL))
[/S]
;=====
;=====
;===== CRLcats
;=====
; recherche MERCEDES 200 modèle 1995
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=MERCEDES,
vehicle-model=2060, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,

```

```

vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=1995))}
[S]
wan(saloon, wanted)
mak(saloon,
    MERCEDES(country>germany,country>eur
ope))
[/S]
;=====
;=====
;===== CRLcats
;=====
; Besoin d'une voiture ne dépassant pas
1500 dinars en bon état et économique
première main
; {a:give-information+concept(action=,
vehicle-spec=(car, vehicle-make=, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
[/S]
;=====
;=====
;===== CRLcats
;=====
; Recherche voiture KIA
; {a:give-information+concept(action=e_buy,
vehicle-spec=(car, vehicle-make=KIA,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, KIA(country>korea))
[/S]
;=====
;=====
;===== CRLcats
;=====
; recherche ^ l'achat BMW modèle 86 ou KIA
modèle 90 ^ 96
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=BMW, vehicle-
model=, vehicle-age=96, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(operator=disjunct,
[(md=86), (md=90)]))}
[S]
wan(saloon, wanted)
mak(saloon,
    B.M.W(country>germany,country>europe
))
yea(saloon, 96)
[/S]
;=====
;=====
;===== CRLcats
;=====
; recherche CHEVROLET Caprice paiement en
espèces
; {a:give-information+concept(action=e_buy,

```

```

vehicle-spec=(, vehicle-make=CHEVROLET,
vehicle-model=Caprice, vehicle-age=,
vehicle-price=, vehicle-color=, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, CHEVROLET(country>Ameria))
mod(saloon, Caprice)
[/S]
;=====
;=====
;===== CRLcats
;=====
; recherche pickup NISSAN
; {a:give-information+concept(action=e_buy,
vehicle-spec=(pickup, vehicle-make=NISSAN,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(pickup, wanted)
mak(pickup, NISSAN(country>japan))
[/S]
;=====
;=====
;===== CRLcats
;=====
; je souhaite acheter MERCEDES modèle 82
toutes options
; {a:give-
information+disposition+vehicle(disposition
=(desire, who=i), action=acher, vehicle-
spec=(, vehicle-make=MERCEDES, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=82))}
[S]
mak(saloon,
    MERCEDES(country>germany,country>eur
ope))
[/S]
;=====
;=====
;===== CRLcats
;=====
; recherche pickup NISSAN modèle 95-96-97
; {a:give-information+concept(action=e_buy,
vehicle-spec=(pickup, vehicle-make=NISSAN,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=95))}
[S]
wan(pickup, wanted)
mak(pickup, NISSAN(country>japan))
[/S]
;=====
;=====
;===== CRLcats
;=====
; recherche voiture acompte 1000 et le
reste par tranches

```

```

; {a:give-information+concept(action=e_buy,
vehicle-spec=(car, vehicle-make=, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
[/S]
;=====
;===== CRLcats
;=====
; Besoin d'une voiture ne dépassant pas
2000 dinars en bon état et économique en
essence
; {a:give-information+concept(action=,
vehicle-spec=(car, vehicle-make=, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
[/S]
;=====
;===== CRLcats
;=====
; recherche voiture DAEWOO acompte 1000 et
le reste par tranches
; {a:give-information+concept(action=e_buy,
vehicle-spec=(car, vehicle-make=DAEWOO,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, DAEWOO(country>korea))
[/S]
;=====
;===== CRLcats
;=====
; recherche Lada station modèle 91-92-93-
94-95
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=Lada, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=91))}
[S]
wan(saloon, wanted)
mak(saloon, Lada)
[/S]
;=====
;===== CRLcats
;=====
; recherche Golf m 89 90 ou KIA 94 95
contrTMle complet
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=KIA, vehicle-
model=Golf, vehicle-age=89, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, KIA(country>korea))
mod(saloon,
Golf(country>germany,country>europe,
make>VOLKS)
yea(saloon, 89)
[/S]
;=====
;===== CRLcats
;=====
; je veux acheter une voiture acompte 500
dinars et le reste par tranches mensuelles
de 100 dinars
; {a:give-
information+disposition+vehicle(disposition
=(desire, who=i), action=acher, vehicle-
spec=(car, vehicle-make=, vehicle-model=,
vehicle-age=100, vehicle-price=100,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
yea(saloon, 100)
pri(saloon, 100)
[/S]
;=====
;===== CRLcats
;=====
; recherche voiture Golf
; {a:give-information+concept(action=e_buy,
vehicle-spec=(car, vehicle-make=, vehicle-
model=Golf, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mod(saloon,
Golf(country>germany,country>europe,
make>VOLKS)
[/S]
;=====
;===== CRLcats
;=====
; recherche KIA Sephia
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=KIA, vehicle-
model=Sephia, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, KIA(country>korea))
mod(saloon,
Sephia(country>koreaF,make>KIA))

```

```

[/S]
;=====
;===== CRLcats
;=====
; recherche voiture MITSUBISHI boite
normale et en bon état
; {a:give-information+concept(action=e_buy,
vehicle-spec=(car, vehicle-make=MITSUBISHI,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, MITSUBISHI(country>japan))
[/S]
;=====
;===== CRLcats
;=====
; recherche voiture MITSUBISHI Galant
modèle 1988 ^ 1991 boite manuelle et en bon
état
; {a:give-information+concept(action=e_buy,
vehicle-spec=(car, vehicle-make=MITSUBISHI,
vehicle-model=Galant, vehicle-age=1991,
vehicle-price=, vehicle-color=, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=1988))}
[S]
wan(saloon, wanted)
mak(saloon, MITSUBISHI(country>japan))
mod(saloon,
Galant(country>japan,make>MITSUBISHI
))
yea(saloon, 1991)
[/S]
;=====
;===== CRLcats
;=====
; recherche HYUNDAI Elantra ^ 3800 dinars
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=HYUNDAI,
vehicle-model=Elantra, vehicle-age=3800,
vehicle-price=3800, vehicle-color=,
vehicle-condition=, vehicle-assurance=,
vehicle-controle=, vehicle-air-condition=,
vehicle-size=, vehicle-shape=, vehicle-
motor-type=, vehicle-hand=, vehicle-
nationality=, vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, HYUNDAI(country>korea))
mod(saloon,
Elantra(country>korea,make>HYUNDAI))
yea(saloon, 3800)
pri(saloon, 3800)
[/S]
;=====
;===== CRLcats
;=====
; recherche HONDA Civic ou MITSUBISHI
Lancer modèle 97 ^ 99 acompte 3000 dinars
et le reste par tranches mensuelles
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=HONDA,
vehicle-model=Civic, vehicle-age=99,
vehicle-price=, vehicle-color=, vehicle-

```

```

condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=97))}
[S]
wan(saloon, wanted)
mak(saloon, HONDA(country>japan))
mod(saloon,
Civic(country>japan,make>HONDA))
yea(saloon, 99)
[/S]
;=====
;===== CRLcats
;=====
; recherche voiture MITSUBISHI Lancer
modèle de 97 ^ 99 avec un acompte de 4000
dinars et le reste par tranches mensuelles
; {a:give-information+concept(action=e_buy,
vehicle-spec=(car, vehicle-make=MITSUBISHI,
vehicle-model=Lancer, vehicle-age=97,
vehicle-price=4000, vehicle-color=,
vehicle-condition=, vehicle-assurance=,
vehicle-controle=, vehicle-air-condition=,
vehicle-size=, vehicle-shape=, vehicle-
motor-type=, vehicle-hand=, vehicle-
nationality=, vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, MITSUBISHI(country>japan))
mod(saloon,
Lancer(country>japan,make>MITSUBISHI
))
yea(saloon, 97)
pri(saloon, 4000)
[/S]
;=====
;===== CRLcats
;=====
; recherche ^ l'achat MERCEDES 81 contrTMle
complet et toutes options
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=MERCEDES,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon,
MERCEDES(country>germany,country>eur
ope))
[/S]
;=====
;===== CRLcats
;=====
; recherche HYUNDAI ^ 3800 dinars
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=HYUNDAI,
vehicle-model=, vehicle-age=3800, vehicle-
price=3800, vehicle-color=, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, HYUNDAI(country>korea))
yea(saloon, 3800)

```

```

pri(saloon,      3800)
[/S]
;=====
;=====
;===== CRLcats
;=====
; recherche PEUGEOT 307 SALFA
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=PEUGEOT,
vehicle-model=3067, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon,      wanted)
mak(saloon,
      PEUGEOT(country>France,country>europ
e))
[/S]
;=====
;=====
;===== CRLcats
;=====
; recherche BMW toutes options modèle 98 ^
2003
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=BMW, vehicle-
model=, vehicle-age=2003, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=98))}
[S]
wan(saloon,      wanted)
mak(saloon,
      B.M.W(country>germany,country>europ
e))
yea(saloon,      2003)
[/S]
;=====
;=====
;===== CRLcats
;=====
; recherche BMW 740 toutes options modèle
98 ^ 2003
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=BMW, vehicle-
model=7406, vehicle-age=2003, vehicle-
price=2003, vehicle-color=, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=98))}
[S]
wan(saloon,      wanted)
mak(saloon,
      B.M.W(country>germany,country>europ
e))
yea(saloon,      2003)
pri(saloon,      2003)
[/S]
;=====
;=====
;===== CRLcats
;=====
; recherche KIA Sephia 95 acompte et le
reste par tranches
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=KIA, vehicle-
model=Sephia, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=),
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon,      wanted)
mak(saloon,
      Sephia(country>koreF,make>KIA))
[/S]
;=====
;=====
;===== CRLcats
;=====
; J'ai une voiture Land Rover je veux la
vendre
; {a:give-
information+disposition+vehicle(disposition
=(desire, who=i), action=e_sell, vehicle-
spec=(car, vehicle-make=Rover, vehicle-
model=Land Rover, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
sal(saloon,      sale)
mak(saloon,
      Rover(country>England,country>europ
e))
mod(saloon,      Land
Rover(country>England,country>europ
e))
[/S]
;=====
;=====
;===== CRLcats
;=====
; recherche voiture HYUNDAI ATOS avec
facilités de paiement
; {a:give-information+concept(action=e_buy,
vehicle-spec=(car, vehicle-make=HYUNDAI,
vehicle-model=ATOS, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon,      wanted)
mak(saloon,      HYUNDAI(country>korea))
mod(saloon,
      Atos(country>koreF,make>HYUNDAI))
[/S]
;=====
;=====
;===== CRLcats
;=====
; recherche BMW
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=BMW, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon,      wanted)
mak(saloon,
      B.M.W(country>germany,country>europ
e)

```

```

))
[/S]
;=====
=====
;===== CRLcats
=====
; recherche JEEP ou BM ou NISSAN modèle ^
partir de 95 contact 95777953
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=JEEP, vehicle-
model=, vehicle-age=95, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, JEEP(country>Ameria))
yea(saloon, 95)
[/S]
;=====
=====
;===== CRLcats
=====
; recherche voiture BM modèle 75 = 77
; {a:give-information+concept(action=e_buy,
vehicle-spec=(car, vehicle-make=BM,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=75))}
[S]
wan(saloon, wanted)
mak(saloon,
B.M.W(country>germany, country>europ
e))
[/S]
;=====
=====
;===== CRLcats
=====
; je veux une voiture MERCEDES Berline
modèle
; {a:give-
information+disposition+vehicule(disposition
=(desire, who=i), action=, vehicle-
spec=(car, vehicle-make=MERCEDES, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
mak(saloon,
MERCEDES(country>germany, country>eur
ope))
[/S]
;=====
=====
;===== CRLcats
=====
; recherche PEUGEOT 206
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=PEUGEOT,
vehicle-model=206, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,

```

```

vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon,
PEUGEOT(country>France, country>europ
e))
mod(saloon, 206)
[/S]
;=====
=====
;===== CRLcats
=====
; recherche PEUGEOT
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=PEUGEOT,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon,
PEUGEOT(country>France, country>europ
e))
[/S]
;=====
=====
;===== CRLcats
=====
; recherche PEUGEOT 207
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=PEUGEOT,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon,
PEUGEOT(country>France, country>europ
e))
[/S]
;=====
=====
;===== CRLcats
=====
; recherche datsun modèle de 80 ^ 82 moteur
de 1200 ^ 1400 CC
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=datson,
vehicle-model=, vehicle-age=80, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=1400,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, DATSUN(country>japan))
mot(saloon, 1400)
yea(saloon, 80)
[/S]
;=====
=====
;===== CRLcats
=====
; recherche ^ l'achat voiture modèle ^
partir de 90 ^ 3000 dinars avec facilités
de paiement
; {a:give-information+concept(action=e_buy,

```

```

vehicle-spec=(car, vehicle-make=, vehicle-
model=, vehicle-age=90, vehicle-price=3000,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
yea(saloon, 90)
pri(saloon, 3000)
[/S]
;=====
;===== CRLcats
;=====
; recherche pickup TOYOTA modèle environ 80
contrTMle ok
; {a:accept+concept(action=e_buy, vehicle-
spec=(pickup, vehicle-make=TOYOTA, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(pickup, wanted)
mak(pickup, TOYOTA(country>japan))
[/S]
;=====
;===== CRLcats
;=====
; recherche OPEL VECTRA moteur 1600 modèle
^ partir de 94
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=OPEL, vehicle-
model=VECTRA, vehicle-age=94, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon,
OPEL(country>germany,country>europe)
)
mod(saloon,
Vectra(country>germany,country>europ
e,make>OPEL))
yea(saloon, 94)
[/S]
;=====
;===== CRLcats
;=====
; recherche voiture HONDA Civic 2003 en bon
état couleur grise ou dorée
; {a:give-information+concept(action=e_buy,
vehicle-spec=(car, vehicle-make=HONDA,
vehicle-model=Civic, vehicle-age=, vehicle-
price=, vehicle-color=grey, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, HONDA(country>japan))
mod(saloon,
Civic(country>japan,make>HONDA)
)
col(saloon, silver)
[/S]
;=====
;===== CRLcats
;=====
; recherche ^ l'achat voiture modèle ^
partir de 90 acompte 500 dinars et tranches
; {a:give-information+concept(action=e_buy,
vehicle-spec=(car, vehicle-make=, vehicle-
model=, vehicle-age=90, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
yea(saloon, 90)
[/S]
;=====
;===== CRLcats
;=====
; recherche HONDA Civic 96 97 98
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=HONDA,
vehicle-model=Civic, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, HONDA(country>japan))
mod(saloon,
Civic(country>japan,make>HONDA)
)
[/S]
;=====
;===== CRLcats
;=====
; recherche HYUNDAI Excel 1994
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=HYUNDAI,
vehicle-model=Excel, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, HYUNDAI(country>korea))
mod(saloon,
Excel(country>koreF,make>HYUNDAI)
)
[/S]
;=====
;===== CRLcats
;=====
; recherche HONDA Civic 96 97 98
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=HONDA,
vehicle-model=Civic, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}

```



```

vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, HONDA(country>japan))
mod(saloon,
    Civic(country>japan,make>HONDA))
[/S]
;=====
;===== CRLcats
;=====
; recherche HONDA Civic 2003 en bon état
couleur grise ou dorée
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=HONDA,
vehicle-model=Civic, vehicle-age=, vehicle-
price=, vehicle-color=grey, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, HONDA(country>japan))
mod(saloon,
    Civic(country>japan,make>HONDA))
col(saloon, silver)
[/S]
;=====
;===== CRLcats
;=====
; recherche HONDA Civic 2003 en bon état
couleur grise ou dorée
; {a:give-information+concept(action=e_buy,
vehicle-spec=(, vehicle-make=HONDA,
vehicle-model=Civic, vehicle-age=, vehicle-
price=, vehicle-color=grey, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
wan(saloon, wanted)
mak(saloon, HONDA(country>japan))
mod(saloon,
    Civic(country>japan,make>HONDA))
col(saloon, silver)
[/S]
;=====
;===== CRLcats
;=====
; je veux vendre une MERCEDES 280C modèle
2006 jordanienne kilométrage 4000 KM
; {a:give-
information+disposition+vehicule(disposition
=(desire, who=i), action=e_sell, vehicle-
spec=(, vehicle-make=MERCEDES, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-
nationality=jdanienne, vehicle-
mileage=4000), time=(md=2006))}
[S]
sal(saloon, sale)
mak(saloon,
    MERCEDES(country>germany, country>eur
ope))
mil(saloon, 4000)
[/S]
;=====

```

```

=====
;===== CRLcats
;=====
; OPEL kadett pack modèle 90 ou GOLF modèle
90 pack ou CITROEN pack
; {a:give-information+concept(action=,
vehicle-spec=(, vehicle-make=OPEL, vehicle-
model=GOLF, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(operator=disjunct,
[(md=90), (md=90)]))}
[S]
mak(saloon,
    OPEL(country>germany, country>europe)
)
mod(saloon,
    Golf(country>germany, country>europe,
make>VOLKS)
[/S]
;=====
;===== CRLcats
;=====
; A vendre OPEL ASTRA rouge (avec ouverture
centrale vitre et rétro électrique)
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=OPEL, vehicle-
model=ASTRA, vehicle-age=, vehicle-price=,
vehicle-color=red, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
sal(saloon, sale)
mak(saloon,
    OPEL(country>germany, country>europe)
)
mod(saloon,
    Astra(country>germany, country>europe
, make>OPEL))
col(saloon, red)
[/S]
;=====
;===== CRLcats
;=====
; pickup datsun NISSAN diesel modèle 83
assuré un an en bon état ^ vendre, 2000
dinars
; {a:give-information+concept(action=,
vehicle-spec=(pickup, vehicle-make=datsun,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=83))}
[S]
mak(pickup, DATSUN(country>japan))
[/S]
;=====
;===== CRLcats
;=====
; 0795322079, BMW, 1988, 318 en bon état,
6250 dinars contact 0795322 079
; {a:give-
information+concept(numeral=[0795322079])}
[S]

```

```

[/S]
;=====
;=====
;===== CRLcats
;=====
; PEUGEOT 2002 T.O sauf la boite contrTMle
complet pour renseignement 0796702818
; {a:give-
information+concept(exception=discourse,
action=, vehicle-spec=(, vehicle-
make=PEUGEOT, vehicle-model=, vehicle-
age=2002, vehicle-price=, vehicle-color=,
vehicle-condition=, vehicle-assurance=,
vehicle-controle=, vehicle-air-condition=,
vehicle-size=, vehicle-shape=, vehicle-
motor-type=, vehicle-hand=, vehicle-
nationality=, vehicle-mileage=))}
[S]
mak(saloon,
    PEUGEOT(country>France, country>europ
e))
yea(saloon,    2002)
[/S]
;=====
;=====
;===== CRLcats
;=====
; Vente ou échange VOLVO modèle 8 contrTMle
complet sport 2 portes toutes options
; {a:give-information+concept(activity-
spec=sport, time=(md=8))}
[S]
[/S]
;=====
;=====
;===== CRLcats
;=====
; GOLF GTI 89 en bon état originale
première main prix 6500 tel 0647799079
; {a:give-
information+concept(numeral=[89])}
[S]
[/S]
;=====
;=====
;===== CRLcats
;=====
; JEEP Cherokee toutes options couleur
dorée douane payée vitres électriques en
état original pas plus de 60 000 dinars
; {a:accept+concept(action=, vehicle-
spec=(, vehicle-make=JEEP, vehicle-model=,
vehicle-age=60, vehicle-price=, vehicle-
color=, vehicle-condition=, vehicle-
assurance=, vehicle-controle=, vehicle-air-
condition=, vehicle-size=, vehicle-shape=,
vehicle-motor-type=, vehicle-hand=,
vehicle-nationality=, vehicle-mileage=))}
[S]
mak(saloon,    JEEP(country>Ameria))
yea(saloon,    60)
[/S]
;=====
;=====
;===== CRLcats
;=====
; DAEWOO Nubira modèle 1997 doré toutes
options automatique douane payée récemment
6500 dinars
; {a:give-information+concept(action=,
vehicle-spec=(, vehicle-make=DAEWOO,
vehicle-model=Nubira, vehicle-age=,
vehicle-price=, vehicle-color=, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-
type=, vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=1997))}
[S]
mak(saloon,
    Nubira(country>koreF, make>DAEWOO))
[/S]
;=====
;=====
;===== CRLcats
;=====
; range rover 82 renouvelée 95 en bon état
80000 KM 16000 dinars
; {a:give-
information+concept(numeral=[82])}
[S]
[/S]
;=====
;=====
;===== CRLcats
;=====
; RENAULT 5 modèle 82 assurée jusqu'
septembre en bon état 1000 dinars tél
0796232997
; {a:give-information+concept(action=,
vehicle-spec=(, vehicle-make=RENAULT,
vehicle-model=5, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=82))}
[S]
mak(saloon,
    RENAULT(country>France, country>europ
e))
mod(saloon,
    5(country>France, country>europe, make
>RENAULT))
[/S]
;=====
;=====
;===== CRLcats
;=====
; voiture GOLF abs modèle 1998 nouvelle
look contrTMle complet ouverture électrique
; {a:accept+concept(action=, vehicle-
spec=(car, vehicle-make=, vehicle-
model=GOLF, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=1998))}
[S]
mod(saloon,
    Golf(country>germany, country>europe,
make>VOLKS))
[/S]
;=====
;=====
;===== CRLcats
;=====
; A vendre OPEL VECTRA modèle 1990 toutes
options contrTMle complet en bon état
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=OPEL, vehicle-
model=VECTRA, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,

```

```

vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=1990))}
[S]
sal(saloon, sale)
mak(saloon,
    OPEL(country>germany, country>europe)
)
mod(saloon,
    Vectra(country>germany, country>europ
e, make>OPEL))
[/S]
;=====
;===== CRLcats
;=====
; voiture OPEL kadett sport modèle 87
couleur marron en bon état ^ vendre prix
définitif 3000 dinars
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(car, vehicle-make=OPEL, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=brown, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=87))}
[S]
sal(saloon, sale)
mak(saloon,
    OPEL(country>germany, country>europe)
)
col(saloon, 4doors)
[/S]
;=====
;===== CRLcats
;=====
; voiture PEUGEOT 307 modèle 2005 toutes
options couleur grise toit ouvrant 14000 Km
au compteur 15500 dinars
; {a:give-information+concept(action=,
vehicle-spec=(car, vehicle-make=PEUGEOT,
vehicle-model=3067, vehicle-age=, vehicle-
price=, vehicle-color=grey, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=2005))}
[S]
mak(saloon,
    PEUGEOT(country>France, country>europ
e))
col(saloon, silver)
[/S]
;=====
;===== CRLcats
;=====
; voiture PEUGEOT 307 modèle 2005 toutes
options couleur grise toit ouvrant 14000 Km
au compteur 15500 dinars
; {a:give-information+concept(action=,
vehicle-spec=(car, vehicle-make=PEUGEOT,
vehicle-model=3067, vehicle-age=, vehicle-
price=, vehicle-color=grey, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=2005))}
[S]
mak(saloon,
    PEUGEOT(country>France, country>europ

```

```

e))
col(saloon, silver)
[/S]
;=====
;===== CRLcats
;=====
; TOYOTA Celica 1981 boîte automatique 1800
CC en bon état sauf le moteur
; {a:give-
information+concept(exception=discourse,
action=, vehicle-spec=(, vehicle-
make=TOYOTA, vehicle-model=Celica, vehicle-
age=, vehicle-price=, vehicle-color=,
vehicle-condition=, vehicle-assurance=,
vehicle-controle=, vehicle-air-condition=,
vehicle-size=, vehicle-shape=, vehicle-
motor-type=1800, vehicle-hand=, vehicle-
nationality=, vehicle-mileage=))}
[S]
mak(saloon, TOYOTA(country>japan))
mod(saloon,
    Celica(country>japan, make>TOYOTA))
mot(saloon, 1800)
[/S]
;=====
;===== CRLcats
;=====
; voiture DAEWOO 93 vitres électriques
contrTMle parfait couleur verte
; {a:give-information+concept(action=,
vehicle-spec=(car, vehicle-make=DAEWOO,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=green, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
mak(saloon, DAEWOO(country>korea))
col(saloon, green)
[/S]
;=====
;===== CRLcats
;=====
; A vendre voiture taxi NISSAN Sunny modèle
2005 taxe libre pour contact 0795597203
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(car, vehicle-make=NISSAN, vehicle-
model=Sunny, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=2005))}
[S]
sal(saloon, sale)
mak(saloon, NISSAN(country>japan))
mod(saloon,
    Sunny(country>japan, make>NISSAN))
[/S]
;=====
;===== CRLcats
;=====
; A vendre voiture MITSUBISHI LANCER 93
douane payée récemment 5900 dinars ^
débattre
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(car, vehicle-make= Mitsubishi,

```

```

vehicle-model=LANCER, vehicle-age=,
vehicle-price=, vehicle-color=, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
sal(saloon, sale)
mak(saloon, MITSUBISHI(country>japan))
mod(saloon,
    Lancer(country>japan,make>MITSUBISHI
))
[/S]
;=====
=====
;===== CRLcats
=====
; MERCEDES 99 grise 23000 dinars voiture
HONDA CIVIC 2001 marron 10500 dinars
; {a:give-information+concept(action=,
vehicle-spec=(car, vehicle-make=MERCEDES,
vehicle-model=CIVIC, vehicle-age=, vehicle-
price=, vehicle-color=grey, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
mak(saloon,
    MERCEDES(country>germany,country>eur
ope))
mod(saloon,
    Civic(country>japan,make>HONDA))
col(saloon, silver)
[/S]
;=====
=====
;===== CRLcats
=====
; A vendre MERCEDES modèle 2001 couleur
grise toutes options contrTMle complet
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=MERCEDES, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=grey, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=2001))}
[S]
sal(saloon, sale)
mak(saloon,
    MERCEDES(country>germany,country>eur
ope))
col(saloon, silver)
[/S]
;=====
=====
;===== CRLcats
=====
; A vendre VOLVO modèle 1980 et OPEL modèle
1992 contactez 0795902118
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=OPEL, vehicle-
model=1992, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=1980))}
[S]
sal(saloon, sale)
mak(saloon,
    OPEL(country>germany,country>europe)
)
mod(saloon, 1992)
[/S]
;=====
=====
;===== CRLcats
=====
; voiture bmw toutes options sauf la
climatisation avec contrTMle technique
modèle 1989 couleur grise jantes 17
; {a:give-
information+concept(exception=discourse,
action=, vehicle-spec=(car, vehicle-
make=bmw, vehicle-model=, vehicle-age=,
vehicle-price=, vehicle-color=grey,
vehicle-condition=, vehicle-assurance=,
vehicle-controle=, vehicle-air-condition=,
vehicle-size=, vehicle-shape=, vehicle-
motor-type=, vehicle-hand=, vehicle-
nationality=, vehicle-mileage=),
time=(md=1989))}
[S]
mak(saloon,
    B.M.W(country>germany,country>europe
))
col(saloon, silver)
[/S]
;=====
=====
;===== CRLcats
=====
; A vendre voiture américaine modèle 1981
en très bon état
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(car, vehicle-make=, vehicle-
model=1981, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
sal(saloon, sale)
mod(saloon, 1981)
[/S]
;=====
=====
;===== CRLcats
=====
; OPEL kadett modèle 86 moteur 1200CC en
bon état
; {a:give-information+concept(action=,
vehicle-spec=(, vehicle-make=OPEL, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=86))}
[S]
mak(saloon,
    OPEL(country>germany,country>europe)
)
[/S]
;=====
=====
;===== CRLcats
=====
; TOYOTA Corolla modèle 99 couleur verte

```

```

toutes options sauf la boite 8800 dinars
contrTMle complet
; {a:give-
information+concept(exception=discourse,
action=, vehicle-spec=(, vehicle-
make=TOYOTA, vehicle-model=Colla, vehicle-
age=, vehicle-price=, vehicle-color=green,
vehicle-condition=, vehicle-assurance=,
vehicle-controle=, vehicle-air-condition=,
vehicle-size=, vehicle-shape=, vehicle-
motor-type=, vehicle-hand=, vehicle-
nationality=, vehicle-mileage=),
time=(md=99))}
[S]
mak(saloon, TOYOTA(country>japan))
col(saloon, green)
[/S]
;=====
;===== CRLcats
;=====
; voiture DAEWOO Lanos modèle 99 couleur
blanche contrTMle assuré toutes options sauf
la boite
; {a:give-
information+concept(exception=discourse,
action=, vehicle-spec=(car, vehicle-
make=DAEWOO, vehicle-model=Lanos, vehicle-
age=, vehicle-price=, vehicle-
color=whitehe, vehicle-condition=, vehicle-
assurance=, vehicle-controle=, vehicle-air-
condition=, vehicle-size=, vehicle-shape=,
vehicle-motor-type=, vehicle-hand=,
vehicle-nationality=, vehicle-mileage=),
time=(md=99))}
[S]
mak(saloon, DAEWOO(country>korea))
mod(saloon,
Lanos(country>koreF,make>DAEWOO))
[/S]
;=====
;===== CRLcats
;=====
; voiture DAEWOO modèle 99 contrTMle complet
économique couleur jaune vitres électriques
0796864442 Kaled Achour
; {a:give-information+concept(action=,
vehicle-spec=(car, vehicle-make=DAEWOO,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=yellow, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=99))}
[S]
mak(saloon, DAEWOO(country>korea))
col(saloon, yellow)
[/S]
;=====
;===== CRLcats
;=====
; A vendre voiture DAEWOO NUBIRA PREMIUM 2
modèle 2000 8000 dinars propriétaire
0795080690
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(car, vehicle-make=DAEWOO, vehicle-
model=NUBIRA, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=89))}
[S]
sal(saloon, sale)
mak(saloon, HONDA(country>japan))
mod(saloon, Rio(country>koreF,make>KIA))

```

```

vehicle-mileage=), time=(md=2000))}
[S]
sal(saloon, sale)
mak(saloon, DAEWOO(country>korea))
mod(saloon,
Nubira(country>koreF,make>DAEWOO))
[/S]
;=====
;===== CRLcats
;=====
; A vendre KIA Shuma modèle 98 toutes
options contrTMle complet ^ 6800 dinars
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=KIA, vehicle-
model=Shuma, vehicle-age=6800, vehicle-
price=6800, vehicle-color=, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=98))}
[S]
sal(saloon, sale)
mak(saloon, KIA(country>korea))
mod(saloon,
Shuma(country>koreF,make>KIA))
yea(saloon, 6800)
pri(saloon, 6800)
[/S]
;=====
;===== CRLcats
;=====
; A vendre KIA Sephia modèle 1994 toutes
options contrTMle non assuré 3800 dinars
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=KIA, vehicle-
model=Sephia, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=1994))}
{a:negate} {a:give-
information+concept(numeral=[3800])}
[S]
sal(saloon, sale)
mak(saloon, KIA(country>korea))
mod(saloon,
Sephia(country>koreF,make>KIA))
[/S]
;=====
;===== CRLcats
;=====
; A vendre HONDA Rio modèle 89 blanche
toutes options sauf la boite 4000 dinars
; {a:give-
information+concept(exception=discourse,
action=e_sell, vehicle-spec=(, vehicle-
make=HONDA, vehicle-model=Rio, vehicle-
age=, vehicle-price=, vehicle-
color=whitehe, vehicle-condition=, vehicle-
assurance=, vehicle-controle=, vehicle-air-
condition=, vehicle-size=, vehicle-shape=,
vehicle-motor-type=, vehicle-hand=,
vehicle-nationality=, vehicle-mileage=),
time=(md=89))}
[S]
sal(saloon, sale)
mak(saloon, HONDA(country>japan))
mod(saloon, Rio(country>koreF,make>KIA))

```

```

[/S]
;=====
;=====
;===== CRLcats
;=====
; A vendre HONDA CIVIC LXI climatisée
automatique contrTMle complet modèle 96
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=HONDA, vehicle-
model=CIVIC LXI, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=96))}
[S]
sal(saloon, sale)
mak(saloon, HONDA(country>japan))
mod(saloon,
Civic(country>japan,make>HONDA) LXI)
[/S]
;=====
;=====
;===== CRLcats
;=====
; A vendre HONDA CIVIC LXI climatisée
automatique contrTMle complet modèle 96
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=HONDA, vehicle-
model=CIVIC LXI, vehicle-age=, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=96))}
[S]
sal(saloon, sale)
mak(saloon, HONDA(country>japan))
mod(saloon,
Civic(country>japan,make>HONDA) LXI)
[/S]
;=====
;=====
;===== CRLcats
;=====
; MERCEDES modèle 99 classique 200 toutes
options sauf l'ouverture ^ 22000 dinars ou
acompte et tranches bancaires
; {a:give-
information+concept(exception=discourse,
action=, vehicle-spec=(, vehicle-
make=MERCEDES, vehicle-model=2060, vehicle-
age=22000, vehicle-price=22000, vehicle-
color=, vehicle-condition=, vehicle-
assurance=, vehicle-controle=, vehicle-air-
condition=, vehicle-size=, vehicle-shape=,
vehicle-motor-type=, vehicle-hand=,
vehicle-nationality=, vehicle-mileage=),
time=(md=99))}
[S]
mak(saloon,
MERCEDES(country>germany, country>eur
ope))
yea(saloon, 22000)
pri(saloon, 22000)
[/S]
;=====
;=====
;===== CRLcats
;=====
; MERCEDES C180 modèle 1996 grise toutes
options contrTMle complet 16000 dinars
; {a:give-information+concept(action=,
vehicle-spec=(, vehicle-make=MERCEDES,
vehicle-model=, vehicle-age=, vehicle-
price=, vehicle-color=grey, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=1996))}
[S]
mak(saloon,
MERCEDES(country>germany, country>eur
ope))
col(saloon, silver)
[/S]
;=====
;=====
;===== CRLcats
;=====
; A vendre voiture taxi HONDA modèle 2001
taxe libre 23000 dinars
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(car, vehicle-make=HONDA, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=2001))}
[S]
sal(saloon, sale)
mak(saloon, HONDA(country>japan))
[/S]
;=====
;=====
;===== CRLcats
;=====
; A vendre voiture HONDA CIVIC modèle 96
injection moteur 1005 cc assurée et
contrTMle complet
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(car, vehicle-make=HONDA, vehicle-
model=CIVIC, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=1005,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=96))}
[S]
sal(saloon, sale)
mak(saloon, HONDA(country>japan))
mod(saloon,
Civic(country>japan,make>HONDA))
mot(saloon, 1005)
[/S]
;=====
;=====
;===== CRLcats
;=====
; A vendre HONDA CIVIC modèle 2000 toutes
options vitres manuelles couleur noire
0795066927
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=HONDA, vehicle-
model=CIVIC, vehicle-age=, vehicle-price=,
vehicle-color=black, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,

```

```

vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=2000))}
[S]
sal(saloon, sale)
mak(saloon, HONDA(country>japan))
mod(saloon,
    Civic(country>japan,make>HONDA))
col(saloon, bck)
[/S]
;=====
;===== CRLcats
;=====
; A vendre HONDA SONATA modèle 97 toutes
options couleur jaune
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=HONDA, vehicle-
model=SONATA, vehicle-age=, vehicle-price=,
vehicle-color=yellow, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=97))}
[S]
sal(saloon, sale)
mak(saloon, HONDA(country>japan))
mod(saloon,
    Sonata(country>koreF,make>HYUNDAI))
col(saloon, yellow)
[/S]
;=====
;===== CRLcats
;=====
; A vendre HONDA modèle 1998 toutes options
sauf boite contrTMle complet, verte, 5500
dinars
; {a:give-
information+concept(exception=discourse,
action=e_sell, vehicle-spec=(, vehicle-
make=HONDA, vehicle-model=, vehicle-age=,
vehicle-price=, vehicle-color=, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=1998))}
[S]
sal(saloon, sale)
mak(saloon, HONDA(country>japan))
[/S]
;=====
;===== CRLcats
;=====
; A vendre MERCEDES modèle 96 rouge CT OK
13000 dinars première main tél 0795535225
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=MERCEDES, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=red, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=96))}
[S]
sal(saloon, sale)
mak(saloon,
    MERCEDES(country>germany,country>eur
ope))
col(saloon, red)
[/S]

```

```

;=====
;===== CRLcats
;=====
; A vendre FOX FAJIN BEETLE modèle 1999
toutes options sauf la boite rouge douane
payée récemment 10500 dinars
; {a:give-
information+concept(exception=discourse,
time=(md=1999))}
[S]
[/S]
;=====
;===== CRLcats
;=====
; A vendre KIA sport 4*4 m 2003 toutes
options sauf la boite douane non payée 7000
dinars couleur marron
; {a:give-
information+concept(exception=discourse,
action=e_sell, vehicle-spec=(, vehicle-
make=KIA, vehicle-model=, vehicle-age=2003,
vehicle-price=, vehicle-color=, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))} {a:negate} {a:give-
information+concept}
[S]
sal(saloon, sale)
mak(saloon, KIA(country>korea))
yea(saloon, 2003)
[/S]
;=====
;===== CRLcats
;=====
; A vendre KIA Sportage JEEP 4*4 m 2004
toutes options sauf la boite couleur noire
11800 dinars assurance complète
; {a:give-
information+concept(exception=discourse,
action=e_sell, vehicle-spec=(, vehicle-
make=KIA, vehicle-model=Sptage, vehicle-
age=2004, vehicle-price=, vehicle-
color=black, vehicle-condition=, vehicle-
assurance=, vehicle-controle=, vehicle-air-
condition=, vehicle-size=, vehicle-shape=,
vehicle-motor-type=, vehicle-hand=,
vehicle-nationality=, vehicle-mileage=))}
[S]
sal(saloon, sale)
mak(saloon, KIA(country>korea))
col(saloon, bck)
yea(saloon, 2004)
[/S]
;=====
;===== CRLcats
;=====
; A vendre MITSUBISHI LANCER modèle 93
automatique climatisée en bon état le prix
4000 d
; {a:give-
information+price+concept(action=e_sell,
vehicle-spec=(, vehicle-make= Mitsubishi,
vehicle-model=LANCER, vehicle-age=4000,
vehicle-price=, vehicle-color=, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=93))}
[S]
sal(saloon, sale)

```

```

mak(saloon,    MITSUBISHI(country>japan))
mod(saloon,
    Lancer(country>japan,make>MITSUBISHI
))
yea(saloon,    4000)
[/S]
;=====
;===== CRLcats
;=====
; A vendre voiture KIA CIVIC modèle 94
toutes options sauf la boîte contrTMle
complet
; {a:give-
information+concept(exception=discourse,
action=e_sell, vehicle-spec=(car, vehicle-
make=KIA, vehicle-model=CIVIC, vehicle-
age=, vehicle-price=, vehicle-color=,
vehicle-condition=, vehicle-assurance=,
vehicle-controle=, vehicle-air-condition=,
vehicle-size=, vehicle-shape=, vehicle-
motor-type=, vehicle-hand=, vehicle-
nationality=, vehicle-mileage=),
time=(md=94))}
[S]
sal(saloon,    sale)
mak(saloon,    KIA(country>korea))
mod(saloon,
    Civic(country>japan,make>HONDA)
[/S]
;=====
;===== CRLcats
;=====
; A vendre bus KIA BASTA commun blanc
modèle 95 5300 dinars ^ débattre
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(bus, vehicle-make=KIA, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=white, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=95))}
[S]
sal(bus,        sale)
mak(bus,        KIA(country>korea))
col(bus,        white)
[/S]
;=====
;===== CRLcats
;=====
; A vendre GrandCherokee modèle 1994
couleur noire 6 cylindres ^ traction
arrière toutes options
; {a:accept+concept(time=(md=1994))}
[S]
[/S]
;=====
;===== CRLcats
;=====
; A vendre GOLF 1600CC distribution
électrique ABS noire toutes options
; {a:give-information+concept}
[S]
[/S]
;=====
;===== CRLcats
;=====
; A vendre GOLF GTI modèle 89 noire jantes
15 contrTMle original avec ouverture
système 5
; {a:give-information+concept(time=(md=89),
numeral=[15])}
[S]
[/S]
;=====
;===== CRLcats
;=====
; A vendre bmw modèle 94 toutes options
avec cuir sauf la boîte ^ 11000 dinars
; {a:give-
information+concept(exception=discourse,
action=e_sell, vehicle-spec=(, vehicle-
make=bmw, vehicle-model=, vehicle-
age=11000, vehicle-price=11000, vehicle-
color=, vehicle-condition=, vehicle-
assurance=, vehicle-controle=, vehicle-air-
condition=, vehicle-size=, vehicle-shape=,
vehicle-motor-type=, vehicle-hand=,
vehicle-nationality=, vehicle-mileage=),
time=(md=94))}
[S]
sal(saloon,    sale)
mak(saloon,
    B.M.W(country>germany,country>europe
))
yea(saloon,    11000)
pri(saloon,    11000)
[/S]
;=====
;===== CRLcats
;=====
; A vendre AUDI allemande modèle 78 moteur
et boîte neufs ^ traction avant 800 dinars
; {a:give-information+concept(before-
rr=discourse, action=e_sell, vehicle-
spec=(, vehicle-make=AUDI, vehicle-model=,
vehicle-age=, vehicle-price=, vehicle-
color=, vehicle-condition=, vehicle-
assurance=, vehicle-controle=, vehicle-air-
condition=, vehicle-size=, vehicle-shape=,
vehicle-motor-type=, vehicle-hand=,
vehicle-nationality=, vehicle-mileage=),
time=(md=78))}
[S]
sal(saloon,    sale)
mak(saloon,
    AUDI(country>germany,country>europe)
)
[/S]
;=====
;===== CRLcats
;=====
; A vendre OPEL VECTRA modèle 97 toutes
options sauf la boîte et la climatisation
9750 dinars
; {a:give-
information+concept(exception=discourse,
action=e_sell, vehicle-spec=(, vehicle-
make=OPEL, vehicle-model=VECTRA, vehicle-
age=, vehicle-price=, vehicle-color=,
vehicle-condition=, vehicle-assurance=,
vehicle-controle=, vehicle-air-condition=,
vehicle-size=, vehicle-shape=, vehicle-
motor-type=, vehicle-hand=, vehicle-
nationality=, vehicle-mileage=),
time=(md=97))}
[S]
sal(saloon,    sale)
mak(saloon,
    OPEL(country>germany,country>europe)
)

```



```

mod(saloon,
    Vectra(country>germany,country>europ
e,make>OPEL))
[/S]
;=====
;===== CRLcats
;=====
; A vendre BM 520 modèle 93 sans contrTMle
toutes options sauf la boite excellente
10500 dinars
; {a:give-
information+concept(exception=discourse,
action=e_sell, vehicle-spec=(, vehicle-
make=BM, vehicle-model=, vehicle-age=,
vehicle-price=, vehicle-color=, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=93))}
[S]
sal(saloon, sale)
mak(saloon,
    B.M.W(country>germany,country>europ
e))
[/S]
;=====
;===== CRLcats
;=====
; A vendre BMW 520 modèle 2000 couleur
dorée avec cuir blanc 0795610755
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=BMW, vehicle-model=,
vehicle-age=, vehicle-price=, vehicle-
color=white, vehicle-condition=, vehicle-
assurance=, vehicle-controle=, vehicle-air-
condition=, vehicle-size=, vehicle-shape=,
vehicle-motor-type=, vehicle-hand=,
vehicle-nationality=, vehicle-mileage=),
time=(md=2000))}
[S]
sal(saloon, sale)
mak(saloon,
    B.M.W(country>germany,country>europ
e))
col(saloon, white)
[/S]
;=====
;===== CRLcats
;=====
; A vendre BMW modèle 95 contrTMle complet
options complètes avec des options
extraordinaires
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=BMW, vehicle-model=,
vehicle-age=, vehicle-price=, vehicle-
color=, vehicle-condition=, vehicle-
assurance=, vehicle-controle=, vehicle-air-
condition=, vehicle-size=, vehicle-shape=,
vehicle-motor-type=, vehicle-hand=,
vehicle-nationality=, vehicle-mileage=),
time=(md=95))}
[S]
sal(saloon, sale)
mak(saloon,
    B.M.W(country>germany,country>europ
e))
[/S]
;=====
;===== CRLcats

```

```

=====
; A vendre voiture PEUGEOT 307 modèle 2005
6000 KM compteur 1,4l toutes options sauf
la boite 0795321200
; {a:give-
information+concept(exception=discourse,
action=e_sell, vehicle-spec=(car, vehicle-
make=PEUGEOT, vehicle-model=3067, vehicle-
age=, vehicle-price=, vehicle-color=,
vehicle-condition=, vehicle-assurance=,
vehicle-controle=, vehicle-air-condition=,
vehicle-size=, vehicle-shape=, vehicle-
motor-type=, vehicle-hand=, vehicle-
nationality=, vehicle-mileage=6000),
time=(md=2005))}
[S]
sal(saloon, sale)
mak(saloon,
    PEUGEOT(country>France,country>europ
e))
mil(saloon, 6000)
[/S]
;=====
;===== CRLcats
;=====
; A vendre TOYOTA Corolla modèle 1982 boite
automatique en bon état batterie neuve et
fauteuils neufs
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=TOYOTA, vehicle-
model=Colla, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=1982))}
[S]
sal(saloon, sale)
mak(saloon, TOYOTA(country>japan))
[/S]
;=====
;===== CRLcats
;=====
; A vendre GOLF modèle 1998 comme neuve
contrTMle complet ouverture électrique
; {a:give-
information+concept(time=(md=1998))}
[S]
[/S]
;=====
;===== CRLcats
;=====
; A vendre voiture GOLF modèle 92 noire
système d'alarme vitre électrique
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(car, vehicle-make=, vehicle-
model=GOLF, vehicle-age=, vehicle-price=,
vehicle-color=black, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=92))}
[S]
sal(saloon, sale)
mod(saloon,
    Golf(country>germany,country>europ
e,
make>VOLKS)
col(saloon, bck)
[/S]

```

```

;=====
;=====
;===== CRLcats
;=====
; A vendre voiture DAEWOO modèle 97 noire
très économique première main assurée
jusqu'en mai
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(car, vehicle-make=DAEWOO, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=black, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=97))}
[S]
sal(saloon, sale)
mak(saloon, DAEWOO(country>korea))
col(saloon, bck)
[/S]
;=====
;===== CRLcats
;=====
; A vendre voiture DAEWOO espero modèle 92
toutes options sauf la boîte en bon état
3800 dinars
; {a:give-
information+concept(exception=discourse,
action=e_sell, vehicle-spec=(car, vehicle-
make=DAEWOO, vehicle-model=espero, vehicle-
age=, vehicle-price=, vehicle-color=,
vehicle-condition=, vehicle-assurance=,
vehicle-controle=, vehicle-air-condition=,
vehicle-size=, vehicle-shape=, vehicle-
motor-type=, vehicle-hand=, vehicle-
nationality=, vehicle-mileage=),
time=(md=92))}
[S]
sal(saloon, sale)
mak(saloon, DAEWOO(country>korea))
mod(saloon,
espero(country>koreaF,make>DAEWOO))
[/S]
;=====
;===== CRLcats
;=====
; A vendre SUBARU modèle 1983 toutes
options sauf la boîte 0795131719
; {a:give-
information+concept(exception=discourse,
action=e_sell, vehicle-spec=(, vehicle-
make=SUBARU, vehicle-model=, vehicle-age=,
vehicle-price=, vehicle-color=, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=1983))}
[S]
sal(saloon, sale)
mak(saloon, SUBARU(country>japan))
[/S]
;=====
;===== CRLcats
;=====
; A vendre SUZUKI modèle 97 en bon état
très économique 2700 dinars ^ débattre
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=SUZUKI, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=97))}
[S]
sal(saloon, sale)
mak(saloon, SUZUKI(country>japan))
[/S]
;=====
;===== CRLcats
;=====
; A vendre VOLKSWAGEN GOLF modèle 94 toutes
options sauf la boîte et la climatisation
; {a:give-
information+concept(exception=discourse,
action=e_sell, vehicle-spec=(, vehicle-
make=VOLKSWAGEN, vehicle-model=GOLF,
vehicle-age=, vehicle-price=, vehicle-
color=, vehicle-condition=, vehicle-
assurance=, vehicle-controle=, vehicle-air-
condition=, vehicle-size=, vehicle-shape=,
vehicle-motor-type=, vehicle-hand=,
vehicle-nationality=, vehicle-mileage=),
time=(md=94))}
[S]
sal(saloon, sale)
mak(saloon,
VOLKSB.M.W(country>germany,country>e
urope)AGEN)
mod(saloon,
Golf(country>germany,country>europe,
make>VOLKS))
[/S]
;=====
;===== CRLcats
;=====
; A vendre VOLKSWAGEN POLO m 2003 grise
avec climatisation contrTMle complet prix
définitif 10000 dinars
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=VOLKSWAGEN, vehicle-
model=, vehicle-age=2003, vehicle-price=,
vehicle-color=grey, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=air_condition,
vehicle-size=, vehicle-shape=, vehicle-
motor-type=, vehicle-hand=, vehicle-
nationality=, vehicle-mileage=))}
[S]
sal(saloon, sale)
mak(saloon,
VOLKSB.M.W(country>germany,country>e
urope)AGEN)
col(saloon, silver)
fea(saloon, air_cditi)
yea(saloon, 2003)
[/S]
;=====
;===== CRLcats
;=====
; A vendre MAZDA sport modèle 85 toutes
options sauf la boîte 2400 dinars
; {a:give-
information+concept(exception=discourse,
action=e_sell, vehicle-spec=(, vehicle-
make=MAZDA, vehicle-model=, vehicle-age=,
vehicle-price=, vehicle-color=, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,

```

```

vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=85))}
[S]
sal(saloon, sale)
mak(saloon, MAZDA(country>japan))
[/S]
;=====
;===== CRLcats
;=====
; A vendre MITSUBISHI LANCER modèle 90
contrTMle OK ouverture et rétro électriques
prix demandé 4500 dinars
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make= Mitsubishi, vehicle-
model=LANCER, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=90))}
[S]
sal(saloon, sale)
mak(saloon, MITSUBISHI(country>japan))
mod(saloon,
Lancer(country>japan,make>MITSUBISHI
))
[/S]
;=====
;===== CRLcats
;=====
; A vendre MERCEDES C180 modèle 95 noire
automatique contrTMle complet
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=MERCEDES, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=black, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=95))}
[S]
sal(saloon, sale)
mak(saloon,
MERCEDES(country>germany,country>eur
ope))
col(saloon, bck)
[/S]
;=====
;===== CRLcats
;=====
; A vendre MITSUBISHI coupé modèle 99
ouverture automatique 1500 dinars
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=MITSUBISHI, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=99))}
[S]
sal(saloon, sale)
mak(saloon, MITSUBISHI(country>japan))
[/S]
;=====
;===== CRLcats

```

```

=====
; A vendre voiture DAEWOO mod 94 toutes
options sauf la boite, couleur marron
; {a:give-
information+concept(exception=discourse,
action=e_sell, vehicle-spec=(car, vehicle-
make=DAEWOO, vehicle-model=, vehicle-
age=94, vehicle-price=, vehicle-
color=brown, vehicle-condition=, vehicle-
assurance=, vehicle-controle=, vehicle-air-
condition=, vehicle-size=, vehicle-shape=,
vehicle-motor-type=, vehicle-hand=,
vehicle-nationality=, vehicle-mileage=))}
[S]
sal(saloon, sale)
mak(saloon, DAEWOO(country>korea))
col(saloon, 4doors)
yea(saloon, 94)
[/S]
;=====
;===== CRLcats
;=====
; A vendre voiture de sport SUBARU modèle
2001 double turbo couleur très spéciale
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(car, vehicle-make=SUBARU, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=2001))}
[S]
sal(saloon, sale)
mak(saloon, SUBARU(country>japan))
[/S]
;=====
;===== CRLcats
;=====
; A vendre voiture samsung modèle 99 du
concessionnaire en très bon état toutes
options
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(car, vehicle-make=samsung, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=99))}
[S]
sal(saloon, sale)
mak(saloon, samsung(country>korea))
[/S]
;=====
;===== CRLcats
;=====
; A vendre VOLVO 240 avec climatisation
boite automatique 3000 dinars assurance
complète contrTMle m 1983
; {a:give-information+concept}
[S]
[/S]
;=====
;===== CRLcats
;=====
; A vendre voiture cresseda 1900 CC modèle
86 ou échange avec MERCEDES blanche m 82

```

```

; {a:give-
information+concept(action=e_sell, vehicle-
spec=(car, vehicle-make=cresseda, vehicle-
model=, vehicle-age=82, vehicle-price=,
vehicle-color=whitehe, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=1900,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=86))}
[S]
sal(saloon, sale)
mak(saloon, cresseda(country>japan))
mot(saloon, 1900)
yea(saloon, 82)
[/S]
;=====
;===== CRLcats
; A vendre MITSUBISHI LANCER 89 bleu ciel
contrTMle complet douane payée récemment
prix définitif 4800 dinars
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make= Mitsubishi, vehicle-
model=LANCER, vehicle-age=, vehicle-price=,
vehicle-color=blue, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
sal(saloon, sale)
mak(saloon, MITSUBISHI(country>japan))
mod(saloon,
Lancer(country>japan,make>MITSUBISHI
))
col(saloon, blue)
[/S]
;=====
;===== CRLcats
; A vendre voiture MITSUBISHI LANCER modèle
82 en bon état prix définitif 3000 dinars
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(car, vehicle-make=MITSUBISHI,
vehicle-model=LANCER, vehicle-age=,
vehicle-price=, vehicle-color=, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=82))}
[S]
sal(saloon, sale)
mak(saloon, MITSUBISHI(country>japan))
mod(saloon,
Lancer(country>japan,make>MITSUBISHI
))
[/S]
;=====
;===== CRLcats
; A vendre MERCEDES modèle 84 toutes
options sauf la boîte avec climatisation et
ouverture centrale 5350 dinars
; {a:give-
information+concept(exception=discourse,
action=e_sell, vehicle-spec=(, vehicle-
make=MERCEDES, vehicle-model=, vehicle-
age=, vehicle-price=, vehicle-color=,
vehicle-condition=, vehicle-assurance=,
vehicle-controle=, vehicle-air-condition=,
vehicle-size=, vehicle-shape=, vehicle-
motor-type=, vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(operator=disjunct,
[(md=84), (md=5350)]))}
[S]
sal(saloon, sale)
mak(saloon,
MERCEDES(country>germany,country>eur
ope))
fea(saloon, air_cdti)
[/S]
;=====
;===== CRLcats
; A vendre voiture MERCEDES C180 modèle 98
toutes options sauf l'ouverture, couleur
grise 18500 dinars
; {a:give-
information+concept(exception=discourse,
action=e_sell, vehicle-spec=(car, vehicle-
make=MERCEDES, vehicle-model=, vehicle-
age=, vehicle-price=, vehicle-color=grey,
vehicle-condition=, vehicle-assurance=,
vehicle-controle=, vehicle-air-condition=,
vehicle-size=, vehicle-shape=, vehicle-
motor-type=, vehicle-hand=, vehicle-
nationality=, vehicle-mileage=),
time=(md=98))}
[S]
sal(saloon, sale)
mak(saloon,
MERCEDES(country>germany,country>eur
ope))
col(saloon, silver)
[/S]
;=====
;===== CRLcats
; A vendre KIA CIVIC toutes options boîte
automatique
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=KIA, vehicle-
model=CIVIC, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
sal(saloon, sale)
mak(saloon, KIA(country>korea))
mod(saloon,
Civic(country>japan,make>HONDA))
[/S]
;=====
;===== CRLcats
; A vendre NISSAN MAXIMA mod 97 toutes
options cuir ouverture 11200 dinars
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=NISSAN, vehicle-
model=MAXIMA, vehicle-age=97, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,

```

```

vehicle-mileage=));
[S]
sal(saloon, sale)
mak(saloon, NISSAN(country>japan))
mod(saloon,
    MAx(country>France, country>europe, ma
ke>CITROEN)IMA)
yea(saloon, 97)
[/S]
;=====
;=====
;===== CRLcats
;=====
; A vendre voiture HONDA CIVIC mod 1997
boite automatique avec climatisation, 7750
dinars
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(car, vehicle-make=HONDA, vehicle-
model=CIVIC, vehicle-age=1997, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
sal(saloon, sale)
mak(saloon, HONDA(country>japan))
mod(saloon,
    Civic(country>japan, make>HONDA))
yea(saloon, 1997)
[/S]
;=====
;=====
;===== CRLcats
;=====
; A vendre HYUNDAI Accent mod 1995 toutes
options sauf la boite 5000 dinars
; {a:give-
information+concept(exception=discourse,
action=e_sell, vehicle-spec=(, vehicle-
make=HYUNDAI, vehicle-model=, vehicle-
age=1995, vehicle-price=, vehicle-color=,
vehicle-condition=, vehicle-assurance=,
vehicle-controle=, vehicle-air-condition=,
vehicle-size=, vehicle-shape=, vehicle-
motor-type=, vehicle-hand=, vehicle-
nationality=, vehicle-mileage=))}
[S]
sal(saloon, sale)
mak(saloon, HYUNDAI(country>korea))
yea(saloon, 1995)
[/S]
;=====
;=====
;===== CRLcats
;=====
; A vendre HYUNDAI mod 2001 toutes options
sauf la boite, 8900 dinars
; {a:give-
information+concept(exception=discourse,
action=e_sell, vehicle-spec=(, vehicle-
make=HYUNDAI, vehicle-model=, vehicle-
age=2001, vehicle-price=, vehicle-color=,
vehicle-condition=, vehicle-assurance=,
vehicle-controle=, vehicle-air-condition=,
vehicle-size=, vehicle-shape=, vehicle-
motor-type=, vehicle-hand=, vehicle-
nationality=, vehicle-mileage=))}
[S]
sal(saloon, sale)
mak(saloon, HYUNDAI(country>korea))
yea(saloon, 2001)
[/S]
;=====
;=====

```

```

;===== CRLcats
;=====
; A vendre BMW 520 mod 1992 grise toutes
options en bon état, 14000 dinars
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=BMW, vehicle-
model=1992, vehicle-age=1992, vehicle-
price=, vehicle-color=grey, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
sal(saloon, sale)
mak(saloon,
    B.M.W(country>germany, country>europe
))
mod(saloon, 1992)
col(saloon, silver)
yea(saloon, 1992)
[/S]
;=====
;=====
;===== CRLcats
;=====
; A vendre FORD GT neuve mod 2005 moteur
4600 couleur rouge toutes options
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=FORD, vehicle-model=,
vehicle-age=2005, vehicle-price=, vehicle-
color=red, vehicle-condition=, vehicle-
assurance=, vehicle-controle=, vehicle-air-
condition=, vehicle-size=, vehicle-shape=,
vehicle-motor-type=, vehicle-hand=,
vehicle-nationality=, vehicle-mileage=))}
[S]
sal(saloon, sale)
mak(saloon, FORD(country>Ameria))
col(saloon, red)
yea(saloon, 2005)
[/S]
;=====
;=====
;===== CRLcats
;=====
; A vendre Caprice mod 86 modernisée 90
jantes aluminum nouvelle peinture prix
définitif 4600 dinars
; {a:give-information+concept( numeral=[86],
who=she)}
[S]
[/S]
;=====
;=====
;===== CRLcats
;=====
; A vendre KIA Sephia mod 96 verte contrTMle
complet toutes options prix définitif 5250
tél 95918898
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=KIA, vehicle-
model=Sephia, vehicle-age=96, vehicle-
price=, vehicle-color=green, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
sal(saloon, sale)
mak(saloon, KIA(country>korea))
mod(saloon,

```

```

        Sephia(country>koreF,make>KIA))
col(saloon, green)
yea(saloon, 96)
[/S]
;=====
;===== CRLcats
;=====
; A vendre KIA Sephia mod 96 toutes options
avec boite automatique couleur verte
contrTMle complet 5800 dinars
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=KIA, vehicle-
model=Sephia, vehicle-age=96, vehicle-
price=, vehicle-color=green, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-condition=, vehicle-
size=, vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
sal(saloon, sale)
mak(saloon, KIA(country>korea))
mod(saloon,
        Sephia(country>koreF,make>KIA))
col(saloon, green)
yea(saloon, 96)
[/S]
;=====
;===== CRLcats
;=====
; A vendre LAND ROVER DISCOVER mod 2001
toutes options 8 cylindres état original
tél 0796635036
; {a:give-information+concept}
[S]
[/S]
;=====
;===== CRLcats
;=====
; A vendre MERCEDES mod 76 modernisée 84
avec ouverture climatisation automatique
cuir couleur grise
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=MERCEDES, vehicle-
model=, vehicle-age=76, vehicle-price=,
vehicle-color=grey, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
sal(saloon, sale)
mak(saloon,
        MERCEDES(country>germany,country>eur
ope))
col(saloon, silver)
yea(saloon, 76)
[/S]
;=====
;===== CRLcats
;=====
; A vendre HONDA Accord mod 91 couleur
blanche avec climatisation automatique
contrTMle complet
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=HONDA, vehicle-
model=Accd, vehicle-age=91, vehicle-price=,
vehicle-color=whitehe, vehicle-
condition=, vehicle-assurance=, vehicle-
controle=, vehicle-air-
condition=air_condition, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=))}
[S]
sal(saloon, sale)
mak(saloon, HONDA(country>japan))
fea(saloon, air_cdti)
yea(saloon, 91)
[/S]
;=====
;===== CRLcats
;=====
; A vendre MERCEDES m 1989 contrTMle complet
toutes options sauf l'ouverture 0795517487
; {a:give-
information+concept(exception=discourse,
action=e_sell, vehicle-spec=(, vehicle-
make=MERCEDES, vehicle-model=, vehicle-
age=1989, vehicle-price=, vehicle-color=,
vehicle-condition=, vehicle-assurance=,
vehicle-controle=, vehicle-air-condition=,
vehicle-size=, vehicle-shape=, vehicle-
motor-type=, vehicle-hand=, vehicle-
nationality=, vehicle-mileage=))}
[S]
sal(saloon, sale)
mak(saloon,
        MERCEDES(country>germany,country>eur
ope))
yea(saloon, 1989)
[/S]
;=====
;===== CRLcats
;=====
; A vendre LADA SAMARA modèle 88 couleur
marron 900 dinars
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=LADA, vehicle-
model=SAMARA, vehicle-age=, vehicle-price=,
vehicle-color=brown, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=88))}
[S]
sal(saloon, sale)
mak(saloon, Lada)
mod(saloon, SAMARA(make>Lada))
col(saloon, 4doors)
[/S]
;=====
;===== CRLcats
;=====
; A vendre MAZDA noire contrTMlée et assurée
prix définitif 2500 dinars son état est
très bonne
; {a:give-
information+vehicle(action=e_sell, vehicle-
spec=(, vehicle-make=MAZDA, vehicle-model=,
vehicle-age=, vehicle-price=, vehicle-
color=black, vehicle-condition=, vehicle-
assurance=, vehicle-controle=, vehicle-air-
condition=, vehicle-size=, vehicle-shape=,
vehicle-motor-type=, vehicle-hand=,
vehicle-nationality=, vehicle-mileage=))}
[S]

```

```

sal(saloon, sale)
mak(saloon, MAZDA(country>japan))
col(saloon, bck)
[/S]
=====
;===== CRLcats
=====
; A vendre DAEWOO NUBIRA mod 2000 prix
intéressant +0795200986 MITSUBISHI LANCER
2003
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(, vehicle-make=DAEWOO, vehicle-
model=NUBIRA, vehicle-age=2000, vehicle-
price=, vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=)}}
[S]
sal(saloon, sale)
mak(saloon, DAEWOO(country>korea))
mod(saloon,
Nubira(country>koreF,make>DAEWOO))

```

```

yea(saloon, 2000)
[/S]
=====
;===== CRLcats
=====
; A vendre pickup NISSAN diesel modèle 1985
assuré jusqu'au 8/12/2005 contrTMle complet.
; {a:give-
information+concept(action=e_sell, vehicle-
spec=(pickup, vehicle-make=NISSAN, vehicle-
model=, vehicle-age=, vehicle-price=,
vehicle-color=, vehicle-condition=,
vehicle-assurance=, vehicle-controle=,
vehicle-air-condition=, vehicle-size=,
vehicle-shape=, vehicle-motor-type=,
vehicle-hand=, vehicle-nationality=,
vehicle-mileage=), time=(md=1985))}
[S]
sal(pickup, sale)
mak(pickup, NISSAN(country>japan))
[/S]
=====

```

Annexe V. Résultats d'extraction de contenu par la méthode de TA

```

===== UNL
=====
البنزين او ديزل دينار 90050 6500 من سياره
/0795601711
[S]
pri(saloon:00, 400:09)
[/S]
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
المانى او يابانى paiement من بالاقصاط سياره مطلوب
0795777953
[S]
wan(saloon:06, wanted:00)
cou(saloon:06, japan:0X)
cou(saloon:06, germany:18)
yea(saloon:06, 0795777953:1E)
[/S]
=====
;;Time 0.2 Sec
;;Done!
===== UNL
=====
فكترا اوبل سياره مطلوب
[S]
wan(saloon:06, wanted:00)
mak(saloon:06,
      OPEL(country>germany, country>europe)
:0C)
mod(saloon:06,
      Vectra(country>germany, country>europ
e, make>OPEL):0H)
[/S]
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
مقبول وبسعر 99-2000 موديل سياره شراء مطلوب
[S]
wan(saloon:0B, wanted:00)
yea(saloon:0B, 99:0N)
yea(saloon:0B, 2000:0Q)
[/S]
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
100 مع 85 من partir 90050 موديل سياره مطلوب
دفعه بدون شهري دينار
[S]
wan(saloon:06, wanted:00)
yea(saloon:06, 90050:0J)
yea(saloon:06, 85:0Z)
pri(saloon:06, 100:15)
[/S]
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
سياره مطلوب
[S]
wan(saloon:06, wanted:00)
[/S]
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
سياره مطلوب
[S]
wan(saloon:06, wanted:00)
[/S]
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
ترانسپورتر كيا باص مطلوب
commun paiement
fractionné
[S]
wan(bus:06, wanted:00)
mak(bus:06, KIA(country>korea):0A)
fea(bus:06, power steering:0J)
[/S]
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
او هونداي او كيا commun ترانسپورتر باص مطلوب
اقساط الباقى و trois mille بدفعه مع يونج سانج
[S]
wan(bus:06, wanted:00)
fea(bus:06, power steering:0F)
mak(bus:06, KIA(country>korea):0S)
mak(bus:06, HYUNDAI(country>korea):0Z)
[/S]
=====
;;Time 0.2 Sec
;;Done!
===== UNL
=====
دينار 2000 1900 شيروكي جيبي مطلوب
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, JEEP(country>America):06)
mod(saloon:00,
      Cherokee(country>America, make<JEEP):
0A)
yea(saloon:00, 2000:0H)
pri(saloon:00, 1900:0M)
[/S]
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
اوبشن فل بطه classique مرسيدس
[S]
mak(saloon:00,
      MERCEDES(country>germany, country>eur
ope):00)
[/S]

```



```

=====
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
;مرسيدس سيارة مطلوب سي1300 ، سياره a مطلوب اريد;
دينار 85 2500 موديل
[S]
[W]
""
[/W]
[/S]
=====
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
;90050 condition que le امركيه سياره مطلوب;
دينار 1000 قابل غير ne dépasse بس عر
[S]
wan(saloon:06, wanted:00)
cou(saloon:06, Ameria:0D)
pri(saloon:06, 90050:0L)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
;دايو مطلوب;
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, DAEWOO(country>korea):06)
[/S]
=====
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
;93 موديل عادي صرني نيسان مطلوب;
90050 97
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, NISSAN(country>japan):06)
mod(saloon:00,
Sunny(country>japan,make>NISSAN):0C)
mak(saloon:00,
AUDI(country>germany,country>europe)
:0H)
yea(saloon:00, 93:0R)
yea(saloon:00, 90050:0V)
yea(saloon:00, 97:11)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
;95 موديل عادي لكليو رينو مطلوب;
90050 2000
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
RENAULT(country>France,country>europ
e):06)
mod(saloon:00,
Clio(country>France,country>europe,m
ake>RENAULT):0B)
mak(saloon:00,
AUDI(country>germany,country>europe)
:0H)
yea(saloon:00, 95:0R)
yea(saloon:00, 90050:0V)

```

```

yea(saloon:00, 2000:11)
[/S]
=====
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
;سياره هونداي مطلوب;
[S]
wan(saloon:06, wanted:00)
mak(saloon:06, HYUNDAI(country>korea):0B)
[/S]
=====
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
;متسوبيشي سياره مطلوب;
[S]
wan(saloon:06, wanted:00)
mak(saloon:06,
MITSUBISHI(country>japan):0C)
[/S]
=====
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
;92 موديل دبليو ام بي سياره a مطلوب اريد;
[S]
[W]
""
[/W]
[/S]
=====
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
;دايو مطلوب;
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, DAEWOO(country>korea):06)
[/S]
=====
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
;عادي 95 90050 موديل صرني نيسان مطلوب;
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, NISSAN(country>japan):06)
mod(saloon:00,
Sunny(country>japan,make>NISSAN):0C)
yea(saloon:00, 93:0M)
yea(saloon:00, 90050:0Q)
yea(saloon:00, 95:0W)
mak(saloon:00,
AUDI(country>germany,country>europe)
:10)
[/S]
=====
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
;عادي اوبشن فل 94 لمودي سيفك-هوندا مطلوب;
[S]
wan(saloon:00, wanted:00)

```

```

mak(saloon:00, HONDA(country>japan):06)
mak(saloon:00,
    AUDI(country>germany, country>europe)
:10)
[/S]
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
سيفيا لثيا مطلوب
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, KIA(country>korea):06)
mod(saloon:00,
    Sephia(country>korea, make>KIA):0A)
[/S]
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
عادي 97 90050 95 لانوس دايو مطلوب
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, DAEWOO(country>korea):06)
mod(saloon:00,
    Lanos(country>korea, make>DAEWOO):0B)
yea(saloon:00, 95:0H)
yea(saloon:00, 90050:0L)
yea(saloon:00, 97:0R)
mak(saloon:00,
    AUDI(country>germany, country>europe)
:0V)
[/S]
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
سامارا لادا مطلوب
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, Lada:06)
mod(saloon:00, SAMARA(make>Lada):0B)
[/S]
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
سياره مورن دايو مطلوب
[S]
wan(saloon:06, wanted:00)
mak(saloon:06, HYUNDAI(country>korea):0B)
[/S]
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
الفسل سياره مطلوب
[S]
wan(saloon:06, wanted:00)
mod(saloon:06,
    Excel(country>korea, make>HYUNDAI):0C
)
[/S]
=====
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
لادا او شكودا مطلوب
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, SKODA(country>europe):06)
mak(saloon:00, Lada:0F)
[/S]
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
شهرى دينار 4 آلاف بدفعه مع جيب 90050 مطلوب
200 من des للشرء 0795291030
[S]
[W]
wanted:00
[/W]
[/S]
[S]
[W]
90050:07
[/W]
[/S]
[S]
[W]
JEEP(country>Ameria):0H
[/W]
[/S]
[S]
[W]
4:0U
[/W]
[/S]
[S]
[W]
"ل"
[/W]
[/S]
[S]
[W]
Dinar:11
[/W]
[/S]
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
يفضل ويفضل امريكيه سياره مطلوب
[S]
wan(saloon:06, wanted:00)
cou(saloon:06, Ameria:0D)
[/S]
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
كاليبرا اوبل يفضل سيور سياره مطلوب
[S]
wan(saloon:06, wanted:00)
fea(saloon:06, power steering:0D)
mak(saloon:06,
    OPEL(country>germany, country>europe)
:0M)
[/S]
=====
=====

```

```

;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
316 دبل يو ام بي مطلوب
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    B.M.W(country>germany, country>europe
):06)
mod(saloon:00, 316:0I)
[/S]
;=====
;=====
;;Time 0.0 Sec
;;Done!
;===== UNL
;=====
2000 موديل CLK مرسيدس 1'achat ل90050 مطلوب
نم 99
[S]
wan(saloon:00, wanted:00)
mod(saloon:00, 90050:07)
mak(saloon:00,
    MERCEDES(country>germany, country>eur
ope):0L)
yea(saloon:00, 2000:12)
yea(saloon:00, 99:1B)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
316 دبل يو ام بي مطلوب
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    B.M.W(country>germany, country>europe
):06)
mod(saloon:00, 316:0I)
[/S]
;=====
;=====
;;Time 0.0 Sec
;;Done!
;===== UNL
;=====
القدس او 1990 تويوتا باص 1'achat ل90050 مطلوب
numero ت 0795133829
[S]
mak(bus:0L, TOYOTA(country>japan):0P)
yea(bus:0L, 1990:0W)
yea(bus:0L, 0795133829:1J)
[/S]
[S]
mak(bus:0L, TOYOTA(country>japan):0P)
yea(bus:0L, 1990:0W)
yea(bus:0L, 0795133829:1J)
[/S]
[S]
mak(bus:0L, TOYOTA(country>japan):0P)
yea(bus:0L, 1990:0W)
yea(bus:0L, 0795133829:1J)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
92 موديل صرني نيسان 1'achat ل90050 مطلوب
ل90050 95
[S]
mak(saloon:00, NISSAN(country>japan):0L)

```

```

mod(saloon:00,
    Sunny(country>japan, make>NISSAN):0R)
yea(saloon:00, 92:11)
yea(saloon:00, 90050:15)
yea(saloon:00, 95:1B)
[/S]
[S]
mak(saloon:00, NISSAN(country>japan):0L)
mod(saloon:00,
    Sunny(country>japan, make>NISSAN):0R)
yea(saloon:00, 92:11)
yea(saloon:00, 90050:15)
yea(saloon:00, 95:1B)
[/S]
[S]
mak(saloon:00, NISSAN(country>japan):0L)
mod(saloon:00,
    Sunny(country>japan, make>NISSAN):0R)
yea(saloon:00, 92:11)
yea(saloon:00, 90050:15)
yea(saloon:00, 95:1B)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
m5 دبل يو ام بي للبي ع مطلوب
[S]
wan(saloon:00, wanted:00)
sal(saloon:00, sale:06)
mak(saloon:00,
    B.M.W(country>germany, country>europe
):0C)
mod(saloon:00, 5:0P)
[/S]
;=====
;=====
;;Time 0.0 Sec
;;Done!
;===== UNL
;=====
او 98 موديل اوتوماتيك جيير فحوص سيك-هوندا مطلوب
اقساط اول دفعه مع 99
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, HONDA(country>japan):06)
fea(saloon:00, automatic gear:0L)
yea(saloon:00, 98:15)
yea(saloon:00, 99:1B)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
مرسيدس مطلوب
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    MERCEDES(country>germany, country>eur
ope):06)
[/S]
;=====
;=====
;;Time 0.0 Sec
;;Done!
;===== UNL
;=====
فل كامل فحوص 92 موديل LIMINZ دا يو سياره بي ع اردي
اوبشن
[S]
[W]
"
```

```

[/W]
[/S]
=====
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
;دينار 3500 بسعر سياره مطلوب
[S]
wan(saloon:06, wanted:00)
pri(saloon:06, 3500:0H)
pri(saloon:06, 90050:0N)
pri(saloon:06, 3850:0T)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
للشرا سياره
[S]
[W]
saloon:00
[/W]
[/S]
=====
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
مريس يدس مطلوب
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
MITSUBISHI(country>japan):06)
[/S]
=====
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
بحاله 86 ل90050 80 موديل جالنت ميتسوبيشي مطلوب
les اوبشن فل مع جديده
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
MITSUBISHI(country>japan):06)
mod(saloon:00,
Galant(country>japan,make>MITSUBISHI
):0G)
yea(saloon:00, 80:0S)
yea(saloon:00, 90050:0W)
yea(saloon:00, 86:12)
fea(saloon:00, good condition:15)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
مريس يدس سيارة مطلوب
[S]
wan(saloon:06, wanted:00)
mak(saloon:06,
MERCEDES(country>germany, country>eur
ope):0C)
[/S]
=====
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
سياره اعشر مطلوب
[S]
wan(saloon:0B, wanted:00)
[/S]
=====
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
دينار 3850 ل90050 بسعر سياره مطلوب
[S]
wan(saloon:06, wanted:00)
pri(saloon:06, 3500:0H)
pri(saloon:06, 90050:0N)
pri(saloon:06, 3850:0T)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
للشرا سياره
[S]
[W]
saloon:00
[/W]
[/S]
=====
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
مريس يدس مطلوب
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
MERCEDES(country>germany, country>eur
ope):06)
[/S]
=====
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
سيفيا كيا مطلوب
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, KIA(country>korea):06)
mod(saloon:00,
Sephia(country>korea,make>KIA):0A)
[/S]
=====
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
90050 ل سيفيا كيا او افانتي او الكسنت مطلوبونداي
3800 دینار
[S]
wan(saloon:00, wanted:00)
mod(saloon:00,
Ax(country>France, country>europe, mak
e>CITROEN):09)
mak(saloon:00, KIA(country>korea):0V)
mod(saloon:00,
Sephia(country>korea,make>KIA):0Z)
yea(saloon:00, 90050:16)
pri(saloon:00, 3800:1C)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
2005 مريس يدس 1'achat ل90050 مطلوب

```

```

[S]
wan(saloon:00, wanted:00)
mod(saloon:00, 90050:07)
mak(saloon:00,
    MERCEDES(country>germany, country>eur
ope):0L)
yea(saloon:00, 2005:0S)
[/S]
=====
;;Time 0.0 Sec
;;Done!
=====
UNL
=====
دايو سياره مطلوب;
[S]
wan(saloon:06, wanted:00)
mak(saloon:06, DAEWOO(country>korea):0C)
[/S]
=====
;;Time 0.1 Sec
;;Done!
=====
UNL
=====
جرسيديس من وشراء 1990 جوديل فنكترا اوبل من بيخ;
CD200 جوديل 1995
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
    OPEL(country>germany, country>europe)
:07)
mod(saloon:00,
    Vectra(country>germany, country>europ
e, make>OPEL):0C)
mak(saloon:00,
    MERCEDES(country>germany, country>eur
ope):12)
yea(saloon:00, 200:1B)
yea(saloon:00, 1995:1L)
[/S]
=====
;;Time 0.1 Sec
;;Done!
=====
UNL
=====
1995 جوديل 200 مرسيدس مطلوب;
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    MERCEDES(country>germany, country>eur
ope):06)
mod(saloon:00, 200:0D)
yea(saloon:00, 1995:0N)
[/S]
=====
;;Time 0.0 Sec
;;Done!
=====
UNL
=====
جديه بحاله دينار 1500 نتج اوز لا لس ياره بحاجه;
الملك ومن واق تصراديه
[S]
[W]
" "
[/W]
[/S]
=====
;;Time 0.0 Sec
;;Done!
=====
UNL
=====
لثي س ياره مطلوب;

```

```

[S]
wan(saloon:06, wanted:00)
mak(saloon:06, KIA(country>korea):0C)
[/S]
=====
;;Time 0.0 Sec
;;Done!
=====
UNL
=====
أو من 86 جوديل دبليو ام بي l'achat ل90050 مطلوب;
90 جوديل لثي 96
[S]
wan(saloon:00, wanted:00)
mod(saloon:00, 90050:07)
mak(saloon:00,
    B.M.W(country>germany, country>europe)
):0L)
yea(saloon:00, 86:13)
mak(saloon:00, KIA(country>korea):1B)
yea(saloon:00, 90:1L)
yea(saloon:00, 90050:1P)
yea(saloon:00, 96:1V)
[/S]
=====
;;Time 0.1 Sec
;;Done!
=====
UNL
=====
espèces بحاله payment وي فضل شفر كابر س مطلوب;
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    CHEVROLET(country>America):06)
mod(saloon:00, Caprice
Classic(country>America, make>CHEVROLET):09)
[/S]
=====
;;Time 0.0 Sec
;;Done!
=====
UNL
=====
نيسان اب بك مطلوب;
[S]
wan(pick up:06, wanted:00)
mak(pick up:06,
    NISSAN(country>japan):0C)
[/S]
=====
;;Time 0.0 Sec
;;Done!
=====
UNL
=====
الاضافات جمعي 82 جوديل مرسيدس souhaite اريدي;
[S]
[W]
" "
[/W]
[/S]
=====
;;Time 0.0 Sec
;;Done!
=====
UNL
=====
95-96-97 جوديل نيسان با بك مطلوب;
[S]
wan(pick up:06, wanted:00)
mak(pick up:06,
    NISSAN(country>japan):0C)
yea(pick up:06, 95:00)
yea(pick up:06, 96:0R)
yea(pick up:06, 97:0U)

```

```

[/S]
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
اقساط الباقى و 1000 دفعه سياره مطلوب;
[S]
wan(saloon:06, wanted:00)
yea(saloon:06, 1000:0H)
[/S]
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
و جیده بحاله دينار 2000 عن تزييد لا لسياره بحاج;
البنزين في واقصراديه
[S]
[W]
""
[/W]
[/S]
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
اقساط الباقى و 1000 دفعه دايمو سياره مطلوب;
[S]
wan(saloon:06, wanted:00)
mak(saloon:06, DAEWOO(country>korea):0C)
yea(saloon:06, 1000:0M)
[/S]
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
91-92-93-94-95 موديل ستيتشين لادا مطلوب;
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, Lada:06)
yea(saloon:00, 91:00)
yea(saloon:00, 92:0R)
yea(saloon:00, 93:0U)
yea(saloon:00, 94:0X)
yea(saloon:00, 95:10)
[/S]
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
كامل فحص 95 94 اوڤيا 90 89م جولف مطلوب;
[S]
wan(saloon:00, wanted:00)
mod(saloon:00,
    Golf(country>germany,country>europe,
    make>VOLKS WAGEN):06)
yea(saloon:00, 89:0C)
yea(saloon:00, 90:0F)
mak(saloon:00, KIA(country>korea):0K)
yea(saloon:00, 94:00)
yea(saloon:00, 95:0R)
fea(saloon:00, total check:0U)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
اقساط والباقي دينار 500 بدفحه سياره a بشراء أرغب;
دينار 100 شهريه
[S]
[W]
""
[/W]
[/S]
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
جولف سياره مطلوب;
[S]
wan(saloon:06, wanted:00)
mod(saloon:06,
    Golf(country>germany,country>europe,
    make>VOLKS WAGEN):0C)
[/S]
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
سيڤيا كيا مطلوب;
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, KIA(country>korea):06)
mod(saloon:00,
    Sephia(country>korea,make>KIA):0A)
[/S]
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
جیده بحاله و عادي جير بيشتستو سياره مطلوب;
[S]
wan(saloon:06, wanted:00)
mak(saloon:06,
    MITSUBISHI(country>japan):0C)
fea(saloon:06, manual gear:0I)
fea(saloon:06, good condition:0W)
[/S]
=====
;;Time 0.0 Sec
;;Done!
===== UNL
=====
90050ل 1988 موديل جالنت ميتسوبيشي سياره مطلوب;
جیده بحاله و عادي جير 1991
[S]
wan(saloon:06, wanted:00)
mak(saloon:06,
    MITSUBISHI(country>japan):0C)
mod(saloon:06,
    Galant(country>japan,make>MITSUBISHI
    ):0M)
yea(saloon:06, 1988:0Y)
yea(saloon:06, 90050:14)
yea(saloon:06, 1991:1A)
fea(saloon:06, manual gear:1F)
fea(saloon:06, good condition:1Q)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====

```

```

===== UNL wanted:00
===== [/W]
;دينار 3800 90050 ل أفانتي مظلوبونداي [S]
wan(van:0D, wanted:00) [W]
yea(van:0D, 90050:0K) 90050:0D
pri(van:0D, 3800:0Q) [/W]
[/S] [/S]
=====
=====
;;Time 0.1 Sec ;Time 0.0 Sec
;;Done! ;Done!
===== UNL ;===== UNL
=====
97 موديل لانسر متسوبيشي او سيفتيك اموند مظلوب; 307 SALFA بيجو مظلوب;
ل90050 99 بدفغه 4 الف ديوار اقساط والباقي ديوار 4 بدفغه 99 [S]
wan(saloon:00, wanted:00) wan(saloon:00, wanted:00)
mak(saloon:00, HONDA(country>japan):06) mak(saloon:00, PEUGEOT(country>France, country>europ
mod(saloon:00, e):06)
Civic(country>japan, make>HONDA):0C) mod(saloon:00, 307:0B)
mak(saloon:00, MITSUBISHI(country>japan):0L) [/S]
mod(saloon:00, Lancer(country>japan, make>MITSUBISHI) ;=====
):0U) ;Time 0.0 Sec
yea(saloon:00, 97:16) ;Done!
yea(saloon:00, 90050:1A) ;===== UNL
yea(saloon:00, 99:1G) ;=====
yea(saloon:00, 4:1P) ;ديوار 98 موديل ابشن فول يوديل ام بي مظلوب;
[/S] 2003 [S]
wan(saloon:00, wanted:00)
mak(saloon:00, B.M.W(country>germany, country>europe
):06)
yea(saloon:00, 98:0X)
yea(saloon:00, 90050:11)
yea(saloon:00, 2003:17)
[/S]
=====
=====
;;Time 0.1 Sec ;Time 0.1 Sec
;;Done! ;Done!
===== UNL ;===== UNL
=====
97 موديل لانسر متسوبيشي سياره مظلوب; 99 model, with بدفغه 4-5 ديوار اقساط والباقي ديوار 4-5 من بدفغه 99 شريه
ل90050 [S]
wan(saloon:06, wanted:00)
mak(saloon:06, MITSUBISHI(country>japan):0C)
mod(saloon:06, Lancer(country>japan, make>MITSUBISHI)
):0L)
yea(saloon:06, 97:10)
yea(saloon:06, 90050:14)
yea(saloon:06, 99:1A)
yea(saloon:06, 4:1Y)
pri(saloon:06, 5:20)
[/S]
=====
=====
;;Time 0.2 Sec ;Time 0.1 Sec
;;Done! ;Done!
===== UNL ;===== UNL
=====
اوبشن فل و وجميج فحص 81 مرسيدس شراء مظلوب; [S]
wan(saloon:00, wanted:00)
mak(saloon:00, MERCEDES(country>germany, country>eur
ope):0B)
mod(saloon:00, 81:0I)
[/S]
=====
=====
;;Time 0.1 Sec ;Time 0.1 Sec
;;Done! ;Done!
===== UNL ;===== UNL
=====
دينار 3800 90050 ل مظلوبونداي [S]
[W]

```

```

=====
;بي عها بددي روفر لاند سياره سياره عندي
[S]
mod(saloon:05, Land
Rover(country>England, country>europe, make>R
over):0H)
[/S]
[S]
mod(saloon:05, Land
Rover(country>England, country>europe, make>R
over):0H)
[/S]
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
;سياره مونداي مطلوب
[S]
wan(saloon:06, wanted:00)
mak(saloon:06, HYUNDAI(country>korea):0B)
[/S]
;=====
;;Time 0.0 Sec
;;Done!
;===== UNL
=====
;دبل يو ام بي مطلوب
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
B.M.W(country>germany, country>europe
):06)
[/S]
;=====
;;Time 0.0 Sec
;;Done!
;===== UNL
=====
;موديل نيسان او م ب او جيب مطلوب
90050 partir ل لشراء 95
من 95777953
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, JEEP(country>Ameria):06)
mak(saloon:00, NISSAN(country>japan):0L)
yea(saloon:00, 90050:0Y)
yea(saloon:00, 95:1E)
yea(saloon:00, 95777953:10)
[/S]
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
;موديل ام بي سياره مطلوب
75 =77
[S]
wan(saloon:06, wanted:00)
mak(saloon:06,
B.M.W(country>germany, country>europe
):0C)
yea(saloon:06, 75:00)
yea(saloon:06, 77:0S)
[/S]
;=====
;;Time 0.0 Sec
;;Done!
;===== UNL
=====
;موديل برس يدس سياره a اري دب ع
[S]
=====
[W]
sale:04
[/W]
[/S]
;=====
;;Time 0.0 Sec
;;Done!
;===== UNL
=====
;206 بي جو مطلوب
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
PEUGEOT(country>France, country>europ
e):06)
mod(saloon:00, 206:0B)
[/S]
;=====
;;Time 0.0 Sec
;;Done!
;===== UNL
=====
;بي جو مطلوب
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
PEUGEOT(country>France, country>europ
e):06)
[/S]
;=====
;;Time 0.0 Sec
;;Done!
;===== UNL
=====
;207 بي جو مطلوب
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
PEUGEOT(country>France, country>europ
e):06)
mod(saloon:00, 207:0B)
[/S]
;=====
;;Time 0.0 Sec
;;Done!
;===== UNL
=====
;1200 من محرك 82 90050 ل 80 من موديل داتسن مطلوب
س 1400 ، 90050
[S]
wan(saloon:00, wanted:00)
mod(saloon:00,
Atos(country>korea, make>HYUNDAI):07)
yea(saloon:00, 80:0L)
yea(saloon:00, 90050:0P)
yea(saloon:00, 82:0V)
mot(saloon:00, 1200:16)
mot(saloon:00, 90050:1C)
mot(saloon:00, 1400:1K)
[/S]
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
;90050 موديل سياره 1'achat ل 90050
partir من 90 ل 3000 دينار
من بالاقصاط
paieiment
[S]
yea(saloon:0L, 90050:0Y)

```



```

yea(saloon:0L, 90:1E)
yea(saloon:0L, 90050:1I)
pri(saloon:0L, 3000:1O)
[/S]
[S]
yea(saloon:0L, 90050:0Y)
yea(saloon:0L, 90:1E)
yea(saloon:0L, 90050:1I)
pri(saloon:0L, 3000:1O)
[/S]
[S]
yea(saloon:0L, 90050:0Y)
yea(saloon:0L, 90:1E)
yea(saloon:0L, 90050:1I)
pri(saloon:0L, 3000:1O)
[/S]
[S]
=====
;;Time 0.1 Sec
;;Done!
=====
UNL
=====
; فحوص تقريبياً 80 موديل تويوتا بكم مطلوب
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, TOYOTA(country>japan):0A)
[/S]
=====
;;Time 0.1 Sec
;;Done!
=====
UNL
=====
; 90050 موديل 1600 بـ محرك فـكـتـرا اوبـل مـطـلـوب
partir من 94
[S]
wan(saloon:00, wanted:00)
mak(saloon:00,
    OPEL(country>germany, country>europe)
:06)
mod(saloon:00,
    Vectra(country>germany, country>europ
e, make>OPEL):0B)
mot(saloon:00, 1600:0N)
yea(saloon:00, 90050:0Z)
yea(saloon:00, 94:1F)
[/S]
=====
;;Time 0.1 Sec
;;Done!
=====
UNL
=====
; سـلـفـر لـون جـيـدة بـحـالـة 2003 سـيـفـك هـونـدا سـيـارة مـطـلـوب
ذـهـبي، او
[S]
wan(saloon:06, wanted:00)
mak(saloon:06, HONDA(country>japan):0C)
yea(saloon:06, 2003:0N)
fea(saloon:06, good condition:0S)
col(saloon:06, silver:17)
col(saloon:06, goldr:1F)
[/S]
=====
;;Time 0.1 Sec
;;Done!
=====
UNL
=====
; 90050 موديل سياره l'achat 90050 مـطـلـوب
اقساط و دينار 500 بـدفعـة 90 مـن
[S]
yea(saloon:0L, 90050:0Y)
pri(saloon:0L, 90:1E)
pri(saloon:0L, 500:1N)
[/S]

```

```

[S]
yea(saloon:0L, 90050:0Y)
pri(saloon:0L, 90:1E)
pri(saloon:0L, 500:1N)
[/S]
[S]
yea(saloon:0L, 90050:0Y)
pri(saloon:0L, 90:1E)
pri(saloon:0L, 500:1N)
[/S]
=====
;;Time 0.1 Sec
;;Done!
=====
UNL
=====
; 96 97 98 سـيـفـك هـونـدا مـطـلـوب
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, HONDA(country>japan):06)
yea(saloon:00, 96:0H)
yea(saloon:00, 97:0K)
yea(saloon:00, 98:0N)
[/S]
=====
;;Time 0.0 Sec
;;Done!
=====
UNL
=====
; 1994 اكتمل مطلوب هونداي
[S]
[W]
wanted:00
[/W]
[/S]
[S]
[W]
Excel(country>korea, make>HYUNDAI):0C
[/W]
[/S]
[S]
[W]
1994:0H
[/W]
[/S]
=====
;;Time 0.0 Sec
;;Done!
=====
UNL
=====
; 96 97 98 سـيـفـك هـونـدا مـطـلـوب
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, HONDA(country>japan):06)
yea(saloon:00, 96:0H)
yea(saloon:00, 97:0K)
yea(saloon:00, 98:0N)
[/S]
=====
;;Time 0.0 Sec
;;Done!
=====
UNL
=====
; او سـلـفـر لـون جـيـدة بـحـالـة 2003 سـيـفـك هـونـدا مـطـلـوب
ذـهـبي،
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, HONDA(country>japan):06)
yea(saloon:00, 2003:0H)
fea(saloon:00, good condition:0M)
col(saloon:00, silver:11)
col(saloon:00, goldr:19)
[/S]

```

```

;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
; سلفر لون جيدة بحالة 2003 سيفك مودنا مطلوب
ذهبي،
[S]
wan(saloon:00, wanted:00)
mak(saloon:00, HONDA(country>japan):06)
yea(saloon:00, 2003:0H)
fea(saloon:00, good condition:0M)
col(saloon:00, silver:11)
col(saloon:00, goldr:19)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
; 280C مرسيدس a اريديبيج
kilométrage 4000 _
[S]
[W]
sale:04
[/W]
[/S]
;=====
;=====
;;Time 0.0 Sec
;;Done!
;===== UNL
;=====
; 90 موديل جولف او 90 موديل باك هاتش كاديك اوبل
باك هاتش CITROEN او باك هاتش
[S]
mak(saloon:00,
OPEL(country>germany, country>europe)
:00)
mod(saloon:00,
kadet(country>germany, country>europe
, make>OPEL):05)
yea(saloon:00, 90:0Q)
mod(saloon:00,
Golf(country>germany, country>europe,
make>VOLKS WAGEN):0W)
[/S]
;=====
;=====
;;Time 0.2 Sec
;;Done!
;===== UNL
;=====
; بور سنتر (avec Astra أهر) اوبل سياره للبيج
vitre و rétro électrique)
[S]
sal(saloon:06, sale:00)
mak(saloon:06,
OPEL(country>germany, country>europe)
:0C)
col(saloon:06, red:0N)
fea(saloon:06, center lock:0Y)
fea(saloon:06, power steering:13)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
; 83 assuré with an موديل ديزل نيسان داتسون بكفب
دينار 90050 vendre, 2000 جیده بحاله
[S]
mod(pick up:00,
Atos(country>korea, make>HYUNDAI):05)
mak(pick up:00,
NISSAN(country>japan):0B)
yea(pick up:00, 83:0S)
fea(pick up:00, good condition:1A)
yea(pick up:00, 90050:1M)
pri(pick up:00, 2000:20)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
; بحالة 318 دبليو ام 1988, م ب ب 079532279
للشراء دينار 6250 état, قطعت ممتازة
[S]
mak(saloon:00,
B.M.W(country>germany, country>europe)
):0R)
mod(saloon:00, 318:0Z)
fea(saloon:00, good condition:14)
pri(saloon:00, 6250:1S)
[/S]
[S]
mak(saloon:00,
B.M.W(country>germany, country>europe)
):0R)
mod(saloon:00, 318:0Z)
fea(saloon:00, good condition:14)
pri(saloon:00, 6250:1S)
[/S]
[S]
mak(saloon:00,
B.M.W(country>germany, country>europe)
):0R)
mod(saloon:00, 318:0Z)
fea(saloon:00, good condition:14)
pri(saloon:00, 6250:1S)
[/S]
[S]
mak(saloon:00,
B.M.W(country>germany, country>europe)
):0R)
mod(saloon:00, 318:0Z)
fea(saloon:00, good condition:14)
pri(saloon:00, 6250:1S)
[/S]
[S]
mak(saloon:00,
B.M.W(country>germany, country>europe)
):0R)
mod(saloon:00, 318:0Z)
fea(saloon:00, good condition:14)
pri(saloon:00, 6250:1S)
[/S]
[S]
mak(saloon:00,
B.M.W(country>germany, country>europe)
):0R)
mod(saloon:00, 318:0Z)
fea(saloon:00, good condition:14)
pri(saloon:00, 6250:1S)
[/S]
;=====
;=====
;;Time 0.6 Sec
;;Done!
;===== UNL
;=====
; كامل فحص الجير عدا ما 2002 T.O
بيجوبارتي نرفان
pour renseignement 0796702818
[S]
mak(saloon:00,
PEUGEOT(country>France, country>europ

```

```

e):00)
mod(saloon:00, 2002:0F)
fea(saloon:00, manual gear:00)
fea(saloon:00, total check:11)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
;سبور كامل فحص 8 موديل فولفو المبادل او تويوتا 2
sail اوبشن فل
[S]
mak(saloon:00, TOYOTA(country>japan):00)
mak(saloon:00,
VOLVO(country>Sweden, country>europe)
:0J)
yea(saloon:00, 8:0V)
fea(saloon:00, total check:0X)
fea(saloon:00, power steering:17)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
;6500 بس عر أصلي شيبرد جديد بحال GTI 89 جولف
0647799079 ت مباشره المالك من
[S]
mod(saloon:00,
Golf(country>germany, country>europe,
make>VOLKS WAGEN):00)
yea(saloon:00, 89:09)
fea(saloon:00, good condition:0C)
pri(saloon:00, 6500:13)
pri(saloon:00, 0647799079:1R)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
;4 douane payée، لون اوبشن فل شيروكي جي ب
original جديد بحاله كعرباه زجاج
أردني ألفديزار بنك بدون
[S]
mak(saloon:00, JEEP(country>Ameria):00)
mod(saloon:00,
Cherokee(country>America, make<JEEP):
04)
col(saloon:00, goldr:00)
yea(saloon:00, 4:17)
fea(saloon:00, good condition:1L)
[/S]
;=====
=====
;;Time 0.2 Sec
;;Done!
;===== UNL
=====
;اوتوماتيك اوبشن فل doré 1997 موديل زوبيرا دايو
6500 دينار جديد
[S]
mak(saloon:00, DAEWOO(country>korea):00)
mod(saloon:00,
Nubira(country>korea, make>DAEWOO):05
)
fea(saloon:00, automatic gear:13)
fea(saloon:00, new:10)
pri(saloon:00, 6500:1T)
[/S]
;=====
=====

```

```

;;Time 0.1 Sec
;;Done!
;===== UNL
=====
;80000 جديد بحاله 95 renouvelée 82 روفر
دينار 16000
[S]
[W]
""
[/W]
[/S]
;=====
=====
;;Time 0.0 Sec
;;Done!
;===== UNL
=====
;5 رينو 82 jusqu'à septembre بحاله
0796232997 ت دينار 1000
[S]
[W]
RENAULT(country>France, country>europe):00
[/W]
[/S]
[S]
[W]
5:05
[/W]
[/S]
[S]
[W]
checked:0A
[/W]
[/S]
[S]
[W]
good condition:0Y
[/W]
[/S]
[S]
[W]
0796232997:1L
[/W]
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
;1998 nouvelle look موديل abs جولف سياره؛
فروش الكعرباه على فتحه كامل
[S]
mod(saloon:00,
Golf(country>germany, country>europe,
make>VOLKS WAGEN):06)
yea(saloon:00, 1998:0L)
fea(saloon:00, total check:14)
fea(saloon:00, Fatha:1D)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
;الاضافات كامله 1990 موديل فكترا اوبل سياره للبيع؛
جديد بحاله كامل فحص
[S]
sal(saloon:06, sale:00)
mak(saloon:06,
OPEL(country>germany, country>europe)
:0C)
mod(saloon:06,
Vectra(country>germany, country>europ
e, make>OPEL):0H)

```

```

yea(saloon:06, 1990:0T)
fea(saloon:06, full options:0Y)
fea(saloon:06, total check:1D)
fea(saloon:06, good condition:1M)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
جيهه بحاله خمري لون 87 سبورموديل كاديك اوبل سيارة؛
دينار 3000 definitif بسعر للبيح
[S]
mak(saloon:00,
    OPEL(country>germany, country>europe)
:06)
mod(saloon:00,
    kadet(country>germany, country>europe
, make>OPEL):0B)
fea(saloon:00, power steering:0I)
fea(saloon:00, good condition:13)
sal(saloon:00, sale:1E)
pri(saloon:00, 3000:1Z)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
لون الاضافات لكاملة 2005 مديل 307 بي جو سيارة؛
15500 السقف في 14 Km عداد بالسقف فتحة سلفنر
دينار
[S]
mak(saloon:00,
    PEUGEOT(country>France, country>europ
e):06)
mod(saloon:00, 307:0B)
yea(saloon:00, 2005:0K)
fea(saloon:00, full options:0P)
col(saloon:00, silver:18)
fea(saloon:00, Fatha:1D)
yea(saloon:00, 14:1U)
pri(saloon:00, 15500:29)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
لون الاضافات لكاملة 2005 مديل 307 بي جو سيارة؛
15500 السقف في 14 Km عداد بالسقف فتحة سلفنر
ناردي
[S]
mak(saloon:00,
    PEUGEOT(country>France, country>europ
e):06)
mod(saloon:00, 307:0B)
yea(saloon:00, 2005:0K)
fea(saloon:00, full options:0P)
col(saloon:00, silver:18)
fea(saloon:00, Fatha:1D)
yea(saloon:00, 14:1U)
pri(saloon:00, 15500:29)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
بحاله سي 1800 اونكوماتيكي جيير Celica 1981 سويوتا؛
le ،

```

```

[S]
[W]
TOYOTA(country>japan):00
[/W]
[/S]
[S]
[W]
1981:0E
[/W]
[/S]
[S]
[W]
automatic gear:0J
[/W]
[/S]
[S]
[W]
1800:0X
[/W]
[/S]
[S]
[W]
good condition:15
[/W]
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
لون parfait سبويلرات كسبباء زجاج 93 داوي سياره؛
اخضر
[S]
mak(saloon:00, DAEWOO(country>korea):06)
yea(saloon:00, 93:0B)
col(saloon:00, green:1B)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
حره طبعه 2005 مديل صني نيسان taxi سياره للبيح؛
pour للشراء 0795597203
[S]
sal(saloon:06, sale:00)
mak(saloon:06, NISSAN(country>japan):0H)
mod(saloon:06,
    Sunny(country>japan, make>NISSAN):0N)
yea(saloon:06, 2005:0W)
yea(saloon:06, 0795597203:1M)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
===== UNL
=====
93 douane payée لانسر متسوبيشي سياره للبيح؛
للتفاوض قابل دينار 5900 جدي
[S]
sal(saloon:06, sale:00)
mak(saloon:06,
    MITSUBISHI(country>japan):0C)
mod(saloon:06,
    Lancer(country>japan, make>MITSUBISHI
):0L)
yea(saloon:06, 93:0R)
fea(saloon:06, new:17)
pri(saloon:06, 5900:1C)
[/S]
=====
=====

```

```

;;Time 0.1 Sec
;;Done!
=====
=====
سيارة دينار الف 23 المطلوب السعر سلفر 99 مرسيدس
دينار 10500 حمري 2001 سيفك موندا
[S]
mak(saloon:00,
    MERCEDES(country>germany, country>eur
ope):00)
mod(saloon:00, 99:07)
col(saloon:00, silver:0A)
pri(saloon:00, 23:0T.@ALF)
mak(saloon:00, HONDA(country>japan):1C)
yea(saloon:00, 2001:1N)
pri(saloon:00, 10500:1X)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
=====
=====
الاضافات كامل سلفر لون 2001 موديل مرسيدس للبيح
كامل فحص
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
    MERCEDES(country>germany, country>eur
ope):06)
col(saloon:00, silver:0S)
fea(saloon:00, total check:1B)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
=====
=====
1992 موديل اوبل و 1980 موديل فولفو للبيح
contactez 0795902118
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
    VOLVO(country>Sweden, country>europe)
:06)
yea(saloon:00, 1980:0I)
mak(saloon:00,
    OPEL(country>germany, country>europe)
:0P)
yea(saloon:00, 1992:10)
yea(saloon:00, 0795902118:1F)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
=====
=====
فحص مع الكندشن عدا ما اوبشن فل م ب سياره
technique سلفر لون 1989 موديل
[S]
fea(saloon:00, air condition:0T)
col(saloon:00, silver:1V)
yea(saloon:00, 17:23)
yea(saloon:00, 17:26)
[/S]
=====
=====
;;Time 0.2 Sec
;;Done!
=====
=====
جدا جيده بحاله 1981 ديلعو امريكيه سياره للبيح
[S]
sal(saloon:06, sale:00)

```

```

cou(saloon:06, America:0D)
yea(saloon:06, 1981:0Q)
fea(saloon:06, good condition:0V)
[/S]
=====
=====
;;Time 0.0 Sec
;;Done!
=====
=====
UNL
جيده بحاله 1200CC ، 86 موديل كادييت اوبل
[S]
mak(saloon:00,
    OPEL(country>germany, country>europe)
:00)
mod(saloon:00,
    kadet(country>germany, country>europe
, make>OPEL):05)
yea(saloon:00, 86:0H)
fea(saloon:00, good condition:0T)
[/S]
=====
=====
;;Time 0.0 Sec
;;Done!
=====
=====
UNL
عدا ما اوبشن فل اخضر لون 99 موديل كورولا شويوتا
كامل فحص دينار 8800 الجير
[S]
mak(saloon:00, TOYOTA(country>japan):00)
mod(saloon:00,
    Corolla(country>japan, make>TOYOTA):0
7)
col(saloon:00, green:0R)
fea(saloon:00, manual gear:16)
pri(saloon:00, 8800:1J)
fea(saloon:00, total check:1T)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
=====
=====
UNL
assuré فحص ابيض لون 99 موديل لانوس دايو سياره
الجير عدا ما اوبشن فل
[S]
mak(saloon:00, DAEWOO(country>korea):06)
mod(saloon:00,
    Lanos(country>korea, make>DAEWOO):0B)
col(saloon:00, white:0U)
fea(saloon:00, manual gear:1J)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
=====
=====
UNL
jaune لون واقتصاديه كامل فحص 99 موديل دايو سياره
Kaled Achour 0796864442 كهرباء زجاج
[S]
mak(saloon:00, DAEWOO(country>korea):06)
yea(saloon:00, 99:0H)
fea(saloon:00, total check:0K)
[/S]
=====
=====
;;Time 0.1 Sec
;;Done!
=====
=====
UNL
2000 موديل PREMIUM 2 نوبيرا دايو سياره للبيح
8000 دينار propriétaire 0795080690
[S]

```

```

sal(saloon:06, sale:00)
mak(saloon:06, DAEWOO(country>korea):0C)
mod(saloon:06,
    Nubira(country>korea,make>DAEWOO):0H
)
yea(saloon:06, 2:0W)
yea(saloon:06, 2000:14)
pri(saloon:06, 8000:19)
yea(saloon:06, 0795080690:1X)
[/S]
=====
;;Time 0.1    Sec
;;Done!
=====
UNL
=====
;فحص الاضافات لكامله 98 موديل Shuma لكيا للبيج;
دينار 90050 6800 لكامل
[S]
sal(saloon:00, sale:00)
mak(saloon:00, KIA(country>korea):06)
yea(saloon:00, 98:0M)
fea(saloon:00, full options:0P)
fea(saloon:00, total check:14)
yea(saloon:00, 90050:1E)
pri(saloon:00, 6800:1K)
[/S]
=====
;;Time 0.1    Sec
;;Done!
=====
UNL
=====
;فحص اوبشن فل 1994 موديل سيفيا لكيا للبيج;
non assured 3800 دينار
[S]
sal(saloon:00, sale:00)
mak(saloon:00, KIA(country>korea):06)
mod(saloon:00,
    Sephia(country>korea,make>KIA):0A)
pri(saloon:00, 3800:1F)
[/S]
=====
;;Time 0.2    Sec
;;Done!
=====
UNL
=====
;عدا ما اوبشن فل ابيض 89 موديل Rio وندا للبيج;
دينار 4-5 ال جير
[S]
sal(saloon:00, sale:00)
mak(saloon:00, HONDA(country>japan):06)
yea(saloon:00, 89:0M)
col(saloon:00, white:0P)
fea(saloon:00, manual gear:13)
yea(saloon:00, 4:1G)
pri(saloon:00, 5:1I)
[/S]
=====
;;Time 0.1    Sec
;;Done!
=====
UNL
=====
;بور اتوماتيك مكيفه اي اكس ل سيفك وندا للبيج;
96 موديل لكامل فحص
[S]
sal(saloon:00, sale:00)
mak(saloon:00, HONDA(country>japan):06)
fea(saloon:00, air condition:0Q)
fea(saloon:00, automatic gear:0W)
fea(saloon:00, power steering:15)
fea(saloon:00, total check:19)
yea(saloon:00, 96:10)
[/S]
=====
;;Time 0.1    Sec
;;Done!
=====
UNL
=====
;بور اتوماتيك مكيفه اي اكس ل سيفك وندا للبيج;
96 موديل لكامل فحص
[S]
sal(saloon:00, sale:00)
mak(saloon:00, HONDA(country>japan):06)
fea(saloon:00, air condition:0Q)
fea(saloon:00, automatic gear:0W)
fea(saloon:00, power steering:15)
fea(saloon:00, total check:19)
yea(saloon:00, 96:10)
[/S]
=====
;;Time 0.1    Sec
;;Done!
=====
UNL
=====
;عدا ما الاضافات لكامله 200كل اسيك 99 موديل مرسيدس;
بنكويه اقساط و دينار دفعه أو ألف 22 90050
[S]
mak(saloon:00,
    MERCEDES(country>germany,country>euro
ope):00)
yea(saloon:00, 99:0D)
mod(saloon:00,
    Classic(country>germany,country>euro
pe,make>VOLKS WAGEN):0G)
yea(saloon:00, 200:0M)
fea(saloon:00, full options:0Q)
yea(saloon:00, 90050:1D)
yea(saloon:00, 22:1J)
[/S]
=====
;;Time 0.3    Sec
;;Done!
=====
UNL
=====
;فحص الاضافات لكامل سلوفر 1996 موديل c180 مرسيدس;
دينار 16.000 لكامل
[S]
mak(saloon:00,
    MERCEDES(country>germany,country>euro
ope):00)
mod(saloon:00, 180:08)
yea(saloon:00, 1996:0I)
col(saloon:00, silver:0N)
fea(saloon:00, total check:16)
yea(saloon:00, 16:1F)
pri(saloon:00, 000:1I)
[/S]
=====
;;Time 0.1    Sec
;;Done!
=====
UNL
=====
;جره طبعه 2001 موديل مونداي taxi سياره للبيج;
دينار الف 23 المطلوب السع
[S]
sal(saloon:06, sale:00)
mak(saloon:06, HYUNDAI(country>korea):0H)
pri(saloon:06, 2001:0U)
pri(saloon:06, 23:1M.@ALF)
[/S]
=====
;;Time 0.1    Sec

```

```

;;Done!
;===== UNL
;=====
;موديل لكسيف هوندا سيارة للبيح
لكامل فحص و مرخصه مدرس يدس سيارة مطلوب
[S]
sal(saloon:06, sale:00)
mak(saloon:06, HONDA(country>japan):0C)
yea(saloon:06, 96:0T)
wan(saloon:1J, wanted:1D)
yea(saloon:06, 1005:18)
mak(saloon:06,
MERCEDDES(country>germany, country>eur
ope):1P)
fea(saloon:06, checked:1W)
fea(saloon:06, total check:24)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
;زجاج اوبشن فل 2000 موديل سيفك هوندا للبيح
manuelles اسود لون 0795066927
[S]
sal(saloon:00, sale:00)
mak(saloon:00, HONDA(country>japan):06)
col(saloon:00, black:1K)
yea(saloon:00, 0795066927:1P)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
;لون الاضافات لكاملة 97 موديل سوناتا هوندا للبيح
jaune
[S]
sal(saloon:00, sale:00)
mak(saloon:00, HONDA(country>japan):06)
mod(saloon:00,
Sonata(country>korea, make>HYUNDAI):0
C)
yea(saloon:00, 97:0P)
fea(saloon:00, full options:0S)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
;لجبر ماعدا الاضافات جميع 1998 موديل هونداي للبيح
فحص complet, verte, 5500 دينار
[S]
sal(saloon:00, sale:00)
mak(saloon:00, HYUNDAI(country>korea):06)
yea(saloon:00, 1998:0J)
fea(saloon:00, full options:00)
fea(saloon:00, manual gear:14)
pri(saloon:00, 5500:1Y)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
;من دينار 13000 CT OK أحمر 96 موديل مدرس يدس للبيح
لكامل ت 0795535225
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
MERCEDDES(country>germany, country>eur

```

```

ope):06)
yea(saloon:00, 96:0J)
col(saloon:00, red:0M)
pri(saloon:00, 13000:0X)
yea(saloon:00, 0795535225:1L)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
;اوبشن فول 1999 موديل بيتل فاين فولس للبيح
دينار 10500 جديد douane payée أحمر الجير عدا ما
[S]
sal(saloon:00, sale:00)
mod(saloon:00,
Beetle(country>germany, country>europ
e, make>VOLKS WAGEN):0H)
fea(saloon:00, manual gear:18)
col(saloon:00, red:1L)
fea(saloon:00, new:23)
pri(saloon:00, 10500:28)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
;الجير عدا ما اوبشن فل 2003 م 4*4 سبور كيا للبيح
douane non payée 7000 دينار
[S]
sal(saloon:00, sale:00)
mak(saloon:00, KIA(country>korea):06)
fea(saloon:00, power steering:0B)
yea(saloon:00, 4:0F)
yea(saloon:00, 4:0H)
fea(saloon:00, manual gear:0Z)
pri(saloon:00, 7000:1T)
[/S]
;=====
;=====
;;Time 0.2 Sec
;;Done!
;===== UNL
;=====
;اوبشن فل 2004 م 4*4 جيب Sportage لكيا للبيح
شامل تمامين دينار 11800 اسود لون الجير عدا ما
complet
[S]
sal(saloon:00, sale:00)
mak(saloon:00, KIA(country>korea):06)
mak(saloon:00, JEEP(country>Ameria):0J)
yea(saloon:00, 4:0N)
yea(saloon:00, 4:0P)
fea(saloon:00, manual gear:17)
col(saloon:00, black:1O)
pri(saloon:00, 11800:1T)
[/S]
;=====
;=====
;;Time 0.2 Sec
;;Done!
;===== UNL
;=====
;لكنشن توماتيك 93 موديل لانسر متسوبيشي للبيح
د 4000 السعر le جيه بحاله
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
MITSUBISHI(country>japan):06)
mod(saloon:00,
Lancer(country>japan, make>MITSUBISHI
):0F)
yea(saloon:00, 93:0R)

```



```

:06)
cou(saloon:00, germany:0B)
yea(saloon:00, 78:0P)
fea(saloon:00, new:13)
yea(saloon:00, 90050:1B)
pri(saloon:00, 800:1W)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;=====
UNL
;=====
;ما الاضافات لكاملة 97 موديل فكترا اوبل سياره للبيح
دينار 9750 المكيف و الجير عدا
[S]
sal(saloon:06, sale:00)
mak(saloon:06,
OPEL(country>germany, country>europe)
:0C)
mod(saloon:06,
Vectra(country>germany, country>europ
e, make>OPEL):0H)
yea(saloon:06, 97:0T)
fea(saloon:06, full options:0W)
fea(saloon:06, manual gear:1B)
pri(saloon:06, 9750:1X)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;=====
UNL
;=====
;ما الاضافات لكامل صفح بدون 93 موديل 520 ام بي للبيح
دينار 10500 excellente الجير عدا
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
B.M.W(country>germany, country>europe
):06)
mod(saloon:00, 520:0C)
yea(saloon:00, 93:0M)
fea(saloon:00, total check:0U)
fea(saloon:00, manual gear:1C)
pri(saloon:00, 10500:20)
[/S]
;=====
=====
;;Time 0.2 Sec
;;Done!
;=====
UNL
;=====
;مع ذهبى، لون 2000 موديل 520 دبليو ام بي للبيح
0795610755 ابيض جلد فرش
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
B.M.W(country>germany, country>europe
):06)
mod(saloon:00, 520:0I)
col(saloon:00, goldr:10)
col(saloon:00, white:1H)
yea(saloon:00, 0795610755:1M)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;=====
UNL
;=====
;ما الاضافات لكامل فحص 95 موديل دبليو ام بي للبيح
des الاضافات extraordinaires مع
[S]
sal(saloon:00, sale:00)
mak(saloon:00,

```

```

B.M.W(country>germany, country>europe
):06)
yea(saloon:00, 95:00)
fea(saloon:00, total check:0R)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;=====
UNL
;=====
;1,41 كم 2005 مديلي 307 بيجو سياره للبيح
0795321200 الجير عدا ما اوبشن فل
[S]
sal(saloon:06, sale:00)
mak(saloon:06,
PEUGEOT(country>France, country>europ
e):0C)
mod(saloon:06, 307:0H)
yea(saloon:06, 2005:0Q)
yea(saloon:06, 6000:0V)
yea(saloon:06, 1:15)
fea(saloon:06, manual gear:1J)
yea(saloon:06, 0795321200:1W)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;=====
UNL
;=====
;اوتوماتيك جير 1982 موديل كورولا تويوتا للبيح
fauteuils و جديد كوشوك جديد batterie جده بحاله
جديدين
[S]
sal(saloon:00, sale:00)
mak(saloon:00, TOYOTA(country>japan):06)
mod(saloon:00,
Corolla(country>japan, make>TOYOTA):0
D)
yea(saloon:00, 1982:0Q)
fea(saloon:00, automatic gear:0V)
fea(saloon:00, good condition:19)
fea(saloon:00, new:1T)
fea(saloon:00, new:24)
fea(saloon:00, new:2L)
[/S]
;=====
=====
;;Time 0.2 Sec
;;Done!
;=====
UNL
;=====
;جديد كوشوك جديد comme 1998 موديل جولف للبيح
électrique فحص لكامل
[S]
sal(saloon:00, sale:00)
mod(saloon:00,
Golf(country>germany, country>europe,
make>VOLKS WAGEN):06)
yea(saloon:00, 1998:0H)
fea(saloon:00, new:0S)
fea(saloon:00, new:13)
fea(saloon:00, total check:18)
fea(saloon:00, Fatha:1H)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;=====
UNL
;=====
;92 اسود système موديل جولف سياره للبيح
d'alarme vitre électrique
[S]
sal(saloon:06, sale:00)

```

```

mod(saloon:06,
    Golf(country>germany,country>europe,
make>VOLKS WAGEN):0C)
yea(saloon:06, 92:0N)
col(saloon:06, black:0Q)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
من جدا اقتصادية اسود 97 موديل دايو سياره للبيعي
المالك jusqu'en mai
[S]
sal(saloon:06, sale:00)
mak(saloon:06, DAEWOO(country>korea):0C)
yea(saloon:06, 97:0N)
col(saloon:06, black:0Q)
fea(saloon:06, checked:1I)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
عدا ما اوبشن فل 92 موديل espero دايو سياره للبيعي
دينار 3800 جيده بحاله الاجير
[S]
sal(saloon:06, sale:00)
mak(saloon:06, DAEWOO(country>korea):0C)
fea(saloon:06, manual gear:16)
fea(saloon:06, good condition:1J)
pri(saloon:06, 3800:1U)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
جيريال عدا الاضافات لكاملة 1983 سوبارو للبيعي
0795131719
[S]
sal(saloon:00, sale:00)
mak(saloon:00, SUBARU(country>japan):06)
yea(saloon:00, 1983:0D)
fea(saloon:00, full options:0I)
fea(saloon:00, manual gear:0X)
yea(saloon:00, 079:18)
yea(saloon:00, 5131719:1D)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
جدا اقتصادية ممتازة بحالة 97 موديل سوزوكي للبيعي
دينار للتفاوض قابل 2700
[S]
sal(saloon:00, sale:00)
mak(saloon:00, SUZUKI(country>japan):06)
yea(saloon:00, 97:0J)
fea(saloon:00, good condition:0M)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
عدا ما اوبشن فل 94 موديل جولف فاجن فلكس للبيعي
المكيف و الاجير

```

```

[S]
sal(saloon:00, sale:00)
mod(saloon:00,
    Golf(country>germany,country>europe,
make>VOLKS WAGEN):0F)
fea(saloon:00, manual gear:12)
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
مكيف مع سلفر 2003 م POLO فاجن فلكس للبيعي
دينار 10000 ديفينيف بسعر لكامل فحص
[S]
[W]
sale:00
[/W]
[/S]
[S]
[W]
year:0K
[/W]
[/S]
[S]
[W]
2003:0M
[/W]
[/S]
[S]
[W]
silver:0R
[/W]
[/S]
[S]
[W]
"ع"
[/W]
[/S]
[S]
[W]
air condition:0Z
[/W]
[/S]
[S]
[W]
total check:14
[/W]
[/S]
[S]
[W]
Price:1D
[/W]
[/S]
[S]
[W]
10000:1S
[/W]
[/S]
;=====
=====
;;Time 0.1 Sec
;;Done!
;===== UNL
=====
رالجي عدا ما اوبشن فل 85 موديل سيور مازدا للبيعي
دينار 2400
[S]
sal(saloon:00, sale:00)
mak(saloon:00, MAZDA(country>japan):06)
fea(saloon:00, power steering:0D)
fea(saloon:00, manual gear:0Z)
pri(saloon:00, 2400:1C)
[/S]

```

```

;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
; و فتحة OK فحوص 90 موديل لانسر متسوبيشي للبيح
rétro دينار 4500 المطلب السعير لغيره
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
MITSUBISHI(country>japan):06)
mod(saloon:00,
Lancer(country>japan,make>MITSUBISHI
):0F)
yea(saloon:00, 90:0R)
fea(saloon:00, Fatha:11)
pri(saloon:00, 4500:1Z)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
; فحوص انتيكاوتوم اسود 95 موديل c180 مرسيدس للبيح
لكامل
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
MERCEDES(country>germany, country>eur
ope):06)
mod(saloon:00, 180:0E)
yea(saloon:00, 95:00)
col(saloon:00, black:0R)
fea(saloon:00, automatic gear:0Y)
fea(saloon:00, total check:16)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
; ، أوتوماتك فتحه 99 موديل coupé متسوبيشي للبيح
دينار 1500.سي
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
MITSUBISHI(country>japan):06)
yea(saloon:00, 99:0R)
fea(saloon:00, Fatha:0U)
yea(saloon:00, 1500:19)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
; boîte, عدا ما اوبشن فل 94 مديل داويو سياره للبيح
خبري نلو
[S]
sal(saloon:06, sale:00)
mak(saloon:06, DAEWOO(country>korea):0C)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
; دبل 2001 موديل سوبارو نوع رياضييه من سياره للبيح
جدا مهميز لون تيربو
[S]
sal(saloon:06, sale:00)

```

```

mak(saloon:06, SUBARU(country>japan):0Q)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
; 99 ديلجو سامسونج سياره للبيح
concessionaire اوبشن فل جدا جيده بحاله
[S]
sal(saloon:06, sale:00)
yea(saloon:06, 99:0Q)
fea(saloon:06, good condition:1B)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
; 3000 اوتوماتيك جيير مكيف مع 240 فولفو للبيح
1983 م الفحص على شامل تامين دينار
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
VOLVO(country>Sweden, country>europe)
:06)
mod(saloon:00, 240:0C)
fea(saloon:00, air condition:0J)
fea(saloon:00, automatic gear:00)
pri(saloon:00, 3000:12)
fea(saloon:00, total check:10)
yea(saloon:00, 1983:20)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
; او 86 موديل سيس cresseda 1900 سياره للبيح
82 م ابيض مرسيدس مع المبادله
[S]
sal(saloon:06, sale:00)
yea(saloon:06, 1900:0L)
yea(saloon:06, 86:0Z)
mak(saloon:06,
MERCEDES(country>germany, country>eur
ope):1H)
col(saloon:06, white:10)
yea(saloon:06, 82:1V)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
; كامل فحوص 89 bleu ciel لانسر متسوبيشي للبيح
دينار 4800 définitif بسعير حديت
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
MITSUBISHI(country>japan):06)
mod(saloon:00,
Lancer(country>japan,make>MITSUBISHI
):0F)
yea(saloon:00, 89:0L)
fea(saloon:00, total check:0Y)
pri(saloon:00, 4800:24)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!

```

```

;===== UNL
;=====
جديه بحاله 82 موديل لانسر متسوبيشي سياره للبيوع;
دينار definitif 3000 السعز
[S]
sal(saloon:06, sale:00)
mak(saloon:06,
    MITSUBISHI(country>japan):0C)
mod(saloon:06,
    Lancer(country>japan,make>MITSUBISHI
):0L)
yea(saloon:06, 82:0X)
fea(saloon:06, good condition:10)
pri(saloon:06, 3000:1R)
[/S]
;=====
;Time 0.1 Sec
;;Done!
;===== UNL
;=====
مع الجير عدا ما اوبشن فل 84 موديل مرسيدس للبيوع;
دينار 5350 بور سنتر و مكيف
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
    MERCEDES(country>germany, country>eur
ope):06)
fea(saloon:00, manual gear:0V)
fea(saloon:00, air condition:1B)
fea(saloon:00, center lock:1I)
fea(saloon:00, power steering:1N)
pri(saloon:00, 5350:1R)
[/S]
;=====
;Time 0.1 Sec
;;Done!
;===== UNL
;=====
الاضافات لكامله 98 موديل c180 مرسيدس سياره للبيوع;
دينار 18500 سلغر لون l'ouverture
[S]
sal(saloon:06, sale:00)
mak(saloon:06,
    MERCEDES(country>germany, country>eur
ope):0C)
mod(saloon:06, 180:0K)
yea(saloon:06, 98:0U)
fea(saloon:06, full options:0X)
col(saloon:06, silver:1W)
pri(saloon:06, 18500:21)
[/S]
;=====
;Time 0.1 Sec
;;Done!
;===== UNL
;=====
اوتوماتيك جير ايشن فل سيفك كيا للبيوع;
[S]
sal(saloon:00, sale:00)
mak(saloon:00, KIA(country>korea):06)
fea(saloon:00, automatic gear:0N)
[/S]
;=====
;Time 0.0 Sec
;;Done!
;===== UNL
;=====
فتحه جلد اوبشن فل 97 موديل مكسيدم نيسان للبيوع;
ينارد 11200
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
    NISSAN(country>japan):06)
mod(saloon:00,
    Maxima(country>japan,make>NISSAN):0C
)
fea(saloon:00, Fatha:14)
pri(saloon:00, 11200:19)
[/S]
;=====
;Time 0.1 Sec
;;Done!
;===== UNL
;=====
اوتوماتيك جير 1997 موديل سيفك هوندا سياره للبيوع;
ع climatisation, 7750 دينار
[S]
sal(saloon:06, sale:00)
mak(saloon:06, HONDA(country>japan):0C)
yea(saloon:06, 1997:0S)
fea(saloon:06, automatic gear:0X)
pri(saloon:06, 7750:1T)
[/S]
;=====
;Time 0.1 Sec
;;Done!
;===== UNL
;=====
الجير عدا ما اوبشن فل 1995 موديل الكسنت للبيوعونداي;
دينار 5000 ب
[S]
sal(saloon:00, sale:00)
mak(saloon:00, HYUNDAI(country>korea):05)
mod(saloon:00,
    Accent(country>korea,make>HYUNDAI):0
C)
fea(saloon:00, manual gear:11)
pri(saloon:00, 5000:1F)
[/S]
;=====
;Time 0.1 Sec
;;Done!
;===== UNL
;=====
boite, عدا ما الاضافات لكامله 2001 موديل للبيوعونداي;
دينار 8900
[S]
sal(saloon:00, sale:00)
mak(saloon:00, HYUNDAI(country>korea):05)
yea(saloon:00, 2001:0H)
fea(saloon:00, full options:0M)
pri(saloon:00, 8900:1F)
[/S]
;=====
;Time 0.1 Sec
;;Done!
;===== UNL
;=====
ايشن فل سلغر 1992 موديل 520 دبليو ام بي للبيوع;
البحاله état, 14 دينار
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
    B.M.W(country>germany, country>europe
):06)
mod(saloon:00, 520:0I)
yea(saloon:00, 1992:0R)
col(saloon:00, silver:0W)
pri(saloon:00, 14:1L)
[/S]
;=====

```

```

;;Time 0.2 Sec
;;Done!
=====
UNL
=====
; 4600 ، 2005 يلمد جديد لكوشوك جديد GT فورد للبيج;
اوبشن فل أحر لون
[S]
sal(saloon:00, sale:00)
mak(saloon:00, FORD(country>Ameria):06)
fea(saloon:00, new:0E)
fea(saloon:00, new:0P)
yea(saloon:00, 2005:0Z)
col(saloon:00, red:1F)
[/S]
=====
;;Time 0.1 Sec
;;Done!
=====
UNL
=====
; aluminum جنطات 90 مجهزة 86 مديل ويفضل للبيج;
دينار 4600 définitif بسعر nouvelle peinture
[S]
[W]
sale:00
[/W]
[/S]
[S]
[W]
year:0C
[/W]
[/S]
[S]
[W]
86:0H
[/W]
[/S]
[S]
[W]
year:0K
[/W]
[/S]
[S]
[W]
90:0Q
[/W]
[/S]
[S]
[W]
Price:1Q
[/W]
[/S]
[S]
[W]
4600:25
[/W]
[/S]
=====
;;Time 0.4 Sec
;;Done!
=====
UNL
=====
; فل لكامل فحص اخضر 96 مديل سيفيا ليا للبيج;
95918898 ت 5250 définitif سعر اوبشن
[S]
sal(saloon:00, sale:00)
mak(saloon:00, KIA(country>korea):06)
mod(saloon:00,
Sephia(country>korea,make>KIA):0A)
yea(saloon:00, 96:0L)
col(saloon:00, green:00)
fea(saloon:00, total check:0U)
pri(saloon:00, 5250:1G)
pri(saloon:00, 95918898:1X)
[/S]

```

```

=====
;;Time 0.1 Sec
;;Done!
=====
UNL
=====
; استومتيك الجير مع فل 96 مديل سيفيا ليا للبيج;
دينار 5800 لكامل فحص اخضر لون
[S]
sal(saloon:00, sale:00)
mak(saloon:00, KIA(country>korea):06)
mod(saloon:00,
Sephia(country>korea,make>KIA):0A)
yea(saloon:00, 96:0L)
col(saloon:00, green:1C)
fea(saloon:00, total check:1I)
pri(saloon:00, 5800:1R)
[/S]
=====
;;Time 0.1 Sec
;;Done!
=====
UNL
=====
; الاضافات لكاملة 2001 مديل ديسكفري لاندروفر للبيج;
8 original سلفون 0796635036
[S]
sal(saloon:00, sale:00)
mod(saloon:00, Land
Rover(country>England, country>europe, make>R
over):06)
mod(saloon:00,
Discovery(country>England, country>eu
rope, make>Rover):0F)
yea(saloon:00, 2001:0S)
fea(saloon:00, full options:0X)
yea(saloon:00, 0796635036:1Z)
[/S]
=====
;;Time 0.1 Sec
;;Done!
=====
UNL
=====
; مكيف فتحه مع 84 مجهزة 76 مديل جرس يدس للبيج;
سلفر لون جلد فرش اوتوماتيك
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
MERCEDES(country>germany, country>eur
ope):06)
yea(saloon:00, 76:0I)
yea(saloon:00, 84:0R)
fea(saloon:00, Fatha:0X)
fea(saloon:00, air condition:12)
fea(saloon:00, automatic gear:19)
col(saloon:00, silver:1T)
[/S]
=====
;;Time 0.2 Sec
;;Done!
=====
UNL
=====
; مكيف مع ابيض لون 91 مديل الكورد هوندا للبيج;
لكامل فحص اوتوماتيك
[S]
sal(saloon:00, sale:00)
mak(saloon:00, HONDA(country>japan):06)
mod(saloon:00,
Accord(country>japan, make>HONDA):0C)
col(saloon:00, white:0U)
fea(saloon:00, air condition:12)
fea(saloon:00, automatic gear:19)
fea(saloon:00, total check:1H)
[/S]

```

```

;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
عدا ما اوبشن فل كامل فحوص 1989 م مرسيدس للبي ع
0795517487
[S]
sal(saloon:00, sale:00)
mak(saloon:00,
MERCEDDES(country>germany,country>eur
ope):06)
yea(saloon:00, 1989:0F)
fea(saloon:00, total check:0K)
yea(saloon:00, 0795517487:19)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
دينار 900 خمري لون 88 موديل سمارا لادا للبي ع
[S]
sal(saloon:00, sale:00)
mak(saloon:00, Lada:06)
mod(saloon:00, SAMARA(make>Lada):0B)
pri(saloon:00, 900:0Z)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
بس عر ومرخصة contrôlée اسود مازدا للبي ع
définitif 2500 دينار son جديده est bonne
[S]
sal(saloon:00, sale:00)
mak(saloon:00, MAZDA(country>japan):06)
col(saloon:00, black:0C)
fea(saloon:00, checked:0S)
pri(saloon:00, 2500:1D)

```

```

[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
intéressant بس عر 2000 مديل نوبيرا دايو للبي ع
+0795200986 لانسر متسوبيشي
[S]
sal(saloon:00, sale:00)
mak(saloon:00, DAEWOO(country>korea):06)
mod(saloon:00,
Nubira(country>korea,make>DAEWOO):0B
)
pri(saloon:00, 2000:0N)
pri(saloon:00, 0795200986:1A)
mak(saloon:00,
MITSUBISHI(country>japan):1L)
mod(saloon:00,
Lancer(country>japan,make>MITSUBISHI
):1U)
yea(saloon:00, 2003:20)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done!
;===== UNL
;=====
1985 موديل ديزل فسورويل اب بك نيسان للبي ع
assuré فحوص 8/12/2005 السرقت في الى
[S]
sal(saloon:00, sale:00)
mak(saloon:00, NISSAN(country>japan):06)
yea(saloon:00, 1985:10)
yea(saloon:00, 8:1P)
yea(saloon:00, 12:1R)
yea(saloon:00, 2005:1U)
[/S]
;=====
;=====
;;Time 0.1 Sec
;;Done

```

Résumé : Nous nous intéressons à la multilinguisation, ou « portage linguistique » (plus simple que la localisation) des services de gestion de contenu traitant des énoncés spontanés en langue naturelle, souvent bruités, mais contraints par la situation, et constituant toujours un « sous-langage » plus ou moins restreint.

Un service de ce type (soit App) utilise une représentation du contenu spécifique (RC-App) sur laquelle travaille le noyau fonctionnel. Le plus souvent, cette représentation est produite à partir de la langue « native » L1 par un extracteur de contenu (EC-App). Nous avons dégagé trois méthodes de portage possibles, et les avons illustrées par le portage en français d'une partie de CATS, un système de traitement de petites annonces en SMS (en arabe) déployé à Amman, ainsi que sur IMRS, un système de recherche de morceaux de musique dont l'interface native est en japonais et dont seule la RC est accessible. Il s'agit de : (1) localisation « interne », i.e. adaptation à L2 de l'EC donnant EC-App-L2 ; (2) localisation « externe », i.e. adaptation d'un EC existant pour L2 au domaine et à la représentation de contenu de App (EC-X-L2-App); (3) traduction des énoncés de L2 vers L1.

Le choix de la stratégie est contraint par la situation traductionnelle : types et niveau d'accès possibles (accès complet au code source, accès limité à la représentation interne, accès limité au dictionnaire, et aucun accès), ressources disponibles (dictionnaires, corpus), compétences langagières et linguistiques des personnes intervenant dans la multilinguisation des applications.

Les trois méthodes ont donné de bons résultats sur le portage d'arabe en français de la partie de CATS concernant l'occasion automobile. En localisation interne, la partie grammaticale a été très faiblement modifiée, ce qui prouve que, malgré la grande distance entre l'arabe et le français, ces deux sous-langages sont très proches l'un de l'autre, une nouvelle illustration de l'analyse de R. Kittredge. La localisation externe a été expérimentée sur CATS et sur IMRS en adaptant au nouveau domaine considéré l'extracteur de contenu du français écrit initialement par H. Blanchon pour le domaine du tourisme (projet CSTAR/Nespole!), puis en changeant de langue pour IMRS (anglais).

Enfin, le portage par TA statistique a également donné de très bonnes performances, et cela avec un corpus d'apprentissage très petit (moins de 10.000 mots) et un dictionnaire complet. Cela prouve que, dans le cas de sous-langages très petits, la TA statistique peut être de qualité suffisante en partant de corpus 100 à 500 fois moins grands que pour de la langue générale.

Mots-clés : énoncés spontanés, langue naturelle, e-commerce, portage linguistique, sous-langage, extraction de contenu, traduction statistique, corpus.

Abstract: We are interested in the multilinguization, or “linguistic porting” (simpler than localization) of management content services processing spontaneous utterances in natural language, often noisy, but constrained by the situation and constituting a restricted “sublanguage”.

Any service of this type (App) uses a specific content representation (CR-App) on which the functional kernel operates. Most often, this representation is produced from the “native” language L1 by a content extractor (CE-App). We identified three possible methods of porting and have illustrated them by porting to French a part of CATS, a system handling small ads in SMS (in Arabic), deployed in Amman, as well as IMRS, a music retrieval system, where the native natural language interface is in Japanese and only the CR is accessible. These are: (1) “internal localisation”, i.e. adaptation to L2 of the CE, giving CE-App-L2; (2) “external” localization, i.e. adaptation of an existing CE for L2 to the domain and to the App content representation (CE-X-L2-App); (3) translation of utterances from L2 to L1.

The choice of the strategy is constrained by the translational situation: type and level of possible access (complete access to the source code, access limited to the internal representation, access limited to the dictionary, and no access), available resources (dictionaries, corpus), competences in languages and linguistics of persons taking part in the multilinguisation of application.

The three methods gave good results on the Arabic to French porting of the CARS part of CATS. For internal localization, the grammatical part was very little modified, which proves that, despite the great distance between Arabic and French, these two sub-languages are very near one to another. This is a new illustration of R.Kittredge’s analysis. The external localization was experimented with CATS and with IMRS by adapting to the new domain the French content extractor written initially by H. Blanchon for the tourism domain (CSTAR/Nespole! project), and then by changing the language for IMRS (English).

Finally, porting by statistical MT gave also a very good performance, and that with a very small training corpus (less than 10 000 words) and a complete dictionary. This proves that, in the case of very small sub-languages, statistical MT may be of sufficient quality, starting from a corpus 100 to 500 smaller than for the general language.

Keywords: spontaneous utterances, natural language, e-commerce, linguistic porting, sub-language, content extraction, statistical translation, corpora.