

N° d'ordre:

# THÈSE

présentée

devant l'Université de Bretagne Sud

pour obtenir

le grade de : DOCTEUR DE L'UNIVERSITÉ DE BRETAGNE SUD  
Mention SCIENCES POUR L'INGÉNIEUR – INFORMATIQUE

par

Jérôme GOULIAN

École Doctorale :

Laboratoire : VALORIA EA2593 - UNIVERSITÉ DE BRETAGNE SUD

Titre de la thèse :

*Stratégie d'analyse détaillée pour la Compréhension Automatique  
robuste de la Parole*

présentée le 13 décembre 2002 devant la commission d'examen

MM. :	Jean-Marie Jean	PIERREL VÉRONIS	Rapporteurs
MMme. :	Jacques Nadine	SIROUX VIGOUROUX	Examineurs
M. :	Franck	POIRIER	Directeur de thèse
M. :	Jean-Yves	ANTOINE	Co-Directeur de thèse







# Table des matières

Introduction	i
<b>I Comprendre la parole : problématiques et enjeux</b>	<b>1</b>
<b>1 Compréhension automatique de la parole et dialogue homme-machine</b>	<b>3</b>
1.1 Quelles applications?	4
1.2 Quelles difficultés?	7
1.2.1 La parole : quel matériau?	7
1.2.2 Reconnaître - Comprendre - Dialoguer	15
1.3 Reconnaissance, compréhension : quelles démarches?	17
1.3.1 Reconnaître la parole	17
1.3.2 Comprendre la parole	23
1.3.3 Conclusions	31
<b>2 Compréhension : quelles perspectives ?</b>	<b>33</b>
2.1 Bilan de la compréhension sélective	33
2.1.1 Portée de l'analyse et prédictabilité	34
2.1.2 Influence des prétraitements	35
2.1.3 Généricité	36
2.2 Vers un changement de paradigme ?	37
2.2.1 Traitements linguistiques et CHM : une nécessité qui s'impose	37
2.2.2 Rôle de la syntaxe	44
2.2.3 Enrichissement du domaine de la tâche	46
2.3 Objectifs généraux de la thèse	48
<b>II Traitements linguistiques robustes pour la compréhension automatique de la parole – le système ROMUS.</b>	<b>51</b>
<b>3 L'oral – De l'analyse des usages à l'architecture du système proposé</b>	<b>53</b>
3.1 Études descriptives de l'oral	53
3.1.1 Régularités des productions orales	54
3.1.2 Apports pour le traitement automatique	57

3.2	Linguistique de corpus . . . . .	58
3.2.1	Études de corpus et communication orale homme-machine . . . . .	60
3.2.2	Apports pour le traitement automatique . . . . .	66
3.3	Conclusions et architecture du système proposé . . . . .	68
<b>4</b>	<b>Application au renseignement touristique</b>	<b>71</b>
4.1	Caractérisation du domaine . . . . .	72
4.2	Définitions des représentations sémantiques attendues . . . . .	74
<b>5</b>	<b>Segmentation grammaticale des énoncés oraux</b>	<b>79</b>
5.1	Objectifs . . . . .	79
5.2	Analyse robuste de l'oral . . . . .	81
5.2.1	Le TAL robuste . . . . .	81
5.2.2	Adaptation à la segmentation des énoncés oraux . . . . .	83
5.3	Segmentation grammaticale dans le système ROMUS . . . . .	84
5.3.1	Définitions des unités syntagmatiques retenues . . . . .	84
5.3.2	Étiquetage grammatical de l'oral . . . . .	87
5.3.3	Segmentation . . . . .	93
5.4	Conclusions . . . . .	98
<b>6</b>	<b>Analyse sémantico-pragmatique détaillée</b>	<b>101</b>
6.1	Introduction à l'analyse par dépendance . . . . .	101
6.2	Mise en relation des segments dans le système ROMUS . . . . .	103
6.2.1	Objectifs . . . . .	103
6.2.2	Grammaires de lien . . . . .	104
6.2.3	Contraintes exploitées . . . . .	106
6.2.4	Construction de la grammaire de lien . . . . .	109
6.2.5	Algorithme d'analyse et gestion de l'ambiguïté . . . . .	111
6.2.6	Génération des représentations sémantiques finales . . . . .	112
6.3	Quelques exemples de traitement de phénomènes oraux et linguistiques . . . . .	113
6.3.1	Traitement des coordinations et des inattendus structuraux . . . . .	114
6.3.2	Procédés réguliers . . . . .	116
6.3.3	Richesse du domaine applicatif . . . . .	117
6.4	Bilan . . . . .	119
<b>III</b>	<b>Bilan et perspectives</b>	<b>121</b>
<b>7</b>	<b>Évaluation</b>	<b>123</b>
7.1	Campagne DEFI . . . . .	124
7.1.1	Méthodologie d'évaluation . . . . .	124
7.1.2	Limites de la première campagne DEFI . . . . .	125
7.1.3	Résultats et discussion . . . . .	125
7.2	Influence des erreurs de reconnaissance . . . . .	130
7.2.1	Tests à partir d'une dictée vocale . . . . .	130

7.2.2	Typologies des problèmes évalués . . . . .	130
7.2.3	Évaluation . . . . .	131
7.2.4	Conclusions . . . . .	132
<b>8</b>	<b>Conclusions et perspectives</b>	<b>135</b>
8.1	Conclusions . . . . .	135
8.2	Perspectives . . . . .	137
	<b>Bibliographie</b>	<b>139</b>
	<b>Annexes</b>	<b>159</b>
<b>A</b>	<b>Étude des phénomènes d'extraction en français parlé sur deux corpus de dialogue oral finalisé</b>	<b>159</b>
A.1	Corpus étudiés et méthodologie d'analyse . . . . .	160
A.1.1	Corpus Air France et Murol . . . . .	160
A.1.2	Recensement des phénomènes . . . . .	161
A.1.3	Limites méthodologiques de l'étude : discussion . . . . .	163
A.2	Résultats : l'extraction, un phénomène incontournable à l'oral . . . . .	165
A.2.1	Importance relative des phénomènes d'extraction . . . . .	165
A.2.2	Extractions orales et ordre canonique sujet-verbe-objet . . . . .	167
A.2.3	Extractions orales et projectivité . . . . .	169
A.2.4	Fonctions et procédés . . . . .	170
A.3	Extractions et domaine d'application de la CHM orale . . . . .	172
A.3.1	Interactivité et extractions orales . . . . .	172
A.3.2	Variabilité d'occurrence et fonctions concernées . . . . .	173
A.3.3	Variabilité d'occurrence et procédés . . . . .	174
<b>B</b>	<b>Segmentation des énoncés oraux – liste complète des étiquettes et des règles utilisées</b>	<b>177</b>
B.1	Exemples du lexique . . . . .	178
B.2	Règles de désambiguïsation utilisées . . . . .	183
B.3	Liste complète et définitions des groupes élémentaires retenus . . . . .	185
<b>C</b>	<b>Construction semi-automatique de la grammaire de liens</b>	<b>189</b>
C.1	Expertise du domaine d'application . . . . .	190
C.1.1	Les objets du monde de l'application . . . . .	191
C.1.2	Caractéristiques générales . . . . .	191
C.1.3	Liste des objets répertoriés . . . . .	191
C.1.4	Les schémas retenus . . . . .	200
C.2	Liste des étiquettes pragmatiques retenues pour l'analyse des dépendances entre segments . . . . .	206
C.2.1	Étiquettes exprimant les actes dialogiques retenus . . . . .	206
C.2.2	Schémas et principaux concepts arguments associés . . . . .	206

C.2.3	Autres concepts arguments . . . . .	208
C.2.4	Concepts de propriété . . . . .	209
C.2.5	Autres étiquettes . . . . .	210
C.3	Exemple du dictionnaire sémantico-pragmatique . . . . .	211
C.4	Exemple de la grammaire de liens . . . . .	213
<b>D</b>	<b>Évaluation par défi – Énoncés dérivés</b>	<b>221</b>
D.1	Série 1 . . . . .	222
D.2	Série 2 . . . . .	222
D.3	Série 3 . . . . .	223



# Table des figures

1.1	Architecture générale d'un système de dialogue oral homme-machine . . .	16
5.1	Exemple simplifié d'accès au lexique par un transducteur . . . . .	91
5.2	Illustration du processus de segmentation . . . . .	94
5.3	Séquence ordonnée de l'application des transducteurs . . . . .	96
6.1	Exemples de structures élémentaires d'une grammaire de lien . . . . .	105
C.1	Objets Localisables. Les feuilles de l'arbre (en gras) représentent les concepts ; les noeuds dénotent l'ensemble des concepts qu'ils englobent. . .	193
C.2	Objets non localisables (par une adresse en ville). Les feuilles de l'arbre (en gras) représentent les concepts ; les noeuds dénotent l'ensemble des concepts qu'ils englobent. . . . .	195



# Liste des tableaux

1.1	Campagnes d'évaluation ARPA : performances obtenues par le meilleur système de reconnaissance. . . . .	23
2.1	Patterns de correspondance identifiant répétitions et corrections . . . . .	41
3.1	Linguistique de corpus et CHM – Description des corpus étudiés . . . . .	61
3.2	Linguistique de corpus et CHM – Caractérisation des corpus étudiés en fonction du contexte interactif . . . . .	62
4.1	Exemples simplifiés des schémas et concepts considérés dans le domaine d'application du système ROMUS. . . . .	74
5.1	Segmentation en chunk – groupes grammaticaux élémentaires retenus . . . . .	86
5.2	Jeu d'étiquettes grammaticales retenu pour la segmentation. . . . .	90
7.1	Performances globales des systèmes $R_1$ et $R_2$ . . . . .	127
7.2	Taux de performance attestés sur les différentes classes de phénomènes envisagés . . . . .	128
7.3	Classification des sources d'erreurs – système $R_2$ . . . . .	129
A.1	Description des corpus d'étude . . . . .	160
A.2	Fréquence d'apparition des extractions sur chaque corpus (nombre moyen de tours de parole présentant au moins une extraction) . . . . .	166
A.3	Fréquence d'apparition des extractions suivant le statut des locuteurs dans le corpus Air France (nombre moyen de tours de parole présentant au moins une extraction) . . . . .	166
A.4	Répartition des extractions en fonction du sens du détachement (pourcentage moyen de phénomènes d'un sens donné) . . . . .	167
A.5	Répartition des extractions suivant la fonction de l'élément détaché (pourcentage moyen de phénomènes d'un type donné) . . . . .	167
A.6	Répartition des extractions d'éléments sujets en fonction du sens du détachement (pourcentage moyen de phénomènes d'un sens donné) . . . . .	168
A.7	Répartition des extractions d'éléments sujets en fonction du procédé utilisé (pourcentage moyen de phénomènes d'un sens donné . . . . .	169

A.8	Part relative des extractions conservant l'ordre canonique SVO (pourcentage par rapport à l'ensemble des phénomènes observés et rapport à l'ensemble des énoncés) . . . . .	169
A.9	Part relative des extractions non projectives (pourcentage par rapport à l'ensemble des phénomènes observés et par rapport à l'ensemble des énoncés) . . . . .	169
A.10	Répartition des extractions suivant le procédé utilisé (pourcentage moyen de procédés d'un type donné) . . . . .	170
A.11	Répartition des extractions d'éléments arguments en fonction du procédé utilisé (pourcentage moyen de phénomènes d'un sens donné) . . . . .	171
A.12	Répartition des extractions d'éléments modificateurs et associés en fonction du procédé utilisé (pourcentage moyen de phénomènes d'un sens donné) . . . . .	171
A.13	Répartition des doubles-marquages et présentatifs . . . . .	172
A.14	Fréquence d'apparition des extractions sur chaque sous-corpus Murol (nombre moyen de tours de parole présentant au moins une extraction) . . . . .	173
B.1	Groupes élémentaires identifiant les marqueurs des répétitions et des corrections . . . . .	186
B.2	Définitions des groupes grammaticaux élémentaires retenus . . . . .	187
B.3	Quelques exemples des définitions pour les groupes élémentaires retenus concernant les expressions langagières indépendantes de l'application . . . . .	188
C.1	Étiquettes des liens correspondant aux principaux schémas et concepts arguments des actes de demande d'information . . . . .	207
C.2	Étiquettes des liens correspondant aux principaux schémas et concepts arguments des actes de demande de confirmation . . . . .	208
C.3	Étiquettes des liens correspondant aux concepts arguments . . . . .	209
C.4	Étiquettes des liens correspondant aux concepts de propriétés généraux . . . . .	209
C.5	Étiquettes des liens correspondant aux concepts de propriété . . . . .	210
C.6	Autres Étiquettes . . . . .	211

# Introduction

La langue évolue de plusieurs façons ; parfois accidentellement, parfois nécessairement pour répondre à nos besoins et désigner des réalités nouvelles. Les dictionnaires, observateurs des usages de la langue, sont les témoins privilégiés de cette évolution. Le rôle des lexicographes, chargés de leur actualisation, est de faire le point chaque fois qu'un écart devient sensible entre le dictionnaire en tant que texte achevé et l'univers culturel présent. On notera ainsi l'apparition, dans l'édition actualisée du *Nouveau Petit Robert* de mars 2001, de l'entrée suivante, inexistante dans l'édition précédente de 1994 :

**Serveur vocal interactif** : *système informatique connecté à un poste téléphonique à touches à partir duquel un usager peut obtenir de manière interactive des informations*

L'échange d'informations, le "dialogue", entre l'utilisateur et la machine qui y est décrit semble évoquer une forme d'interaction avec la machine assez rudimentaire. L'utilisation du terme *poste téléphonique à touches* suggère une séquence de questions à choix multiples parmi lesquels l'utilisateur navigue en utilisant les touches "multi-fréquences" de son téléphone. Le qualificatif *vocal* ne fait d'ailleurs référence ici qu'à ce média, le téléphone, et en rien à une modalité d'interaction permettant à l'utilisateur de formuler librement ses besoins en "langage naturel". Pourtant, la Communication Homme-Machine (CHM par la suite) atteste depuis ces dernières années de progrès significatifs dont témoigne l'apparition récente d'applications dans lesquelles les modes de communication s'orientent vers un véritable dialogue coopératif en langage naturel. Force est néanmoins de constater que, si leur existence est admise, ces applications ne font pas encore partie de l'univers culturel du grand public dont rendent compte les dictionnaires généraux. Ceci explique peut-être les présupposés qu'ont les usagers eux-mêmes sur les capacités et surtout les limites de tels systèmes. Dans une interaction langagière avec une machine, on remarque, par exemple, qu'il est difficilement concevable pour un utilisateur d'oser prendre une initiative, sans y avoir été préalablement invité par la machine, pour corriger une partie de l'information dont il a perçu qu'elle avait été mal comprise, alors même que de telles initiatives auraient techniquement pu être prises en compte [Karsenty, 2001].

Sous d'autres formes cependant l'interaction transactionnelle effective avec une machine s'est banalisée. Séquences de menus, formulaires, interfaces graphiques ou tactiles sont autant de moyens courants d'interagir efficacement en temps réel avec un système informatique pour réaliser une tâche précise (résoudre un problème, se renseigner,

effectuer une réservation, concevoir, éduquer, etc.). La tendance est d'ailleurs au développement d'interfaces multimodales, intégrant, par exemple, les modalités vocales et tactiles, qui se révéleront peut-être plus conviviales, plus efficaces et plus adaptées dans certains cadres applicatifs.

Il pourrait ainsi sembler réducteur de ne considérer qu'une seule modalité dans le développement des systèmes d'interaction homme-machine. Pourtant, dans le cadre de ce mémoire, nous nous intéressons exclusivement à la communication orale homme-machine. Non que l'usage de la parole soit nécessairement la modalité d'interaction la plus "naturelle" qui soit pour l'utilisateur mais parce qu'elle peut être particulièrement utile dans certaines situations. Elle s'impose d'une part tout simplement dans le cadre d'une interaction par téléphone et nul doute que ce média sera celui utilisé à l'avenir pour que de telles applications soient mises à la disposition du plus grand nombre. Elle peut s'avérer d'autre part essentielle pour des utilisateurs souffrant de handicap moteur, un public âgé ou peu sensibilisé aux interfaces traditionnelles, ou encore dans le cadre de commandes vocales de robots effectuant des actions en milieu hostile, etc.

Le traitement de la langue parlée en situation de dialogue homme-machine est par ailleurs en lui-même un sujet de recherche à part entière. Les systèmes ne doivent pas seulement être en mesure de transcrire ce qui a été dit à partir du signal vocal produit par l'utilisateur, comme dans les systèmes de dictée vocale, mais aussi de permettre à la machine de "comprendre" le sens de l'énoncé prononcé en rapport avec l'application considérée. Ils doivent enfin être capables de conduire et de gérer l'interaction en guidant l'utilisateur, afin qu'il fournisse par exemple des informations supplémentaires, parfois indispensables à la construction de la réponse du système. La communication homme-machine se situe en effet dans une relation opérateur-tâche où la machine doit jouer un rôle collaboratif. La communication dont il sera question dans ce mémoire est dite finalisée : l'usager entre en contact avec la machine dans le but d'effectuer une tâche précise, portant sur un domaine bien défini. Nous traiterons ici de la recherche d'information, la consultation en langage naturel de bases de données dans le domaine du renseignement touristique.

Dans ce contexte, nos travaux s'attachent principalement aux processus mis en œuvre pour la *compréhension* de la parole. Parmi toutes les questions qui restent posées en communication homme-machine, notre recherche concerne plus spécifiquement l'analyse des énoncés oraux, étape essentielle pour la compréhension de l'énoncé et sur laquelle repose l'analyse du dialogue. Pour que l'interaction avec la machine soit un succès, en terme d'aboutissement adéquat de la transaction tout d'abord mais aussi et surtout en terme de convivialité et de simplicité dans le déroulement de l'échange oral, il convient d'autoriser une expression relativement libre de l'utilisateur tout en cherchant une détection fine de ses buts et de ses attentes. Sans perdre de vue les autres niveaux de traitement (reconnaissance et gestion du dialogue), nous pensons qu'une analyse détaillée, fondée sur des informations grammaticales, syntaxiques et sémantiques, et réalisée au cours du processus de compréhension, peut permettre une meilleure prise en considération de la richesse informationnelle des messages oraux spontanés. Nos travaux porteront ainsi principalement sur la dimension purement langagière de la CHM.

Ce mémoire est divisé en trois parties. La première partie est une introduction aux techniques et stratégies mises en œuvre dans le développement des systèmes interactifs vocaux. Le chapitre 1 introduit ce domaine de recherche en faisant le point sur les systèmes opérationnels existants ainsi que sur les principales difficultés liées au traitement de la langue parlée. Après une présentation de l'architecture générale des systèmes traditionnels, nous développons plus particulièrement les différentes stratégies utilisées en reconnaissance de la parole (en amont de la chaîne de traitement) puis en compréhension dans le cadre du domaine d'application considéré. Le chapitre 2 est un résumé des avancées réelles dans ces deux domaines de recherche. Nous y analysons les possibilités et les limites des systèmes actuels et discutons de la place laissée dans ce contexte aux traitements reposant sur des informations linguistiques fines, principalement grammaticales et syntaxiques. Les améliorations que de tels traitements sont susceptibles d'apporter sont ensuite évoquées. Nous abordons en particulier la question de la généralité des méthodes de compréhension traditionnelles, fortement guidées par un domaine applicatif généralement très restreint, et de leur portabilité à des domaines plus riches. Nous nous interrogeons enfin sur quelques questions qui nous semblent essentielles. Comment concilier une analyse linguistique fine des énoncés oraux et la contrainte de robustesse indispensable pour faire face aux "perturbations" attestées dans les productions orales issues de la reconnaissance ? Comment détecter les auto-corrrections des locuteurs ? Faut-il chercher à les corriger ? Quel est le niveau de détail des représentations sémantiques qu'il est souhaitable et possible de produire ? Quelles sont les informations à prendre en compte pour aider au mieux la gestion du dialogue ? Les premières réponses que nous apportons définissent les objectifs que nous nous sommes progressivement fixés dans ce travail de thèse.

Dans la seconde partie, nous présentons le système de compréhension hors-contexte (ROMUS pour **RObust Message Understanding System**) que nous proposons. Le chapitre 3 fait état de la démarche que nous avons adoptée pour le prototypage de notre système (analyse des usages, linguistique de corpus) et pose les fondements de notre approche : une stratégie d'analyse en deux étapes (segmentation syntaxique partielle suivie d'une analyse globale sémantico-pragmatique<sup>1</sup>) permettant l'utilisation du pouvoir structurant de la syntaxe et l'intervention précoce de la sémantique. Ce cahier des charges nous amènera rapidement aux frontières d'un autre domaine de recherche : celui du Traitement Automatique du Langage Naturel (TAL par la suite), traitant principalement de la langue écrite, dont nous nous inspirons dans les choix des techniques et formalismes utilisés. Après la description du domaine d'application que nous avons choisi, le renseignement touristique, et une présentation générale des représentations sémantiques attendues (chapitre 4), nous développons et détaillons successivement (chapitres 5 et 6) chacune des deux grandes étapes d'analyse que nous proposons. Le chapitre 5 traite plus précisément des difficultés de l'étiquetage grammatical des productions orales et présente les unités de segmentation que nous avons retenues. Nous nous inspirons, pour cette étape, reposant sur l'application d'une cascade de transduc-

---

<sup>1</sup>Nous faisons ici référence au domaine de l'application et non à la prise en compte du dialogue.

teurs, des recherches effectuées en TAL robuste. Le chapitre 6 justifie notre choix, pour l'analyse sémantico-pragmatique du formalisme des grammaires de lien. Ce formalisme lexicalisé, proche des grammaires de dépendance, a été initialement conçu pour l'analyse syntaxique de la langue écrite. Nous y décrivons l'organisation et la constitution de notre lexique sémantique. La dernière partie de ce chapitre donne quelques exemples de la façon selon laquelle sont appréhendés les problèmes liés aux phénomènes spécifiques de l'oral : robustesse face aux inattendus mais aussi traitement des procédés linguistiques réguliers attestés en corpus.

La dernière partie analyse les résultats de cette expérimentation par rapport à nos objectifs initiaux. Nous nous appuyons pour cela sur une évaluation de notre système réalisée dans le cadre de la campagne d'évaluation "par défi" des systèmes de compréhension de la parole, initiée par le groupe de travail 5.1 *compréhension robuste* du GDR-I3 du CNRS. Nous présentons une analyse qualitative des résultats obtenus par notre système dans le cadre de cette campagne au chapitre 7. Nous analysons également dans ce chapitre le comportement de notre système face aux erreurs de la reconnaissance de la parole et des aménagements envisagés pour traiter le problème des mots hors vocabulaire ou mal reconnus par la reconnaissance. Enfin, le chapitre 8 dresse un bilan complet des enseignements qui peuvent être dégagés de cette expérimentation. Il y sera question des limites et des perspectives d'évolution du système.



## Première partie

# Comprendre la parole : problématiques et enjeux



## Chapitre 1

# Compréhension automatique de la parole et dialogue homme-machine

L'application des techniques d'analyse des langages formels et la confrontation de la linguistique aux exigences de la modélisation informatique ont conduit en quelques décennies à l'émergence d'un véritable domaine de recherche : la linguistique informatique. Les progrès réalisés en électronique (augmentation de la vitesse des processeurs et des capacités de mémoire des ordinateurs) associés à la disponibilité des ressources linguistiques électroniques ont permis le développement d'applications concrètes, dont la demande s'est progressivement faite croissante. Ces applications, chacune circonscrite à un domaine particulier mais nécessitant des traitements automatiques du langage naturel, écrit ou oral, relèvent aujourd'hui de ce que l'on appelle plus généralement *l'ingénierie des langues* [Cunningham, 1999]. Leur développement a surtout été rendu possible grâce aux efforts et aux progrès réalisés dans la définition d'outils pour le traitement de la langue (algorithmes d'analyse, de génération, d'acquisition de connaissances linguistiques), de modèles formels (grammaires, formalismes lexicalisés) pour représenter ces connaissances, et dans le développement de procédés de création et de gestion de ces ressources. On trouvera dans [Pierrel et Romary, 2000], une introduction complète et un bilan concernant les domaines couverts par l'ingénierie des langues.

Nous nous intéressons dans le cadre de cette thèse à l'un des grands domaines applicatifs de l'ingénierie des langues : le dialogue oral homme-machine. En guise d'introduction, ce chapitre fait tout d'abord le point sur les systèmes opérationnels existants, témoins de l'avancée des recherches dans ce domaine (§ 1.1). Nous verrons que réaliser un système de dialogue homme-machine tirant réellement parti de la souplesse des langues naturelles relève encore pour une grande part de la théorie ; la mise en pratique est loin d'être à l'ordre du jour tant les complexités conjuguées de la langue et du dialogue sont importantes. Nous insistons principalement dans ce chapitre sur les principales difficultés liées au traitement automatique de la langue parlée (§ 1.2), en n'évoquant dans un premier temps que brièvement celles liées à la gestion du dialogue. Après avoir rappelé l'architecture et les principales caractéristiques des systèmes de dialogue homme-machine, la dernière partie de ce chapitre (§ 1.3) évoque ainsi es-

sentiellement les stratégies et méthodes actuellement utilisées en communication orale homme-machine pour *reconnaître* la parole puis pour la *comprendre*, le thème central de ces travaux de recherche.

## 1.1 Quelles applications ?

Les dernières années ont vu l'apparition progressive de systèmes opérationnels permettant de reconnaître et de comprendre le langage parlé. Le cadre de la simple commande vocale a rapidement été dépassé, y compris dans les applications de domotique (contrôle vocal d'appareils électroniques domestiques). Les projets développant des systèmes de contrôle de l'environnement, comme par exemple le projet HOME (dédié principalement aux personnes âgées) [Shao, Tazine, Lamel, Prouts, et Schröter, 1998], cherchent ainsi notamment à intégrer la compréhension de commandes en langue naturelle c'est-à-dire non limitées à de simples mots-clefs.

Nous nous intéresserons ici aux systèmes permettant non seulement de reconnaître et de comprendre la parole, mais aussi de conduire un dialogue avec l'utilisateur. Ces systèmes, souvent qualifiés de *systèmes conversationnels* [Zue, 1997], ont pour but d'aider les utilisateurs à accéder à des informations ou à résoudre des problèmes spécifiques en s'appuyant en partie ou totalement sur une interface vocale. Ils doivent ainsi être capables de "comprendre" les énoncés des utilisateurs et d'y répondre, éventuellement verbalement. Un grand nombre de serveurs vocaux interactifs ont vu le jour ces dernières années. Les applications les plus fréquentes sont celles pour lesquelles l'interaction se fait via le téléphone. Elles ont d'abord principalement concerné l'accueil et le routage téléphonique (détermination des motifs de l'appel, redirection automatique de l'appel vers le correspondant ou une personne capable de résoudre le problème évoqué, prise de message en l'absence du correspondant, etc.) [Ehrlich et al., 1997; Gorin, Riccardi, et Wright, 1997; Lokbani et White, 1999] et la consultation d'annuaires comme celui des services audiotel (emploi et météo) mis en place par le CNET [Sadek, 1996]. D'autres applications permettent aux utilisateurs d'accéder à des sources d'informations dans un domaine défini. Elles ont concerné par exemple l'aide à la navigation entre différents lieux (restaurants, hôtels, banques, ...) situés dans une même zone géographique. Dans ce cadre, le système permet un accès aux informations concernant les distances, les directions à prendre, le temps des trajets entre les différents sites ainsi que des informations portant sur les sites eux-mêmes (adresses, numéros de téléphone, ...). Cela a été le cas du système multilingue VOYAGER<sup>1</sup>[Glass et al., 1995].

Ce type d'application, la recherche d'informations finalisée, a été particulièrement développé ces dernières années des deux côtés de l'Atlantique. Aux États-Unis, le projet ATIS (Air Travel Information System) [Price, 1990] financé par la DARPA<sup>2</sup> a permis le développement et l'évaluation [Pallett et al., 1994] de nombreux systèmes dans le domaine du renseignement et de la planification de voyages aériens : PHOENIX [Is-

<sup>1</sup>Dans le cas des systèmes multilingues, les travaux sont fondés sur l'hypothèse qu'une représentation sémantique commune peut-être extraite quelle que soit la langue, au moins dans des domaines restreints.

<sup>2</sup>Defense Advanced Research Projects Agency, à l'époque ARPA.

sar et Ward, 1993], PEGASUS [Zue et al., 1994], CHRONUS [Levin et Pieraccini, 1995], HUM [Miller, Stallard, Bobrow, et Schwartz, 1996] mais aussi GALAXY [Goddeau et al., 1994]. Ce dernier système, distribué, permettant l'exploitation de supports d'informations hétérogènes (bases de données spécifiques, Web, ...) a pour particularité de travailler sur des domaines multiples : le voyage aérien, l'aide à la navigation précédemment citée et les renseignements météorologiques. Développé à sa suite, le système JUPITER [Zue et al., 1997], dédié uniquement à l'univers de la météorologie, repose sur la même architecture distribuée.

En Europe, les projets ESPRIT-SUNDIAL (Speech Understanding and Dialogue) [Peckham, 1993] et RAILTEL (Railway Telephone Information Service) puis ARISE (Automatic Railway Information System in Europe) ont permis de façon similaire le développement de systèmes d'accès automatique par téléphone à des renseignements concernant principalement le domaine ferroviaire. À titre d'exemple, le système ARISE du LIMSI [Lamel et al., 2000], faisant suite au système du même laboratoire développé dans le cadre du projet RAILTEL [Bennacef, Devillers, Rosset, et Lamel, 1996], délivre les horaires, les réductions et services disponibles et effectue des simulations de prix et de réservations. À la différence des projets américains, l'accent a été mis sur l'aspect multilingue de ces services. La plupart des prototypes ont été mis en service auprès du grand public dans le cadre de procédures d'évaluation. Citons le système EVAR [Gallwitz et al., 1998] de l'université d'Erlangen pour les chemins de fer allemands, le système de Philips [Aust, Oerder, Seide, et Steinbiss, 1995] et ses versions néerlandaise [Os, Boves, Lamel, et Baggia, 1999] et française<sup>3</sup> [Pérénou, Calmès, Lavelle, et Tronel, 1998].

D'autres applications se distinguent par ailleurs des précédentes par la nature plus coopérative de l'interaction qu'elles imposent. Ainsi, dans le système TRAINS [Allen, Miller, Ringger, et Sikorski, 1996], dédié à la planification des transports ferroviaires de marchandises, l'objectif de l'interaction n'est pas seulement d'accéder à des informations mais aussi de chercher à optimiser les différents itinéraires. Dans un autre contexte, celui de la communication médiatisée par ordinateur, comme dans le projet VERMOBIL [Bub et Schwimm, 1996], le système sert de médiateur entre deux interlocuteurs dans un cadre de prise de rendez-vous dont il assure la traduction. Le système doit être capable de se substituer à un locuteur à tout moment dès que celui-ci est incapable de traduire certains mots de son message<sup>4</sup>. C'est ce domaine (planification et négociation de rendez-vous) qui a servi de domaine d'expérimentation pour le système de traduction multilingue JANUS [Waibel et al., 1996]. Les versions plus récentes de JANUS portent sur les domaines liés à la planification de voyages (trajets, réservation hôtelière, etc.) [Levin et al., 2000] dans le cadre des projets C-STAR et maintenant NESPOLE! [Rossato, Blanchon, et Besacier, 2002].

Les systèmes que nous avons présentés jusqu'à présent ne font intervenir en entrée qu'une modalité d'interaction : la parole. Si l'interface vocale se justifie pleinement dans le cadre d'une interaction par téléphone ou des conditions particulières privant les uti-

---

<sup>3</sup>La plateforme DEMON a été développée par l'IRIT.

<sup>4</sup>Dans ce projet national allemand de traduction vocale en face à face, la langue cible est l'anglais.

lisateurs d'une utilisation aisée des modalités tactiles et visuelles (handicap, conduite d'un véhicule<sup>5</sup>, ...), la communication dite multimodale peut être dans certains cas plus conviviale [Oviatt, 1996]. Ainsi le système GEORAL [Siroux, Guyomard, Jolly, Multon, et Remondeau, 1995], ayant pour objectif de donner des informations touristiques sur la région de Lannion, ainsi que GALAXY-II [Seneff, Lau, et Poliforni, 1999], utilisent le mode tactile et vocal. De la même manière, le projet ESPRIT MASK (Multimodal Multimedia Service Kiosk) [Temem, Lamel, et Gauvain, 1999] a eu pour but la conception d'un kiosque d'information et de distribution, dans le domaine ferroviaire, avec une interface conviviale combinant les modalités tactiles et vocales. Une évaluation de ce kiosque, conduite en 1998 à la gare Saint-Lazare, est prometteuse notamment en ce qui concerne le temps nécessaire, réduit par rapport aux kiosques traditionnels, pour mener à bien la transaction. On pourra se reporter à [Lamel, Bennacef, Gauvain, Dartigues, et Temem, 2002] pour les détails de cette évaluation et en particulier pour l'étude différentielle des choix des modalités attestés en fonction de la nature de l'information recherchée (consultation des horaires, des prix ou réservation de billets). L'intégration et la gestion de différentes modalités d'interaction posent des problèmes que nous n'aborderons pas dans le cadre de cette thèse. Nous nous placerons en effet uniquement dans le cadre d'une interaction purement orale.

L'éventail des systèmes que nous avons décrits ici, loin d'être exhaustif<sup>6</sup>, est cependant très représentatif des domaines d'applications envisagés. Comme nous pouvons le constater, à l'inverse d'une conversation de salon, le dialogue homme machine est étroitement finalisé : il vise un échange d'informations en vue de la réalisation d'une tâche nettement spécifiée. Le vocabulaire jugé nécessaire pour atteindre les objectifs fixés par ces applications varie entre 1500 et 2000 mots parmi lesquels figurent les noms des villes ou des lieux concernés<sup>7</sup>.

Nos travaux de thèse s'inscrivent également dans le contexte de la communication homme-machine *finalisée*. Ils concerneront plus précisément la recherche d'informations dans le domaine du renseignement touristique, domaine d'application sur lequel nous reviendrons au chapitre 4. Dans la suite de ce document, nous ferons une distinction entre le *domaine de l'application* et le *domaine de la tâche*. Le domaine de la tâche est un sous-ensemble du domaine d'application. Il décrit l'ensemble des requêtes, qui pourront être traitées par le système au sein du domaine de l'application. Par exemple, dans le domaine ferroviaire, certains des systèmes présentés précédemment ne considèrent que la tâche relative à la consultation des *horaires* des trains et ne peuvent traiter des demandes concernant les tarifs ni effectuer des réservations. Cette distinction est importante car, comme nous le verrons, le développement de tels systèmes requiert, à un

---

<sup>5</sup>Le projet européen VODIS par exemple a eu pour but d'élaborer un système d'aide à la navigation guidé par la voix à l'intérieur des véhicules.

<sup>6</sup>Pour un aperçu plus large des applications en reconnaissance et compréhension de la parole, on pourra se reporter à [Néel, Chollet, Lamel, Minker, et Constantinescu, 1996].

<sup>7</sup>Ce nombre varie, selon les applications, entre 200 (PEGASUS) et 1200 (EVAR) en conservant une taille globale du vocabulaire à 2000 mots. Les applications concernant le renseignement ferroviaire comportent à l'heure actuelle environ 500 noms de gare.

moment ou un autre, la prise en compte du domaine concerné et de la (ou des) tâche(s) visée(s).

Après ce tour d’horizon des systèmes existants, les paragraphes suivants font un bilan des stratégies et des techniques utilisées pour mener à bien l’interaction. Nous commencerons par évoquer les difficultés qui se posent dans le traitement des requêtes orales exprimées librement.

## 1.2 Quelles difficultés ?

Le traitement automatique de la langue parlée est un problème délicat qui a suscité de nombreuses recherches depuis les années 1960. Nous exposons dans ce paragraphe les principales difficultés auxquelles doivent faire face les systèmes devant intégrer des processus de traitements automatiques de la parole. Nous nous concentrerons dans un premier temps sur les caractéristiques du matériau que nous devons traiter : les productions orales spontanées. Nous insisterons ensuite sur les spécificités induites par la situation très particulière qu’est la communication homme-machine.

### 1.2.1 La parole : quel matériau ?

Le Traitement Automatique du Langage Parlé (TALP par la suite) concerne tout d’abord le domaine du traitement du signal. Les systèmes de TALP travaillent en entrée sur un signal acoustique vocal non décodé. Une étape de reconnaissance, de transformation du signal vocal enregistré en une suite de mots, est donc nécessaire. Les principaux problèmes sont posés par la variabilité du signal acoustique :

- variabilité *intrinsèque* autorisant des structures acoustiques différentes pour un même contenu phonétique,
- et variabilité *contextuelle* autorisant inversement des formes acoustiques proches pour des contenus phonétiques différents.

Ces problèmes de variabilité relèvent du domaine de la reconnaissance des formes. Certaines approches ont ainsi envisagé la reconnaissance de la parole sous la forme d’un décodage acoustico-phonétique analytique à base de règles [Méloni, 1982]. Les augmentations indispensables de la taille des vocabulaires à reconnaître, augmentant par là-même la variabilité intra- et inter- locuteurs, ont toutefois conduit au début des années 1980 à un changement radical de paradigme [Huckvale, 1996] au profit d’approches globales stochastiques, sur lesquelles nous reviendrons au paragraphe 1.3.1. On est ainsi passé de systèmes capables de reconnaître en mots isolés un seul locuteur prononçant un lexique d’une cinquantaine de mots, à des systèmes capables de reconnaître en parole continue n’importe quel locuteur utilisant un vocabulaire de 60 000 mots et plus.

Le traitement de la parole ne se limite cependant pas à ces travaux d’identification du signal. Contrairement aux systèmes de dictée vocale [Hunt, 1998], où l’on se contente de transcrire ce qui a été dit, ce qui tient lieu d’énoncé reconnu doit être analysé. Or, cette analyse n’est pas évidente. En particulier, les techniques d’analyse syntaxique

développées pour l'écrit se révèlent impuissantes face aux énoncés oraux. Ainsi que le notait Luzzati :

(Cit.1.1) “*En dehors même des problèmes liés au traitement du signal de parole, le traitement automatique d'énoncés **non normés** tels que les énoncés oraux ne nous paraît guère envisageable à l'aide des méthodes classiques.*”  
(développées pour le langage écrit) [Luzzati, 1989] p. 12

Les langages écrit et parlé sont en effet issus de deux conditions de production très différentes<sup>8</sup> et l'oral se caractérise par la présence de phénomènes venant "perturber" les régularités syntaxiques sur lesquelles reposent en général les stratégies d'analyse de l'écrit. Leur présence laisse également penser, en première approximation, que l'adaptation à l'oral des techniques conçues pour l'écrit est elle-même difficile.

Dans le paragraphe suivant, nous donnons une première idée de ces principaux phénomènes. Nous adoptons dans un premier temps (§ 1.2.1.1) un schéma comparatif strict par rapport à la langue écrite en distinguant les phénomènes spécifiques de l'oral, c'est-à-dire absents de l'écrit<sup>9</sup>, des phénomènes linguistiques dont l'utilisation y est tout simplement très fréquente. Nous abandonnerons ce schéma au chapitre 3 où nous analyserons de manière plus détaillée, et sans prendre cette fois la langue écrite comme la norme de référence, les principales études linguistiques du français parlé. Cette première description des productions orales est très générale et ne concerne pas uniquement les situations dialogiques. L'influence de la situation particulière de la CHM est abordée au paragraphe 1.2.1.2.

### 1.2.1.1 Le français parlé

Les premiers phénomènes à prendre en compte proviennent du caractère spontané des productions orales. On retrouve dans ces productions la trace même de leur élaboration. Outre les courtes hésitations et les nombreuses reprises, il est en effet important de constater qu'environ un tiers du temps de parole des locuteurs (qu'ils soient engagés dans un dialogue ou non) est destiné à des commentaires<sup>10</sup> très divers portant sur ce qu'ils sont en train de dire [Blanche-Benveniste, 1990]. Le *dire* et la *façon de le dire* sont donc extrêmement liés dans le discours oral et ces commentaires, emmêlés à la production, sont parfois très difficilement démêlables, même prosodiquement<sup>11</sup>. On

---

<sup>8</sup>Il n'existe pas de frontière claire entre écrit et oral. On ne peut pas opposer complètement ces deux productions langagières [Biber, 1986; Blanche-Benveniste et Bilger, 1999]. Les linguistes se refusent d'ailleurs à établir une quelconque comparaison entre "écrit" et "oral" [Gadet, 1989; Blanche-Benveniste, 1990], la transition entre les deux productions devant plutôt être appréhendée sous la forme d'un continuum de genres [Biber, Johansson, Leech, Conrad, et Finegan, 1999; Blanche-Benveniste et Bilger, 1999].

<sup>9</sup>du moins peu présents pour certains d'entre eux.

<sup>10</sup>Ces commentaires vont de la recherche par les locuteurs des "bonnes dénominations" (des formules du type "*comment ça s'appelle déjà*" par exemple) utilisant ou non des procédés d'approximation variés (les mots *truc*, *chose*, ...) à des remarques parfois très longues de constante évaluation de leur propre discours.

<sup>11</sup>Blanche-Benveniste, au cours de son exposé lors de la quatrième journée d'étude de l'ATALA sur les grammaires de dépendance, Paris, 16 mai 1998



distingue différents types de phénomènes dus à la spontanéité de l'interaction [Blanche-Benveniste, 1990; Block et Schachtl, 1995]. Nous les présentons ci-après en adoptant un point de vue purement descriptif. Les problèmes qu'ils posent à l'analyse automatique seront analysés aux chapitres 2 et 3.

- **les hésitations.** Elles correspondent aux manifestations les plus courantes des procédés dits de "recherche de dénomination" [Blanche-Benveniste, 1990]. Les plus simples vont de la courte pause, silencieuse, dans la prononciation à des pauses plus longues remplies éventuellement par des interjections (*euh, hum, ...*) ou des appuis du discours (*bon, donc* ou *alors* par exemple). L'utilisation de ces mots pour la structuration du discours est très variable selon les locuteurs [Luzzati, 1989]. Les hésitations peuvent revêtir cependant des formes plus complexes allant de l'inachèvement à l'incise que nous évoquons par la suite. Elles peuvent s'accompagner par ailleurs de répétitions ou de corrections.
- **les inachèvements.** Hormis les interruptions mutuelles des interlocuteurs dans une situation dialogique, on les rencontre lors :
  - de brusques changements de la stratégie du locuteur [Morel, 1989]. Celui-ci peut ne pas achever son énoncé parce qu'il a perçu que son interlocuteur a compris son intention. Dans le cas contraire, l'énoncé est en général suivi d'une incise qui explicite les raisons de ce changement. L'énoncé inachevé est alors éventuellement reformulé pour résoudre certains problèmes communicatifs. Dans l'exemple (E.1.1), le locuteur envisage d'indiquer un itinéraire puis se ravise :
 

(E.1.1) *alors à ce moment là euh vous allez prendre euh / si vous voulez alors je vous donne (...)*<sup>12</sup> [corpus Murol.1<sup>13</sup>]
  - de la recherche volontaire de la collaboration du partenaire dans le dialogue :
 

(E.1.2) *j'ai pas très bien noté où il se trouvait à gauche ou* –<sup>14</sup>  
[corpus Murol.1]
  - de commentaires effectués à voix haute sur les actions réalisées conjointement au discours. Dans l'exemple (E.1.3), le locuteur essaye de corriger avec une souris, un plan affiché sur un moniteur :
 

(E.1.3) *ah mince j'ai vraiment des problèmes avec* – [corpus Murol.1]
  - de pannes lexicales où le locuteur, faute de pouvoir accéder dans l'instant à un lexique approprié, laisse un syntagme en blanc :
 

(E.1.4) *ça colorait légèrement le bouillon pour qu'il soit* –  
[corpus Viandox [Blanche-Benveniste, 1990]]

Ces inachèvements ou amorces avortées posent essentiellement des problèmes de

---

<sup>12</sup>Les points de suspension indiquent que l'énoncé se poursuit normalement. Le / précise l'endroit où a lieu le changement de stratégie.

<sup>13</sup>Le corpus Murol, enregistré et transcrit au CLIPS/ICP [Bessac et Caelen, 1995], réunit deux interlocuteurs simulant une conversation téléphonique entre un touriste et l'office de tourisme local.

<sup>14</sup>Nous marquons par – l'endroit où l'énoncé reste inachevé.

complétude à l'analyse automatique.

- **les répétitions.** Nous regroupons sous ces termes tous les phénomènes dans lesquels le locuteur répète ou corrige une portion de son énoncé. Nous parlerons pour ces derniers cas de répétitions correctives ou d'auto-corrrections. La répétition peut porter sur un mot, un groupe de mots ou une formulation qu'il a utilisée. Le procédé de répétition peut être appréhendé dans un premier temps comme un processus inconscient destiné à meubler l'attente induite par la recherche de la bonne dénomination. Les dernières paroles sont alors répétées jusqu'à pouvoir reprendre le fil de l'énoncé. La répétition corrective est quant à elle souvent utilisée pour corriger une anticipation erronée. Le locuteur peut également se reprendre lorsqu'il doute *a posteriori* de l'adéquation de la formulation qu'il vient d'employer. Il ne s'agit ainsi pas seulement de la correction d'erreurs, ou de lapsus phonétiques, mais également de procédés permettant de clarifier ses attentes [Okada et Otsuka, 1993].

On distingue généralement deux types de répétitions selon la présence ou l'absence d'éléments entre le ou les mots répétés ou repris. Ces *marques* peuvent être soit des interjections mettant en évidence la recherche de dénomination, soit des appuis du discours (par exemple *oui, donc* pour les répétitions et *non* ou *pardon* pour les répétitions correctives), soit encore une incise de longueur très variable. Les répétitions non marquées sont en général caractérisées par un inachèvement ou une amorce lexicale aussitôt reprise. Si les répétitions correctives présentent des difficultés similaires à la répétition, leur interprétation est toutefois plus délicate surtout dans les cas où elles ne sont pas marquées par une particule négative (*pas, non* par exemple). Remarquons enfin que dans la plupart des cas, ces répétitions ne reprennent pas exactement la partie initiale. C'est le cas très fréquent à l'oral du phénomène d'enrichissement lexical où l'on adjoint au groupe de mots répétés un ensemble de termes permettant de le préciser. Il peut s'agir d'une simple qualification comme dans les exemples (E.1.5) et (E.1.6) ou d'un enrichissement complet d'un lexème (E.1.7). La limite entre répétition et répétition corrective est dans ce cas encore plus délicate.

(E.1.5) *tu places **une porte une double porte** au milieu en bas de la pièce* [corpus Levelt.5<sup>15</sup>]

(E.1.6) (...) *un carré **à gauche non quand même pas tout à fait à gauche*** [corpus Levelt.5]

(E.1.7) *vous tombez sur **une petite avenue un petit chemin un petit chemin de graviers** et vous tombez sur le musée (...)* [corpus Murol.1]

---

<sup>15</sup>Ce corpus a été enregistré à l'ICP [Ozkan, 1994]. Il s'agit d'une expérience de faux magicien d'Oz destinée à étudier les procédés de description de configurations spatiales employées par l'utilisateur d'un système de dessin. Le terme de magicien d'Oz désigne une expérience de simulation dans laquelle un individu, le compère, joue le rôle du système informatique. On distingue les expériences du vrai magicien, où le locuteur n'est pas au courant de l'existence du compère, des situations de faux magicien où il est précisé au locuteur qu'un être humain remplace la machine.

- **les incises**. Elles marquent une parenthèse dans le déroulement du discours. Blanche-Benveniste les considère comme :

(Cit.1.2) “*des séquences qui s’introduisent dans le discours, sous un autre registre intonatif, sans participer à l’organisation de l’énoncé, mais sans le perturber pour autant*”. [Blanche-Benveniste, 1990] p. 145

Dans le cadre qui nous intéresse, on peut ainsi qualifier d’incise toute partie de l’énoncé – présente à quelque endroit que ce soit et pouvant casser la régularité syntaxique de ce dernier – qui ne relève pas de l’acte de langage principalement véhiculé. Les incises de recherche de dénomination, les plus simples, ont avant tout un rôle d’appui du discours. D’autres peuvent toutefois correspondre à des productions cohérentes qui peuvent concerner le message véhiculé (E.1.8) ou être de simples digressions. Les locuteurs n’ont aucun mal à “retomber” sur le fil syntaxique de leur énoncé et ce phénomène est très fréquent dans les productions orales. La difficulté d’analyse des énoncés à incises réside dans la grande diversité des lieux d’intrusion et de leur longueur très variable (E.1.9 et E.1.10).

(E.1.8) *on revient au carré 3 c’est-à-dire le carré 3 c’est celui d’en dessous du carré de départ voilà et on va faire la même chose* [corpus Levelt.5]

(E.1.9) *je vais vous indiquer un petit peu comment vous allez procéder c’est assez simple quand vous allez sortir de la gare (...)* [corpus Murol.1]

(E.1.10) *alors vous prenez la l’avenue de Grande Bretagne et alors à ce moment là il faudrait que vous gariez votre vélo parce que vous arrivez à dans une zone piétonniè... piétonne donc à ce moment là euh la rue Lakanal est piétonne donc l’avenue de Grande Bretagne (...)* [corpus Murol.1]

Nous reviendrons au chapitre 3 sur l’apparente liberté d’allure du discours oral induite par les phénomènes que nous venons d’évoquer. Nous verrons en particulier que les études descriptives spécifiquement menées sur les modes de production de l’oral ont permis de mettre en évidence une complexité structurelle et une certaine logique de production des énoncés oraux. La liberté de l’oral dans l’interaction est en réalité, pour reprendre une expression de Kerbrat-Orecchioni, une liberté *surveillée* [Kerbrat-Orecchioni, 1999].

Ces travaux descriptifs ont également démonté certains mythes comme celui de la pauvreté ou de la simplicité de l’oral. Certes, un locuteur s’exprimant à l’oral a par exemple souvent recours à l’intonation pour véhiculer son message. La communication orale directe lui permet par ailleurs une certaine approximation informative qu’il peut ajuster à tout moment en fonction des réactions de son interlocuteur. Mais ces possibilités ne se traduisent pas par une diminution de la richesse informative des énoncés oraux<sup>16</sup>. D’une manière générale, le locuteur peut choisir la façon de s’exprimer qui

<sup>16</sup>ni d’ailleurs par une diminution de la complexité des organisations pouvant être mises en œuvre. À titre d’exemple on pourra se reporter aux études de [Halliday, 1985].

lui sera la moins coûteuse et la plus appropriée par rapport à ses objectifs. Il utilise dans cette optique, avec une forte régularité, des procédés, dont certains ne sont pas spécifiques à l'oral, que nous évoquons ci-après.

– **Extractions et doubles-marquages : effets d'insistance**

Afin de marquer une insistance sur une partie de son énoncé, les locuteurs font un usage fréquent à l'oral de constructions où l'ordre standard des mots – le schéma canonique sujet/verbe/compléments pour les assertions par exemple – est modifié. Un élément de l'énoncé est ainsi détaché (extrait) du reste de l'énoncé. On distingue les détachements à gauche (antépositions) des détachements à droite (postpositions) [Gadet, 1992] respectivement illustrés par les exemples (E.1.11) et (E.1.12) :

(E.1.11) *si**on** **cette salle de musculation** il faut peut-être réserver*  
[corpus Murol.1]

(E.1.12) (...) *parce que c'est en sens unique **le boulevard Voltaire***  
*hein* [corpus Murol.4]

Souvent l'élément déplacé est repris par un clitique ou un pronom qui le remplace à la position standard. C'est le cas de l'exemple (E.1.12) précédent où le sujet, postposé, est remplacé par un pronom démonstratif et de l'exemple (E.1.13) où l'argument *la cave* est repris par un clitique :

(E.1.13) ***la cave** on l'a déjà beaucoup rétrécie* [[Blanche-Benveniste, 1990]]

Nous nommerons ces procédés, à la suite de [Blanche-Benveniste, 1990], **doubles-marquages** et nous appellerons **inversions** les procédés impliquant une modification d'ordre qui ne se manifeste par aucune autre marque linguistique (E.1.14) :

(E.1.14) ***jusqu'à 21h30** euh vous pouvez manger au restaurant*  
[corpus Murol.1]

D'autres constructions permettent d'introduire explicitement la partie de l'énoncé détachée. Il s'agit par exemple du dispositif de pseudo-clivage [Roubaud, 2000] et principalement des extractions à **présentatifs** construites à partir d'un introductif utilisant le verbe être (*c'est*) ou avoir (*il y a, j'ai, ...*) et suivi d'une (fausse) subordonnée introduite par *qui* ou *que* (E.1.15) et (E.1.16) :

(E.1.15) *c'est bien le renseignement que vous vouliez avoir*  
[corpus Air-France.2<sup>17</sup>]

(E.1.16) *j'ai le docteur Burchis qui a demandé à me prendre dans son service* [corpus Durand [Blanche-Benveniste, 1990]]

Nous reviendrons au chapitre 3 sur la question de l'ordre variable des mots, importante pour le traitement automatique des énoncés oraux.

<sup>17</sup>Le corpus Air-France, recueilli par l'équipe de M.A. Morel à l'université de la Sorbonne Nouvelle, puis retravaillé par P. Nerzie dans le cadre du projet DALI [Sabah, 1994], réunit un ensemble de conversations téléphoniques entre un centre de réservation aérienne et différents clients, des particuliers ou des personnels d'agence de voyage.

– **Énumérations, ellipses et anaphores**

Ces phénomènes grammaticaux ne sont pas spécifiques à l’oral. Nous les citons ici uniquement en raison de leur fréquence d’utilisation et des problèmes qu’ils poseront à l’analyse (cf. chapitre 3).

- **les énumérations.** À une même place syntaxique on trouve une liste de composants de même type (verbes, adjectifs, noms) qui ne sont pas explicitement liés entre eux par une coordination (E.1.17, E.1.18). Ce procédé, très fréquent et naturel à l’oral, peut parfois surprendre à l’écrit.

(E.1.17) *je lui apprenais à lire à écrire* [[Blanche-Benveniste, 1990]]

(E.1.18) *oui il y a des réductions famille nombreuse étudiant et carte vermeille* [corpus Murol.1]

- **les ellipses.** Ce sont des phénomènes conscients qui évitent au locuteur des répétitions fastidieuses. D’une manière générale, elles caractérisent un manque dans l’énoncé. La partie omise reprend implicitement un élément qui a été évoqué auparavant. C’est le cas de la relation syntaxique portée par le verbe *travailler* dans l’exemple (E.1.19<sup>18</sup>) qui est rappelée sans être explicitement répétée dans la deuxième partie de l’énoncé.

(E.1.19) *il travaille le soir elle le matin* [corpus Murol.1]

Souvent, un simple terme : *aussi, jamais, non* suffit à rappeler implicitement la partie que l’on veut éviter de répéter (E.1.19 et E.1.20).

(E.1.20) *ça dépend si vous louez ou **pas*** [corpus Murol.1]

(E.1.21) *vous avez cette église très très réputée les vitraux **surtout*** [corpus Murol.1]

Ce procédé autorise une grande concision, c’est pourquoi on le retrouve beaucoup à l’oral. Dans les situations dialogiques en particulier, un locuteur, en réponse à une question, peut se contenter de préciser l’élément qui manquait à son partenaire reprenant ainsi implicitement une partie de l’énoncé de son interlocuteur. L’ellipse peut se rencontrer également dans les reprises où elle se traduit par la répétition d’une information fournie par l’interlocuteur [Pierrel, 1988]. Il s’agit d’un procédé d’écho, qui sert de marque d’approbation ou de simple maintien du contact phatique.

- **les anaphores.** Les locuteurs en font fréquemment usage dans les situations dialogiques : utilisation de déictiques et reprises, sous forme de pronoms, des éléments lexicaux prononcés au fil de l’interaction par leurs interlocuteurs. La nécessité de leur analyse est renforcée en raison de la présence des incises et des extractions évoquées précédemment dans lesquelles le rappel de l’élément lexical déplacé ou évoqué avant digression l’est souvent sous forme de pronom.

<sup>18</sup>Dans ces exemples, la partie non répétée est soulignée.

Nous regrouperons désormais sous le terme de *phénomènes spécifiques de l'oral* à la fois les phénomènes dus au caractère spontané de l'interaction orale ainsi que les procédés réguliers fréquemment utilisés que nous venons d'évoquer. L'éventail que nous avons donné n'est pas complet. Nous avons passé sous silence un certain nombre de constructions utilisées par les locuteurs qui attestent de "libertés" prises par rapport à l'écrit standard : élision des marques de négation, omission des conjonctions nécessaires au discours rapporté, coordinations non marquées où les liens logiques sont remplacés implicitement par l'intonation (parataxe), constructions des relatives, etc. Citons encore la diversité des formes dans les moyens d'exprimer l'interrogation. Si les formes par intonation ("*il se trouve où ?*" par exemple) prédominent largement à l'oral [Gadet, 1989], on atteste également de formes très particulières comme par exemple l'insistance en *c'est ... que* : "*c'est quand qu'il vient ?*". Les phénomènes que nous avons décrits nous permettent toutefois d'avoir une première appréciation des difficultés qui se posent pour l'analyse automatique des énoncés oraux.

La plupart des exemples précédents sont tirés de corpus de conversations homme-homme<sup>19</sup> qui portent sur des domaines assez différents. Or, la question de l'influence de la situation interactionnelle sur les productions orales mérite d'être étudiée. Elle fait l'objet du paragraphe suivant.

### 1.2.1.2 Le français spontané en CHM

Par rapport aux phénomènes évoqués précédemment, on ne sera pas étonné, par exemple, que les situations dialogiques très interactives, où les locuteurs sont enclins à se disputer la prise de parole<sup>20</sup>, soient caractérisées par une plus forte proportion d'inachèvements non rattrapés et de répétitions de fragments de discours – le leur ou celui de leur interlocuteur – que cela ait un rapport direct ou non avec l'information à communiquer.

Diverses études différentielles (voir par exemple [Morel, 1989]) ont montré que nous étions à même de modifier, sans effort notable, notre comportement langagier suivant la situation dans laquelle nous nous trouvons. Biber souligne [Biber, 1986] que les résultats parfois contradictoires des études sur les distinctions entre écrit et oral, proviennent précisément de l'absence de prise en compte des modes d'expression, forts distincts (aussi bien à l'écrit qu'à l'oral) que nous mettons en œuvre suivant les situations. À partir d'analyses factorielles discriminantes, il a classé différents types de corpus selon quelques facteurs, comme par exemple le degré d'interactivité que nous venons de citer. On distingue en général trois facteurs de modification d'attitude : *la familiarité* de la situation et *l'interactivité* mais aussi *la nature de la tâche* que les locuteurs ont à remplir [Morel, 1989; Letellier-Zarshenas, Nicolas, Goulian, et Antoine, 1999]. Cette dernière dimension est particulièrement importante. Dans des tâches très répétitives par exemple, comme celle du corpus Levelt qui consiste uniquement à ajouter les mêmes éléments (carrés, traits, etc.) afin d'enrichir un dessin, la proportion des énoncés elliptiques peut facile-

<sup>19</sup> Avec la nuance apportée note 15 page 10, pour le corpus Levelt.

<sup>20</sup> Ces situations se caractérisent par des interruptions mutuelles nombreuses et des énoncés qui se chevauchent souvent.

ment atteindre le tiers des productions. Les locuteurs utilisent en effet tout simplement les constructions orales qui vont leur permettre de tirer profit au mieux du caractère répétitif des opérations.

Le dialogue homme-machine, quant à lui, entraîne de la même manière des modifications de notre comportement langagier. La communication avec une machine ne constitue pas, en effet, un événement anodin. Nos productions orales sont fortement influencées par cette dernière. En particulier, les locuteurs semblent utiliser des énoncés relativement figés sensés correspondre aux attentes du système et s'efforcent d'être le plus informatif possible. La longueur de l'incise de l'exemple (E.1.10) que nous avons donné page 11, issu de la transcription d'un dialogue homme-homme simulé dans une situation familière et très interactive, n'est ainsi pas réaliste en situation de dialogue finalisé avec une machine, du moins dans l'état actuel des recherches dans ce domaine, où il semble que la "simplification" du discours soit due à un comportement mimétique du locuteur [Morel, 1988].

Il est ainsi impératif, dans le cadre qui nous concerne, d'étudier la fréquence de l'apparition des phénomènes et de l'utilisation des procédés en fonction du domaine d'application et de la situation orale interactionnelle envisagée avec la machine. Nous étudierons ces aspects en détail au chapitre 3. Rappelons toutefois que si cette modification du comportement langagier est une réalité, il serait illusoire de ne chercher à ne traiter que des standardisations forcées de l'expression des locuteurs. D'une part, ceci ne serait envisageable que pour des situations excessivement restreintes [Morel, 1988]. D'autre part, même dans les dialogues orientés par la tâche, la restriction de l'ensemble des phénomènes langagiers à prendre en compte ne semble pas évidente [Sabah et al., 1997], les énoncés oraux conservant une grande richesse structurelle. C'est pourquoi, en dépit des différences entre les situations de dialogue homme-machine et de dialogue homme-homme, l'étude de ces dernières est de notre point de vue également essentielle à la poursuite des recherches dans ce domaine.

### 1.2.2 Reconnaître - Comprendre - Dialoguer

Après avoir analysé rapidement les principales difficultés de nature linguistique auxquelles doivent faire face les systèmes de traitement automatique de la parole, nous présentons dans ce paragraphe l'architecture générale, souvent modulaire, des systèmes conversationnels. L'architecture globale des systèmes de dialogue homme-machine dans le domaine de la recherche d'informations est résumée figure 1.1.

Ces systèmes comprennent ainsi :

- un module de **reconnaissance de la parole** dont le but est de transcrire le signal vocal donné en entrée en un message orthographique. Il est composé de deux composants principaux, le modèle acoustique et le modèle de langage, que nous détaillons ci-après,
- un module de **compréhension de la parole** dont le but est d'extraire les informations pragmatiques (domaine d'application) rendant compte des actes de langage exprimés par l'utilisateur (requête sur le domaine de la tâche, réponses au système, demande de reformulation, etc. cf. chapitre 4). On peut distinguer

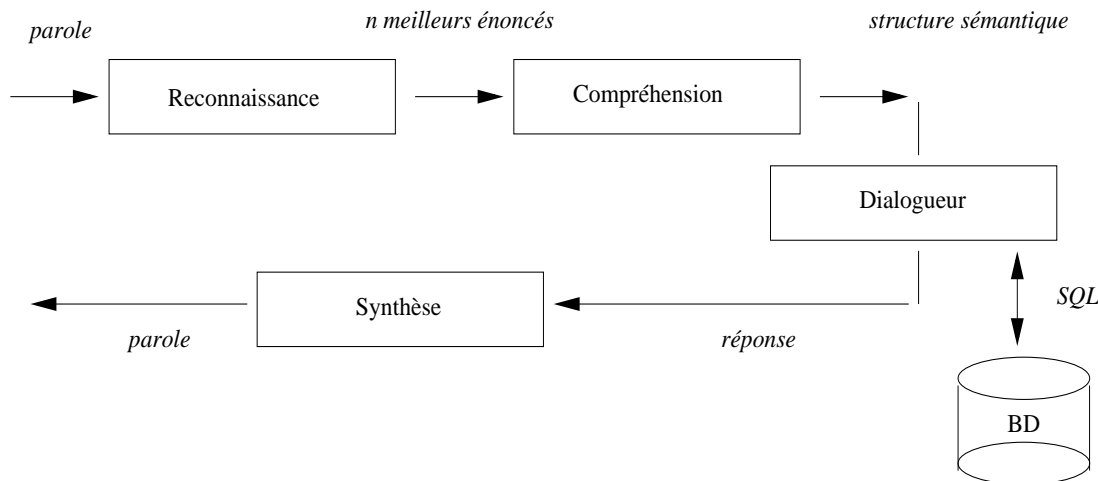


FIG. 1.1 – Architecture générale d'un système de dialogue oral homme-machine

deux grandes étapes dans les processus visant à comprendre un énoncé : la compréhension *hors contexte* et la compréhension *contextuelle*. La compréhension hors contexte, sujet principal de notre thèse, consiste à comprendre l'énoncé sans tenir compte, dans un premier temps, de ce qui a été dit précédemment dans le dialogue. Ce module doit composer avec les phénomènes spécifiques de l'oral déjà évoqués mais aussi avec les erreurs de reconnaissance éventuelles,

- un module de **gestion du dialogue** qui supporte le contexte dialogique et l'historique de l'interaction. Il initie les actions à effectuer suite aux requêtes des utilisateurs, que le sens de ces requêtes ait été correctement identifié ou non. Le gestionnaire du dialogue est non seulement un lien entre la base de données et l'utilisateur mais aussi le coordinateur de ce qui se passe, le reflet des capacités du système pour l'utilisateur (types d'interaction auxquels il peut s'attendre ; demandes qu'il peut formuler). Il doit gérer l'échange "sur l'instant et dans la durée" [Rosset, 2000]. Son développement dépend de la tâche (représentation et structuration de l'univers de référence), des connaissances disponibles (connaissances linguistiques, modèle de l'utilisateur, etc.) mais aussi des choix faits au préalable sur les caractéristiques du dialogue (directif, coopératif, etc.).
- un module de **génération de la réponse** qui convertit l'information à délivrer à l'utilisateur sous forme sonore (synthèse vocale), textuelle ou graphique. Nous ne nous intéresserons pas dans le cadre de ce travail à la synthèse vocale. Le lecteur pourra toutefois utilement se reporter à [Alessandro et Tzoukermann, 2001] pour un point complet sur les avancées réalisées dans ce domaine en liaison avec le traitement automatique des langues.

L'élaboration de l'interaction entre le système et l'utilisateur nécessite des mécanismes plus complexes que ne le laisse supposer cette juxtaposition de modules. Le



gestionnaire de dialogue occupe une place centrale dans le système. À partir généralement de la compréhension effectuée sur des énoncés transcrits automatiquement, il doit gérer l'interprétation contextuelle (dans l'historique du dialogue) et l'avancement de l'interaction. Il sera toutefois principalement question dans la suite de ce chapitre des stratégies d'analyse et des techniques utilisées pour la mise en œuvre des deux premiers modules : la reconnaissance et le module de compréhension. Concernant les différents aspects de la modélisation et de la gestion du dialogue on pourra se reporter par exemple à [Bilange, 1992] et [Sabah et al., 1997], à [Rosset, 2000] pour un état de l'art des stratégies de dialogue envisagées dans les systèmes d'interrogation de bases de données et à [Webber, 1999] pour les aspects computationnels liés au traitement automatique du langage.

### 1.3 Reconnaissance, compréhension : quelles démarches ?

Nous introduisons successivement dans cette partie les différentes techniques utilisées en reconnaissance et en compréhension de la parole. On notera dès à présent que ces deux processus ont parfois été réalisés conjointement [Pieraccini et al., 1992], cette intégration étant préconisée plus récemment par [Furui, 2000]. Cette décomposition en deux étapes est toutefois représentative de la plupart des systèmes existants. Elle nous permettra de mieux caractériser les principaux paradigmes sur lesquels reposent ces deux niveaux de traitement qui concernent des problèmes différents. Elle permettra en outre d'identifier les rapprochements technologiques qui ont pu s'opérer, principalement pour éviter un fonctionnement aveugle vis-à-vis des erreurs de reconnaissance.

#### 1.3.1 Reconnaître la parole

Comme nous l'avons déjà signalé, l'extrême variabilité du signal de parole (voir par exemple [Olive, Greenwood, et Coleman, 1993] pour une introduction détaillée à ces problèmes) ont conduit dès le début des années 1980 à l'adoption d'approches stochastiques globales pour les méthodes de décodage acoustico-phonétique au détriment des approches analytiques [Méloni, 1982; Mercier, Bigorgne, Miclet, Le Guennec, et Querré, 1990], qui se sont révélées difficiles à mettre en œuvre et instables. La reconnaissance *globale* consiste à segmenter le signal en entités (mots, phonèmes, diphtongues, etc.) qui sont stockées dans un dictionnaire de références acoustiques. Il s'agit donc d'un traitement global sur le signal qui cherche à identifier la séquence de mots la plus probable d'avoir donné lieu au signal acoustique observé, à partir de références acoustiques préalablement apprises. La modélisation acoustique (c'est-à-dire la correspondance entre les unités phonétiques – phonèmes éventuellement en contexte – et leur réalisation acoustique) y est réalisée sous forme probabiliste et repose désormais, qu'il s'agisse de la reconnaissance de mots-clefs ou de la reconnaissance de parole continue, sur l'emploi généralisé des *modèles de Markov cachés* (HMM pour *Hidden Markov Models*) [Rabiner, 1989]. Un HMM se définit formellement par un quintuplet  $(\Omega, \Gamma, A, B, \Pi)$  où :

- $\Omega$  est un ensemble fini de symboles d'observation  $o_m$ ,

- $\Gamma$  est l'ensemble d'états  $s_i$  du modèle,
- $A = \{a_{ij}\}$  sont les probabilités de transition entre les états  $s_i$  et  $s_j$ ,
- $B = \{b_{mj}\}$  sont les probabilités d'émission de l'observation  $o_m$  dans l'état  $s_j$ ,
- $\Pi$  est la distribution des probabilités d'états initiaux.

Progresser dans le modèle signifie changer d'état toutes les unités de temps  $t$ . Dès que l'on entre dans un état  $s_j^t$ , une observation  $o_m^t$  est générée avec une probabilité  $b_{mj}$ . La transition entre l'état  $s_i$  et l'état  $s_j$  s'effectue selon la probabilité  $a_{ij}$ . L'émission de tous les symboles de  $\Omega$  est considérée comme possible dans tous les états. Le problème de la reconnaissance est d'aligner temporellement le signal d'entrée échantillonné<sup>21</sup> sur ce modèle de référence stochastique. Le principe de décodage consiste, à partir d'une suite d'observations  $O$  en entrée, à déterminer dans le modèle de Markov de référence, la séquence d'états  $S$  optimale permettant de la générer<sup>22</sup>. Un des formalismes permettant de résoudre ce problème de manière efficace est par exemple *l'algorithme de Viterbi* [Viterbi, 1967; Rabiner et Juang, 1986]. Notons que l'application des HMM à la reconnaissance de la parole met généralement en jeu des modèles à deux niveaux, c'est-à-dire des modèles dont les observations de chaque état sont modélisées également par un modèle de Markov<sup>23</sup>.

Une des questions principales pour l'application des modèles de Markov cachés réside dans le développement du modèle (choix du nombre d'états, unités qu'ils représentent, utilisation d'un modèle discret ou continu, etc.) et surtout dans son apprentissage. Cette phase cherche à optimiser les paramètres du modèle sur un ensemble de séquences d'entraînement. L'optimisation de ces paramètres doit permettre de décrire au mieux la manière dont chaque séquence observée est générée. Le corpus d'apprentissage doit être aussi représentatif du problème que possible. On notera que les modèles acoustiques sont en général appris sur des corpus de textes lus c'est-à-dire sans présence des phénomènes dus à la spontanéité des productions orales. Quelques travaux se sont pourtant attachés à introduire des modèles acoustiques tenant compte explicitement des pauses non silencieuses (remplies par des appuis du discours ou des interjections) [Butzberger, Murveit, et Weintraub, 1992]. Les travaux actuels se concentrent sur le problème du traitement des données acoustiques en milieu bruité [Veth, Cranen, et Boves, 2001].

Les limites du décodage acoustico-phonétique sont nombreuses. D'une part, la modélisation acoustique ne peut rendre compte de la variabilité de la production de la parole. La collecte d'échantillons pour les corpus d'apprentissage est en effet très difficile (nécessité par exemple de disposer de plusieurs occurrences de chaque mot dans le corpus d'entraînement), les effets perturbateurs du contexte sur les unités phonémiques sont en conséquence uniquement approchés. À ce problème s'ajoute entre autres celui

---

<sup>21</sup>Le signal est en général représenté sous formes de vecteurs acoustiques comprenant la fréquence fondamentale, l'amplitude et les coefficients cepstraux ou LPC (Linear Predictive Coding) [Haton, Pierrel, Perennou, et Gauvain, 1991; Rabiner et Juang, 1993].

<sup>22</sup>Seule la séquence d'observation  $O$  est connue et la (meilleure) séquence d'états  $S$  sous-jacente qui a permis de la générer est dite *cachée*. D'où le nom de *Hidden Markov models*.

<sup>23</sup>On trouvera une description plus complète des modèles de Markov cachés dans [Rabiner, 1989] ou [De Mori, 1998].

de l'appariement graphèmes-phonèmes (homonymes hétérophones). D'autre part, il est impossible de considérer toutes les combinaisons de mots possibles et il convient de maîtriser la combinatoire de recherche.

Ainsi le décodage ne peut se mener de façon fiable en ne considérant uniquement que des données acoustiques. On utilise ainsi, pour guider le décodage, des connaissances linguistiques. Ces connaissances se limitent généralement à l'intégration de certaines prévisions sur l'ordre de la suite de mots de l'énoncé<sup>24</sup>. Cela suffit à resserrer directement la recherche (sélection de mots candidats) et à rendre plus pertinents les scores sur lesquels repose le décodage.

Une des solutions consiste à décomposer la probabilité d'une suite de mots étant donnée une observation acoustique en deux contributions indépendantes : une contribution acoustique et une contribution linguistique. Ainsi si l'on cherche la suite de mots  $W$  ayant la plus grande vraisemblance étant donnée l'observation acoustique  $O$ , le problème consiste à trouver l'estimation  $\hat{W}$  qui maximise la probabilité *a posteriori*  $P(W|O)$  :

$$\hat{W} = \arg \max_W P(W|O) \quad (1.1)$$

Suivant la loi de Bayes, il vient :

$$P(W|O) = \frac{P(W)P(O|W)}{P(O)} \quad (1.2)$$

Ce qui revient à maximiser (la distribution  $P(O)$  étant jugée équiprobable) :

$$\hat{W} = \arg \max_W P(W)P(O|W) \quad (1.3)$$

Le second terme correspond au *modèle acoustique* que nous avons déjà évoqué. La contribution linguistique, la probabilité  $P(W)$  d'une suite de mots, est estimée par un *modèle de langage* [Roukos, 1996]. Une estimation fiable et complète de  $P(W)$  est impossible. On fait ainsi reposer les modèles de langage utilisés pour guider la reconnaissance sur une estimation statistique, à partir de grands corpus de textes représentatifs des énoncés à reconnaître, des probabilités d'occurrence d'un mot donné *connaissant ses  $N$  prédécesseurs*<sup>25</sup>.

---

<sup>24</sup>Nous verrons toutefois au chapitre suivant que les informations linguistiques prises en compte peuvent être plus riches.

<sup>25</sup>Il faut noter qu'une autre solution consiste à contraindre le décodage acoustico-phonétique par une grammaire (séquence de mots autorisés par la grammaire). Ces techniques présentent des inconvénients majeurs : pas d'acquisition automatique de la grammaire, problème de couverture du langage reconnu et fortes contraintes imposées aux locuteurs par la grammaire. Des traitements, comme l'utilisation de séquences arbitraires dans la grammaire (expérimenté dans le système commercialisé NUANCE par exemple) permettent de limiter ces problèmes mais les avantages d'un modèle de langage statistique, dérivé à partir de corpus, sur le travail colossal de description nécessaire dans le premier cas, ont conduit à un abandon progressif des méthodes par grammaires. Notons cependant que des grammaires hors-contexte représentées sous forme d'automates à états finis ont montré leur efficacité sur des petits vocabulaires. De plus, ces grammaires à états finis peuvent être également stochastiques, les probabilités associées aux règles étant apprises sur corpus [Jelinek et Lafferty, 1991]. Elles demeurent toutefois désormais davantage utilisées en conjonction avec des modèles *N-gram* que seules [Stolcke et Segal, 1994].

Ces modèles, appelés modèles *N-gram*, réduisent ainsi le contexte des mots précédents pris en compte à une longueur  $N$  donnée (la taille de la fenêtre de l'historique utilisée est en générale égale à 2 ou 3 mots, plus rarement 4). Dans le cas des modèles *trigram* ( $N = 3$ ), la probabilité  $P(W)$  est ainsi :

$$P(W) = \prod_{i=2}^N P(w_i | w_{i-2}, w_{i-1}) \quad (1.4)$$

Cette modélisation linguistique minimale pour maîtriser la combinatoire des hypothèses du décodage acoustique peut sembler limitée. Toutefois, les performances obtenues par les systèmes markoviens guidés par un tel modèle de langage attestent de l'efficacité de cette hypothèse réductrice. L'utilisation conjointe de connaissances linguistiques de plus haut niveau dans un modèle stochastique risque de se traduire par des difficultés supplémentaires de résolution de conflits entre ces connaissances, qui peuvent être incertaines ou antagonistes notamment lorsque le vocabulaire à reconnaître est étendu. Il s'avère que l'utilisation des seules probabilités des fréquences d'occurrence permet d'obtenir une prédiction moins sensible aux éventuelles erreurs en un point donné de l'énoncé, en particulier lorsque les caractéristiques lexicales, syntaxiques et sémantiques du discours diffèrent entre les corpus d'entraînement et les corpus de tests. Il a par ailleurs été montré qu'une stratégie d'analyse linguistique fondée sur le graphe des hypothèses acoustiques est moins performante que la même stratégie fondée sur un graphe de mots représentant les hypothèses du système de reconnaissance, filtrées suivant les probabilités d'un modèle de langage [Noord, Bouma, Koeling, et Nederhof, 1999].

Il n'en reste pas moins que la modélisation markovienne du langage reste limitée à des dépendances entre mots et des rétroactions du contexte très simples. Une des conséquences de cette limitation concerne par exemple les erreurs fréquentes d'accords en genre et en nombre attestées. Il faut remarquer que ces erreurs d'accords relèvent de régularités modélisables par un nombre limité de règles [Smaïli, Zitouni, Charpillat, et Haton, 1997].

Par ailleurs, la couverture et la qualité des corpus d'apprentissage est essentielle. Or la nature des corpus eux-mêmes constitue un premier obstacle. Les différences entre les corpus écrits et les corpus de parole que nous avons évoquées rendent les premiers peu adaptés en théorie pour la détermination du modèle de langage. En contrepartie, plusieurs réserves ont été émises quant à la constitution d'échantillons *représentatifs* de transcriptions de corpus oraux [Habert, 2000; Habert, Nazarenko, et Salem, 1997], essentiellement en raison des différents domaines d'application et contextes d'interaction utilisés pour leur recueil. Ainsi, les travaux actuels sur les modèles de langage stochastiques portent davantage sur la recherche d'un compromis optimal entre la couverture du vocabulaire (précision de l'estimation des paramètres du modèle) et la capacité de généralisation (prédiction dans d'autres contextes), tout en conservant un taux de reconnaissance acceptable. En d'autres termes, il s'agit de produire une "bonne" estimation des probabilités d'apparition d'un mot conditionnellement à un contexte donné, malgré les faiblesses potentielles du corpus d'apprentissage, les limitations du paradigme *N-*

*gram* et les différences éventuelles entre les situations de test et d'apprentissage.

Plusieurs solutions ont été envisagées pour faire face à ces deux principaux problèmes :

1. **Augmentation de la couverture.** Chercher à augmenter la couverture lexicale, et celle de la modélisation *N-gram*, a pour but principal de remédier au problème de la faible représentativité d'une séquence de *N* mots dans le corpus et son corollaire : le traitement des mots inconnus ou *hors-vocabulaire*. Ces derniers induisent des erreurs de substitution qui se répercutent sur les mots suivants, le modèle de langage n'opérant plus sur le contexte adéquat. Pour éviter ces erreurs, il est impératif d'optimiser la taille du vocabulaire pour diminuer ce nombre de mots *hors-vocabulaire* mais ce faisant, la confusion acoustique potentielle entre deux mots se voit augmentée. Il a été montré d'autre part que quelle que soit la taille des corpus d'apprentissage le système devra immanquablement faire face à des mots dits *hors-vocabulaire* [Hetherington et Zue, 1991]. Plusieurs approches ont été abordées pour améliorer le traitement des mots inconnus. Pour un aperçu des techniques employées en reconnaissance<sup>26</sup>, on pourra se reporter à [Bousquet-Vernhettes, 2002].
2. **Modèles *N-classes*.** Ces travaux visent à regrouper en classes les mots semblant se comporter de manière similaire dans les corpus de façon à affiner la prédiction des modèles de langage pour les mots rares. Ces approches ont des performances en général moins bonnes que les modèles *N-gram* mais permettent de réduire la taille des modèles. On notera toutefois qu'une variante intéressante a été explorée par [Jardino et Adda, 1993; Jardino et Beaujard, 2000]. Dans cette approche les classes de mots sont apprises sur corpus sans prendre en compte de connaissances syntaxiques ou sémantiques; les rapprochements en classes y sont fondés principalement sur des contraintes grammaticales ou morphologiques et sont effectués par un algorithme "aveugle" de minimisation de perplexité. D'autres méthodes, fondées sur ce même principe de recouplement d'informations, consistent à utiliser en concurrence plusieurs modèles *N-gram* appris sur différents ensembles [Iyer et Ostendorf, 1999].
3. **Modification de l'étendue de la couverture *N-gram*.** Il s'agit de faire "varier" la taille de la fenêtre de l'historique pris en compte [Niesler et Woodland, 1996], c'est-à-dire de pouvoir utiliser d'autres informations que l'ordre linéaire pour faire reposer la prédiction des mots sur des dépendances longue-distance. Ces techniques reposent essentiellement sur des considérations syntaxiques ou sémantiques, utilisées conjointement avec le paradigme *N-gram*. Deux types d'approches peuvent être distinguées. Les premières visent l'intégration de la structure syntaxique dans le modèle probabiliste de prédiction [Jelinek et Chelba, 1999; Zhang, Black, et Finch, 1999] ou [Roark, 2002] pour les travaux les plus récents dans ce domaine. L'analyse syntaxique intégrée est utilisée pour modifier le modèle stan-

---

<sup>26</sup>[Bousquet-Vernhettes, 2002] traite également de ce problème en ce qui concerne la compréhension.

dard  $N$ -gram. Il s'agit ainsi de pouvoir tenir compte à la fois d'une prédiction fondée sur les  $(N - 1)$  étiquettes grammaticales et les  $(N - 1)$  têtes lexicales des sous-arbres caractérisés par l'analyse syntaxique incrémentale réalisée, plutôt que de reposer uniquement sur les  $(N - 1)$  mots<sup>27</sup>. Les secondes cherchent simplement à utiliser un contexte de prédiction plus large sans s'intéresser à la structure de l'énoncé. Elles tirent principalement parti des dépendances longue-distance entre paires de mots attestant de corrélations sémantiques significatives dans le domaine considéré, afin de modifier les probabilités  $N$ -gram [Bellegarda, 2001]. Les classes sémantiques utilisées peuvent être très générales. Dans l'approche proposée par [Gillet et Ward, 1998] par exemple, elles concernent les dates et les heures modélisées par une grammaire hors-contexte stochastique.

4. **Modèles de langage adaptatifs.** Ces approches cherchent à limiter la diminution des performances due aux différences entre les corpus d'apprentissage et de test. Elles requièrent une modification dynamique des probabilités du modèle de langage en fonction du contexte d'application. Parmi ces approches citons les modèles à mémoire *cache* [Kuhn et De Mori, 1990].

Enfin d'autres approches concernent essentiellement le développement d'algorithmes d'estimation robuste des paramètres des modèles (techniques de lissage, maximum d'entropie par exemple) [Bellegarda, 2001].

Deux directions générales émergent dans les tentatives d'améliorer les modèles de langage. Les premières se focalisent sur les données et la façon d'améliorer leur exploitation dans les modèles stochastiques utilisés. Les secondes cherchent à utiliser des informations supplémentaires : structures syntaxiques, schémas sémantiques ou spécificités du domaine pour améliorer la conception des modèles de langage. Les recherches dans ce domaine sont encore en plein développement. On notera cependant que l'ensemble de ces tentatives n'ont conduit ces dernières années qu'à de faibles améliorations par rapport aux modèles classiques trigram [Brill, Florian, Henderson, et Mangu, 1998; Bellegarda, 2001].

En résumé, l'adéquation de la modélisation statistique du langage à la problématique de la reconnaissance de la parole n'est plus à démontrer. L'intérêt de ces approches n'est cependant manifeste que si l'on dispose à la fois de corpus étendus susceptibles de garantir la généralité de l'apprentissage et de systèmes informatiques puissants capables d'explorer les larges espaces de recherche mis en jeu par l'analyse probabiliste. Depuis quelques années, les systèmes stochastiques de reconnaissance multi-locuteurs présentent des performances remarquables<sup>28</sup> attestées par plusieurs campagnes d'éva-

---

<sup>27</sup>La qualité de ces approches repose principalement sur la robustesse des analyseurs syntaxiques utilisés et en particulier sur l'hypothèse implicite que la probabilité maximale soit affectée à la l'analyse syntaxique correcte [Bellegarda, 2001]. Le modèle probabiliste nécessite la mise en œuvre d'importantes techniques d'optimisation pour faire face à une combinatoire de recherche très élevée [Chelba et Jelinek, 1998; Chelba, 2000].

<sup>28</sup>On notera toutefois que les taux d'erreurs sont très nettement supérieurs (près de 40%) en parole spontanée sur des domaines d'application non finalisée.

luation (tableau 1.1 d'après [Pallett et al., 1994]).

Campagne	Élocution	Application	taux d'erreur (mots)
ATIS	Parole spontanée	finalisée (2000 mots)	3%
Buisness News	Textes lus	non finalisée (>20 000 mots)	7,2%

TAB. 1.1 – Campagnes d'évaluation ARPA : performances obtenues par le meilleur système de reconnaissance.

Précisons pour terminer qu'un système de reconnaissance ne se contente pas de délivrer en sortie uniquement *une* séquence de mots reconnus, mais les  $N$  *meilleures séquences* filtrées selon le modèle de langage. Ces séquences peuvent éventuellement être représentées sous forme d'un treillis ou un graphe de mots rendant compte des transitions lexicales possibles et de leurs scores associés.

Les travaux actuels en reconnaissance portent sur les systèmes multi-locuteurs, nécessitant de larges corpus d'apprentissage et dont les performances se dégradent dès lors que les conditions acoustiques des tests changent. Un intérêt grandissant s'est développé autour des systèmes adaptatifs [Tapias Merino, 2001], tentant de combiner les avantages des systèmes mono et multi-locuteurs. Ces systèmes utilisent des modèles de reconnaissance généraux lors d'une première utilisation du système par un individu. Lors des utilisations suivantes, une détection de certaines caractéristiques du locuteur ou de la situation d'interaction permet le rappel du modèle correspondant à ce locuteur ou le choix de modèles spécifiques, les mieux adaptés<sup>29</sup> à la situation présente. Ces systèmes sont étudiés pour pallier des problèmes importants dus à la variabilité entre locuteurs (style, volume, vitesse d'élocution<sup>30</sup>, locuteurs étrangers, accents) ou les problèmes entraînés par les canaux de transmission du signal (microphones de sensibilités différentes, interaction téléphonique, etc.) [Tapias Merino, 2001]. D'autres travaux portent sur la conception de modèles acoustiques robustes en milieu bruité [Veth et al., 2001]. En ce qui concerne les avancées des recherches sur la reconnaissance grand vocabulaire, on pourra se reporter au numéro spécial de *Computer Speech and Language* [Gauvain, De Mori, et Lamel, 2002] traitant de ces aspects. Un état de l'art complet pourra en outre être trouvé dans [Junqua et Haton, 1996].

### 1.3.2 Comprendre la parole

Les modules de compréhension cherchent à attribuer une représentation sémantique aux énoncés prononcés. En dialogue homme-machine finalisé, il ne s'agit pas seule-

<sup>29</sup>La plupart de ces systèmes utilisent des techniques de compensation au niveau du modèle acoustique en fonction des caractéristiques détectées.

<sup>30</sup>Ce problème est particulièrement important. Martínez rapporte par exemple que le taux d'erreur de la reconnaissance est doublé voire triplé en fonction de la seule variation de la vitesse d'élocution [Martínez, Tapias, Álvarez, et León, 1997].

ment de comprendre la requête du locuteur (la nature des informations qu'il cherche à obtenir, la tâche qu'il cherche à effectuer) mais également de pouvoir utiliser les représentations sémantiques construites pour *entreprendre une action pertinente* dans le contexte dialogique. La difficulté réside dans l'évaluation de cette pertinence. Les enjeux de la communication imposent ainsi l'intégration d'une représentation du dialogue, soit explicite, soit implicite, permettant de gérer l'avancement du dialogue et de spécifier la réponse du système. Ils imposent par ailleurs la prise en compte de l'historique de l'interaction. Ce dernier impératif nécessite l'exploitation des représentations sémantiques construites pour chaque énoncé lors de l'étape de compréhension. Cette étape d'analyse des énoncés oraux cherche à identifier et à représenter aussi précisément que possible l'*acte de langage* véhiculé par l'énoncé ainsi que les informations qui lui sont associées. Nous définirons plus précisément cette notion au chapitre 4 par rapport au domaine d'application que nous avons choisi. D'une manière très générale, dans le cas d'une application de recherche d'information, on peut distinguer les actes de demande d'information (requêtes sollicitées par l'utilisateur concernant le domaine de la tâche) et les actes de langage portant sur le dialogue en cours (ouverture et clôture du dialogue, accord ou désaccord, situation d'incompréhension, demande de clarification, etc.).

Dans cette partie, il sera essentiellement question des stratégies et des techniques utilisées pour réaliser cette étape de compréhension hors-contexte<sup>31</sup>. Comme nous l'avons déjà souligné, l'analyse des énoncés oraux doit faire face à plusieurs problèmes : la nature des productions orales spontanées et les erreurs de reconnaissances éventuelles. Concernant ce second point, afin de clarifier l'exposé, nous ferons l'hypothèse que les stratégies présentées traitent en entrée une seule séquence de mots reconnus, la meilleure parmi les différentes hypothèses de reconnaissance. Nous nous contenterons d'évoquer, quand cela est le cas, les techniques utilisées pour tenir compte de l'ensemble des hypothèses délivrées par le module de reconnaissance. Pour des compléments sur le processus de compréhension de la parole, les stratégies et les différents formalismes que nous allons évoquer, on pourra se reporter à [Allen, 1998].

### 1.3.2.1 Stratégies de compréhension : fondements

L'analyse du langage est envisagée traditionnellement selon des approches reposant sur la syntaxe c'est-à-dire fondées sur une analyse syntaxique *complète*, cherchant à rendre compte des relations entre tous les mots de la phrase. En analyse automatique de la parole, il est apparu que de telles approches (exploitées notamment par [Seneff, 1989] pour le système TINA par exemple) pouvaient faire décroître rapidement les performances en présence des mots inconnus, des constructions spécifiques à l'oral, des erreurs de reconnaissance et surtout des phénomènes dus à la spontanéité de l'interaction [De Mori, 1994]. Aussi, l'idée de réaliser une analyse syntaxique complète de l'énoncé a été progressivement abandonnée. La syntaxe n'a pas pour autant totale-

---

<sup>31</sup>Nous soulignerons toutefois les apports (ou les faiblesses) de chacune des approches que nous présenterons, vis-à-vis de la compréhension en contexte.



ment disparu de certains systèmes de compréhension de la parole<sup>32</sup>, où les écarts par rapport à l'écrit sont supportés à l'aide de mécanismes de corrections additionnels ou de stratégies d'analyse partielle. Les besoins de robustesse face aux perturbations de l'oral ont pris ainsi le pas sur les limitations éventuelles de méthodes qui, ignorant les contraintes posées par la syntaxe, pourraient ne pas être capables de traiter correctement des constructions linguistiques complexes, voire présenter des faiblesses sur des constructions simples mais nécessitant une connaissance de la structure de l'énoncé.

Ainsi, il est possible aujourd'hui d'identifier un dénominateur commun entre la plupart des stratégies mises en œuvre pour la compréhension de la parole. Ces approches cherchent en effet tout d'abord à *isoler* ou à *repérer* dans l'énoncé les unités (ou les parties de l'énoncé) jugées porteuses d'information pertinente. Sur la base d'informations rendant compte du *domaine d'application* considéré, ces unités sont ensuite combinées ou reliées entre elles pour obtenir une représentation cohérente. Les différences entre les systèmes proviennent de la nature de ces unités qui peuvent être sémantiques ou syntaxiques, ainsi que de la nature des connaissances mises en jeu pour leur identification et leur traitement. La diversité des approches que nous allons évoquer réside ensuite dans les modèles et formalismes employés et, dans une moindre mesure, dans la forme des représentations sémantiques générées<sup>33</sup>. Dans tous les cas l'analyse est ou doit pouvoir être partielle, dans le sens où elle ne cherche pas à tenir compte de tous les mots de l'énoncé, et s'appuie, à un niveau ou à un autre, sur des informations liées au domaine de l'application.

Nous répertorions dans les paragraphes suivants les principales approches ayant été proposées en compréhension. Nous évoquons tout d'abord les approches guidées principalement par la sémantique. Cette voie qualifiée de *sémantique globale* par De Mori [De Mori, 1994] ou de *sélective* par Luzzati [Luzzati, 1989], terme que nous reprendrons dans la suite, est celle qui profite le plus du caractère finalisé du dialogue. La stratégie de compréhension est entièrement orientée par la tâche. Les partisans de ces approches affirment en effet que la connaissance complète de la structure d'une phrase n'est pas nécessaire dans un dialogue finalisé restreint à des tâches très ciblées. Ainsi pour des tâches de renseignement horaire dans le domaine ferroviaire, Luzzati résume :

(Cit.1.3) “*Tout repose en somme sur un postulat : un quart du lexique suffit pour accéder au sens*” [Luzzati, 1989], p. 271

Nous évoquons ensuite les approches qui se caractérisent par l'utilisation de méthodes d'analyse syntaxique robustes pour la définition ou le repérage des unités de sens. Nous les qualifierons de méthodes par *exploration d'îlots syntaxiques*. Dans cette même catégorie, nous regroupons les méthodes d'analyse “flexibles” dans lesquelles une analyse

<sup>32</sup>Nous reviendrons en particulier sur les développements intéressants du système TINA que nous venons d'évoquer, utilisés par de nombreux systèmes de dialogue homme-machine.

<sup>33</sup>On pourra trouver dans [Kuhn et De Mori, 1997] un aperçu des représentations du sens envisagées : cadres sémantiques, concepts, structures de cas, expressions logiques associées aux nœud d'un arbre syntaxico-sémantique, etc.

syntaxique complète est tout d'abord envisagée. En cas d'échec de cette analyse, des techniques par identification d'îlots, syntaxiques ou sémantiques, sont alors utilisées.

### 1.3.2.2 Méthodes sélectives

Dans ces approches, la compréhension est limitée à la recherche du sens dit *utile* de l'énoncé. Il s'agit donc de méthodes fondées sur l'identification de séquences-clés, appelées le plus souvent *segments conceptuels*, à partir desquels une structure sémantique prédéfinie va pouvoir être instanciée<sup>34</sup>. Un premier exemple illustrant ce type d'approche peut être donné par les systèmes ([Bennacef, Bonneau-Maynard, Gauvain, Lamel, et Minker, 1994; Minker, 1996] par exemple) utilisant un modèle fondé sur la théorie des *grammaires de cas* [Fillmore, 1968]. Cette théorie a été étendue par la définition du formalisme de *frame grammars* [Bruce, 1975] pour être applicable à tout système faisant usage de concepts. Ce formalisme permet de construire une représentation sémantique sous forme de *schéma*. Le concept du schéma est identifié par plusieurs mots de référence dans la phrase. Les attributs du schéma sont instanciés à l'aide de marqueurs qui représentent des contraintes syntaxiques locales. L'analyse sémantique permet d'identifier dans chaque requête le ou les concepts attendus (liés à l'application) et de les valider par un jeu de contraintes. Les concepts significatifs de l'application et les mots de référence doivent être prédéfinis<sup>35</sup>. Dans l'exemple ci-dessous adapté de [Minker, 1999] concernant le domaine ATIS,

(E.1.22) je voudrais **aller** euh *de* Philadelphie à San Fransisco en passant *par* Dallas

le concept <vol> est identifié grâce au mot 'aller', les attributs requis de ce concept (*ville-départ*, *ville-arrivée* et *escale*) sont instanciés grâce aux marqueurs en italique. Le schéma sémantique correspondant est donné par la structure :

<vol>	
	<i>ville-départ</i> : Philadelphie
	<i>ville-arrivée</i> : San Fransisco
	<i>escale</i> : Dallas

Le principal travail lors de l'élaboration de tels analyseurs réside dans la définition des concepts et des mots de référence et, dans le cas d'analyseurs par règles, dans l'écriture des règles cohérentes de la grammaire. Ceci est une tâche importante, difficile mais réalisable pour des domaines applicatifs restreints. Cette définition, ne reposant que sur le domaine applicatif est d'ailleurs plus ou moins indépendante de la langue. De plus, il est possible en portant l'analyseur dans une autre application (mais d'un domaine assez proche toutefois), de conserver de nombreux concepts et mots de références. Enfin, et

<sup>34</sup>Certains systèmes ont poussé cette démarche d'extraction du sens "utile" par rapport à l'application, à l'extrême en ne considérant que la détection de *mots-clés* et non pas de séquences [Minami et al., 1995; Yamada et al., 1995]. Nous ne traiterons pas de ces systèmes qui ne présentent un intérêt que pour des tâches extrêmement ciblées et n'autorisent qu'un dialogue (oral) très directif.

<sup>35</sup>La détermination des segments se fait de manière descendante lors de la modélisation. L'analyse peut être quant à elle ascendante ou descendante.

c'est son intérêt, cette méthode présente une certaine robustesse face aux énoncés oraux, ceux-ci n'étant pas analysés dans leur globalité.

Dans l'analyseur PHOENIX [Ward, 1990, 1991; Issar et Ward, 1993], la grammaire sémantique sélective est représentée par une grammaire hors-contexte associée à une structure de schéma. Les parties gauches de la grammaire définissent les concepts sémantiques. Les parties droites décrivent chacune des façons d'exprimer le concept au moyen d'items lexicaux (terminaux) et de concepts (non terminaux). Dans l'exemple suivant, l'étoile indique un terminal optionnel :

$$\begin{aligned} \langle \text{MEET} \rangle &\longrightarrow (*\text{if } *we \text{ meet}) | (\text{meeting})... \\ \langle \text{SUGGEST} \rangle &\longrightarrow (\langle \text{SUGGESTION} \rangle \langle \text{MEET} \rangle) \end{aligned}$$

La grammaire est compilée sous forme de Réseaux de Transitions Récursifs (RTN). Elle est écrite de manière à ce que les éléments des phrases puissent être isolés ou bien intégrés dans une requête entière. Chaque fois qu'un élément sémantique est reconnu, il est intégré aux schémas auquel il appartient. Ces derniers peuvent ainsi être "remplis" dans un ordre indéterminé et les différentes interprétations possibles (les différents schémas instanciés) sont analysés en parallèle. L'analyse permet en outre de "sauter" les mots inconnus ou les perturbations dans la production orale. La production de schémas incomplets est admise. Dans une première étape, l'analyseur recherche l'interprétation avec le maximum de mots d'appariement. S'il n'existe pas de solution unique, il recherche le schéma le moins fragmenté. Un mécanisme de score permet de résoudre les ambiguïtés résiduelles. La grammaire étant compilée dans un ensemble d'automates spécifiant toutes les interprétations sémantiques, chacun faisant référence à d'autres automates, la taille du système est réduite de manière significative. Remarquons enfin que cette méthode a été conçue pour traiter directement le graphe des *N-meilleures* hypothèses issues de la reconnaissance [Issar et Ward, 1993; Gillet et Ward, 1998].

### Approches par règles *versus* approches stochastiques

Les deux approches que nous venons d'évoquer utilisent des analyseurs à base de règles. La compréhension sélective sur des domaines restreints se prête également bien à une modélisation stochastique. On parle alors de *décodage conceptuel*. Ce décodage repose sur l'hypothèse qu'il existe une correspondance séquentielle entre les séquences de mots et les séquences de concepts [Pieraccini, Levin, et Vidal, 1993]. Le processus de compréhension se ramène alors à une étape de maximisation de la probabilité *a posteriori* d'une séquence de segments conceptuels connaissant la séquence de mots [Levin et Pieraccini, 1995; Pérénou et al., 1998; Bousquet-Vernhettes, 2002]. On notera que ce décodage nécessite d'affecter un concept particulier (concept vide) aux mots supposés n'apporter aucune information pertinente dans l'énoncé.

Cette modélisation a l'avantage de pouvoir s'intégrer aisément avec les systèmes de reconnaissance pour lesquels les méthodes stochastiques ont été appliquées avec succès. Elle permet la prise en compte des scores des hypothèses de reconnaissance. Ainsi, dans la première version du système CHRONUS [Pieraccini et al., 1992] développé par AT&T sur le domaine ATIS, l'étape de compréhension (modélisation par modèles

de Markov cachés) est réalisée en même temps que la reconnaissance<sup>36</sup>. Dans le cas du système de Philips [Aust et al., 1995], qui utilise une grammaire hors-contexte stochastique, le graphe de mots issu de la reconnaissance est transformé en un graphe de concepts contenant toute l'information du graphe d'origine jugée significative par rapport à la grammaire sémantique.

Une modélisation par modèles de Markov cachés a été expérimentée par le LIMSI dans le cadre du projet MASK ainsi que sur les domaines ATIS et ARISE<sup>37</sup> puis comparée avec une méthode sélective par règles [Minker, 1999]. Les performances de l'analyse sémantiques sur le domaine ATIS sont identiques pour les deux méthodes.

Le principal avantage de ces approches réside ainsi plutôt dans le temps réduit nécessaire au développement de ces systèmes. Un premier corpus d'apprentissage est en général étiqueté manuellement à partir de la définition des concepts sémantiques liés au domaine<sup>38</sup>. Une procédure itérative de *bootstrap* permet ensuite l'étiquetage semi-automatique de corpus plus importants jusqu'à permettre une bonne estimation des paramètres stochastiques du modèle<sup>39</sup>. L'ensemble des étiquettes des concepts sémantiques peut être également progressivement enrichi de cette façon [Minker, 1999]. Des travaux plus récents [Maynard et Lefèvre, 2002] envisagent de la même manière la sélection automatique sur corpus des marqueurs de concepts.

Une vision globale de la réalisation des concepts sémantiques peut être ainsi obtenue facilement. Une bonne appréciation de cette connaissance nécessite au contraire un long travail d'expertise au moment de la conception d'un jeu de règles. En contrepartie, on retrouve les principaux problèmes de toute approche impliquant un apprentissage : celui de la représentativité du corpus entraînant une modélisation limitée des événements faiblement représentés, et celui du nécessaire compromis entre la couverture (précision de l'estimation des paramètres du modèle) et la capacité de généralisation. Le volume de données nécessaire pour atteindre une certaine robustesse face à une augmentation du domaine est de plus difficilement appréciable *a priori*.

### Formalismes utilisés

Notons enfin la diversité des formalismes utilisés pour mettre en œuvre ces approches sélectives, à titre d'exemples : arbres de classification (SCTs : *String Classification Trees*) acquis automatiquement à partir d'une description sémantique et pragmatique minimale

---

<sup>36</sup>En dépit de sa pauvreté linguistique, CHRONUS a présenté les meilleures performances en terme de compréhension lors de la dernière campagne d'évaluation ARPA sur le domaine ATIS [ARPA, 1995].

<sup>37</sup>Par rapport aux deux autres domaines, la représentation sémantique tient compte, outre le sens de l'énoncé par rapport à la tâche, des éléments pertinents du dialogue.

<sup>38</sup>Quand il existe déjà un analyseur par règles sur le même domaine applicatif ou sur un domaine proche, cette première étape peut être entièrement automatique.

<sup>39</sup>Dans le cas de l'utilisation de grammaires hors-contexte stochastiques [Aust et al., 1995], la grammaire est générée manuellement et les probabilités sont acquises automatiquement sur un grand volume de données.

du domaine [Kuhn et De Mori, 1995], automate prédéfini des états du dialogue<sup>40</sup> [Gorin et al., 1997; Rose et al., 2001], ou encore cascade de transducteurs<sup>41</sup> [Castellanos, Vidal, Varó, et Oncina, 1998]. On pourra se reporter à [Allen, 1998] pour des compléments sur le processus de compréhension de la parole, les stratégies et les différents formalismes que nous avons évoqués.

### 1.3.2.3 Exploration d'îlots syntaxiques - Analyse syntaxique robuste

Dans les approches sélectives, la syntaxe n'est pas ou relativement peu présente. Elle n'intervient principalement qu'à un niveau très local dans la prise en compte des marqueurs de concepts par exemple. D'autres méthodes sélectives ont toutefois fait le choix de faire reposer leur analyse sur des connaissances linguistiques plus poussées.

C'est le cas par exemple du système HUM [Miller et al., 1996]. Dans ce modèle, l'analyse est fondée sur un modèle probabiliste. On suppose que les séquences de mots  $W$  sont générées par une grammaire sémantique probabiliste. L'analyseur sémantique doit déterminer non plus la meilleure séquence de concepts mais les  $n$  meilleurs arbres sémantiques ayant généré la séquence de mots  $W$ . Les mots du corpus d'apprentissage sont ainsi annotés à la fois par une étiquette sémantique (unités de base du sens par rapport au domaine applicatif) et une étiquette syntaxique permettant d'établir les relations syntaxiques entre les unités sémantiques. Dans le domaine ATIS les mots *from* et *Atlanta* seront par exemple étiquetés respectivement par **departure/prep** et **city/npr**. Dans la représentation arborescente générée, le nœud immédiatement supérieur mettant en relation ces deux termes sera étiqueté **departure/pp**. Les mots n'ayant pas de sens particulier dans le domaine applicatif (déterminants par exemple) ne reçoivent qu'une étiquette syntaxique. Ce caractère syntaxico-sémantique de la représentation permet de concevoir un schéma d'étiquetage bien structuré et consistant. Le décodeur recherche les éléments à la fois syntaxiquement et sémantiquement corrects. On notera que cette modélisation détaillée se traduit par une augmentation de la complexité et de la taille du modèle. L'utilisation de restrictions syntaxiques se retrouve dans la stratégie ascendante du système sélectif TRAINS [Allen et al., 1996] et a été explorée dans le cadre de l'application du modèle DOP (Data Oriented Parsing) à la problématique de l'analyse des énoncés oraux en dialogue homme-machine [Bod, 1998].

### Stratégies mixtes

Une stratégie similaire d'intégration de contraintes syntaxiques est réalisée dans le système TINA du MIT. L'analyseur sémantique de ce système [Seneff, 1992b] utilise une grammaire hors-contexte, augmentée de structures de traits pour représenter des contraintes syntaxiques et sémantiques, qui génère des arbres d'analyses. L'énoncé est

<sup>40</sup>L'accent est mis dans [Rose, Yao, Riccardi, et Wright, 2001] sur l'intégration du traitement des mots hors-vocabulaire.

<sup>41</sup>Le processus de compréhension est appréhendé comme une simple étape de transformation entre la séquence de mots et un langage formel dépendant de l'application et tenant lieu de représentation sémantique. Les transducteurs sont acquis automatiquement sur des ensembles d'exemples en supposant la correspondance entre les deux niveaux.

analysé entièrement à partir de sa forme syntaxique [Seneff, 1989]. La grammaire est transformée automatiquement en un automate probabilisé qui permet d'avantager les constructions les plus fréquentes. Devant l'insuffisance de cette analyse complète sur l'oral spontané, cette approche a été modifiée en intégrant une stratégie d'analyse robuste par *chart* étendu (analyse tabulaire). Cette analyse permet le recouvrement des groupes analysés correctement dans l'énoncé, en cas d'échec de l'analyse complète [Seneff, 1992a]. Cette approche se fonde ainsi sur les combinaisons d'îlots correctement reconnues en relâchant certaines contraintes d'analyses. On peut voir ces techniques comme un compromis entre la qualité des résultats (utilisation de principes linguistiques) et la robustesse (relâchement de contraintes locales pour faire progresser l'analyse). Ce système de compréhension est utilisé dans tous les systèmes du MIT (voir par exemple [Glass et al., 1995]).

[Seneff, 1998] insiste sur la nécessaire coopération de la syntaxe et de la sémantique et sur l'encodage de ces deux niveaux dans les règles de la grammaire tout en conservant un espace de connaissance maintenable. On retrouve cette préoccupation dans le système EVAR [Gallwitz et al., 1998; Boros, Aretoulaki, Gallwitz, Nöth, et Niemann, 1997] qui utilise un analyseur ascendant par *chart* [Mecklenburg, Heisterkamp, et Hanrieder, 1995]. Si l'analyse reste principalement guidée par la sémantique, elle permet toutefois l'utilisation de contraintes syntaxiques. Les connaissances linguistiques sont représentées au moyen de Grammaires Catégorielles d'Unification (UCG) [Calder, Klein, et Zeevat, 1988].

Le système TINA est apparu comme le meilleur système de compréhension à base de connaissance lors de la campagne d'évaluation ATIS [ARPA, 1995] en présentant un taux d'erreur de compréhension pure<sup>42</sup> similaire à celui de CHRONUS. Il a par ailleurs inspiré de nombreux autres modules de compréhension, comme celui du système WAXHOLM [Carlson et Hunnicutt, 1996] ou DELPHI [Stallard et Bobrow, 1992]. Les développements récents et intéressants à partir du système TINA concernent l'acquisition automatique des mots hors-vocabulaire pendant l'interaction [Chung, 2000].

### Filtrage linguistique des hypothèses de reconnaissance

La contribution potentielle de l'analyse syntaxico-sémantique est double. Il s'agit à la fois d'initier le processus de compréhension en fournissant une analyse structurée de l'énoncé et d'achever le processus de reconnaissance en fournissant des indices pertinents pour retrouver l'hypothèse correcte. Les travaux que nous évoquons ici s'intéressent en effet, en plus de la phase de compréhension proprement dite, à l'analyse grammaticale du treillis de mots hypothèses. C'est d'ailleurs principalement cette dernière préoccupation qui a guidé le choix de l'utilisation de contraintes syntaxiques.

Ces méthodes reposent généralement sur la technique dite des îlots de confiance, souvent mise en œuvre par les analyseurs ascendants : il s'agit de chercher à analyser les segments non ambigus, quelles que soient leurs positions dans la phrase. Des

---

<sup>42</sup>C'est-à-dire sans la prise en compte de la reconnaissance de la parole.

algorithmes tabulaires particuliers ont été développés pour effectuer cette recherche préliminaire d'îlots. C'est le cas de l'algorithme *Head Corner* où durant une première phase ascendante, une catégorie est choisie pour chaque règle de réécriture pour guider la prédiction d'un îlot. Cette catégorie est celle qui correspond à la tête sémantique du syntagme<sup>43</sup>. Cette stratégie [Noord, 1997a] a été appliquée dans le système OVIS avec d'excellents résultats [Nederhof, Bouma, Koeling, et Noord, 1997] permettant de faire émerger la meilleure hypothèse de reconnaissance du treillis de mots hypothèses [Noord, 2001].

On retrouve une stratégie identique fondée sur un analyseur tabulaire ascendant dans le système GEMINI [Dowding et al., 1993]. Une analyse grammaticale fine des énoncés reconnus y est menée. Cette analyse est complète quand cela est possible ou en îlots si nécessaire. Les règles des grammaires sont accompagnées de mises à jour de formes logiques qui traduisent la sémantique de l'énoncé. Cette stratégie inclut des techniques de réparation des phénomènes oraux spontanés sur laquelle nous reviendrons au chapitre suivant. Notons que ces deux systèmes requièrent une description complète de l'univers syntaxique et sémantique des énoncés. Une solution complémentaire consiste à appliquer des mécanismes de réparation spécifiques [Rosé et Lavie, 1997, 2001] à la suite d'une analyse robuste par *chart* [Lavie, 1996].

### 1.3.3 Conclusions

Les procédures d'extraction du sens sont diverses, selon les tâches visées, les modes de représentations adoptés et les paradigmes implémentés. On peut toutefois identifier une succession d'étapes communes : l'isolement, dans la chaîne d'entrée, d'unités représentant une cohérence (syntaxique ou sémantique), l'activation par ces unités de structures de plus haut niveau puis la consolidation de la représentation obtenue dans le contexte dialogique. Les stratégies développées doivent faire face à deux types de perturbations : les limitations du décodage acoustico-phonétique et des modèles de langage *N-gram* et les perturbations induites par le caractère spontané de l'oral. La présence de ces perturbations rend difficile une analyse linguistique complète des énoncés, du moins avec les modèles développés pour le langage écrit. Une première solution pour éviter ce problème consiste à profiter du caractère finalisé de la communication homme-machine en développant des stratégies sélectives essentiellement guidées par le domaine d'application. Il s'agit alors d'extraire le sens global de l'énoncé, représenté sous forme de traits considérés comme pertinents. Comprendre se ramène ainsi à extraire des valeurs de rubriques prédéfinies. Une autre approche serait d'essayer de ne pas perdre le pouvoir prédictif et structurel de la syntaxe. Ces stratégies s'orientent ainsi vers l'utilisation de techniques d'analyse ascendantes robustes tolérantes aux problèmes évoqués.

Quels que soient les choix et les modélisations adoptés, les méthodes stochastiques, qui ont montré leur efficacité pour la reconnaissance de la parole, ont aujourd'hui une

---

<sup>43</sup>L'algorithme *Left Corner* est un cas particulier de l'algorithme *Head Corner* dans lequel la catégorie tête est systématiquement la première rencontrée à gauche.

position dominante dans la plupart des niveaux de traitement de la langue parlée. Il s'agit essentiellement d'éviter le travail fastidieux qu'est la conception manuelle d'un jeu de règles de grammaire. L'analyse probabiliste permet en outre de garantir une certaine robustesse vis-à-vis des énoncés déviants. Le principal problème réside toutefois dans le volume de données nécessaire à l'apprentissage et à la réduction de la taille des modèles. À la difficulté de la constitution des corpus oraux s'ajoute celle de leur représentativité.

Si ces modèles stochastiques, notamment les approches de décodage conceptuel, s'intègrent bien aux modèles markoviens utilisés en reconnaissance, il est intéressant de constater que l'origine de la plupart des stratégies d'analyses robustes que nous avons évoquées vient précisément de la nécessité d'exploiter des contraintes linguistiques pour pallier les limitations des modèles de langage et aider le filtrage des hypothèses de la reconnaissance. La plupart de ces techniques ont ainsi été conçues pour traiter le treillis de mots issu du module de reconnaissance.

Enfin, il est frappant de constater que la plupart de ces approches ignorent le matériau sur lequel elles travaillent, appréhendé bien souvent comme un ensemble aléatoire de variations à une norme de référence : l'écrit.

La situation actuelle de la CHM orale semble ainsi paradoxale : d'un côté on dispose de systèmes de plus en plus efficaces sur un type de tâches bien défini ; de l'autre, en ignorant les phénomènes linguistiques qui caractérisent le langage oral en situation de dialogue homme-machine, la question de la pertinence réelle des paradigmes développés peut légitimement être posée. Cette interrogation fait l'objet du chapitre suivant.



## Chapitre 2

# Compréhension : quelles perspectives ?

Ce chapitre regroupe quelques réflexions dégagées à partir de l'examen des applications et des techniques actuelles en traitement automatique de la langue parlée en situation de dialogue oral homme-machine. Nous y analysons les facteurs qui, à notre avis, plaident pour une intégration de traitements linguistiques dans le processus de compréhension. Nous ne remettons pas en cause les stratégies employées jusqu'ici. Les conclusions que nous tirons reflètent simplement quelques unes des évolutions qui nous semblent essentielles pour tendre vers des progrès significatifs dans ce domaine de recherche. Elles justifient l'orientation que nous avons progressivement adoptée au cours de nos travaux de recherche et peuvent de ce fait être légitimement débattues.

### 2.1 Bilan de la compréhension sélective

Avec l'apparition des systèmes interactifs vocaux grand public opérationnels, la question de la robustesse des stratégies employées pour la reconnaissance et la compréhension de la parole est devenue au cœur des recherches du domaine. L'application réelle de tels systèmes a souligné la nécessité de développer des méthodes leur permettant d'être de moins en moins sensibles aux diverses perturbations pouvant survenir au cours de l'interaction : signal sonore perturbé temporairement par une source parasite ; utilisation de mots en dehors du vocabulaire du système ; erreurs de reconnaissance (substitutions, omissions et insertions) ; hésitations, corrections ou reprises du locuteur etc. La robustesse concerne ainsi tous les niveaux de traitement de la langue parlée. L'analyse doit avant tout pouvoir produire de l'information partielle en cas d'analyses incomplètes et essayer d'accepter les principaux phénomènes de la langue orale. Cette nécessité, à laquelle s'ajoute le problème de l'ambiguïté, a été à l'origine des méthodes sélectives guidées par la sémantique. Ces techniques ont été appliquées avec un succès incontestable sur des domaines applicatifs restreints.

### 2.1.1 Portée de l'analyse et prédictabilité

Une critique générale des approches sélectives est que l'utilisateur ne peut s'écarter des postulats de compréhension du système sans provoquer des problèmes de communication.

Ce problème concerne tout d'abord la portée de l'analyse réalisée. Comme le fait remarquer Seneff [Seneff, 1998], il est quasiment impossible, même pour un domaine très restreint, d'écrire une grammaire qui couvre toutes les formulations envisageables. Ainsi il n'est pas question de construire la structure sémantique complète de l'énoncé. Une première conséquence est que certaines nuances de la requête de l'utilisateur sont clairement passées sous silence [Colineau, 1997]. On peut citer par exemple les répercussions des adverbes comme *éventuellement* sur la réponse à la requête du locuteur, qui ne sont pas prises en compte.

La seule observation de séquences restreintes d'entités linguistiques, qui ne peut rendre compte de la structure de l'énoncé, a également des conséquences sur la complexité des énoncés pouvant être traités. On touche ici aux limites d'une modélisation du langage reposant uniquement sur une correspondance séquentielle entre les mots et les concepts liés au domaine applicatif. Cette modélisation constitue un compromis technique sous optimal pour de nombreuses classes d'applications. Nous avons déjà donné une illustration de ce phénomène en évoquant la gestion très approximative des accords par les systèmes de dictée vocale, dus à une modélisation imparfaite des dépendances longue distance. Ces limites concernent également les niveaux supérieurs de traitement et s'étendent au problème du sens en contexte. [Minker, 1999] souligne que l'alignement des étiquettes sémantiques dans une structure de surface s'est révélé inapproprié pour les relations distantes. Il illustre son propos en prenant l'exemple d'erreurs dues à des marqueurs non adjacents aux arguments des concepts qu'ils auraient du identifier et celui des structures sémantiques imbriquées. On remarquera que les approches sélectives sont par ailleurs souvent incapables de traiter les énoncés contenant des requêtes multiples ou des contraintes multiples [Minker, 1999].

Le second problème concerne les écarts éventuels entre les constructions syntaxiques acceptées par le système et les constructions pouvant être utilisées par les locuteurs. Comme nous l'avons déjà évoqué, la réussite des approches sélectives stochastiques réside principalement dans la représentativité des corpus choisis pour l'apprentissage. Or il est difficile d'obtenir une bonne représentation de tous les phénomènes<sup>1</sup>. Que faire de ces phénomènes ou constructions faiblement représentés ? La principale faiblesse de ces approches vient du fait qu'il est impossible d'en avoir une caractérisation précise. Ces formulations peuvent pourtant répondre à des mécanismes spécifiques de l'oral (comme les dislocations par exemple) dont la prise en compte pourrait être utile pour le traitement de ces énoncés. En ignorant le matériau sur lequel elles travaillent, les

---

<sup>1</sup>Nous verrons par ailleurs (cf. chapitre 3) que le recueil de corpus de dialogues oraux représentatifs des usages en fonction de la situation d'interaction concernée et du domaine d'application constituent déjà en soi un premier obstacle.

méthodes sélectives sont dans l'incapacité de diagnostiquer les sources d'erreurs comme de mettre en œuvre des stratégies permettant d'y remédier. Ainsi on ne peut bien souvent que se contenter d'imposer un dialogue directif aux utilisateurs pour y faire face : on les contraint à reprendre leur formulation.

Certes, on pourra argumenter que l'utilisation de ces constructions sera rare en pratique. Leur rejet par le système, contrainte imposée au locuteur, est toutefois, de notre point de vue, un frein majeur à l'élaboration d'une interaction naturelle et collaborative, à l'évolution de nos comportements langagiers face à la machine et en conséquence à l'utilisation par le grand public de cette modalité d'interaction.

### 2.1.2 Influence des prétraitements

La plupart des techniques sélectives stochastiques utilisent une étape de prétraitement des énoncés lors de l'analyse. Il s'agit le plus souvent d'une analyse lexicale permettant une normalisation de l'entrée et indispensable pour réduire la taille du modèle<sup>2</sup>. Une analyse morphologique est réalisée, permettant la lemmatisation des formes lexicales rencontrées, remplaçant les valeurs de la base de données par des catégories (numéro identifiant un vol par exemple), regroupant des mots liés sémantiquement dans l'application (noms de villes, noms d'avions). Les critères de regroupement peuvent ne pas reposer sur la morphologie mais uniquement sur quelques classes lexicales significatives dans le corpus d'apprentissage. C'est au cours de cette étape que les mots jugés non pertinents sont identifiés.

Cette analyse traite généralement les mots hors de leur contexte. Ce prétraitement n'est ainsi pas sans conséquences. On notera en particulier que l'étiquetage d'un mot dans une catégorie "non pertinente" peut conduire à des erreurs. [Minker, 1999] donne l'exemple d'un élément de phrase "(...) à 19 heures j'ai 1 (...)" qui après la suppression de *j'ai* est décodé en heure et minute de départ. Par ailleurs, l'ambiguïté sémantique restreinte des domaines applicatifs traditionnellement envisagés permet d'éviter le risque d'effectuer un prétraitement erroné des mots pouvant être utilisés différemment selon le contexte<sup>3</sup>. On peut toutefois s'interroger sur la validité de cette approche dans des domaines applicatifs plus riches<sup>4</sup>.

Ces étapes de prétraitements permettent également de supprimer les redondances d'information induites par les répétitions. Par exemple, [Issar et Ward, 1993] éliminent les mots identiques à ceux qui les précèdent. En cas de correction ils ne considèrent que la dernière information fournie. L'objectif de ce filtrage est également guidé par la nécessité de réduire la taille des modèles. Encore une fois, ce filtrage aveugle peut poser problème. La suppression des éléments marquant les hésitations, par exemple, risque de faire succéder dans l'énoncé des unités de sens n'ayant rien en commun comme dans

---

<sup>2</sup>Notons que la taille du modèle augmente avec la richesse des représentations sémantiques que l'on cherche à fournir. L'équilibre entre le nombre de paramètres et la qualité du modèle est ainsi plus difficile à atteindre.

<sup>3</sup>De telles erreurs ont été attestées par Minker sur le domaine ATIS [Minker, 1999] p.91

<sup>4</sup>On notera par ailleurs que cette analyse est assez dépendante de l'application donc peu portable.

les cas des inachèvements. Ceci peut avoir un impact sur le décodage, fondé sur la correspondance linéaire entre mots et concepts. De même, rien ne garantit que des erreurs supplémentaires ne soient pas artificiellement introduites lors des traitements concernant les corrections. Nous verrons qu'un filtrage de ces phénomènes n'est pas si évident que cela puisse laisser paraître (cf. paragraphe 2.2.1.2).

Soulignons enfin à ce sujet que les méthodes sélectives, dont l'objectif était précisément la robustesse à ce type de phénomènes, se révèlent limitées : les répétitions et les corrections représentent une des sources principales d'erreurs de ces systèmes [Minker, 1999]. Il semble en effet difficile de trouver une loi statistique qui s'ajuste convenablement aux phénomènes oraux dus à la spontanéité [Antoine et Genthial, 1999]. Ces problèmes ont été masqués dans les premières campagnes d'évaluation ARPA car beaucoup d'énoncés (sur le corpus ATIS 2 en particulier) ont été considérés dès le départ comme hors de propos ou difficiles à interpréter automatiquement et ont été retirés des évaluations. On notera qu'à l'exception de [Minker, 1999], peu de travaux en compréhension sélective font état de la couverture réelle de ces phénomènes et des impacts des étapes de prétraitement sur l'analyse.

### 2.1.3 Généricité

Les erreurs attestées ne proviennent pas seulement du décodage ou des étapes de prétraitement. Elles sont également introduites par les concepteurs sous forme de règles de grammaire inadaptées ou d'étiquettes sémantiques inappropriées [Minker, 1999]. Si la modélisation stochastique permet de s'affranchir du coût de la conception des grammaires et d'obtenir rapidement des corpus étiquetés (approche par *bootstrap* déjà évoquée), le problème du choix d'appariement lors de la conception entre mots et étiquettes sémantiques, dont il faut veiller à la consistance, reste entier. Ce problème n'est pas spécifique aux méthodes sélectives. Il prend cependant une importance particulière dans ces approches, où l'extraction du sens repose uniquement sur les concepts liés au domaine applicatif et leur réalisation (contraintes locales). On peut raisonnablement s'interroger sur la validité de cette hypothèse pour la gestion d'une ambiguïté sémantique supérieure à celle des domaines restreints traditionnellement envisagés<sup>5</sup>.

D'une manière générale, si une compréhension sélective semble suffire aux besoins d'un dialogue finalisé portant sur une tâche réduite et bien spécifiée, la portabilité de cette approche minimaliste (d'un point de vue linguistique) à des domaines plus complexes constitue une question largement ouverte [Hirschman, 1998]. En accordant une certaine importance à l'analyse syntaxique, les méthodes d'exploration *d'îlots syntaxiques* spéculent moins sur la capacité de la tâche d'interprétation. En permettant la réutilisabilité d'une partie de l'analyse, elles restent ainsi (relativement) moins dépendantes du domaine applicatif.

---

<sup>5</sup>Nous reviendrons sur l'augmentation éventuelle de l'ambiguïté au chapitre 4 portant sur le domaine d'application que nous avons choisi.

Si la portabilité d'un système sélectif stochastique, à un domaine proche, est aisée<sup>6</sup> [Minker, 1999], on notera qu'une augmentation de la richesse du domaine se traduit par une augmentation de la taille des modèles. En conséquence, l'équilibre entre la complexité du modèle et la pertinence de l'information doit être optimisé. Il est intéressant de constater que le passage de l'application ATIS à ARISE<sup>7</sup> dans le cadre d'une modélisation markovienne [Minker, 1999] nécessite le recours à des prétraitements prenant en compte le contexte de l'énoncé et une interpolation de ce contexte dans le modèle. Cette prise en compte du contexte s'avère toutefois parfois insuffisante, notamment pour la portée de la négation et les relations sémantiques distantes.

Cette question de généralité ne se limite pas à la seule problématique de la réutilisabilité du système. Rien ne garantit, de notre point de vue, qu'une compréhension sélective suffise aux besoins d'une interaction homme-machine réellement collaborative et non plus limitée à une recherche d'information très ciblée. Il est en particulier difficile d'appréhender quelles seront les conséquences sur l'analyse réalisée de l'enrichissement du domaine applicatif. La prolifération des règles d'appariement et des heuristiques à mettre en œuvre pour la gestion des conflits entre règles risquent de conduire à des ambiguïtés dues au manque de précision linguistique (présence des quantifieurs, portée de la négation, référence par exemple). Nous y reviendrons au paragraphe 2.2.3.

## 2.2 Vers un changement de paradigme ?

Dans la suite de ce chapitre, nous illustrons le besoin de l'intégration de traitements linguistiques dans les processus de traitement de la langue parlée. L'évolution des travaux concernant notamment l'interfaçage des modules de reconnaissance et de compréhension et le traitement des phénomènes spécifiques à l'oral semblent déjà remettre en question le consensus qui s'était développé en faveur des approches sémantiques au sein de la communauté scientifique. Nous analysons ensuite les conséquences d'une augmentation de la richesse du domaine.

### 2.2.1 Traitements linguistiques et CHM : une nécessité qui s'impose

#### 2.2.1.1 Interface entre reconnaissance et compréhension

Un premier exemple de cette nécessité peut être donné par l'examen des techniques qui se sont imposées pour dépasser les limites des modèles de langage *N-gram* utilisés en reconnaissance de la parole. Le chapitre précédent a mis en évidence le besoin de faire intervenir des informations linguistiques, qu'il s'agisse d'améliorer ces modèles en les rendant plus discriminants ou de chercher à ré-ordonner les différentes hypothèses résultantes. Les évolutions majeures des recherches récentes dans ce domaine proviennent

---

<sup>6</sup>Dans une méthode stochastique, le développement du système se limite à l'étiquetage du corpus d'apprentissage. La difficulté vient de la conception de la topologie du modèle et de la consistance du jeu d'étiquettes sémantiques défini.

<sup>7</sup>L'application ARISE regroupe 64 concepts contre 38 pour l'application ATIS [Minker, 1999; Maynard et Lefèvre, 2002].

sans doute du rapprochement significatif qui a été observé ces dernières années entre les communautés du traitement de la langue parlée et du traitement du langage naturel [Antoine et Genthial, 1999; Cunningham, 1999].

La première stratégie d'hybridation forte, que nous avons déjà évoquée au chapitre précédent, consiste à intégrer des informations syntaxiques dans l'apprentissage des modèles *N-gram*. Le calcul des paramètres du modèle est ainsi effectué en fonction des probabilités d'une grammaire hors-contexte stochastique. L'utilisation d'un tel modèle permet de mieux exploiter des connaissances générales et structurantes de la langue. Les recherches sur ces grammaires [Booth et Thompson, 1973] sont plus anciennes que celles des premières applications des modèles markoviens. De nombreux systèmes intègrent ainsi une analyse syntaxique avec un modèle de langage *N-gram* pour aider le processus de décodage [Goddeau et Zue, 1992; Kita, Morimoto, Ohkura, Sagayama, et Yano, 1994]. D'autres approches proposent des alternatives plus complexes, l'intégration de probabilités ayant été étendue aux formalismes lexicalisés actuels [Lafferty, Sleator, et Temperley, 1992; Schabes, 1992; Chelba et Jelinek, 1998].

Comme nous l'avons déjà souligné (cf. paragraphe 1.3.1), cette stratégie consistant à améliorer la pertinence des *N*-meilleures hypothèses grâce à des informations syntaxiques et sémantiques intervenant tôt dans le processus de reconnaissance n'a pas encore donné lieu à des progrès significatifs par rapport aux modèles *N-gram*. Cela peut s'expliquer par le fait qu'étant donné que les connaissances linguistiques sont peu présentes dans les modèles *N-gram*, une erreur de reconnaissance en un point donné de l'énoncé ne compromet pas la suite du décodage. D'autre part, les approches les plus complètes [Jelinek et Chelba, 1999; Zhang et al., 1999] requièrent des modèles probabilistes très complexes. Ainsi, si la nécessité d'utiliser des informations supplémentaires (structures syntaxiques mais aussi prosodiques [Hunt, 1994] voire sémantiques) est admise, la meilleure façon d'incorporer ces traitements linguistiques dans le processus de reconnaissance n'est toutefois pas encore une question tranchée [Seneff, 1998]. Le foisonnement des recherches dans ce domaine témoigne toutefois bien de leur importance pour améliorer la robustesse et la précision des prédictions d'un mot étant donné le contexte d'observation et ainsi limiter la multiplicité des hypothèses de reconnaissance.

Une autre stratégie consiste à appliquer une analyse linguistique sur le treillis de mots issu de la reconnaissance<sup>8</sup>. Il s'agit de filtrer les différentes hypothèses *a posteriori* tout en amorçant les étapes supérieures du traitement. C'est ce double objectif qui a été à l'origine des approches d'analyse robuste utilisées en compréhension et qui a motivé l'intégration de traitements linguistiques fins. L'analyse par *chart*, en particulier, constitue un moyen optimal pour mettre en concurrence de façon non déterministe les différentes hypothèses de reconnaissance. Les différentes approches envisagées [Ehrlich et Hanrieder, 1996; Dowding et al., 1993; Noord et al., 1999] utilisent des grammaires d'unification<sup>9</sup> pour représenter les connaissances linguistiques nécessaires à l'analyse

<sup>8</sup>On notera également l'existence de techniques intermédiaires d'*intégration progressive* [Murveit, Butzberger, Digalakis, et Weintraub, 1993] qui appliquent en étapes successives des traitements linguistiques de plus en plus fins sur les sorties de la reconnaissance.

<sup>9</sup>Grammaires catégorielles d'unification (UCG) [Calder et al., 1988] pour [Ehrlich et Hanrieder,

du graphe ; ce formalisme permettant l'intégration aisée de contraintes syntaxiques et sémantiques. L'analyseur cherche d'abord à trouver un chemin dans le graphe auquel puisse être attachée une analyse complète étant donnée la grammaire. Quand ce n'est pas le cas, l'analyse tire profit des analyses partielles présentes dans le *chart*. Des techniques spécifiques sont alors parfois nécessaires pour améliorer l'efficacité de l'analyse. [Noord et al., 1999], par exemple, emploient une analyse *Head Corner* centrée sur les informations *a priori* les plus pertinentes pour diminuer le nombre d'hypothèses. Le choix de la meilleure analyse s'effectue en appliquant un certain nombre d'heuristiques. Il s'agit le plus souvent d'une pondération entre différents scores : score acoustique déjà présent dans le graphe, score de "complétude" syntaxique, et score pragmatique de vraisemblance de l'analyse pour [Ehrlich et Hanrieder, 1996] et [Dowding et al., 1993]. [Noord, 2001] détermine une relation de préférence entre les analyses partielles en prenant en compte le nombre de "sauts" dans l'analyse, le nombre d'îlots d'analyse dont la projection maximale est une catégorie majeure (NP, PP, etc.), le score acoustique et un score donné par un modèle de langage trigram associé au chemin considéré dans le graphe de mots.

Les résultats de l'évaluation de cette dernière approche, exprimés en terme de précision conceptuelle (*Concept Accuracy*<sup>10</sup>) peuvent être trouvés dans [Zanten, Bouma, Sima'an, Noord, et Bonnema, 1999]. Lorsque le graphe de mots est filtré par un modèle de langage trigram, [Noord, 2001] obtient une précision conceptuelle de 82.6%. D'une manière générale, ces méthodes ne permettent pas d'améliorer la précision de la reconnaissance c'est-à-dire que l'analyse grammaticale effectuée sans prendre en compte les scores d'un modèle de langage ne permet pas une désambiguïsation du graphe de mots meilleure que dans le cas de l'utilisation d'un modèle *N-gram* seul. En revanche, à partir d'un graphe préalablement filtré par un modèle *N-gram*, elle est indispensable pour mettre en concurrence l'ensemble des hypothèses et d'en déterminer, en extrayant toutes les informations partielles exploitables linguistiquement, l'analyse sémantique la plus précise. Par ailleurs, ces expériences démontrent que des analyses linguistiques détaillées d'énoncés oraux peuvent être appliquées de façon performante [Noord et al., 1999].

### 2.2.1.2 Inattendus structuraux

Considérons à présent les difficultés induites par la présence des hésitations, répétitions et corrections dans les productions orales. Comme nous l'avons mentionné au chapitre précédent, ces phénomènes sont généralement considérés comme un bruit qu'il convient au mieux d'ignorer. Les modèles markoviens offrent une capacité d'adaptation autour des exemples ayant servi à l'apprentissage, ce qui garantit une certaine robustesse vis-à-vis de ces énoncés déviants. C'est pourquoi peu de travaux ont cherché à

---

1996], Grammaires à Clauses Définies (DCG) [Pereira et Warren, 1980] proches des HPSG en ce qui concerne l'approche proposée par [Noord et al., 1999].

<sup>10</sup>La précision conceptuelle est définie par  $CA = 1 - \frac{SU_s + SU_i + SU_d}{SU}$  où  $SU$  est le nombre total d'unités sémantiques de référence et  $SU_s$ ,  $SU_i$ ,  $SU_d$  dénotent respectivement le nombre de substitutions, d'insertions et de suppressions nécessaires pour faire correspondre l'analyse fournie à celle de référence.

modéliser ces événements<sup>11</sup>.

Comme le souligne [Price, 1996], si ces phénomènes sont plus fréquents dans des dialogues homme-homme qu'en situation de dialogue homme-machine, cela risque d'être différent dès lors que nous serons plus familiarisés avec ce type d'interaction et que les locuteurs se focaliseront davantage sur leurs buts (par rapport à leurs objectifs communicationnels) que sur la forme même de leur production orale. On peut noter à ce propos que dans le domaine de l'interaction médiatisée, comme dans le projet VERMOBIL [Bub et Schwimm, 1996], caractérisée par une situation de négociation élevée, plus de 20% des énoncés attestent la présence d'auto-corrections [Ruland, Rupp, Spilker, Weber, et Worm, 1998].

Nous examinons dans cette partie les techniques *correctives*, cherchant à rattraper explicitement ces perturbations en mettant en place des procédures de détection, d'identification et de corrections associées. Cela passe par une étude, une catégorisation et une modélisation de ces phénomènes. Les travaux que nous évoquons sont révélateurs de la difficulté de la tâche et des connaissances qu'elle nécessite.

La plupart de ces travaux sont inspirés de l'expérience, plus ancienne, proposée par [Hindle, 1983]. Ils sont fondés sur une modélisation de ces phénomènes illustrée par l'exemple suivant issu de [Heeman et Allen, 1999] :

$$(E.2.1) \quad \begin{array}{ccccccc} \textit{we'll pick up} & \textit{a tank of} & & \textit{uh} & \textit{the tanker of} & \textit{oranges} & \\ & \textit{reparandum} & \uparrow & \textit{zone} & \textit{alteration} & & \\ & & \textit{ip} & \textit{d'édition} & & & \end{array}$$

La partie du discours que le locuteur cherche à corriger est le *reparandum*, *ip* (ou point d'interruption) représente la fin du *reparandum*, l'élément remplacé est nommé *altération* et la *zone d'édition* est la frontière entre les deux éléments. Cette dernière peut être simplement intonative ou marquée par un mot précis : appuis du discours ou interjections. Les objectifs des traitements sur la base de cette modélisation sont non seulement de pouvoir détecter ces phénomènes mais aussi de les corriger<sup>12</sup> en déterminant l'étendue des *reparanda* et des zones d'édition de remplacement<sup>13</sup>.

Les travaux de [Bear, Dowding, et Shriberg, 1992] reposent ainsi sur l'utilisation de *patterns* de correspondance très génériques caractérisant le *reparandum* et l'*altération* sur la base d'informations lexicales et syntaxiques. Le tableau 2.1 donne quelques

<sup>11</sup>Rappelons qu'ils constituent une des sources d'erreurs principale des méthodes sélectives [Minker, 1999].

<sup>12</sup>On notera à ce propos que le but recherché n'est pas obligatoirement d'éliminer de l'énoncé les *reparanda* et les zones d'édition. [Heeman et Allen, 2001] soulignent précisément que ces éléments peuvent contenir des éléments informatifs importants pour les anaphores par exemple.

<sup>13</sup>Notons que ces travaux s'intéressent également à la détection des changements intonatifs dans l'énoncé ; plusieurs études ayant montré que cette information prosodique pouvait réduire l'ambiguïté syntaxique [Bear et Price, 1990]. Précisons qu'il n'y a pas actuellement de réel consensus sur l'utilisation de la prosodie. Malgré le fait que le contour intonatif ne soit sans doute pas la meilleure unité pour la modélisation du dialogue, il semble toutefois permettre de mieux déterminer les intentions du locuteur [Traum et Heeman, 1997].



exemples de tels patterns permettant d'identifier répétitions et corrections. Un mot étiqueté  $M_i$  indique qu'il reprend le mot de la position  $i$  (les mots doivent être exactement identiques), une étiquette  $R_i$  indique un mot de remplacement (même catégorie grammaticale) à celui de la position  $i$ ,  $|$  représente la zone d'édition et  $X$  les mots neutres c'est-à-dire n'intervenant pas dans le phénomène [Bear, Dowding, Shriberg, et Price, 1993].

<i>Patterns</i>	Exemples
$M_1 M_1$	<i>what what are the fares</i> $M_1 \quad   \quad M_1$
$M_1 X_1...X_iM_1$	<i>show me flights uh well flights</i> $M_1 \quad   \quad X \quad M_1$ <i>show me flights daily flights</i> $M_1 \quad   \quad X \quad M_1$
$M_1R_1 M_1R_1$	<i>... fly from Boston from Denver</i> $M_1 \quad R_1 \quad   \quad M_1 \quad R_1$
$R_1M_1X R_1M_1$	<i>... fly from Boston or uh to Boston</i> $R_1 \quad M_1 \quad X \quad   \quad R_1 \quad M_1$

TAB. 2.1 – Patterns de correspondance identifiant répétitions et corrections

Le principal problème de cette approche est la surgénérativité c'est-à-dire l'étiquetage erroné en tant que correction d'une succession de mots. Ainsi, si cette approche donne d'assez bons résultats pour la détection de ces phénomènes, elle ne peut pas être directement utilisée pour la correction. C'est le cas de l'exemple "*show me flights daily flights*" dans lequel il serait abusif de supprimer le mot *daily* porteur de sens.

Ainsi dans le système GEMINI [Dowding et al., 1993] une analyse complète est d'abord réalisée. C'est seulement si elle échoue, que la technique de correction que nous avons évoquée est appliquée. Une seconde analyse sur l'énoncé corrigé est alors de nouveau tentée. Ceci reste cependant encore insuffisant et les auteurs insistent pour la prise en compte de connaissances de plus haut niveau (contraintes syntaxiques plus riches que la simple catégorie grammaticale ou contraintes sémantiques) pour espérer pouvoir corriger sans risques ces phénomènes.

D'autres approches ont été envisagées dans cette même optique. [Nakatani et Hirschberg, 1994; Shriberg, Bates, et Stolcke, 1997] détectent les corrections par un modèle utilisant principalement des indices acoustiques et prosodiques. Les indices les plus pertinents selon les résultats de l'évaluation de l'approche de [Nakatani et Hirschberg, 1994] semblent être la durée des pauses entre les mots, la présence de fragments de mots<sup>14</sup> et la correspondance lexicale (modèle *trigram*). [Stolcke et Shriberg, 1996] ont proposé un modèle de langage permettant la prise en compte de certaines répétitions simples. Pour

<sup>14</sup>Nous émettons toutefois quelques réserves sur l'exploitation de cet indice, rendu disponible grâce à une transcription des corpus oraux. Or, comme nous l'avons vu au paragraphe 1.3.1, dans les systèmes de reconnaissance traditionnels, une amorce de mots suffisamment longue est traduite par le mot du vocabulaire le plus approchant.

les énoncés susceptibles de contenir une correction, le contexte de prédiction d'un mot par le modèle de langage est modifié. Ce modèle, restreint à des phénomènes simples est une des premières tentatives d'intégration de ces phénomènes dans un modèle de langage stochastique.

Des travaux plus récents [Heeman et Allen, 2001] prônent une modélisation de ces événements spontanés, dès l'étape de reconnaissance, via un modèle de langage statistique approprié. L'objectif y est clairement de résoudre à la fois les problèmes de la détection et de la correction de ces phénomènes pour faciliter les traitements syntaxiques et sémantiques ultérieurs. L'annotation des corpus suit, pour la modélisation des inattendus, les conventions de [Bear et al., 1993] que nous venons d'évoquer. Elle intègre toutefois, en plus de l'identification des *reparanda* et des zones d'édition, une catégorisation du type de la correction : *faux départs* (amorces puis abandon d'une partie de l'énoncé), *reprises* (existence d'une correspondance entre l'élément initial et l'élément repris), et *hésitations* (discontinuités dans l'énoncé marquées par un ou plusieurs mots). Cette modélisation est issue d'une analyse descriptive poussée de corpus oraux<sup>15</sup>. Les auteurs proposent de faire reposer la prédiction d'un mot sur des indices acoustiques, lexicaux et sur un étiquetage grammatical. Ils soulignent l'importance de cet étiquetage pour la caractérisation fine des phénomènes modélisés et dans le but d'aider les niveaux de traitement supérieurs. L'incorporation des étiquettes grammaticales au sein du modèle de langage nécessite dans cette approche de redéfinir le problème de la reconnaissance présenté au paragraphe 1.3.1. Soit  $P$  une séquence d'étiquettes grammaticales. Le rôle du module de reconnaissance est maintenant de résoudre :

$$\begin{aligned}\widehat{WP} &= \arg \max_{WP} P(WP|A) \\ &= \arg \max_{WP} P(A|WP)P(WP)\end{aligned}\tag{2.1}$$

$P(A|WP)$  est le modèle acoustique<sup>16</sup>.  $P(WP)$  est le modèle de langage tenant compte à la fois de la séquence de mots et de la séquence d'étiquettes grammaticales associées. Trois variables supplémentaires sont intégrées dans le modèle de langage pour la détection des inattendus. Elles tiennent compte de quelques indices permettant de les identifier : une variable  $R_i$  caractérise le type de la correction ; une variable  $I_i$  dénote la présence d'une marque intonative terminant un segment et une variable  $E_i$  caractérise les marqueurs lexicaux éventuels entre les mots  $W_{i-1}$  et  $W_i$  (codage de la délimitation de la zone d'édition). Cette dernière variable prend quatre valeurs possibles suivant l'appartenance des mots  $W_{i-1}$ ,  $W_i$  et  $W_{i+1}$  à la zone d'édition. L'exemple suivant représente un énoncé étiqueté pour l'apprentissage tenant compte des variables  $R_i$  et  $E_i$  :

(E.2.2) "(...) there at four a.m. PUSH oh ET sorry POP Mod at eleven a.m.  
[Heeman et Allen, 2001]

<sup>15</sup>Le corpus utilisé, TRAINS [Heeman et Allen, 1995] est un corpus homme-homme de dialogues finalisés incluant 34 locuteurs différents et portant sur 20 tâches distinctes.

<sup>16</sup>Il peut être approché par  $P(A|W)$ .

Dans cet exemple, Mod indique la présence d'une reprise, PUSH indique le début de la zone d'édition ( $W_{i-1} : a.m.$  n'appartient pas à la zone d'édition alors que c'est le cas pour  $W_i : oh$ ), ET indique que les deux mots *oh* et *sorry* font partie de la zone d'édition et POP indique de la même façon la fin de la zone. Le but de la reconnaissance est ainsi de trouver la séquence de mots, étiquettes grammaticales associées, type de correction éventuel, marqueurs des zones d'édition et marques intonatives, la plus probable étant donné le signal acoustique. Le modèle de langage nécessite ainsi l'estimation de cinq distributions de probabilité, chacune d'entre elles intégrant dans son contexte mots, étiquettes grammaticales et les trois variables citées précédemment. Le lecteur pourra se reporter à [Heeman et Allen, 2001] pour les techniques d'estimation, complexes, utilisées pour un tel modèle.

Les travaux de ces mêmes auteurs pour permettre une correction appropriée sont par ailleurs très intéressants. Ils reposent essentiellement sur l'identification du *reparandum* et de l'élément repris. Ils insistent sur la nécessaire prise en compte de tous les indices établissant une correspondance entre ces deux groupes, que ce soit au niveau du mot, à un niveau syntaxique ou intonatif. Il s'agit de prédire au mieux le premier mot du groupe qui sera repris. Il est intéressant de remarquer l'exploitation de la continuité entre le contexte précédent le *reparandum* et l'*altération*. Les techniques utilisées sont les mêmes que pour la détection : le modèle de langage est enrichi de trois nouvelles variables : une variable pour la prédiction du début du *reparandum*, une variable établissant des correspondances *syntaxiques* entre un mot du *reparandum* et un mot du groupe repris et une dernière sur le type de la correspondance (identité sur le mot et l'étiquette grammaticale ou uniquement sur leurs catégories grammaticales).

Les résultats de leur expérience indiquent un taux de rappel de détection et de correction de ces phénomènes de 66% avec une précision de 74%. On notera toutefois des différences significatives entre les types de corrections. Ainsi le taux d'erreur sur les faux départs est près de 98% alors qu'il est de 57% sur l'ensemble des phénomènes. On retrouve le même problème que dans l'expérience menée par [Bear et al., 1992] où des énoncés corrects sont faussement détectés et abusivement corrigés en tant que reprises.

Comme la plupart des approches cherchant à améliorer les modèles de langage, on ne constate qu'une amélioration de 1% pour le reclassement des hypothèses de reconnaissance par rapport à l'utilisation d'un modèle classique *N-gram*<sup>17</sup>. La modélisation des inattendus entraîne en revanche une meilleure précision de l'étiquetage grammatical (près de 9%) ce qui souligne l'importance de la modélisation de ces phénomènes pour la conduite d'une analyse détaillée des énoncés oraux. De la même manière que [Bear et al., 1992], les auteurs concluent à la nécessaire prise en compte d'informations syntaxiques et sémantiques plus riches pour améliorer la détection et surtout la correction de ces phénomènes de façon à établir de meilleures correspondances entre le *reparandum* et le groupe repris [Heeman et Allen, 1999, 2001].

---

<sup>17</sup>Cette approche n'a pas pour le moment été intégrée à un module de reconnaissance, mais utilisée sur un graphe de mots hypothèses [Heeman et Allen, 2001].

Cette notion de repérage des *reparanda*, des zones d'édition et des éléments repris, a été par ailleurs exploitée dans le projet VERMOBIL. En utilisant des informations grammaticales et prosodiques, le graphe issu de la reconnaissance est augmenté de nouveaux arcs permettant d'éviter les segments correspondant éventuellement à une correction ou une hésitation [Ruland et al., 1998]. Aucune réparation n'est toutefois réellement effectuée à ce niveau. Le graphe augmenté contient toujours les informations initiales. Il s'agit ici simplement d'ajouter des indices pour le traitement de ces phénomènes par les niveaux supérieurs de l'analyse.

### 2.2.2 Rôle de la syntaxe

Les travaux précédemment évoqués ont souligné l'intérêt de la prise en compte d'informations de haut niveau pour le filtrage des hypothèses de reconnaissance et la correction des inattendus structuraux. En ce qui concerne ces derniers phénomènes, des stratégies correctives ont été proposées à partir d'analyseurs syntaxiques conçus pour l'écrit.

Dans l'approche de [Delisle, Boufaden, et Moulin, 1998], par exemple, les énoncés oraux sont analysés selon deux stratégies qui interagissent l'une avec l'autre. Un premier module réalise une analyse syntaxique en constituants reposant sur la théorie du gouvernement et du liage [Chomsky, 1981] et appliquant une stratégie robuste de type *GLR\** [Lavie, 1996]. L'analyse conduite dans le second module cherche à réparer ce qui est considéré comme les distorsions d'un écrit altéré. L'approche proposée met en jeu des heuristiques de recouvrement qui ne remettent pas en cause l'analyse précédente lorsque l'énoncé ne comporte pas de déviations. Ces heuristiques cherchent à la fois à rétablir la syntaxe de l'écrit (règles de mouvements des constituants pour rétablir un ordre standard lors de dislocations par exemple), à ignorer les phénomènes dus à la spontanéité et à corriger les reprises<sup>18</sup>. L'ordre d'application de ces règles de transformation est déterminant. Le traitement des procédés complexes (extractions, relatives, ellipses, anaphores) nécessitent ainsi, quand ils sont combinés avec des distorsions, des étapes de prétraitements complexes. On notera par ailleurs que l'heuristique ne retenant que le dernier constituant répété est parfois mise en échec, notamment dans les cas des enrichissement lexicaux. D'une manière générale, ces heuristiques permettent de trouver une solution qui s'écarte de façon minimale d'un ensemble de contraintes syntaxiques génériques. Elles nécessitent en contrepartie la définition de conditions d'application complexes qui se sont vite révélées inefficaces dès lors que l'énoncé traité "subissait" plus d'une de ces "distorsions" en même temps [Boufaden, Delisle, et Moulin, 1998].

Une stratégie similaire peut être trouvée dans [Lopez, 1999]. À la suite d'une ana-

---

<sup>18</sup>Pour ces derniers phénomènes, deux types d'heuristiques sont mis en œuvre. L'heuristique de *recouvrement du bruit* consiste à ignorer les unités lexicales qui ne peuvent être insérées dans le syntagme en cours de construction, ni représenter le début d'un nouveau syntagme ou représenter à elles seules un syntagme. Il s'agit principalement d'ignorer les interjections ou les répétitions simples de déterminants par exemple. L'heuristique de *recouvrement des répétitions* consiste à ne retenir comme constituants valides dans l'analyse que ceux répétés en dernier.

lyse syntagmatique normale à l'aide des LTAG<sup>19</sup> cherchant à caractériser des îlots syntaxiques étendus dans une stratégie ascendante, des règles de recouvrement sont appliquées. À la différence notable de l'approche précédente, l'application de ces règles n'est pas destructrice. Après avoir appliqué l'une de ces règles, l'expansion normale des îlots obtenus est tentée jusqu'à obtenir une analyse complète. Si ce n'est pas le cas, on tente à nouveau l'application d'une autre règle et ainsi de suite. Les règles s'appliquent aux items adjacents non rattachables et de manière prioritaire aux segments les plus longs. Cette approche est nettement plus intéressante que la précédente. D'une part, l'utilisation du formalisme lexicalisé choisi permet plus facilement de capter des phénomènes linguistiques complexes. D'autre part, l'analyse tabulaire effectuée permet d'obtenir des îlots grammaticalement corrects les plus étendus qui ne seront pas détruits lors de l'application des règles de recouvrement.

Un autre avantage de cette dernière approche est que l'analyse réalisée permet un traitement fin des constructions syntaxiques utilisées à l'oral. En effet, les règles de recouvrement appliquées ne sont pas simplement des heuristiques tentant de réparer une analyse syntaxique : elles résultent d'une modélisation détaillée des phénomènes de l'oral (cf. chapitre 3) qui ne prend pas en compte un quelconque jugement de "grammaticalité" par rapport à l'écrit. Ces constructions ne peuvent être appréhendées par une méthode sélective stochastique qu'à la condition implicite qu'une combinaison particulière de concepts soit suffisamment représentée dans le corpus d'apprentissage. Or, comme le rappellent [Sabah et al., 1997], le domaine d'application ne restreint en rien la complexité des phénomènes linguistiques ni ne réduit le nombre de ces phénomènes et de ces constructions. L'approche proposée par [Lopez, 1999] permet ainsi de traiter les ellipses, dont on sait qu'elles sont fréquentes à l'oral, notamment dans les situations d'interaction très répétitives, et dont le traitement ne saurait se passer d'informations syntaxiques et sémantiques.

Les constructions mettant en jeu une variation de l'ordre des mots de l'énoncé donnent un autre exemple de la nécessité de prendre en compte des informations syntaxiques. Ces constructions mettent en effet généralement en échec les méthodes sélectives. Par exemple, les systèmes développés pour l'allemand ne peuvent se passer de la syntaxe [Engel, 2000]. C'est le cas du système de dialogue finalisé EVAR [Gallwitz et al., 1998] et plus nettement dans le projet de traduction VERMOBIL qui utilise entre autres une analyse détaillée fondée sur le formalisme HPSG et une analyse syntaxique partielle en *chunks* [Engel, 2000]. Le français parlé est également concerné, dans une moindre mesure, par ce problème de variabilité de l'ordre des mots en raison des procédés d'extractions fréquemment utilisés à l'oral. Cet aspect sera plus précisément détaillé au chapitre 3.

Enfin, [Colineau, 1997] insiste sur la nécessité d'une analyse linguistique pour la désambiguïisation de certaines marques importantes pour l'identification de l'acte de

---

<sup>19</sup>Grammaires Lexicalisées d'Arbres Adjoints [Abeillé, 1991].

dialogue. Elle fait en effet l'hypothèse que l'intention du locuteur est présente dans l'énoncé sous forme d'une combinaison de marques linguistiques. Ces marques peuvent être exploitées pour guider l'interprétation sous réserve qu'elles soient analysées en combinaison avec des informations syntaxiques moins ambiguës. Un des exemples donné est celui des différents rôles que peut prendre la particule interrogative *est-ce-que* : requête (*est-ce-qu'on peut prendre le métro pour aller ...*), question fermée (*est-ce-que vous êtes sûr ?*), ou question alternative (*qu'est-ce-que c'est ce restaurant ?*).

Plus généralement, l'analyse syntaxique s'impose pour détecter plus finement les buts et les attentes de l'utilisateur et tendre ainsi vers une communication réellement collaborative. Ce problème reste masqué dans les systèmes restreints à une tâche bien spécifique. Nous avons vu qu'il était difficile de se rendre compte de la capacité des méthodes sélectives à interpréter correctement des requêtes multiples et souligné leurs faiblesses en cas de contraintes multiples [Minker, 1999]. De même, l'analyse linguistique devient indispensable dès que l'on cherche à interpréter les énoncés correctifs des locuteurs et à identifier la portée des rectifications qu'ils apportent (requête entière, remise en cause partielle d'une contrainte, précision, etc.).

### 2.2.3 Enrichissement du domaine de la tâche

Les différentes approches que nous avons passées en revue font ressortir assez nettement l'importance de la syntaxe, et plus généralement d'une analyse linguistique plus détaillée. En utilisant directement des connaissances de haut niveau, les méthodes sélectives permettent de passer outre les principales difficultés d'une telle analyse : ambiguïtés grammaticales, ambiguïtés de segmentation, "agrammaticalités", incomplétude des énoncés, etc. Elles nécessitent néanmoins en contrepartie une connaissance complète et suffisamment contrainte du schéma d'interprétation du domaine d'application. Aussi il est légitime de s'interroger sur la portabilité des approches sélectives à des domaines applicatifs plus riches [Hirschman, 1998]. La question est de savoir si une analyse essentiellement guidée par l'application est en mesure d'apporter des solutions convenables pour le développement d'applications plus complexes c'est-à-dire susceptibles d'entraîner une ambiguïté sémantique supérieure et des difficultés supplémentaires dans le traitement des phénomènes précédemment cités.

Si ce problème a été soulevé [Zue, 1997; Hirschman, 1998; Seneff, 1998], peu de travaux ont permis d'y apporter des éléments de réponse. Les techniques utilisées par les systèmes portant sur des domaines moins restreints sont toutefois révélatrices des besoins de l'analyse. Dans le projet d'interaction médiatisée VERMOBIL par exemple [Alexandersson, Reithinger, et Maier, 1997; Kipp, Alexandersson, et Reithinger, 1999], la richesse du domaine nécessite la mise en œuvre en concurrence de plusieurs techniques d'analyse : une analyse linguistique détaillée fondée sur le formalisme HPSG [Pollard et Sag, 1994], une analyse reposant sur une grammaire hors-contexte probabilisée, une analyse partielle en *chunks*<sup>20</sup> [Abney, 1991], et un décodeur d'actes de dialogue modélisé

<sup>20</sup>Segments minimaux non récursifs (cf. chapitre 5).

par modèles de Markov cachés. Chacune d'entre elles permet d'atteindre un compromis différent entre robustesse et précision de l'analyse. Elles produisent des représentations sémantiques dans le même formalisme, qui peuvent ainsi être combinées [Ruland et al., 1998].

Dans le domaine de la génération automatique de résumé de dialogues oraux informels c'est-à-dire non orientés par la tâche, [Zechner et Waibel, 1998; Zechner, 1998] mettent en œuvre un étiquetage grammatical, une analyse en *chunks* et des techniques spécifiques correctives pour les perturbations de l'oral. Ces techniques sont fondées exclusivement sur des considérations grammaticales et syntaxiques. Cette approche totalement indépendante de l'application a donné de meilleurs résultats que les méthodes traditionnelles en particulier dans les domaines les moins ciblés [Zechner, 2001].

Dans le système de traduction JANUS, l'accent a été de la même manière porté sur la nécessité de l'utilisation de la syntaxe pour autoriser l'adaptabilité du système à des domaines différents et la couverture des phrases les moins conventionnelles du domaine [Levin et al., 2000]. Il s'agit ici de remédier à l'incomplétude inévitable des corpus d'apprentissage. De nouvelles règles sont en effet obligatoires pour traiter les constructions syntaxiques autorisées pour chaque nouveau concept. La stratégie préconisée est ainsi de tendre au maximum vers une partie de l'analyse indépendante du domaine d'application.

D'autres problèmes risquent de ne plus pouvoir être contournés avec l'augmentation de la couverture des domaines applicatifs. Nous pensons en particulier à la gestion des mots inconnus du système. La contrainte minimale face à ce problème est que la présence de ces mots ne doit pas gêner l'analyse. Les méthodes sélectives stochastiques permettent de répondre de façon intéressante à ce problème. Elles autorisent en général une prédiction de la classe sémantique de ces mots (notamment des noms propres utiles à l'application), reposant sur les éléments manquants des schémas instanciés. Ces techniques ne sont toutefois viables encore une fois que dans des systèmes très ciblés, où le nombre de catégories ambiguës dans lesquelles est possible la classification d'un mot inconnu est peu élevé. Or, les ambiguïtés du choix de la classe d'affectation risquent d'augmenter avec l'enrichissement du domaine. [Zue, 1997] insiste sur la nécessité de détecter ces mots inconnus sur la base d'indices linguistiques ; une solution optimale étant de permettre leur acquisition automatique (incluant leurs propriétés linguistiques). Quelques travaux intéressants ont été expérimentés dans ce domaine [Chung, 2000].

Un saut qualitatif nous semble ainsi essentiel pour répondre aux enjeux d'une communication collaborative. Au vu des quelques observations précédentes, nous pensons que l'extraction du sens ne peut pas être uniquement fondée sur des critères dépendants du domaine applicatif ni reposer uniquement sur une modélisation stochastique ne permettant pas le traitement des phénomènes rares. Nous suivons ainsi l'avis de G. Van Noord :

(Cit.2.1) *“The grammatical approach may become essential as soon as the application is extended in such a way that more complicated grammatical constructions need to be recognized. In that case, simple concept spotting may*

*not be able to correctly process all constructions, whereas the capabilities of the grammatical approach extend much further.” [Noord et al., 1999]*

Précisons toutefois que le dialogue oral restreint à des domaines très ciblés nous semble déjà concerné par ces problèmes, même si ceux-ci y sont davantage masqués.

## 2.3 Objectifs généraux de la thèse

Comme nous l’avons vu, la compréhension hors-contexte des énoncés oraux en situation de dialogue homme-machine se heurte à deux problèmes principaux :

- les phénomènes spécifiques de l’oral spontané,
- et les erreurs éventuelles du module de reconnaissance.

Ces problèmes peuvent être contournés dans le cadre d’une communication homme-machine finalisée restreinte à un domaine applicatif particulier. La recherche de robustesse a ainsi pris le pas sur les limitations éventuelles de méthodes qui, perdant les contraintes de la syntaxe, pourraient ne pas être capables de traiter correctement des constructions linguistiques complexes. Cette thèse s’inscrit dans la problématique des actions de recherche visant à développer des traitements linguistiques robustes pour la communication homme-machine. Sur la base de l’état de l’art du chapitre précédent et des réflexions que nous avons développées dans ce chapitre, nous nous fixons deux objectifs principaux :

- mener une analyse détaillée des énoncés oraux tout en respectant la contrainte de robustesse imposée. Nous espérons montrer l’intérêt d’une telle approche pour une interaction homme-machine qui ne soit plus seulement limitée à une recherche d’information très finalisée. Nous pensons qu’une détection plus fine des buts et attentes de l’usager repose en partie sur la richesse des informations pouvant être extraites dès l’étape de compréhension de l’énoncé. Les requêtes multiples doivent pouvoir être autorisées. L’analyse proposée devra par ailleurs être à même de traiter les constructions syntaxiques les plus fréquentes à l’oral et les phénomènes dus à la spontanéité,
- enfin, répondre en partie à la question de la généralité des méthodes utilisées en compréhension en tendant vers une approche reposant en partie sur des connaissances indépendantes du domaine d’application.

Notre laboratoire ne disposant pas de module de reconnaissance, nous ne travaillerons pas dans le cadre de cette thèse sur un graphe de mots hypothèses mais sur un seul énoncé. Comme nous l’avons vu, cette limitation n’est pas satisfaisante. Cette contrainte nous a toutefois permis dans un premier temps de pouvoir évaluer les capacités de l’analyse proposée. Nous évaluerons la pertinence de l’approche en prenant en compte des variations dues au module de reconnaissance au chapitre 7 § 7.2.3 (tests à partir d’un système de dictée vocale) et discuterons des perspectives d’intégration pour le traitement d’un graphe de mots.



De la même manière, nous nous limiterons dans un premier temps à la compréhension hors-contexte des énoncés. Les problèmes liés à l'interprétation en contexte et à la référence ne doivent toutefois pas être ignorés. Le développement de solutions adéquates à ces problèmes complexes est cependant en lui-même un sujet de recherche à part entière auquel nous ne prétendons pas répondre complètement dans le cadre de cette thèse. Notre objectif se limitera dans ce domaine à tenter d'extraire de l'énoncé les représentations sémantiques les plus informatives possibles pour la gestion du dialogue et l'initiation de sous-dialogues pertinents. Nous veillerons de plus à ne pas chercher à interpréter abusivement les intentions du locuteur.

Afin d'assurer un maximum de robustesse au système global tout en offrant à l'utilisateur une interaction la plus naturelle possible, nous avons choisi d'adopter une approche *centrée utilisateur*. Il nous semble en effet impératif de chercher à caractériser le plus précisément possible les différentes constructions syntaxiques et phénomènes spécifiques utilisés à l'oral ainsi que les phénomènes dus à la spontanéité de l'interaction, trop souvent considérés uniquement comme des "déviations agrammaticales" par rapport à l'écrit.

Nous avons ainsi choisi de nous appuyer, pour le prototypage de notre système et nous guider dans le choix de la stratégie à adopter, sur un examen préliminaire des études descriptives de l'oral et la linguistique de corpus. Dans le chapitre suivant, nous présentons les travaux sur lesquels nous nous appuyons, les études de corpus que nous avons réalisées et les principales conclusions que nous avons pu en dégager.



## Deuxième partie

Traitements linguistiques robustes  
pour la compréhension automatique  
de la parole – le système ROMUS.



## Chapitre 3

# L’oral – De l’analyse des usages à l’architecture du système proposé

Ce chapitre retrace la démarche que nous avons adoptée pour le prototypage de notre système. Afin de guider nos choix nous avons orienté nos recherches dans deux directions complémentaires :

- l’examen des études descriptives de l’oral. Nous considérons en effet que l’analyse automatique des énoncés oraux doit passer par une prise en compte explicite des spécificités de l’oral. Nous verrons en particulier que les phénomènes spécifiques de l’oral n’impliquent pas forcément de l’aléatoire mais présentent des régularités sur lesquelles une analyse automatique peut utilement reposer.
- la réalisation et l’exploitation de quelques études de corpus oraux. Ces études ont deux objectifs principaux. Il s’agit d’une part d’évaluer les fréquences relatives des principaux phénomènes en fonction de la situation d’interaction envisagée et du domaine applicatif concerné. D’autre part, il s’agit de caractériser, quantitativement et qualitativement, la réalisation de quelques uns de ces phénomènes afin d’examiner leurs conséquences en matière de traitement automatique.

### 3.1 Études descriptives de l’oral

Les linguistes ont peu étudié la langue parlée [Gadet, 1989]. Dans une démarche rationaliste chomskyenne, des modèles grammaticaux sophistiqués ont été produits pour étudier l’écrit standard, en rejetant toutes ses variations, qu’elles soient dues à des facteurs internes à la langue ou extralinguistiques<sup>1</sup> (variations sociales, régionales, situationnelles, etc. [Gadet, 1989]).

(Cit.3.1) *“la langue parlée, celle qui paraît spontanée, familière, est perçue comme étant sans grammaire, sans règles, car ces règles, acquises tôt dans*

---

<sup>1</sup>Nous nous intéressons ici simplement aux phénomènes linguistiques en situation d’interaction homme-machine finalisée, sans prendre en compte ceux dus aux variations sociales ou régionales.

*la vie, sont suffisamment inconscientes et communes pour devoir échapper à la perception réflexive*” [Kerleroux et Gadet, 1998]

C'est ainsi que les phénomènes qui ressortent de façon saillante des énoncés oraux et absents de l'écrit (du moins non spontanés<sup>2</sup>) ont été “normalisés” (ou parfois caricaturés) par les grammairiens. Il s'agissait en d'autres termes de caractériser les énoncés perçus comme corrects plutôt que les énoncés effectivement utilisés. Pour s'intéresser à la syntaxe de la langue orale, il convient de se fonder sur les usages réels des locuteurs et non sur des critères *a priori* de correction. Les études dont nous nous inspirons se caractérisent ainsi avant tout par un respect des données empiriques et un rejet des jugements d'acceptabilité.

### 3.1.1 Régularités des productions orales

Comme nous l'avons vu au paragraphe 1.2.1.1, le langage oral se caractérise par deux types distincts de phénomènes : ceux induits par le caractère spontané de l'interaction et l'ensemble des procédés utilisés avec régularité répondant le plus souvent à un principe d'économie ou d'engagement dans l'interaction. Nous ne reviendrons pas dans ce paragraphe sur la description des procédés réguliers dont plusieurs études linguistiques [Gadet, 1989; Blanche-Benveniste, 1990] nous donnent une connaissance approfondie. Ils nous intéresseront toutefois dans la suite de chapitre du point de vue de leur importance relative dans les productions dialogiques. Pour mémoire, nous avons relevé les procédés suivants :

- les extractions et d'une manière générale les phénomènes entraînant une modification de l'ordre syntagmatique standard,
- les énumérations,
- les ellipses,
- les anaphores,
- le phénomène de parataxe<sup>3</sup>.

La difficulté de la prise en compte des autres phénomènes, ceux dus à la spontanéité de l'interaction, dans un modèle grammatical vient de leur non prédictabilité. Or, l'analyse de ces phénomènes révèle en fait des régularités spécifiques. Pour nous en convaincre, nous nous appuyerons sur les travaux descriptifs de [Blanche-Benveniste, 1990].

---

<sup>2</sup>Blanche-Benveniste [Blanche-Benveniste, 1990] fait explicitement le rapprochement entre deux matériaux : celui des transcriptions de l'oral et celui des brouillons de l'écrit. Comme nous l'avons déjà souligné, il n'existe pas de frontière claire entre oral et écrit, cette transition devant plutôt être appréhendée sous la forme d'un continuum de genres [Biber et al., 1999; Blanche-Benveniste et Bilger, 1999].

<sup>3</sup>Ce phénomène désigne la présence au sein d'un même énoncé de plusieurs propositions non liées explicitement par un connecteur. Les énoncés oraux sont ainsi agencés selon une juxtaposition de segments sans forcément de mots de liaison entre eux [Bellenger, 1979] ; ceux-ci étant compensés le plus souvent par la prosodie.

Les travaux du GARS<sup>4</sup>, autour de C. Blanche-Benveniste, ont envisagé l'analyse des énoncés oraux en deux niveaux. L'analyse des productions orales, dans lesquelles on retrouve la trace même de leur élaboration, est tout d'abord envisagée selon les axes *syntagmatique* et *paradigmatique*. L'axe syntagmatique correspond à une analyse grammaticale des syntagmes dans leur successivité. Dans l'exemple suivant on peut par exemple isoler cinq syntagmes :

(E.3.1) *à ce moment là / vous / traversez / la rue / à pieds* [corpus Murol.1]

Le niveau syntaxique se fonde chez [Blanche-Benveniste, 1990] sur les catégories grammaticales classiques comme le verbe, le nom ou l'adjectif. Ce déroulement selon l'axe syntagmatique peut être brisé en plusieurs endroits à l'oral. On explique alors ces enchaînements particuliers par des variations se déroulant sur l'axe paradigmatique<sup>5</sup>. L'exemple suivant illustre l'exploitation de ces deux axes de développement. La lecture s'effectue de gauche à droite et de haut en bas, l'axe syntagmatique est représenté horizontalement et l'axe paradigmatique verticalement :

(E.3.2)

<i>à ce moment là</i>	<i>vous</i>	
	<i>vous pouvez</i>	
	<i>eh</i>	<i>vous pouvez</i>
		<i>vous traversez la rue à pieds</i> [corpus Murol.1]

Ces allées et venues incessantes sur les axes syntagmatique et paradigmatique permettent de mieux appréhender l'organisation des phénomènes "perturbateurs" (répétitions, reprises, etc.). Nous reproduisons sous cette représentation les exemples (E.1.5, E.1.7, et E.1.10) que nous avons pris au paragraphe 1.2.1.1 dans lesquels nous avons respectivement identifié deux phénomènes d'enrichissement lexical et une longue incise (cf. page 11).

(E.1.5) *tu places une porte*  
*une double porte au milieu en bas de la pièce*

(E.1.7) *vous tombez sur une petite avenue*  
*un petit chemin*  
*un petit chemin de graviers et vous tombez sur le musée*

(E.1.10) *alors vous prenez* la  
*l'avenue de Grande Bretagne [ incise ]*  
*donc l'avenue de Grande Bretagne*

L'exploitation paradigmatique se traduit par une instanciation de plusieurs syntagmes à la même position syntaxique (place dans l'ordre syntagmatique) dans l'énoncé. Il est également typique de la réalisation d'un certain nombre de procédés dont l'énumération (exemple E.1.18 du paragraphe 1.2.1.1) :

<sup>4</sup>Groupe Aixois de Recherche en Syntaxe. Ces travaux sont poursuivis actuellement par l'équipe DELIC -Description Linguistique Informatisée sur Corpus- de l'université de Provence.

<sup>5</sup>L'axe paradigmatique permet d'explorer des rapports associatifs sur la base de relations diverses (d'après la définition de Saussure) [Moeschler et Auchlin, 1997]

(E.1.18)          *oui il y a des réductions*          *famille nombreuse*  
    *étudiant*  
    *et carte vermeille*

Sur l’axe paradigmatic se manifestent ainsi divers phénomènes allant du “bre-douillage” à l’énumération. Les hésitations, induites le plus souvent par une recherche de dénomination, peuvent de cette manière être interprétées comme un piétinement sur une seule et même place syntaxique. La possibilité de parcourir d’avant en arrière l’axe syntagmatic permet de plus aux locuteurs de corriger, de préciser ou de remédier à une anticipation erronée d’un élément particulier de l’énoncé<sup>6</sup>, mais aussi de revenir sur des portions assez longues de leur discours.

Enfin, le deuxième niveau d’analyse proposé par [Blanche-Benveniste, 1990], celui de la *macro-syntaxe*, fait appel à des critères sémantiques et prosodiques pour la combinaison des syntagmes identifiés<sup>7</sup>.

Nous retiendrons de ce modèle un point essentiel : si le déroulement syntagmatic peut être brisé ou interrompu à un endroit quelconque de l’énoncé, il apparaît que la reprise s’effectue toujours **au début du syntagme** que le locuteur souhaite corriger ou enrichir, de façon à rétablir l’axe syntagmatic. Ce n’est ainsi donc pas par hasard si l’une des catégories de mots les plus souvent répétés soit par exemple les déterminants [Candéa, 2000]<sup>8</sup>. Soulignons également que la présence d’une longue incise comme celle de l’exemple (E.1.10), entraîne très souvent la répétition du syntagme prononcé avant digression.

Ce modèle ne rend compte que d’un aspect limité des régularités des énoncés oraux. Cette régularité minimale mise en exergue est toutefois importante pour le traitement de ce qu’on appellera dans la suite de ce document les *inattendus structuraux*. Ce terme permet de rendre compte à la fois de la non prédictibilité de l’apparition de ces phénomènes au cours de la production orale<sup>9</sup>, et également de la structure sous-jacente à leur réalisation. Il nous permettra par ailleurs de pouvoir différencier sans ambiguïté ces

<sup>6</sup>[Blanche-Benveniste, 1990] souligne que les locuteurs tiennent en réserve les syntagmes qu’ils viennent de dire aussi bien que ceux qu’ils projettent de dire. Les anticipations portent sur des suites de morphèmes ou sur des syntagmes entiers.

<sup>7</sup>Une disposition des paradigmes des syntagmes dans des *grilles* spécifiques de l’oral, permet de retrouver des constructions complètes et de définir les rapports de dépendance entre les différents syntagmes.

<sup>8</sup>[Candéa, 2000] et [Henry, 2002] attestent de fréquences respectivement 6 et 5 fois plus importantes des répétitions des “mots-outils” participant simplement à la structuration de la langue (déterminants, prépositions, pronoms personnels, etc.) en comparaison avec celles des “mots pleins” (adjectifs, noms, etc.). Dans cette dernière catégorie, ce sont les adverbes, puis les adjectifs qui sont majoritairement repris [Henry, 2002].

<sup>9</sup>Nous faisons ici abstraction des causes pouvant entraîner ces phénomènes : hésitations devant des mots peu fréquents ou peu prévisibles, recherche de dénomination, nécessité de clarification etc. D’un point de vue purement descriptif, ces interruptions ou retours en arrière dans l’axe syntagmatic sont susceptibles d’intervenir à tout moment dans la production du discours.



phénomènes des procédés réguliers fréquemment utilisés à l'oral dont nous avons déjà parlé.

### 3.1.2 Apports pour le traitement automatique

Les travaux que nous venons d'évoquer semblent suggérer une première analyse superficielle d'identification grammaticale des syntagmes présents dans l'énoncé. Les éventuelles variations sur l'axe paradigmatique seraient ensuite prises en compte de façon à rétablir l'axe syntagmatique en établissant les relations existant entre ces syntagmes. [Blanche-Benveniste, 1990] souligne que ces aspects ne relèvent pas de la compétence d'un niveau syntaxique. Ils font appel à des informations de nature sémantique (et prosodique) pour être en mesure de rendre compte des relations de dépendance complexes observées à l'oral.

En ce qui concerne la réalisation des inattendus structuraux, cette description selon les axes syntagmatique et paradigmatique peut être rapprochée de leur modélisation en terme de *reparandum*, *zone d'édition* et éléments repris [Levelt, 1983; Shriberg, 1994]. Le modèle évoqué, fondé sur la reprise d'unités syntagmatiques, permet toutefois de notre point de vue d'expliquer quelques faiblesses des approches de prétraitement par *patterns* [Bear et al., 1992]. Nous avons déjà souligné l'incapacité de ces approches à assurer un filtrage correctif approprié de l'énoncé sans recourir à des traitements linguistiques de haut niveau. La détection elle-même des inattendus structuraux par ces techniques nous semble sujette à caution. Les *patterns* utilisés dans ces approches reposent en effet sur des correspondances grammaticales au niveau des mots et non au niveau du syntagme. Pour éviter une surgénérativité trop importante, la taille prise en compte pour caractériser le *reparandum* et les éléments repris ne doit pas être supérieure à deux mots [Bear et al., 1992]. C'est la raison pour laquelle elles ne peuvent rendre compte sans risques des phénomènes de répétitions portant sur des unités syntagmatiques plus larges comme dans le cas des enrichissements lexicaux par exemple. Ces méthodes ne semblent être ainsi valides que pour la détection de répétitions simples limitées à la reprise d'un ou de deux mots c'est-à-dire le plus souvent des "mots-outils" qui, soit débutent un syntagme, soit en constituent un à eux seuls. Or, même dans ces cas apparemment simples, la surgénérativité n'est pas évitable.

Un exemple peut en être donné à partir de l'étude de corpus menée par [Henry, 2002]. Dans cette étude, la méthode par *patterns* a été utilisée sur un corpus de français parlé de près de 1 million de mots<sup>10</sup> pour le repérage de répétitions brutes c'est-à-dire les répétitions de mots identiques séparés éventuellement par un autre mot. Sur près de 30 000 phénomènes ainsi détectés il apparaît que seulement 50% d'entre eux correspondent réellement à des répétitions "de performance" [Henry, 2002]. Les autres phénomènes sont dus par exemple à des constructions pronominales, à la succession de deux unités syntaxiques différentes ou à des dislocations avec double-marquage que

---

<sup>10</sup>Il s'agit du corpus *Corpaix*, établi par le Groupe Aixois de Recherche en Syntaxe, actuellement DELIC, de l'université de Provence.

seule une analyse linguistique plus riche serait en mesure d’identifier<sup>11</sup>. Les exemples (E.3.3) et (E.3.4) tirés de [Henry, 2002] attestent de ces deux derniers cas :

(E.3.3) “*alors la danse fait beaucoup pour nous nous aimons danser*”  
[corpus Corpaix, APOSTROPHE]

(E.3.4) “(…) *nous Marseillais nous resterons attachés particulièrement à nos marins-pompiers*” [corpus Corpaix, BIRNIE]

La régularité dans la manière dont les locuteurs se reprennent est cependant utilisée plus explicitement dans les travaux de [Heeman et Allen, 2001]. Dans leur approche, les auteurs introduisent, pour guider la prédiction du premier mot repris, de plus fortes probabilités *a priori* sur les mots qui débutent les groupes grammaticaux étant donnée la nature grammaticale du *reparandum* supposé. Ainsi dans l’exemple suivant :

(E.3.5) 
$$\begin{array}{ccc} the\ two\ boxcars\ of\ orange\ juice\ should & & \dots \\ & \uparrow & \\ & ip & \end{array}$$

l’apparition à la suite du point d’interruption d’un déterminant comme *the*, du modal *should* ou la reprise de *of*<sup>12</sup> est considérée comme plus probable que la reprise d’un nom par exemple. Pour la correction de ces phénomènes sont par ailleurs prises en compte des correspondances grammaticales entre le *reparandum* et les éléments de réparation. Ces travaux visent toutefois à l’amélioration des modèles de langage *N-gram* et ces informations sont utilisées conjointement avec plusieurs autres distributions de probabilités (aspects prosodiques, lexicaux), parfois contradictoires et dépendant de la représentativité du corpus d’apprentissage. La similarité syntaxique entre le *reparandum* et le syntagme repris n’est ainsi pas complètement exploitée. Si les travaux de [Heeman et Allen, 2001] sont les plus complets en ce qui concerne la modélisation des inattendus structuraux, nous pensons à l’image de [Brill et al., 1998] et de [Noord et al., 1999] que la prise en compte de ces régularités doit être envisagée dans une étape de post-traitement à la reconnaissance plutôt qu’être intégrée au modèle de langage.

## 3.2 Linguistique de corpus

Nos travaux de thèse ont été l’occasion d’orienter les travaux de notre équipe de recherche en matière d’études qualitatives et quantitatives de corpus oraux. En ce qui nous concerne, le but principal de ces études est de fournir une caractérisation utile au prototypage de notre système de compréhension. Nous résumons ici les études auxquelles nous avons participé dont on pourra trouver le détail dans [Antoine et Goulian, 1999; Letellier-Zarshenas et al., 1999; Antoine et Goulian, 2001a, 2001b]. Une version

<sup>11</sup>Dans cette étude, les catégories grammaticales des mots ne sont pas prises en compte à la différence de l’approche de [Bear et al., 1992]. Toutefois, cette information ne pourrait suffire à elle seule pour éviter les problèmes de surgénérativité sur les exemples évoqués.

<sup>12</sup>D’après les travaux de [Levelt, 1983] et les études de corpus menées par [Heeman et Allen, 2001], l’apparition de *of* reste possible mais une probabilité plus faible par rapport à *the* ou *should* lui sera affectée.

courte de [Antoine et Goulian, 2001a] regroupant la méthodologie adoptée et le détail des principaux résultats pourra être consultée en annexe A. Nous insistons plus particulièrement sur leurs implications dans le cadre qui nous intéresse.

Nous avons déjà souligné la difficulté d’obtenir des échantillons suffisamment représentatifs des énoncés oraux en fonction de la situation d’interaction envisagée. La constitution de vastes corpus pour le dialogue homme-machine est déjà en soi un premier obstacle<sup>13</sup>. Avant d’aborder les études de notre équipe de recherche auxquelles nous avons participé, nous présentons brièvement les différentes stratégies utilisées pour le recueil des corpus oraux. Nous soulignons les intérêts et les limites qu’elles présentent pour l’analyse des usages des locuteurs. Trois méthodologies sont traditionnellement envisagées :

- **l’utilisation de corpus *pilotes de dialogue oral***. Il s’agit de corpus réels homme-homme enregistrés dans une situation proche du domaine applicatif envisagé. Leur exploitation nécessite une étape transcription délicate et coûteuse. Si ces corpus présentent un intérêt indéniable dans l’analyse des usages et des besoins réels de l’application [Caelen, 1997], ils constituent une idéalisation peu opérationnelle pour la conception des systèmes.
- **l’utilisation de corpus réels homme-machine**. La technique de *bootstrap* vise à tirer parti d’un premier prototype, grossier, de système opérationnel utilisé pour le recueil de corpus d’interaction réelle entre un usager humain et ce prototype. Ce corpus servira pour entraîner un second prototype et ainsi de suite. Or, sachant que le locuteur adapte son comportement langagier de façon à se conformer de manière plus ou moins consciente aux limitations du système, on peut s’interroger sur la “naturalité” des constructions langagières recueillies à partir du premier prototype et surtout sur leurs conséquences sur les formulations qui seront acceptées à terme par le système.
- **l’utilisation de corpus *simulés***. La technique du magicien d’Oz permet le recueil de dialogues dans lequel le système est simulé, le plus souvent à l’insu de l’usager, par un utilisateur humain (le *compère*). Cette technique est peut-être celle qui permet d’obtenir les corpus les plus proches des conditions réelles, pouvant contenir des hésitations, des corrections et tenant compte de la modification du comportement langagier. Toutefois elles demandent un travail de conception très fin pour arriver à une simulation réaliste. Elles peuvent ainsi difficilement conduire à une simulation rendant compte de toutes les situations d’interaction [Fraser et Gilbert, 1991].

De notre point de vue, ces trois approches nous apparaissent complémentaires même si, dans la pratique, chaque concepteur semble plutôt privilégier une méthode spécifique.

---

<sup>13</sup>On notera à ce propos qu’alors que le retard du français en matière de ressources linguistiques est déjà très sensible [Véronis, 2000], la situation semble encore plus critique en matière de corpus de dialogue oral. On peut toutefois espérer que le récent projet ASILA – Action Spécifique Interaction Langagière et Apprentissage, <http://www.loria.fr/projets/asila> – du CNRS, qui vise entre autres la collection d’un nombre significatif de corpus de dialogues francophones puisse combler ce retard.

Nous pensons que l'utilisation conjointe de ces approches (corpus pilote pour appréhender le domaine, magicien d'Oz pour la réalisation d'un premier prototype, technique de *bootstrap* pour les améliorations suivantes) est en effet à même de faciliter la conception des systèmes.

Rappelons toutefois que les principales utilisations de ces corpus en CHM concernent l'apprentissage des composants langagiers des systèmes de dialogues (modèle de langage par exemple). Or cet apprentissage suit le plus souvent une procédure (automatique ou non) itérative de *bootstrap*<sup>14</sup>. Cette approche par essai-erreur ne laisse pas de place à une analyse détaillée des usages mis en jeu dans la situation considérée. Ceci explique également que notre connaissance des caractéristiques linguistiques du dialogue oral finalisé reste encore relativement limitée<sup>15</sup> et surtout peu exploitée en CHM.

Les travaux que nous retraçons ici portent ainsi principalement sur des corpus pilotes homme-homme. Ils doivent permettre de cerner les pratiques des utilisateurs qui devraient être modélisées dans le système. Cette idéalisation du dialogue homme-machine par l'interaction humaine a ses limites. L'utilisation de tels corpus nous paraît pourtant essentielle pour tenter de ne pas imposer *a priori* trop de contraintes d'élocution aux usagers et garantir l'utilisabilité du système [Bernsen, Dybkjaer, et Dybkjaer, 1996; Dybkjaer et Bernsen, 2000]. Cette approche nous paraît aussi justifiée que celles consistant à s'en remettre aux capacités de l'être humain, étudiées par [Morel, 1989], à adapter son comportement langagier en situation de dialogue avec une machine.

### 3.2.1 Études de corpus et communication orale homme-machine

Nous avons choisi de concentrer nos travaux sur deux aspects importants. Le premier point concerne l'influence quantitative du contexte d'interaction sur l'utilisation des différents procédés et la présence des inattendus structuraux. Le second point s'intéresse plus particulièrement aux phénomènes linguistiques impliquant une variabilité de l'ordre des mots de l'énoncé. La modification de l'ordre des mots est en effet un problème pour le traitement automatique des énoncés qu'on ne saurait ignorer dans la perspective qui nous intéresse.

#### 3.2.1.1 Influence du contexte interactif

L'objectif de cette analyse préalable des usages vise à mesurer les proportions relatives d'apparition des principaux phénomènes spécifiques de l'oral en fonction des caractéristiques de l'interaction. De telles études concernent d'ailleurs aussi bien l'oral que l'écrit où la distribution, en fréquence, de l'utilisation des constructions syntaxiques est très dépendante du genre du texte concerné [Biber, 1993]. Le choix de notre équipe

---

<sup>14</sup>Cette procédure est indépendante de la nature du corpus (pilote, magicien d'Oz) utilisé pour la constitution du premier prototype.

<sup>15</sup>Les principaux apports dans ce domaine concernant le français proviennent des études différentielles de [Morel, 1989] et des travaux réalisés dans le cadre du projet DALI [Sabah, 1997].

de recherche dans ce domaine a été de confronter trois situations interactives différentes issues de deux corpus homme-homme (Air-France et Murol) et d'un corpus recueilli par une expérience de faux magicien d'Oz<sup>16</sup> (Levelt). Le tableau 3.1 rappelle brièvement les caractéristiques de ces corpus.

Corpus	Air-France	Murol	Levelt
Nb de dialogues	103	9	5
Nb de tours de parole	5149	1078	706
Nb de mots	49703	13500	–
Tâche	Conversations téléphoniques entre un centre de réservation aérienne et différents clients	Conversations téléphoniques <i>simulées</i> entre un syndicat d'initiative et un touriste	Conception architecturale assistée par ordinateur
Spécificités	Les interlocuteurs s'interrompent et peuvent parler en même temps.  Le dialogue reste cependant contenu et policé (réserve de la part de l'hôtesse).	Apparition de situations de <i>négociation</i> favorisés par les scénarii.  Dialogue peu coopératif : le client cherche à se renseigner sur des activités sportives ; l'hôtesse a pour mission d'encourager des visites culturelles de la ville.  Plusieurs sous-tâches spécifiques doivent être expérimentées.	Le système, simulé par magicien d'Oz se contente de réagir aux directives du locuteur.

TAB. 3.1 – Linguistique de corpus et CHM – Description des corpus étudiés

D'un point de vue de l'interaction, ils peuvent se distinguer selon plusieurs critères. Nous en avons retenus trois, présentés dans le tableau 3.2.

En l'absence d'une unité de segmentation des productions orales clairement définie [Gadet, 1989; Blanche-Benveniste, 1990]<sup>17</sup>, et dans l'optique d'un traitement automatique, nous avons choisi le tour de parole comme unité de comptabilisation, même si ce choix peut être débattu [Colineau, 1997]. Dans la suite, nous parlerons indifféremment

<sup>16</sup>Rappelons que l'on distingue les expériences du vrai magicien, où le locuteur n'est pas au courant de l'existence du compère, des situations de faux magicien où il est précisé au locuteur qu'un être humain remplace la machine.

<sup>17</sup>[Blanche-Benveniste, 1990] et [Gadet, 1989] rejettent toute idée de segmentation des discours oraux en unités phrastiques. [Gadet, 1989] suggère de se fonder sur la notion d'énoncé sans toutefois pouvoir en apporter une définition.

	Degré de finalisation	Degré d'interactivité	Répétitivité
<b>Air-France</b>	très élevé	élevé	très faible
<b>Murol</b>	modéré	très élevé	faible
<b>Levelt</b>	très élevé	très faible	très élevée

TAB. 3.2 – Linguistique de corpus et CHM – Caractérisation des corpus étudiés en fonction du contexte interactif

d'énoncé ou de tour de parole. Les fréquences d'apparition auxquelles nous faisons référence sont calculées en terme de pourcentage d'énoncés porteurs du phénomène considéré. La dispersion des données est analysée en terme d'écart type de classes, chaque classe correspondant à un dialogue. Selon les cas de figure, les phénomènes observés ont été détectés semi-automatiquement ou entièrement manuellement. Différents experts ont procédé par la validation croisée des phénomènes détectés et comptabilisés. Les premiers travaux effectués ont porté principalement sur les inattendus structuraux (répétitions et corrections), les ellipses et les antépositions. Ces premières études ont uniquement concerné un sous ensemble limité de chacun des corpus étudiés. On pourra en trouver le détail dans [Antoine et Goulian, 1999] et [Letellier-Zarshenas et al., 1999]. Nous nous contentons ici de regrouper les quelques conclusions que nous en avons tirées en fonction des critères interactifs retenus.

- **Degré de finalisation.** Le degré de finalisation des différents corpus a avant tout un impact sur les proportions des ellipses. On a pu ainsi observer un taux relativement élevé d'ellipses dans les corpus Air-France et Levelt (resp. 27,6% et 18,7%) par opposition au corpus Murol (9,4%). L'analyse en écart type a montré que cette différence est statistiquement significative. Dans le cas du corpus Air-France, la tâche de réservation est très finalisée et se résume pour le client à l'obtention d'un nombre de paramètres limités (prix et horaires des vols). La tâche est simple et les intentions des locuteurs sont claires, ce qui facilite l'usage des structures elliptiques. En outre, l'hôtesse reprend souvent une partie des informations afin de meubler l'attente ; ce qui peut expliquer la différence observée entre ce corpus et le corpus Levelt dans lequel le système se contente de réagir aux directives de l'utilisateur. En revanche, le corpus Murol, moins finalisé, donne lieu à des échanges sensiblement plus riches et variés. Les intentions des locuteurs varient fréquemment, ce qui explique ce taux moins élevé. L'influence du degré de finalisation par rapport aux autres facteurs est par ailleurs assez difficile à apprécier sur les autres phénomènes. Ce critère ne semble ainsi pas pouvoir justifier une quelconque restriction des phénomènes langagiers à prendre en compte, même sur des domaines restreints.
- **Répétitivité.** La nature très opérative et répétitive de la tâche du corpus Levelt a également une influence notable sur les proportions d'ellipses. On observe un usage croissant des structures elliptiques au fil du discours, l'utilisateur ne préci-

sant finalement que les paramètres qui varient d'une action à l'autre. Par ailleurs, la redondance des actions réalisables par le système (ajout d'un nouvel élément de dessin, repositionnement, etc.) incite également l'utilisateur à mettre en avant, d'une manière éventuellement elliptique, les paramètres de ses commandes. On retrouve la motivation principale de l'antéposition et du double-marquage, à savoir l'insistance sur une information précise. On notera cependant que l'extrême répétitivité n'implique pas un recours systématique à ces procédés. Une diversité très importante des usages a été en effet observée selon les locuteurs (une fréquence minimale d'antéposition de 5,8% contre une fréquence maximale de 69,4%). On retrouve ces résultats chez [Antoine, 1994] qui note de plus un doublement de la fréquence des énumérations dans ce corpus avec également une forte disparité selon les locuteurs.

- **Degré d'interactivité.** Contrairement aux deux facteurs précédents, le degré d'interactivité influe principalement sur la proportion des inattendus structuraux. Un taux significativement plus élevé de ces phénomènes est en effet observé dans le corpus Murol par opposition aux deux autres (35% contre 14% pour Air-France et 9% pour Levelt). La forte interactivité du corpus Murol se traduit en effet par de fréquentes interruptions. Le caractère peu coopératif du scénario, de même que la richesse de la tâche sont la cause de cette interactivité. Il en résulte des échanges très spontanés qui sont accompagnés de multiples répétitions et corrections. Ces résultats<sup>18</sup> recourent ceux de [Heeman et Allen, 2001] qui attestent, sur le corpus TRAINS (corpus de dialogues finalisés homme-homme composés de 5535 tours de parole<sup>19</sup> et impliquant 20 scénarii de négociation différents, un taux de 54% des tours de parole d'au moins 10 mots contenant au minimum une correction. A l'opposé le corpus Air-France se traduit par une interactivité plus faible en raison de la nature plus contenue du dialogue, alors qu'il n'y a pas d'interaction dans le corpus Levelt. Il n'en reste pas moins que l'élocution reste spontanée et donc source d'erreurs et d'hésitation, la fréquence d'apparition de ces phénomènes ne pouvant pas être négligée.

Ces conclusions sont sans doute largement intuitives et prévisibles. En dépit du caractère limité des études que nous venons de présenter (provenant essentiellement de la taille des corpus pris en compte) nous insistons néanmoins sur l'importance de cette démarche dans le cadre du prototypage des systèmes, point sur lequel insistent également désormais certains chercheurs qui avaient jusque là ignoré en partie ces aspects [Zue, 1997]. Nous avons ainsi souhaité aller plus loin dans notre démarche en nous intéressant aux phénomènes d'extractions entraînant une modification de l'ordre des mots dans l'énoncé. Ces phénomènes sont importants pour le traitement automatique et ont été peu étudiés (du moins en terme de fréquence d'apparition dans les productions orales) par rapport aux phénomènes précédemment évoqués. Nous développons cette étude plus complète au paragraphe suivant.

---

<sup>18</sup>Lors de cette première étude, seulement 1639 tours de parole du corpus Murol ont été considérés.

<sup>19</sup>Les énoncés de clôture du dialogue ont été ignorés [Heeman et Allen, 2001].

Remarquons enfin que ces travaux ont été complétés par une étude qualitative portant sur la caractérisation des répétitions et des reprises. Ces inattendus peuvent en effet être classés en deux catégories selon la présence ou l'absence de mots entre le répétable et le répété [Blanche-Benveniste, 1990]. Sur le corpus Murol, nous avons ainsi recensé 4 fois plus de répétitions et de reprises de syntagmes *non marquées* [Goulian, 1998] que d'inattendus *marqués*. L'étude de ces derniers phénomènes nous a permis de répertorier un ensemble de 14 marques lexicales régulières allant de l'interjection aux appuis du discours et s'apparentant, pour certains d'entre eux (*en fait, enfin, alors*) aux *connecteurs* de rectification, de reformulation et d'explication introduits dans [Colineau, 1997].

### 3.2.1.2 Variabilité de l'ordre des mots en dialogue oral finalisé

La question de l'ordre variable des mots et son corollaire, celui des dépendances discontinues est un sujet central de débat concernant à la fois la théorie linguistique et le traitement automatique des langues. Ainsi, elle a constitué à la suite des travaux de [Tesnière, 1959] un des arguments majeurs des grammaires de dépendance face aux grammaires de constituants [Covington, 1994; Hudson, 2000]. Le traitement des structures discontinues constitue de même un des fondements d'un formalisme tel que les grammaires syntagmatiques dirigées par la tête [Pollard et Sag, 1994] et a été également étudié dans le cadre des grammaires d'arbres adjoints [Rambow et Joshi, 1994].

D'une manière générale, on peut distinguer deux niveaux de variabilité de l'ordre des mots dans l'énoncé [Holan, Kubon, Oliva, et Plátek, 2000] :

- une variabilité *faible* autorisant une position variable des constituants (continus)<sup>20</sup> de l'énoncé. Ces mouvements n'induisent aucune discontinuité dans la structure de dépendance de l'énoncé :

(E.3.6) *bon sinon pour Rio Sao Paulo je pense qu'il y a pas mal de vols* [corpus Air-France]

- et une variabilité *forte* qui se traduit par un relâchement des contraintes de continuité autorisant la production d'énoncés qualifiés de *non projectifs*<sup>21</sup>. Cette variabilité forte peut être observée à l'oral en français comme le montre l'exemple suivant où la présence de l'adverbe temporel *maintenant* casse la continuité entre la relative *qui est nouveau* de son antécédent :

(E.3.7) *vous savez on a un tarif encore plus intéressant sur Londres maintenant qui est nouveau* [corpus Air-France]

<sup>20</sup>Se reporter à [Holan et al., 2000] pour une définition non syntagmatique de cette variabilité faible (*freedom of constituent order within a continuous head domain*).

<sup>21</sup>C'est par exemple le cas de l'extraction des *wh-questions* en anglais : "*who do you think that Mary claims that Sarah likes*" [Hudson, 2000] que l'on retrouve par exemple en français sur des exemples de relativisation à dépendance non bornée : "*la voiture avec laquelle j'ai eu un accident*", "*la dame sur le mari de laquelle je pense que tu peux compter*" [Kahane, 2000a].



Cette distinction entre variabilité forte et faible semble *a priori* primordiale. On sait que la variabilité forte s'observe en premier lieu sur les langages dits à ordre variable (russe, tchèque par exemple) tandis qu'à l'opposé, les langues à ordre fixe sont essentiellement concernées par la variabilité faible. Le français est considéré comme une langue à ordre fixe [Covington, 1994]. À notre connaissance, cette conclusion ne concerne cependant que le français écrit. Il nous semble donc nécessaire de procéder à une analyse détaillée de la variabilité sur le dialogue oral spontané avant de lui appliquer les conclusions portant sur le français écrit. Plusieurs études linguistiques [Blanche-Benveniste, 1990; Gadet, 1989] nous donnent une connaissance approfondie sur la variabilité de l'ordre des mots en français parlé. Les différents procédés qui président aux phénomènes d'extraction sont ainsi bien identifiés. À de rares exceptions, ces études ne présentent cependant qu'un caractère descriptif et explicatif, sans rendre compte de l'importance relative de ces différents phénomènes dans les productions orales. En ce qui concerne le traitement automatique des énoncés, la quantification précise des phénomènes d'extraction nous intéresse selon deux aspects :

1. D'une part, les systèmes de dialogue oral actuels n'ont généralement qu'une vision très limitée de la structure de l'énoncé sur lequel ils travaillent. En ce sens, la question de la projectivité des énoncés oraux ne s'est jamais réellement posée. Il nous semble ainsi important d'évaluer la proportion d'énoncés discontinus dans les productions orales afin de nous renseigner sur les formalismes à utiliser.
2. D'autre part, on note que la variabilité faible pose parfois problème à des systèmes sélectifs utilisés dans un cadre applicatif très restreint [Minker, 1999], travaillant sur des énoncés oraux présentant une ambiguïté structurelle limitée<sup>22</sup>. Or, la généralisation du dialogue oral à des cadres applicatifs plus riches risque de se traduire par une augmentation sensible de cette ambiguïté<sup>23</sup>. Dans ce contexte, il est ainsi légitime de s'interroger sur la part réelle de ces phénomènes dans les productions orales dialogiques pour évaluer les éventuels problèmes posés par la variabilité faible que devront résoudre les modèles développés. La considération de la structure des énoncés et de l'ordre des mots ou des syntagmes s'impose en particulier pour l'analyse fine que nous préconisons.

On pourra trouver le détail de l'étude effectuée en annexe A. Nous avons caractérisé chacun des phénomènes selon plusieurs dimensions : le sens de l'extraction (antéposition ou postposition), le type de procédé mis en jeu (inversions, double-marquage, présentatifs ...), la fonction de l'élément détaché (sujets, arguments gouvernés par le verbe, modificateurs<sup>24</sup> et associés<sup>25</sup>) et leurs conséquences en matière de non-projectivité. Chacune de ces dimensions est illustrée en annexe A. Nous avons procédé pour la comp-

<sup>22</sup>Nous ne considérons pas ici l'ambiguïté due aux résultats issus de la reconnaissance de la parole.

<sup>23</sup>Notons par exemple l'augmentation très sensible de la perplexité des modèles de langages entre le domaine ATIS (renseignement aérien) et l'application *Broadcast News* (informations générales). Nous reviendrons sur ce point dans le chapitre 4 par rapport au domaine d'application que nous avons choisi.

<sup>24</sup>Il s'agit des éléments gouvernés par le verbe mais ne pouvant être considérés comme appartenant à la valence de ce dernier.

<sup>25</sup>Les associés peuvent être définis comme des compléments de phrase [Blanche-Benveniste, 1990].

tabilisation de ces phénomènes par validation croisée sur les deux corpus pilotes Murol et Air-France. Une analyse systématique de la *validité statistique* des résultats obtenus a été par ailleurs menée (test de la variance de Fischer, test de la moyenne de Student [Dudewicz et Mishra, 1988] et test non paramétrique de Wilcoxon-Mann-Whitney dans les cas où la validité était moins tranchée).

Nous avons abouti aux conclusions suivantes :

- Au vu des observations réalisées, la question de la variabilité forte de l'ordre des mots ne semble pas centrale dans le cadre du dialogue oral homme-machine : les extractions à structure non projective ne concernent en effet qu'un nombre très marginal d'énoncés. Au total, les détachements conduisant à des énoncés non projectifs représentent moins de 0.5% des énoncés de nos corpus, ces résultats restant remarquablement stables d'un corpus à l'autre.
- La fréquence d'occurrence moyenne des extractions sur nos deux corpus montre que ce phénomène peut-être très répandu à l'oral. Dans le corpus Murol, un quart des énoncés comprend un élément détaché. La variabilité faible de l'ordre des mots est ainsi un problème que l'on ne saurait ignorer en CHM.
- L'observation de la répartition des extractions suivant la fonction syntaxique de l'élément déplacé est toutefois intéressante. On remarque que les extractions concernent avant tout les éléments libres de l'énoncé (modificateurs et associés) et la fonction sujet dont le déplacement est majoritairement effectué par antéposition avec double-marquage. Ces déplacements n'altèrent en rien l'ordre Sujet-Verbe-Objet standard : la grande variabilité d'ordonnement observé en français oral ne saurait s'affranchir aisément de certaines contraintes fondamentales.
- Nos travaux suggèrent une influence très limitée de la tâche sur les phénomènes d'extraction. Le degré de finalisation de la tâche peut avoir cependant un impact indirect en nécessitant une prise en compte fine de l'ordonnement linéaire dans le traitement de l'énoncé.

### 3.2.2 Apports pour le traitement automatique

Ce genre d'étude nécessite un travail très conséquent de dépouillage et d'analyse de corpus. Pour donner toute sa mesure (analyse de l'influence de la tâche et d'autres facteurs de variabilité), il nécessite en outre le croisement de résultats issus de multiples corpus, dans la perspective d'une analyse factorielle de données. De notre point de vue, les quelques exemples d'analyse des usages que nous avons présentés justifient cependant cet investissement par les enseignements qu'il peut apporter :

1. Nos observations suggèrent d'une part une influence très limitée du degré de finalisation de l'application sur l'étendue des phénomènes linguistiques à prendre en compte. Cette influence se traduit uniquement en terme de fréquence plus ou moins importante des procédés utilisés par les locuteurs pour tirer au mieux parti des caractéristiques de la tâche concernée. L'analyse des usages permet ainsi d'avoir une connaissance précise des phénomènes les plus utilisés. La forte dis-

parité observée dans l'utilisation de ces procédés selon les locuteurs souligne par ailleurs que les procédés moins utilisés en moyenne ne peuvent pas pour autant être négligés. Leur prise en compte s'avère indispensable, quel que soit le domaine applicatif, si on souhaite éviter de rejeter *a priori* certaines formes d'élocution. Le degré d'interactivité, quant à lui, semble être un facteur plus important de variabilité de l'ordre des mots comme de présence des inattendus structuraux. On note toutefois une fréquence d'apparition non négligeable de ces derniers phénomènes dans les situations les moins interactives restreintes à des domaines très spécifiques. Ce constat, associé à la détection incontournable des énoncés correctifs des locuteurs (et de la portée de la correction effectuée) renforce la nécessité du traitement de ces phénomènes.

2. D'autre part, la linguistique variationniste de corpus permet une caractérisation linguistique utile au prototypage des systèmes et présentant un caractère prédictif pour la conduite des recherches futures dans le domaine. L'utilité *immédiate* d'un tel diagnostic n'est certes pas évidente dans le cas des systèmes actuels, fondés sur des approches sélectives et/ou stochastiques dans lesquelles les phénomènes linguistiques sont implicitement représentés. Elle nous semble cependant nécessaire à la généralisation de la CHM orale à des domaines plus riches qui requièrent la mise en œuvre de techniques d'analyses plus détaillées. En matière de traitement automatique, l'étude des phénomènes d'extraction qui nous a servi d'exemple illustratif nous renseigne sur deux points principaux :

- Le faible taux de structures discontinues observé indique que le traitement de la variabilité forte ne saurait être un critère de choix des modèles appelés à être utilisés à l'avenir. L'analyse des énoncés oraux peut ainsi reposer sur un formalisme *projectif*.
- Nos observations suggèrent en revanche que le traitement de la variabilité faible constitue une problématique prégnante en CHM orale. Remarquons que les associés et les modifieurs sont les plus sujets à détachement. Les associés remplissent en effet souvent un rôle de marqueur pragmatique particulièrement sensible à la mise en relief. Ce rôle est rarement utilisé, sauf éventuellement pour l'étude des modalités, par les systèmes de dialogue oral actuels qui se contentent d'extraire le "sens utile" de l'énoncé, c'est-à-dire l'ensemble des éléments permettant de générer une requête d'interrogation de la base de données<sup>26</sup>. De ce point de vue, l'étude des détachements des associés est intéressante dans l'optique d'une modélisation améliorée du dialogue. De la même manière celle des modifieurs est importante dans un objectif de compréhension fine des énoncés.

Remarquons enfin que ces études restent limitées (cf. Annexe A § A.1.3). Les extractions par exemple ne sont pas les seules sources de discontinuités à l'oral. Une analyse de corpus portant sur l'ensemble des procédés potentiellement non-projectifs (incises et reprises dans une moindre mesure) serait utile pour quantifier ce problème. De la même manière, l'utilisation des indices que constituent les marques des inattendus structuraux répertoriées nécessitent des études complémentaires. Nous comptons poursuivre

<sup>26</sup>Cette information utile est le plus souvent portée par le verbe, le sujet et les arguments.

nos études sur l'extraction sur plusieurs nouveaux corpus correspondant à des tâches de renseignement touristique (domaine d'application que nous avons choisi et que nous décrivons dans le chapitre suivant) portant sur des dialogues homme-homme réels et non plus simulés comme dans le cas du corpus Murol, de renseignement administratif et de réservation hôtelière.

### 3.3 Conclusions et architecture du système proposé

Les aspects linguistiques que nous avons pris en compte ne rendent compte que d'un aspect limité et spécifique des énoncés oraux<sup>27</sup>. La théorie des *grilles* de Blanche-Benveniste que nous avons exploitée nous permet toutefois une connaissance minimale de l'oral qui nous suggère une analyse des énoncés oraux en deux étapes :

1. Une première étape d'analyse partielle minimale permettant la caractérisation des segments grammaticaux dans l'énoncé. Le fait que les inattendus structuraux interviennent à des frontières syntaxiques majeures de l'énoncé nous suggère de ne pas négliger le pouvoir structurant de la syntaxe au niveau du syntagme. L'idée qu'il existe des noyaux grammaticaux sur lesquels peut reposer une analyse grammaticale se retrouve dans un certain nombre d'autres travaux. C'est le cas par exemple, sans considération particulière de l'oralité, de la notion de périphérie de [Chanod, 1993] et des travaux de [Abney, 1991] (analyse en *chunks*) que nous proposons d'adapter<sup>28</sup> à notre problématique.
2. Une seconde étape permettant de mettre en relation ces différents segments sur des critères sémantico-pragmatiques. Il s'agit en particulier de pouvoir rétablir l'axe syntagmatique pour le traitement de l'ensemble des phénomènes impliquant des variations selon l'axe paradigmatique (inattendus structuraux, énumérations, ...). Le traitement de ces phénomènes s'appuyera sur les indices dégagés par les études des usages effectuées. Celles-ci nous indiquent par ailleurs que cette analyse doit être tolérante à la variabilité faible de l'ordre des syntagmes dans l'énoncé mais peut reposer sur des formalismes projectifs.

L'architecture du système que nous proposons s'apparente ainsi à celles des outils d'analyse (syntaxique) robuste développés pour le TAL et organisés de la même manière en général en deux grandes étapes [Ejerhed, 1993] : caractérisation de structures syntaxiques minimales puis analyse plus profonde sur la base de ces structures. Nous développons et justifions nos choix des formalismes et des techniques utilisés pour chacun de ces deux niveaux de traitement respectivement aux chapitres 5 et 6.

Cette méthodologie d'analyse devrait permettre par ailleurs de s'affranchir des étapes de prétraitement "aveugles" des inattendus structuraux par *patterns*, qui nous semblent

<sup>27</sup>Nous n'avons pas en particulier évoqué les travaux de l'équipe de Genève (autour de C. Kerbrat-Orecchioni), de Gand (autour de D. Willems) ni les travaux en sociolinguistique menés à Nanterre (F. Gadet).

<sup>28</sup>La motivation de [Abney, 1991] pour l'utilisation de segments grammaticaux est issue des travaux en psycholinguistique de [Gee et Gorsjean, 1983].

hasardeuses du point de vue de la détection (difficulté d'appréciation de la couverture de ces prétraitements et de leur sensibilité sur les niveaux supérieurs d'analyse) et peu fiables en ce qui concerne la "réparation" de ces phénomènes sans l'utilisation de contraintes syntaxiques et sémantiques.

Enfin, l'architecture proposée devrait permettre de répondre en partie à la question de la généralité des approches adoptées en compréhension en faisant reposer entièrement la première partie de l'analyse sur des considérations grammaticales et syntaxiques indépendantes de l'application. L'exploitation de connaissances grammaticales et syntaxiques combinée à une intervention précoce de la sémantique lors de la deuxième partie de l'analyse est par ailleurs essentielle à l'objectif d'analyse détaillée que nous nous sommes fixés.

Avant de présenter en détail les deux étapes de notre analyse, le chapitre suivant décrit le domaine d'application que nous avons choisi : le renseignement touristique ainsi que le niveau de détail des représentations sémantiques que nous souhaitons obtenir.



## Chapitre 4

# Application au renseignement touristique

Nous décrivons brièvement dans ce chapitre le domaine d’application que nous nous sommes choisi : le renseignement touristique. Ce choix a été guidé par nos objectifs de compréhension détaillée, l’idée étant d’appliquer notre système à un domaine finalisé sensiblement plus complexe que les domaines (horaires de trains ou d’avions) envisagés jusqu’à présent. L’un des buts de notre approche est de montrer la pertinence d’une analyse détaillée des énoncés oraux sur des tâches éventuellement multiples et peu structurées *a priori*. De ce point de vue, le domaine du renseignement touristique présente en première approximation des caractéristiques intéressantes parmi lesquelles :

- il est susceptible de présenter *a priori* une ambiguïté sémantique supérieure à un domaine tel que ARISE par exemple. Alors que dans cette dernière application une requête visant à obtenir des horaires ne peut concerner que les trains, cette même requête s’applique dans notre cas à de nombreux objets : cinémas, monuments, commerces, etc. et peut se décliner en horaire d’ouverture ou de fermeture,
- l’utilisateur a la possibilité d’exprimer des contraintes multiples (critères cumulés de prix, de proximité, de prestations pour le choix d’un hôtel par exemple), ces contraintes étant éventuellement relâchées (“*dans le centre ou à la rigueur près de la gare*”). Une appréciation plus fine des buts de l’usager doit passer par la prise en compte de ces contraintes.

Par ailleurs, les débuts de nos travaux ont coïncidé avec l’intégration de notre équipe de recherche à l’*Action de Recherche Concertée B2* de l’AUF<sup>1</sup> concernant précisément l’évaluation des systèmes de compréhension sur une tâche de renseignement touristique. Dans ce cadre, nous avons eu accès à un corpus d’amorce *bootstrap* sur le domaine du renseignement touristique autour de la gare Saint-Lazare à Paris. Un premier système prototype informant les usagers sur les moyens de transport dans le quartier, sur les hôtels, restaurants, magasins, cinémas, monuments, musées et les services disponibles en gare a été développé dans ce cadre à partir de 1000 énoncés décrivant le domaine.

---

<sup>1</sup>Agence Universitaire de la Francophonie, anciennement AUPELF-UREF.

400 énoncés ont ensuite été enregistrés à partir de ce prototype. Le corpus mis à notre disposition (corpus PARISCORP) comporte ainsi 1400 énoncés représentatifs du domaine, comportant en moyenne environ 9 mots<sup>2</sup>. La taille du lexique employé, assez limitée, est d'environ 1000 mots. On pourra trouver une description de l'ensemble du corpus dans [Bonneau-Maynard et Devillers, 1998] et sur le protocole d'enregistrement réalisé dans [Rosset, Lamel, Bennacef, Devillers, et Gauvain, 2000]. Nous avons déjà souligné les limites de tels corpus d'amorces *bootstrap* pour une analyse des usages. Ce corpus nous a ainsi uniquement été utile afin d'établir une première expertise sur l'univers de la tâche. Cette expertise a été réalisée manuellement à partir de ce corpus puis enrichie progressivement à partir de la connaissance intuitive que nous avons de ce domaine. Elle nous a permis la caractérisation et la structuration des connaissances liées à l'univers sémantique de référence en vue de la réalisation d'un premier prototype de notre système. La caractérisation du domaine et la définition d'une taxonomie satisfaisante n'a cependant rien d'évident et notre analyse a connu de nombreuses évolutions avant d'atteindre une certaine stabilité. Nous présentons dans le paragraphe suivant la caractérisation à laquelle nous avons abouti pour la réalisation de notre système.

Notons enfin que le domaine du renseignement touristique a été choisi dans le cadre du récent projet *Technolangue MEDIA*<sup>3</sup> auquel nous participerons.

## 4.1 Caractérisation du domaine

L'objectif du travail d'expertise du domaine réalisé est double. Il s'agit tout d'abord, outre la caractérisation du lexique utilisé dans le domaine, d'identifier les différentes connaissances sémantiques nécessaires à l'application c'est-à-dire :

- les objets manipulés dans l'application et les propriétés permettant leur caractérisation. Un hôtel par exemple peut être caractérisé par une **catégorie** (*un hôtel 3 étoiles*) ou par les différents **services** qu'il propose (*demi-pension* par exemple). Nous appellerons *concepts de propriété* ces relations sémantico-pragmatiques<sup>4</sup>, telle que **catégorie** existant entre l'objet du monde de l'application *hôtel* et l'élément de l'énoncé *3 étoiles*. Dans le cas où une base de données existe, cette catégorisation correspond aux objets de la base et à leurs attributs,
- les *concepts* plus généraux décrivant le sujet même des requêtes pouvant être exprimées de l'utilisateur (domaine de la tâche), que nous qualifierons de *schémas* (par exemple une demande sur les *horaires*, les *tarifs*, une *réservation*, etc.), ainsi que les arguments éventuellement impliqués par ces requêtes (une demande de *tarif* doit par exemple porter sur un objet du monde de l'application dont on peut évaluer le prix). Nous appellerons ces derniers concepts, *concepts-arguments*<sup>5</sup>,

<sup>2</sup>À titre comparatif, on notera que les énoncés d'une application portant sur les horaires de train comportent en moyenne 4 à 5 mots [Bousquet-Vernhettes, 2002].

<sup>3</sup>Méthodologie d'Évaluation automatique de la compréhension hors et en contexte du DIALOGUE (début du projet prévu à partir d'octobre 2002).

<sup>4</sup>Nous faisons ici référence au domaine d'application et non à la prise en compte du dialogue.

<sup>5</sup>On notera qu'à la différence des approches sélectives, nos concepts arguments et nos concepts de



- enfin, les différents actes de dialogue impliqués dans notre contexte. Les requêtes que nous venons d'évoquer correspondent par exemple à un acte de dialogue de *demande d'information*, mais l'énoncé de l'utilisateur peut simplement exprimer l'accord ou le désaccord (acte de *réponse*), introduire ou clore le dialogue par exemple. Un énoncé peut impliquer simultanément plusieurs actes de dialogue.

Ces connaissances ont ensuite été organisées de façon à obtenir une structuration sémantique globale du domaine. Il s'agit en particulier de classer les objets du domaine en fonction de leurs comportements par rapport aux différents schémas établis, c'est-à-dire d'identifier les ensembles d'objets pouvant être impliqués dans une demande d'information concernant les *horaires*, les *tarifs*, une *durée*, etc. Les ensembles ainsi définis doivent être suffisamment génériques pour pouvoir être facilement réutilisés lors de l'introduction de nouveaux schémas. Cette étape permet également de caractériser les différentes valeurs que peuvent prendre nos concepts. Celles-ci correspondent soit à une liste de valeurs ('simple' et 'double' en ce qui concerne le concept de propriété `TypeChambre` se rattachant à l'objet *chambre* (d'hôtel) par exemple) ou un ensemble de concepts ou d'objets du domaine.

Nous avons retenu 8 actes de dialogue principaux :

- l'acte d'ouverture du dialogue,
- l'acte de fermeture du dialogue,
- l'acte de réponse (*d'accord*),
- l'acte de correction (*non je parle du musée du Louvre*),
- l'assertion (*je suis à la gare, c'est trop cher*),
- les demandes d'explicitation (*pouvez vous répéter, préciser*),
- la demande d'information (*à quelle heure ouvre le musée du Louvre*),
- et la demande de confirmation<sup>6</sup> (*le prix c'est bien 20 euros*)

La caractérisation et la classification du domaine à laquelle nous avons abouti couvre 36 schémas et 151 concepts (43 objets, 59 concepts de propriété et 49 concepts-arguments) dont nous donnons quelques exemples dans le tableau 4.1. L'annexe C.1 donne un aperçu plus global des schémas et concepts identifiés.

Cette catégorisation est loin de rendre compte complètement du domaine. À titre d'exemple, dans le cadre des projets de traduction C-STAR [Levin et al., 2000] et NESPOLE! [Rossato et al., 2002], le domaine concerné (renseignement et planification de voyages), proche du notre, on dénombre 26 actes de dialogue, 100 concepts de "niveau supérieur" (à rattacher à nos schémas et objets) et plus de 150 concepts de "niveau inférieur" (à rattacher à nos concepts de propriétés et concepts-arguments). Notre caractérisation nous a semblé toutefois satisfaisante pour le développement d'un

---

propriété expriment des *relations* entre segments éventuellement récursives. La compréhension n'est en ce sens pas sélective mais simplement guidée par la tâche.

<sup>6</sup>Cet acte de dialogue regroupe à la fois les demandes de confirmation par rapport à une information donnée précédemment au cours du dialogue et les demandes d'information correspondant à des requêtes fermées ("*est-ce que le musée est loin d'ici*" par exemple).

Schémas	CONCEPTS		
	<i>objets</i>	<i>propriétés</i>	<i>arguments</i>
Acte : demande-information sélection, tarif,  localisation, distance, trajet,  horaire,  durée,  réservation-touristique, ...	restaurant  hôtel  monument  cinéma, musée commerces film	Spécialité Rapidité-service Catégorie Servives  Moyen-transport  Version Genre-film	Objet-select Critère-prix, Critère-distance Objet-tarif Objet-loc, Référence Lieu-départ Lieu-arrivée Objet-horaire Type-horaire Etablissement-référent Référence Objet-résa Etablissement-résa
Acte : demande-confirmation disponibilité, ...	place réduction	Type-réduction	Etablissement-référent

TAB. 4.1 – Exemples simplifiés des schémas et concepts considérés dans le domaine d’application du système ROMUS.

premier prototype et permettre la confrontation de notre stratégie d’analyse à un domaine applicatif plus riche que ceux des systèmes actuels. À titre de comparaison par rapport aux systèmes de compréhension sélectifs, l’environnement CACAO développé à l’IRIT sur une application de renseignement concernant les horaires de trains [Bousquet-Vernhettes, 2002] comporte une trentaine de concepts<sup>7</sup>. Le système ARISE du LIMSI, destiné à la réservation de billets de trains et à l’obtention de renseignements ferroviaires (trajets, horaires, types de trains, tarifs, prestations, possibilités de réduction, etc.) [Maynard et Lefèvre, 2002] comptabilise quant à lui au total 64 concepts pour 8 schémas distincts.

## 4.2 Définitions des représentations sémantiques attendues

Nous avons choisi de représenter la structure sémantique finale d’un énoncé sous forme de structures de traits c’est-à-dire un ensemble de paires attribut-valeur dont les valeurs peuvent être simples (chaîne de caractères, nombre, ...) ou représenter une structure de traits. Ce choix est indépendant de la stratégie d’analyse adoptée. La conversion du résultat de notre analyse (cf. chapitre 6 § 6.2.6) dans ce format a été guidée uniquement par le souci de faciliter l’étape d’évaluation. Cette représentation offre une grande

<sup>7</sup>Cette comptabilisation tient compte à la fois de concepts illocutoires généraux et de concepts référentiels décrivant la tâche. Ces derniers sont minoritaires (une douzaine de concepts), voir [Bousquet-Vernhettes, 2002], chapitre 3, p 56-65.

lisibilité et rend possible la comparaison automatique aisée des structures générées avec des structures de référence<sup>8</sup>.

Nous décrivons dans ce paragraphe de manière très générale les éléments qu'il nous semble important d'intégrer dans les représentations sémantiques. Ces choix se fondent sur l'expertise du domaine d'application réalisée et tient compte des objectifs d'analyse détaillée que nous nous sommes fixés.

Toutes les représentations comprennent les éléments suivants :

1. une indication de l'acte de dialogue (ou des actes de dialogue) impliqué(s) dans l'énoncé,
2. et les concepts arguments associés aux différents actes impliqués. L'expertise du domaine nous a permis de distinguer pour chaque acte de dialogue et chaque schéma (dans les cas de demande d'information ou de confirmation) les concepts arguments devant *obligatoirement* être instanciés. La représentation sémantique correspondant à l'énoncé "*d'accord merci au revoir*" sera ainsi :

<acte> : réponse Type-réponse : OK Politesse : oui <acte> : fermeture-dialogue
---

De la même manière une demande d'information de type *horaire* nécessite l'instanciation du concept-argument **Objet-horaire** spécifiant l'objet sur lequel porte la requête. Le type de l'horaire souhaité (ouverture ou fermeture par exemple) est quant à lui facultatif. Dans le cas où les informations *obligatoires*<sup>9</sup> ne sont pas présentes dans l'énoncé une valeur particulière (*ctxt*) permet de préciser que l'information doit être présente dans l'historique du dialogue :

<acte> : demande-information Type-information : horaire Objet-horaire : musée Référence : Louvre	<acte> : demande-information Type-information : horaire Objet-horaire : <i>ctxt</i>
<i>quels sont les horaires du musée du Louvre</i>	<i>quels sont les horaires</i>

<acte> : demande-information Type-information : horaire Type-horaire : fermeture Objet-horaire : musée Référence : <i>ctxt</i>
--

<sup>8</sup>On notera que c'est un format de représentation sous forme de liste attribut-valeur (structures de traits non récursives) qui a été choisi dans le cadre du projet d'évaluation MEDIA.

<sup>9</sup>L'instanciation des concepts-arguments obligatoires est indépendante de l'analyse de l'énoncé c'est-à-dire que leur absence dans l'énoncé n'est pas préjudiciable pour l'identification du schéma. L'information est ajoutée lors de la conversion du résultat de l'analyse dans le format de représentation adopté.

## à quelle heure ferme le musée

Nous précisons aux chapitres suivants les informations grammaticales supplémentaires que nous intégrons éventuellement dans les structures générées.

Pour terminer cette introduction générale sur les structures que nous souhaitons générer, notons enfin que nous avons choisi de représenter explicitement dans les représentations sémantiques deux informations particulières qui nous semblent utiles pour le déroulement du dialogue :

- Notre objectif de compréhension hors-contexte ne doit pas chercher à *interpréter* abusivement l'énoncé. L'analyse que nous proposons (cf. chapitre 6) peut rester ambiguë. Dans ce cas, nous avons choisi de représenter explicitement<sup>10</sup> **les ambiguïtés** résiduelles. Certaines de ces ambiguïtés pourront être résolues ultérieurement en utilisant l'historique du dialogue. Nous pensons par exemple aux ambiguïtés portant sur le type d'un acte de demande d'information dans le cas d'un énoncé tel que "*l'hôtel le plus proche*". En l'absence de tout contexte, cet énoncé peut correspondre entre autre à une requête de sélection (identification de l'établissement le plus proche), ou d'une demande de tarif sur cet établissement si l'échange porte jusque là sur les tarifs hôteliers. L'ambiguïté évoquée dans cet exemple semble pouvoir être facilement résolue. Dans d'autres cas en revanche, l'identification de la portée de l'ambiguïté dès ce niveau de l'analyse peut s'avérer très utile pour guider le gestionnaire de dialogue. Considérons par exemple l'énoncé suivant : "*quels sont les horaires du musée Grévin du musée du Louvre*". En l'absence d'informations prosodiques, il est impossible de décider s'il s'agit d'une énumération ou d'une auto-correction du locuteur, la coordination *et* comme les marques lexicales éventuelles de correction (*non, pardon, ...*) pouvant ne pas avoir été reconnues par le module de reconnaissance de la parole. Cette ambiguïté sera représentée explicitement dans la structure sémantique. En s'appuyant sur cette information, une stratégie dialogique pourrait décider de donner la réponse pour le deuxième élément (le musée du Louvre) et ensuite d'interroger l'utilisateur sur ses attentes éventuelles concernant le musée Grévin.
- Une analyse complète des énoncés oraux n'est pas toujours envisageable. La deuxième étape d'analyse que nous proposons peut être partielle, c'est-à-dire qu'il est possible qu'une représentation sémantique soit générée en ignorant un ou plusieurs éléments de l'énoncé. Nous avons choisi d'indiquer dans la représentation finale **le degré de complétude** de l'analyse ayant conduit au résultat<sup>11</sup>. Ce choix nous semble intéressant pour faire face aux problèmes dus aux mots inconnus du lexique sémantique et aux mots mal reconnus par le module de reconnaissance. Nous examinerons l'influence de leur présence sur notre analyse au chapitre 7<sup>12</sup>.

<sup>10</sup>Nous utilisons pour cela un marqueur spécifique portant sur les valeurs des concepts ambigus (cf. par exemple chapitre 6 §6.3.1.2).

<sup>11</sup>Nous précisons cette notion dans les chapitres suivants. D'une manière générale, il s'agit d'indiquer le nombre de mots et de syntagmes ignorés.

<sup>12</sup>La présence de mots inconnus ou mal reconnus ne doivent pas empêcher l'analyse. Dans notre système, ils conduisent généralement à une représentation incomplète de l'énoncé.

L'information du degré de complétude de l'analyse peut s'avérer utile pour indiquer par exemple qu'un concept-argument *obligatoire* de l'acte de dialogue détecté et absent de la représentation n'est pas obligatoirement à rechercher dans le contexte dialogique mais que son absence résulte peut-être de l'utilisation d'un nom propre inconnu du lexique sémantique par exemple ou d'une erreur de reconnaissance. D'une manière générale, elle permet d'indiquer qu'une information potentiellement importante (contraintes, restrictions, objet de requête, etc.) a pu être ignorée ou expliquer une incohérence entre ce qui a été compris et ce qui a été dit auparavant. Le système peut alors éventuellement orienter les demandes à formuler au locuteur, lui suggérant de ne préciser qu'une partie de l'énoncé ou de lui faire savoir l'ensemble des contraintes comprises afin qu'il les confirme.

Les deux chapitres suivants présentent les deux étapes d'analyse que nous proposons. Les connaissances sémantico-pragmatiques issues de l'expertise du domaine d'application interviendront uniquement dans la deuxième étape de l'analyse. Nous précisons la manière dont nous les utilisons au chapitre 6.



## Chapitre 5

# Segmentation grammaticale des énoncés oraux

### 5.1 Objectifs

La stratégie de compréhension proposée repose sur une analyse des énoncés oraux en deux étapes. L'objectif principal de la première étape est d'initier le processus de compréhension en s'appuyant sur les régularités structurelles minimales du français parlé identifiées au chapitre 3. Elle consiste ainsi en une segmentation partielle des énoncés oraux cherchant à caractériser les unités syntagmatiques intervenant dans le processus de production des inattendus structuraux et de certains procédés réguliers (variations selon l'axe paradigmatique cf. chapitre 3 § 3.1). Ces unités correspondent également aux éléments impliqués dans les procédés d'extractions étudiés précédemment (variabilité faible cf. chapitre 3 § 3.2.1.2). La seconde étape, que nous décrivons au chapitre suivant, cherchera à mettre en relations ces différents segments sur des critères sémantico-pragmatique.

Les syntagmes minimaux auxquels fait référence [Blanche-Benveniste, 1990] correspondent à la notion de *chunks* dans la littérature anglaise [Abney, 1991].

(Cit.5.1) “*The typical chunk consists of a single content word surrounded by a constellation of function words, matching a fixed template*” [Abney, 1991]

La définition qu'Abney donne des *chunks* (appelés également constituants minimaux non récursifs) est motivée par les études psycholinguistiques de [Gee et Gorsjean, 1983] fondées sur l'hypothèse de pertinence cognitive de ces groupements syntagmatiques dans la production de la parole. Selon cette étude, ces unités, appelées *performance structures*, rendent compte des pauses et des changements intonatifs intervenant dans la production orale<sup>1</sup>. Ainsi que le résume [Abney, 1991] :

---

<sup>1</sup>L'utilisation que nous faisons en compréhension de la parole de la segmentation en *chunks* n'a pour objectifs que de tenir compte de régularités attestées dans les productions orales, indépendamment de tout jugement sur cette théorie.

(Cit.5.2) “*I begin with an intuition : when I read a sentence, I read it a chunk at a time*”. [Abney, 1991]

Il illustre son propos en présentant une segmentation possible de cette phrase : [I begin] [with an intuition] : [when I read] [a sentence], [I read it] [a chunk] [at a time]

La notion de *chunk* peut ainsi être rapprochée de la notion japonaise plus ancienne de *bunsetsu* qui définit des unités prosodiques de la production orale<sup>2</sup>.

L'idée d'utiliser ces unités syntagmatiques minimales a été exploitée dans différents domaines : en reconnaissance de la parole [Hirose, Minematsu, Hashimoto, et Iwano, 2001], pour l'étiquetage syntaxique de corpus “tout-venant” [Abeillé, Clément, et Kinyon, 2000] ou pour l'analyse syntaxique [Abney, 1996; Ejerhed, 1996; Giguet, 1998; Vergne, 2000]. Ces différentes approches se distinguent par les définitions quelque peu divergeantes attribuées aux unités de segmentation utilisées.

Dans le cadre qui nous intéresse, nous définirons le syntagme minimal comme une unité constituée d'une série de mots contigus et regroupés autour d'une *tête lexicale* (nom, verbe, adjectif, adverbe). Toutes les relations syntaxiques internes au syntagme ont la particularité d'être calculables de manière positionnelle et non ambiguë. Nous précisons cette définition générale au paragraphe 5.3.1 où nous présentons les unités de segmentation que nous avons retenues.

L'analyse que nous proposons est une analyse syntaxique **de surface** en ce sens qu'il ne s'agit pas de chercher à donner une description de la structure syntaxique complète de l'énoncé mais uniquement de le découper en une séquence d'îlots homogènes. Ces îlots seront ensuite considérés comme les unités de base des étapes supérieures de l'analyse.

Cette première étape répond à deux objectifs principaux :

1. Elle est générique c'est-à-dire entièrement portable d'un domaine applicatif à l'autre. Nous faisons ainsi reposer l'identification des unités minimales considérées sur des critères grammaticaux indépendants de l'application<sup>3</sup> (cf. § 5.3.2.2). Par ailleurs, nous avons fait le choix de repérer et d'analyser dès cette étape les expressions dont la formulation est indépendante de l'application (expression des heures, des dates, des prix, etc.). Nous rejoignons en ce sens les travaux menés au MIT sur des domaines d'application multiples, pour lesquels [Seneff, 1998] insiste sur la nécessité de capitaliser l'analyse des concepts communs à plusieurs domaines applicatifs.
2. Elle constitue un compromis acceptable entre nos objectifs d'analyse détaillée et

<sup>2</sup>Dans les années 70-80, Bernard Vauquois parlait de “groupes élémentaires”.

<sup>3</sup>On notera par ailleurs que la notion de *chunk* semble être stable entre les langues, ce qui garantit une certaine portabilité de la stratégie adoptée à une autre langue. La validité multilingue de la segmentation proposée a été montrée par [Déjean, 1998].



de robustesse. Cette étape ne doit pas seulement effectuer la segmentation de l'énoncé mais également fournir une caractérisation détaillée de la structure du syntagme (présence d'articles définis, indéfinis, adjectifs possessifs, etc. cf. 5.3.2.2). En contrepartie, pour assurer la robustesse de l'analyse, la segmentation doit reposer sur des contraintes minimales. Les éventuelles erreurs de reconnaissance, notamment les erreurs d'accord en genre et en nombre, nous interdisent par exemple l'utilisation de critères morphologiques pour guider la segmentation.

## 5.2 Analyse robuste de l'oral

D'une manière générale, cette étape est une première garantie de la robustesse de l'analyse proposée face aux phénomènes spécifiques de l'oral. L'analyse syntaxique de surface que nous proposons a été guidée par l'étude des régularités attestées dans les productions orales. Elle se retrouve également en général dans l'architecture des outils d'analyse syntaxique robuste développés pour le TAL [Ejerhed, 1993]. Nous définissons la notion de "robustesse" au paragraphe suivant.

### 5.2.1 Le TAL robuste

Les premiers travaux en TAL robuste datent des années 70. Depuis, ce domaine a fait l'objet de nombreux projets de recherche dès lors que de larges ensembles de données écrites et orales ont été disponibles. L'essor des méthodes stochastiques [Charniak, 1993] a mis en évidence les difficultés des approches traditionnelles à décrire et à traiter, par des modèles linguistiques formels intéressants par ailleurs, l'ensemble très diversifié des variations autour des phénomènes linguistiques et des constructions syntaxiques attestées en corpus. Cette variété de constructions est particulièrement importante dans les textes ou corpus dits "tout-venant" c'est-à-dire les textes libres hétérogènes non dédiés à un style, un genre, ou un domaine particulier. On trouvera dans [Chanod, 2001], une illustration de cette diversité de formes sur l'exemple des contraintes d'accord en genre et en nombre.

La robustesse est ainsi une question de compromis entre couverture et profondeur de l'analyse réalisée, liée à la notion de *relâchement de contraintes* (ou de contraintes minimales) nécessaire pour rendre compte à la fois des constructions standards et de toutes celles qui sortent des caractéristiques linguistiques générales de la langue concernée. En d'autres termes :

(Cit.5.3) "*Robustness is about exploring all constructions humans actually produce, be they grammatical, conformant to formal models, frequent or not.*"

[Chanod, 2001]

Dans l'analyse des textes hétérogènes par exemple, il s'agit d'atteindre une large couverture des phénomènes linguistiques traités sans perdre en précision et sans se restreindre à un domaine particulier [Abeillé et Blache, 2000].

Les principales caractéristiques d'une analyse robuste sont les suivantes [Abney, 1994] :

- l'analyse est *incrémentale* en ce sens qu'elle est effectuée graduellement : les structures non ambiguës sont d'abord analysées et le résultat de cette première étape d'analyse est utilisé pour faciliter l'étape suivante. En général, chaque étape concerne la caractérisation d'un type particulier de structure ou le calcul d'une relation syntaxique spécifique ;
- l'analyse est généralement *déterministe* (une seule analyse est produite à chaque étape<sup>4</sup>). Le but n'est pas d'explorer toutes les ambiguïtés structurelles ni de fournir toutes les analyses possibles mais plutôt d'extraire l'analyse la plus plausible d'une phrase donnée. Un analyseur robuste doit ainsi contrôler l'explosion combinatoire (le plus souvent en appliquant une série d'heuristiques) liée à ce type d'approche ;
- la formalisation des phénomènes linguistiques traités se fonde sur l'exploitation et l'étude systématique des corpus et n'est pas ou peu guidée par une théorie linguistique ; la grammaire peut être soit décrite par un ensemble de règles [Hindle, 1983; Hobbs et al., 1997; Tapanainen et Järvinen, 1997] soit obtenue automatiquement par apprentissage sur corpus annoté [Church, 1988; Daelemans, Buchholz, et Veenstra, 1999] ;
- l'analyseur doit pouvoir fournir une analyse des phrases<sup>5</sup> sur n'importe quels corpus, aussi hétérogènes soient-ils [Carroll et Briscoe, 1996]. En d'autres termes, l'analyse doit toujours fournir une sortie même si cette dernière n'est pas toujours la meilleure ou est incomplète.

Parmi les analyseurs syntaxiques robustes par règles<sup>6</sup>, on peut distinguer trois types d'approches :

1. les approches reposant sur une construction progressive de structures de plus en plus étendues (accroissement d'îlots de certitude<sup>7</sup> ; on parle également d'approche constructiviste.). Ainsi l'analyse syntaxique réalisée dans le système CASS [Abney, 1991, 1996] résulte d'une succession d'analyses partielles déterministes : étiquetage grammatical, *chunking* progressif, identification des propositions ("simple clause" [Abney, 1996]) ;
2. les approches fondées sur des formalismes lexicalisés, principalement inspirés de la notion de grammaires de dépendance [Tesnière, 1959; Kahane, 2000b, 2001]<sup>8</sup>. Un exemple peut être donné par le système FDP [Tapanainen et Järvinen, 1997]

---

<sup>4</sup>Certains analyseurs peuvent produire plusieurs solutions, c'est le cas par exemple lors d'ambiguïtés dans les rattachements prépositionnels.

<sup>5</sup>L'analyse robuste est également concernée par l'incomplétude des lexiques utilisés. Nous n'aborderons qu'en partie cet aspect (cf. chapitre 8).

<sup>6</sup>Nous écartons volontairement de cette description rapide les analyseurs stochastiques.

<sup>7</sup>On notera que la notion de *chunk* est plus généralement inspirée du formalisme PSG (*Phrase Structure Grammar*) [Bloomfield, 1933]. Ce formalisme repose sur le principe théorique selon lequel les mots d'une phrase peuvent progressivement être groupés en des structures de plus en plus étendues (principe de constituants immédiats)

<sup>8</sup>Nous reviendrons sur les grammaires de dépendance au chapitre suivant.

qui repose sur le formalisme des grammaires de contraintes [Karlsson, Voutilainen, Heikkilä, et Anttila, 1995]. Il s'agit d'une approche réductionniste. L'idée sous-jacente est de considérer dans un premier temps toutes les constructions syntaxiques possibles (l'information est représentée dans le lexique par l'ensemble des étiquettes syntaxiques envisageables pour un mot donné) puis d'éliminer certaines de ces constructions par l'application de contraintes locales<sup>9</sup>. De tels analyseurs n'ont pas recours à une segmentation intermédiaire en syntagmes ;

3. enfin les systèmes hybrides utilisant une segmentation en *chunks* suivi d'une analyse des dépendances entre les segments identifiés [Grefenstette, 1992; Giguet, 1998; Aït-Mokhtar et Chanod, 1997; Basili, Pazienza, et Zanzotto, 1999].

Notre analyseur rentre dans cette dernière catégorie. Dans le cadre qui nous intéresse, l'analyse syntaxique de surface prend une importance particulière. Elle assure en partie la généralité de l'analyse qui ne repose pour cette étape que sur un minimum de contraintes ne faisant pas intervenir de connaissances liées au domaine d'application et permet de s'affranchir des répétitions les plus fréquentes : celles des mots débutant les syntagmes (déterminants par exemple). Elle permet en outre de produire de l'information partielle utilisable en cas d'analyses incomplètes lorsqu'un mot d'une partie de l'énoncé a par exemple été mal reconnu.

### 5.2.2 Adaptation à la segmentation des énoncés oraux

Les analyseurs robustes que nous venons d'évoquer ont été développés pour réaliser une analyse *syntactique* de textes. Même si ces objectifs ne correspondent pas complètement à ceux que nous nous sommes fixés dans le cadre de la compréhension orale, il convient cependant de s'interroger sur leurs apports éventuels dans le cadre qui nous intéresse. En particulier, la question de l'utilisation d'une partie de ces analyseurs (l'étape de segmentation) à notre problématique mérite d'être posée. L'examen des stratégies utilisées pour cette première étape semble indiquer que ces analyseurs ne sont pas directement exploitables pour l'analyse des énoncés oraux.

Dans la plupart des systèmes en effet [Aït-Mokhtar et Chanod, 1997; Giguet, 1998], la segmentation s'effectue en marquant le début et la fin des syntagmes minimaux considérés. La caractérisation de la fin des syntagmes repose principalement sur les indices véhiculés par les marques de ponctuation. Le rôle de la ponctuation est ainsi essentiel pour le chunking [Jones, 1994; Vergne, 2000]. Or, nous ne disposons pas de ces indices à l'oral. Ainsi, la plupart des analyseurs, conçus pour traiter des textes hétérogènes, rejettent explicitement les corpus de transcriptions de l'oral de leur évaluation en raison précisément de l'absence de ponctuation<sup>10</sup> [Gala Pavia, 2001]. Par ailleurs, ces analyseurs utilisent abondamment des critères morphologiques pour la caractérisation

<sup>9</sup>L'analyse se ramène ainsi à une tâche d'étiquetage associée à un formalisme descriptif riche. On retrouve cette idée dans l'analyseur LDA – Lightweight Dependency Analyser – [Srinivas, 1997] utilisant une méthode de désambiguïsation probabiliste fondée sur les *supertags* [Bangalore et Joshi, 1999] avant l'analyse des dépendances syntaxiques.

<sup>10</sup>Jacques Vergne suggère l'exploitation d'indices prosodiques en lieu et place des informations véhi-

des syntagmes. On sait que l'on constate souvent à l'oral un certain relâchement de ces critères (erreurs de conjugaison, concordance des temps, etc.) [Bellenger, 1979]. À ces problèmes s'ajoutent les erreurs de reconnaissance particulièrement fréquentes sur l'accord en genre et en nombre. Ces informations ne sont donc pas entièrement fiables en compréhension orale.

Même si ces analyseurs ont précisément été définis pour s'adapter à tout type de texte (genre, style, etc.), leur évaluation montre que leurs performances respectives tendent à décroître dès lors qu'ils sont testés sur des domaines spécialisés différents que ceux pour lesquels ils ont été développés [Gala Pavia, 2001]. Or les constructions syntaxiques attestées sur les corpus de texte utilisés pour la conception de ces analyseurs robustes (principalement des corpus journalistiques ou des manuels techniques) diffèrent sensiblement (en fréquence relative) de celles attestées dans les corpus oraux. [Monceaux et Robba, 2002] ont montré par exemple la faiblesse de ces analyseurs pour le traitement des questions, peu représentées dans les corpus de textes mais essentielles en dialogue oral finalisé.

Ainsi, ces analyseurs ne peuvent pas facilement être adaptés ni utilisés pour le traitement des énoncés oraux. Nous avons de plus des besoins spécifiques (identification des appuis du discours et des marques lexicales des répétitions et des corrections par exemple cf. § 3.2.1.1.) qui nous ont conduit à proposer un analyseur de surface spécifique.

## 5.3 Segmentation grammaticale dans le système ROMUS

### 5.3.1 Définitions des unités syntagmatiques retenues

Nous avons choisi de retenir trois types principaux d'unités syntagmatiques minimales :

1. les premiers correspondent aux **groupes grammaticaux classiques** : groupe nominal, groupe adjectival, groupe verbal, groupe adverbial et groupe prépositionnel. À notre connaissance et ainsi que le note [Kahane, 2002, page 65], les constituants non récursifs sont absents des travaux de linguistique théorique. Il n'existe ainsi pas de définitions unifiées de ces structures. À l'exception des groupes prépositionnels, les unités minimales grammaticales que nous avons retenues suivent les définitions adoptées par [Abney, 1991] : la tête lexicale autour de laquelle est articulé le syntagme est toujours l'élément situé le plus à droite<sup>11</sup>. Ainsi nous excluons les adjectifs postposés au nom dans le cas des groupes nominaux. “*un petit restaurant*” constituera ainsi une seule unité alors que “*une chambre double*” sera segmenté en deux unités : le groupe nominal *une chambre* et le groupe ad-

---

culées par la ponctuation. Or, cette correspondance est loin d'être attestée [Blanche-Benveniste, 1990]. Comme le notent [Valli et Véronis, 1999] ce qui est ponctué à l'écrit peut fort bien se réaliser sans pause, alors que des pauses orales peuvent n'avoir aucune correspondance à l'écrit.

<sup>11</sup>À l'intérieur d'un *chunk*, l'ordre des mots est rigide.

jectival *double*. Ce choix a été justifié par les études de corpus réalisées sur les corpus PARISCORP et Murol. Nous avons comptabilisé un nombre non négligeable d'énoncés dans lesquels les adjectifs ne sont pas sémantiquement reliés au nom qui les précède. C'est le cas par exemple de l'énoncé suivant : “*je voudrais réserver une chambre avec **douche double***” où le locuteur a marqué une pause silencieuse entre *douche* et *double*.

Le cas des groupes prépositionnels est particulier. Dans la plupart des analyseurs effectuant une segmentation en *chunk*, les rattachements prépositionnels ne sont pas effectués<sup>12</sup>. Nous avons fait le choix inverse. Les études que nous avons menées sur le corpus Murol indiquent en effet que la préposition fait généralement partie du syntagme repris en cas de répétitions et de corrections. Cette hypothèse mériterait toutefois d'être confirmée par des études portant sur des corpus de taille plus importante. Nous évaluerons l'influence de ce choix au chapitre 7.

Enfin, les analyseurs traditionnels n'analysent généralement pas la structure interne des segments identifiés. Ils se contentent d'étiqueter et de délimiter les unités syntagmatiques dans l'énoncé<sup>13</sup> [Megyesi, 2002]. Nous verrons que notre analyse incrémentale nous permet d'identifier précisément la structure interne du syntagme. Par ailleurs, nous cherchons à distinguer lors de cette étape la complexité de la structure verbale (groupe verbal infinitif, présence à droite du verbe d'un modal, d'un adjectif interrogatif<sup>14</sup>, etc.)

Les syntagmes grammaticaux que nous avons choisi de retenir sont répertoriés dans le tableau 5.1. Remarquons enfin que les mots grammaticaux de liaison (coordinations, conjonction de subordination, pronoms relatifs, etc.) constituent des syntagmes minimaux à eux seuls.

2. les seconds regroupent les **expressions langagières** indépendantes de l'application : les heures, les dates, les adresses et les unités de mesure (distance et prix). Ces segments ne sont pas articulés autour d'une tête lexicale particulière. Les groupes prépositionnels associés à ces segments sont étiquetés spécifiquement dès cette étape. L'étiquette GPDDATE identifie par exemple un segment comportant une préposition suivie de l'expression d'une date (*avant le 3 juin, pour demain, à Noël*, etc.).
3. enfin les derniers sont spécifiques à l'**oral** et correspondent aux marques lexicales des hésitations et des corrections notés respectivement HES et COR. Les marques

<sup>12</sup>La préposition constitue un segment à elle seule (à l'image des coordinations et des pronoms relatifs par exemple). Le rattachement de la préposition à ses arguments (groupe nominal, verbal ou adjectival) est délégué à l'étape supérieure de l'analyse.

<sup>13</sup>À l'exception des travaux de [Brants, 1999].

<sup>14</sup>Ce dernier choix fait porter sur l'étiquette du segment une information liée à la nature de la relation entre l'adjectif interrogatif et le verbe. Toutefois nos études de corpus indiquent que de telles suites constituent une unité entièrement reprise lors des hésitations.

Syntagmes	Caractérisation	Exemples
Groupe Adverbial GADV	adverbe éventuellement précédé de certains adverbess qualifiant la tête lexicale ( <i>pas, plus, très, etc.</i> )	<i>déjà, dedans, très rapidement, trop tard, pas tellement</i>
Groupe Adjectival GADJ  GADJ <sub>num</sub>	adjectif précédé éventuellement de ces mêmes adverbess qualificatifs ou d'un superlatif adjectif numéral	<i>cher, pas cher, trop cher le moins cher</i>
Groupe nominal GN  GN <sub>NPROPRE</sub>	nom précédé par un article ou un adjectif possessif ou démonstratif, suivi éventuellement d'un adjectif numéral et d'un ou plusieurs groupes adjectivaux Sont également considérés les pronoms personnels et démonstratifs nom propre seul	<i>le musée, les trois derniers hôtels,  le, la, en, etc. ceci, ça Grévin</i>
Groupe verbal GV GV <sub>PPR</sub> GV <sub>PP</sub> GV <sub>INF</sub> GV <sub>MOD</sub>  GV <sub>PPMOD</sub> GV <sub>INT</sub>	verbe conjugué éventuellement pronominal participe présent auxiliaire suivi d'un participe passé verbe à l'infinitif éventuellement pronominal verbe à l'infinitif précédé d'un ou plusieurs modaux (ou assimilés) conjugués auxiliaire suivi du participe passé d'un modal verbe conjugué précédé d'un pronom, adverbe ou adjectif interrogatifs	<i>coûte me rappelle marchant ai vu manger voudrais réserver aimerais pouvoir faire aurais voulu quels sont combien coûte</i>
Groupe prépositionnel GP	préposition suivie d'un groupe nominal, d'un groupe adjectival, d'un groupe verbal ou d'un pronom personnel suivi d'un groupe verbal infinitif (pour certaines prépositions uniquement)	<i>de fermeture pour pas cher en marchant sans prendre pour y aller</i>
mots de liaison syntaxiques	pronom interrogatif, adjectif interrogatif pronom réfléchi, pronom relatif, conjonction négation, coordination	

TAB. 5.1 – Segmentation en chunk – groupes grammaticaux élémentaires retenus

d'hésitation correspondent à une suite contiguë<sup>15</sup> d'interjections, d'appuis du discours (*euh, enfin, bon, etc.*) ou d'expressions figées attestées sur corpus (*comment dire* par exemple). Une suite contiguë de ces éléments comportant en plus une marque de négation (*non, pardon*) ou une locution telle que *en fait* par exemple est étiquetée comme une marque de correction.

Au total nous retenons ainsi 33 unités minimales différentes<sup>16</sup>. La caractérisation de ces unités reposera sur des informations grammaticales. Une première étape d'étiquetage grammatical des énoncés oraux est ainsi réalisée.

### 5.3.2 Étiquetage grammatical de l'oral

Contrairement à l'usage de l'étiquetage grammatical dans l'analyse syntaxique traditionnelle, notre étiquetage n'est pas un prétraitement fin sensé réduire l'ambiguïté lexicale à son minimum pour que l'analyse syntaxique puisse être calculée en temps raisonnable. À l'image de [Valli et Véronis, 1999], nous ne considérons pas non plus l'étape d'étiquetage comme un modèle ou une référence d'analyse linguistique, mais comme un outil dont le but est ici de fournir un bon découpage en syntagmes (et un bon typage de ces syntagmes) des énoncés oraux issus d'un module de reconnaissance. L'étiqueteur doit ainsi fournir une catégorisation grammaticale suffisante et acceptable des mots de l'énoncé pour garantir ces objectifs. Nous évaluons dans un premier temps les apports que peuvent présenter dans ce contexte les étiqueteurs morpho-syntaxiques conçus pour l'écrit.

#### 5.3.2.1 Étiquetage de l'oral à partir d'étiqueteurs conçus pour l'écrit

L'étiquetage de l'oral pose des problèmes spécifiques et a peu été étudié jusqu'ici<sup>17</sup>. Ceci peut s'expliquer en partie par le manque de corpus oraux d'envergure (pré-étiquetés ou non) en français à partir desquels des étiqueteurs tenant compte des particularités des corpus oraux pourraient être conçus.

Dans la plupart des étiqueteurs conçus pour l'écrit<sup>18</sup>, la désambiguïsation est effectuée sur la prise en compte d'un contexte local très réduit. Ce caractère local de l'analyse a un effet bénéfique sur la robustesse de l'étiquetage, atout considérable pour l'étiquetage de l'oral. Une expérience d'étiquetage de l'oral à partir d'un étiqueteur conçu pour l'écrit (CORDIAL<sup>19</sup>) a été ainsi menée par André Valli et Jean Véronis [Valli et Véronis, 1999]. Cette expérience a été réalisée sur une transcription d'oral d'une quinzaine de minutes choisie en raison de sa représentativité en matière de phénomènes spécifiques de

<sup>15</sup>Un seul de ces éléments constitue à lui seul une marque d'hésitation.

<sup>16</sup>La liste complète des unités minimales retenues peut être consultée en annexe B.

<sup>17</sup>Citons toutefois les travaux sur l'oral de [Mertens, 2002] dans lesquels l'étiquetage et l'analyse syntaxique sont réalisés par un analyseur tabulaire robuste constructiviste.

<sup>18</sup>Pour un état de l'art des étiqueteurs morpho-syntaxiques, on pourra se reporter à [Paroubek et Rajman, 2000].

<sup>19</sup>L'étiqueteur utilisé, CORDIAL est distribué par la société Synapse Développement.

l'oral. Elle indique des performances globales identiques à la plupart de celles publiées sur l'écrit<sup>20</sup>.

Ainsi, les règles de désambiguïsation utilisées à l'écrit semblent pouvoir être appliquées sur l'oral. Par ailleurs, les caractéristiques des productions orales n'ont pas une influence majeure sur la qualité de l'étiquetage<sup>21</sup>. Nous avons pourtant choisi de ne pas utiliser un étiqueteur conçu pour l'écrit. D'une part, comme le notent [Valli et Véronis, 1999], la robustesse de l'analyseur CORDIAL face aux perturbations l'oral s'explique vraisemblablement par sa nature initiale de correcteur orthographique et grammatical. Une telle robustesse ne se retrouverait pas forcément sur les étiqueteurs traditionnels utilisant en particulier, pour la désambiguïsation, les marques de ponctuation. D'autre part, cette expérience a été menée sur de l'oral transcrit respectant les contraintes d'accords en genre et en nombre. Il est difficile de pouvoir apprécier *a priori* l'influence d'erreurs de reconnaissance cumulées concernant ces aspects sur la qualité de l'étiquetage.

Par ailleurs, le jeu d'étiquettes utilisé par CORDIAL, qu'il n'est pas possible de maîtriser, n'est pas adapté à nos objectifs<sup>22</sup>. Nous pensons en particulier à l'absence d'étiquette spécifique pour l'emploi phatique de formes comme *bon, voilà, quoi*, etc. très courantes à l'oral. L'identification de l'emploi particulier de ces formes est relativement importante pour la caractérisation des pauses remplies en cas de répétitions et de correction. Il est également important, par exemple, de faire la distinction entre les emplois de la forme *quoi*, fréquente à l'oral comme élément phatique et dans le cadre de notre domaine applicatif pour l'expression de questions (*les horaires c'est quoi*).

Enfin, l'analyseur CORDIAL, comme beaucoup d'étiqueteurs, utilise à la fois des règles de désambiguïsation explicites et des informations statistiques acquises sur des corpus de texte. Il ne fournit ainsi que l'étiquette la plus probable en cas d'ambiguïté. Or une précision élevée de l'étiquetage n'est pas un critère essentiel dans l'approche ascendante que nous proposons. Il est plutôt préférable de ne pas chercher à trop désambiguïser dès cette étape afin d'éviter que les erreurs ne se répercutent sur les niveaux supérieurs de l'analyse. Nous verrons par ailleurs que la segmentation, comme l'étape de rattachement sur des critères sémantico-pragmatiques, permettent progressivement de lever une partie de ces ambiguïtés résiduelles.

Nous avons ainsi décidé de développer un étiqueteur minimal, adapté à nos besoins et présentant un compromis acceptable entre taux de précision et taux de décision<sup>23</sup>.

<sup>20</sup>Les performances sont exprimées en terme de pourcentage d'étiquettes *acceptables*, c'est-à-dire que ne sont comptabilisées que les erreurs flagrantes sur lesquelles un consensus peut être établi. Sur l'écrit ces performances sont de l'ordre de 95% à 98%. On se reportera à [Valli et Véronis, 1999] pour une discussion sur la difficulté de l'interprétation de ces pourcentages.

<sup>21</sup>Cette expérience a rapporté une baisse de seulement 0.9% des performances par rapport à un étiquetage similaire de la même transcription réécrite (suppression des hésitations, faux départs et ajout de la ponctuation).

<sup>22</sup>Concernant la tâche de segmentation proprement dite, ce jeu d'étiquettes est sur-spécifié. Cette sur-spécification n'est toutefois pas un problème majeur, une étape de normalisation des étiquettes pouvant être mise en œuvre pour éviter de faire intervenir un trop grand nombre d'étiquettes dans le processus de segmentation.

<sup>23</sup>Nous n'excluons pas à terme d'évaluer l'analyseur CORDIAL sur des énoncés oraux issus d'un système de dictée vocale ni de l'adapter (traitement des sorties) à notre problématique ou de l'utiliser en parallèle à notre étiqueteur pour obtenir facilement une caractérisation plus fine. Nous pensons en



### 5.3.2.2 Jeu d'étiquettes retenu

Tout système d'étiquetage adopte un inventaire d'étiquettes particulier. Le choix du jeu d'étiquettes retenu est extrêmement variable d'un système à l'autre en fonction par exemple des choix linguistiques, du détail de l'analyse morphologique visé ou de la langue traitée. Pour le français, le nombre d'étiquettes peut dépasser 300 (dans le cas du projet d'évaluation GRACE par exemple [Adda, Mariani, Paroubek, Rajman, et Lecomte, 1999]). Nous avons retenu un jeu d'étiquettes minimal suffisant pour notre tâche de segmentation.

Les critères morphologiques d'accords en genre et en nombre, n'étant pas fiables, ne sont pas pris en compte. Nous conserverons toutefois la distinction pluriel/singulier portée par les déterminants et les pronoms personnels et démonstratifs. Ce choix est guidé par notre volonté de fournir une caractérisation assez fine de la structure interne du syntagme. Les déterminants, en particulier, sont les éléments des énoncés issus de la reconnaissance les plus sûrs pour véhiculer cette information. De la même manière, nous conservons un certain nombre de catégories mineures qui ne nous sont utiles que pour l'identification de la structure interne du groupe nominal. Nous ferons ainsi une distinction entre adjectif possessif, adjectif démonstratif, article indéfini et défini, alors qu'au cours du processus de segmentation ces différents éléments ont tous le rôle identique de déterminant.

Par ailleurs, nous introduisons des catégories artificielles concernant les adverbes. Il s'agit de faire intervenir, au moment de la segmentation, le moins d'information lexicale possible pour l'identification des groupes grammaticaux. Nous avons besoin, par exemple, de distinguer les adverbes de quantité (*très, trop, assez, etc.*) des autres adverbes et parmi eux derniers les formes *plus* et *moins* dans leur usage comme superlatifs<sup>24</sup>.

Notre jeu d'étiquettes comporte ainsi 34 étiquettes grammaticales (40 en comptant la distinction pluriel/singulier sur 6 d'entre elles) répertoriées dans le tableau 5.2.

### 5.3.2.3 Étiquetage par cascades de transducteurs

Le processus d'étiquetage est réalisé en plusieurs étapes successives appliquant une cascade de transducteurs :

- Une première étape permet le repérage de certaines locutions courantes. L'application du transducteur permet de regrouper les mots impliqués dans ces locutions afin qu'elles puissent correspondre à une unité lexicale. Nous regroupons ainsi au total une centaine de locutions : les expressions les plus fréquentes du domaine (*au revoir, s'il vous plaît, bon marché*), les nombres, les locutions adverbiales (*à peu près, tant pis, bien sûr, en fait, ...*), prépositionnelles (*à partir de, jusqu'à, ...*)

---

particulier aux temps des verbes que nous ne traitons pas pour l'instant complètement.

<sup>24</sup>Ce choix est discutable d'un point de vue linguistique mais permet de mieux répondre à nos objectifs. Aucune ambiguïté supplémentaire n'est introduite par cette spécialisation portant sur quelques formes d'une même catégorie et nos objectifs de généralité sont préservés.

Étiquette	Correspondance	Étiquette	Correspondance
adj	adjectifs	v <sub>pp</sub>	participes passés
adj <sub>int</sub>	adjectifs interrogatifs	aux <sub>pp</sub>	auxiliaire au participe passé
num	numéraux	mod <sub>pp</sub>	modaux au participe passé
adv	adverbes “classiques”	v <sub>ppr</sub>	participes présents
adv <sub>q</sub>	adverbes de quantité <i>très, trop, ...</i>	mod <sub>ppr</sub>	modaux au participe présent
adv <sub>qs</sub>	adverbes de quantité pouvant être utilisés comme superlatifs <i>plus, moins</i>	aux <sub>ppr</sub>	auxiliaires au participe présent
neg	adverbes de négation	v <sub>inf</sub>	verbes à l’infinitif
conj <sub>coo</sub>	conjonctions de coordination	mod <sub>inf</sub>	modaux à l’infinitif
art <sub>ind</sub> [s p]	articles indéfinis	nom	noms communs
art <sub>d</sub> [s p]	articles définis	nomp	noms propres
adj <sub>po</sub> [s p]	adjectifs possessifs	pr <sub>pers</sub> [s p]	pronoms personnels
adj <sub>dem</sub> [s p]	adjectifs démonstratifs	pr <sub>dem</sub> [s p]	pronoms démonstratifs
prep	prépositions	pr <sub>ref</sub>	pronoms réfléchis
v	verbes conjugués	pr <sub>rel</sub>	pronoms relatifs
aux	auxiliaires conjugués	pr <sub>int</sub>	pronoms interrogatifs
mod	modaux (ou assimilés) conjugués	conj <sub>sub</sub>	conjonctions de subordination
		interj	interjections
		x	éléments phatiques

TAB. 5.2 – Jeu d’étiquettes grammaticales retenu pour la segmentation.

et les conjonctions (*afin que*). Elle permet également de repérer quelques mots composés fréquents et des structures particulières (*est ce que* par exemple). Deux passes sont réalisées. La première ne tient pas compte du contexte, la seconde utilise le contexte droit ou gauche pour éviter un regroupement erroné de quelques locutions prépositionnelles (*proche de* par exemple dans un énoncé comme *l’hôtel le plus proche de la gare*)<sup>25</sup>.

- La seconde permet l’accès au lexique et l’attribution à chaque mot de l’énoncé de l’ensemble de ses étiquettes. Notre lexique est composé d’environ 45 000 formes fléchies. Nous avons constitué ce lexique à partir d’un sous-ensemble de deux lexiques électroniques diffusés librement<sup>26</sup>. Nous y avons intégré une partie des mots les plus fréquents comptabilisés par Jean Véronis sur un corpus de français parlé d’un

<sup>25</sup>Ce repérage des locutions est très grossier. Nous nous contentons d’identifier les locutions non ambiguës les plus fréquentes. Nous envisageons à terme l’utilisation d’un étiqueteur tel que CORDIAL qui permet de régler de façon fine le regroupement des mots en locutions.

<sup>26</sup>Nous avons utilisé le lexique de l’ABU (Association des Bibliophiles Universels) contenant quelques 300 000 entrées lexicales (formes fléchies) associées à leurs étiquettes grammaticales et la base de données “Lexique” contenant 55 000 lemmes (129 000 formes fléchies) calculés à partir de deux sources : les pages internet françaises indexées par le moteur de recherche *FastSearch* et un recueil de textes publiés après 1950, gérés par l’ATILF (Analyses et Traitements Informatiques du Lexique Français). Ces ressources sont respectivement accessibles en ligne aux adresses <http://www.abu.cnam.fr/DICO> et <http://www.lexique.org>

million de mots<sup>27</sup>. Nous l'avons par ailleurs enrichi des "troncations" courantes à l'oral telles que *ciné*, *resto*, *appart*, *impec*, etc. et de quelques anglicismes (payer *cash* par exemple)<sup>28</sup>. Ce lexique est ainsi assez significatif de la marque lexicale de l'oral<sup>29</sup>. À noter que seule une trentaine de noms propres (noms de rue, titres de film, ...) ont été intégrés afin de pouvoir évaluer l'ensemble du système.

Le lexique a été compilé en un transducteur dont l'alphabet est composé d'un seul caractère (principalement lettres, espaces, et chiffres), réalisant la fonction identité sur les mots de l'énoncé et ajoutant en fin de mots la liste des étiquettes grammaticales de la forme reconnue. Un exemple très simple de ce transducteur est donné par la figure 5.1, représentant les formes *le*, *la*, *les* et *des* et une partie de leurs étiquettes associées (article, pronom personnel et préposition, représentées respectivement par les chiffres 1, 2 et 3). La forme *le* dans un énoncé sera ainsi transformée en la structure *le/1#2* (le # indiquant une ambiguïté grammaticale pour cette forme). Le transducteur complet comporte environ 10 000 états et 20 000 transitions.

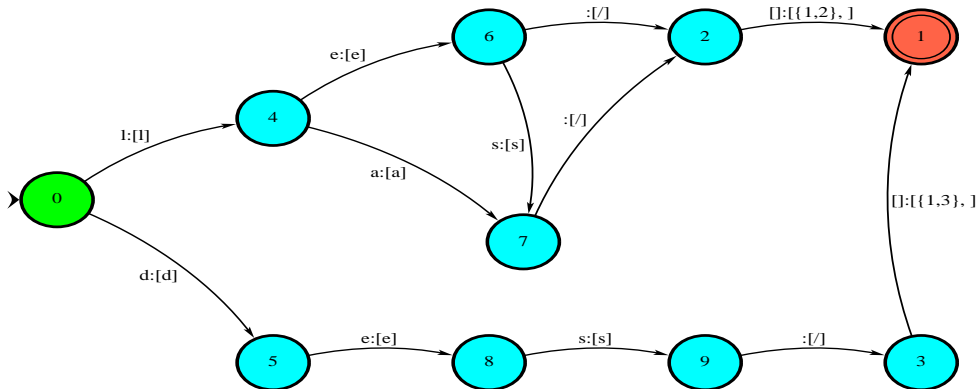


FIG. 5.1 – Exemple simplifié d'accès au lexique par un transducteur

Remarquons que le caractère finalisé de l'application et la tâche de segmentation considérée, nous permet de ne pas représenter dans le lexique toutes les ambiguïtés qu'autorise le jeu d'étiquettes retenu. C'est le cas principalement des mots constituant à eux seuls un syntagme et difficiles à étiqueter. Ainsi le mot *comme* par exemple n'est considéré dans notre lexique qu'en tant que conjonction de subor-

<sup>27</sup>Le corpus de référence est *Corpaix*. Ces données sont accessibles depuis <http://www.up.univ-mrs.fr/veronis>

<sup>28</sup>À titre anecdotique, nous y avons également inclus un certain nombre de termes argotiques et familiers d'après [Dontchev, 2000].

<sup>29</sup>La langue orale se fonde sur un ensemble familier de mots (ce que l'on appelle le *lexique utile*), de l'ordre de 2000 mots en français.

dination. De la même manière, nous ne traitons pas l'ambiguïté entre conjonction de subordination et pronom relatif de la forme *que*. Cette ambiguïté est difficile à lever par la plupart des étiqueteurs car elle nécessite la prise en compte d'un contexte large. Nous avons ainsi décidé d'utiliser une étiquette ambiguë "introduc- teur subordonnant" pour représenter cette ambiguïté pour la forme *que*<sup>30</sup>. Ce sera le rôle de l'analyse sémantico-pragmatique de traiter par la suite les différentes relations que ce syntagme peut entretenir avec les autres éléments de l'énoncé. Remarquons enfin qu'une étiquette particulière est affectée aux mots absents de notre lexique.

- Enfin, la dernière étape applique quelques règles de désambiguïstation contextuelles fiables (utilisées à l'écrit<sup>31</sup>) fondées sur des informations lexicales ou grammati- cales. Nous levons par exemple l'ambiguïté de la forme *le* entre article défini et pronom personnel (*le* est un pronom personnel s'il est précédé de l'adverbe de né- gation *ne* ou s'il est suivi d'un verbe conjugué). Un autre exemple peut être donné par la forme *quel* présentant une ambiguïté entre adjectif interrogatif et pronom interrogatif (la règle considérée affecte à cette forme l'étiquette d'adjectif interro- gatif si elle est suivie d'un nom)<sup>32</sup>. 5 transducteurs sont appliqués succesivement, chacun d'entre eux encodant plusieurs règles dont les contextes d'application ne se recouvrent pas.

Chacun des transducteurs a été compilé à partir d'expressions régulières au moyen de l'utilitaire FSA [Noord, 1997b]. Ils ont été rendus déterministes<sup>33</sup> en appliquant l'algo- rithme décrit dans [Roche et Schabes, 1997a] et sont représentés de manière compacte en C d'après la technique décrite dans [Dacuik, 1998, page 43]<sup>34</sup>. L'utilisation des trans- ducteurs et cette représentation optimisée fournit un accès au lexique et un étiquetage en temps linéaire.

#### 5.3.2.4 Évaluation

L'évaluation de l'étiqueteur a été réalisée sur 1200 énoncés récoltés lors de la cam- pagne d'évaluation DEFI du Groupe de Travail *compréhension robuste* du GDR-I3 du

---

<sup>30</sup>la forme *que* n'est ainsi présente dans notre lexique qu'en tant que pronom interrogatif et "intro- duc- teur subordonnant".

<sup>31</sup>l'expérience précédemment évoquée ayant démontré leur adéquation pour le traitement de l'oral.

<sup>32</sup>Pour la désambiguïstation de certaines formes spécifiques à l'oral, nous nous sommes reportés à quelques travaux de linguistique de corpus attestant de l'emploi de ces formes à l'oral. Ainsi nous avons consulté [Chanet, 2001] pour l'étude des différents emplois de la forme *quoi* à l'oral, pour laquelle il est impératif de distinguer dans notre cadre applicatif ses emplois comme pronom interrogatif et comme élément phatique.

<sup>33</sup>Un automate déterministe est un automate dans lequel il y a au plus une transition à partir d'un état pour un symbole donné. Cette définition est valable pour les transducteurs et est indépendante du nombre de sorties qu'ils peuvent générer étant donnée une chaîne d'entrée. On pourra se reporter à [Autebert, 1994] pour une introduction théorique aux automates à états finis et à [Roche et Schabes, 1997b] pour leur utilisation en traitement automatique du langage.

<sup>34</sup>L'implémentation retenue (*perfect hash tables*) permet une réduction de l'espace mémoire utilisé. À titre d'exemple, le transducteur encodant le lexique a une taille d'environ 450 Ko.

CNRS [Antoine et al., 2002]. Il s'agit d'énoncés oraux simulés, dérivés à partir d'énoncés représentatifs du domaine en y intégrant des phénomènes spécifiques de l'oral. Nous reviendrons sur la production de ces énoncés au chapitre 7. Quelques exemples pourront être trouvés en annexe D. Nous avons évalué notre étiqueteur en terme de taux de *précision* et de *décision* conformément à la méthodologie d'évaluation de l'action GRACE [Adda et al., 1999].

- la *précision* mesure la proportion d'étiquetages corrects parmi les étiquetages stricts (c'est-à-dire non ambigus) ;
- la *décision* mesure la proportion d'étiquetages stricts parmi l'ensemble de tous les étiquetages (stricts et ambigus)

Sur les 1200 énoncés testés nous obtenons un taux de précision de 97.5% et un taux de décision de 80.4%. On retrouve les mêmes erreurs qu'à l'écrit pour certaines formes difficiles à traiter. D'autres erreurs proviennent des répétitions, introduisant une perturbation du contexte local. Ces erreurs sont principalement présentes en cas de répétitions successives d'une même forme : c'est par exemple le cas pour la répétition de pronoms relatifs dont certains peuvent être étiquetés comme pronoms interrogatifs.

Il est difficile d'interpréter ces pourcentages et ils peuvent difficilement être comparés avec les performances des étiqueteurs conçus pour des objectifs différents des nôtres et avec un jeu d'étiquettes de granularité elle-même différente. Le taux de précision obtenu résulte ainsi des libertés (acceptables dans notre contexte) que nous avons prises par rapport à certaines formes traditionnellement considérées comme difficiles. Le taux de décision, assez faible, est quant à lui conforme à notre souhait de laisser aux niveaux supérieurs de l'analyse le soin de lever les ambiguïtés restantes.

### 5.3.3 Segmentation

#### 5.3.3.1 Segmentation par cascades de transducteurs

Comme pour l'étiquetage, le processus de segmentation réalisé résulte de l'application d'une séquence finie et ordonnée de transducteurs. Nous utilisons ces transducteurs pour introduire des marqueurs de délimitation autour des instances des syntagmes que nous cherchons à identifier.

L'analyse est incrémentale. Il s'agit de pouvoir identifier le plus précisément possible la structure interne des syntagmes. Par exemple la segmentation de l'énoncé *dans un très bon restaurant* est illustrée figure 5.2<sup>35</sup>. Cette stratégie nous permet ainsi d'identifier le groupe prépositionnel et de conserver les informations liées à sa structure interne : il est constitué d'une préposition et d'un groupe nominal ; ce groupe nominal inclut un groupe adjectival qualifiant le nom, lui-même qualifié par un groupe adverbial.

Chacun des groupes élémentaires, les groupes grammaticaux et les expressions langagières considérées (date, heure, adresse, etc.) est décrit au moyen d'expressions régulières compilées en transducteurs. Comme pour la tâche d'étiquetage, nous utilisons

<sup>35</sup>dans un souci de lisibilité, nous ne représentons pas les étiquettes grammaticales.

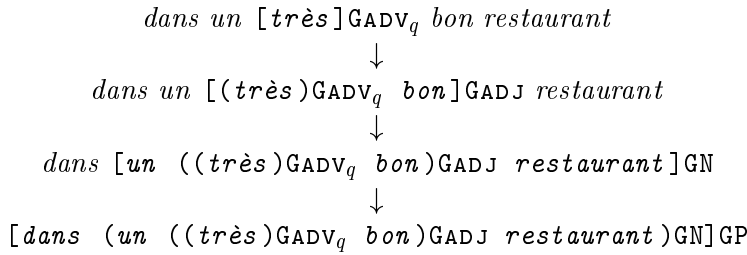


FIG. 5.2 – Illustration du processus de segmentation

l'utilitaire FSA [Noord, 1997b]. Cet outil permet l'utilisation des opérateurs réguliers fondamentaux et la conception aisée d'opérateurs spécifiques<sup>36</sup>. On pourra se reporter à [Karttunen, Chanod, Grefenstette, et Schiller, 1996; Chanod et Tapanainen, 1996; Roche, 1997; Noord et Gederman, 1999] pour un aperçu de l'utilisation et de l'intérêt des expressions régulières et des transducteurs pour le traitement automatique des langues, en particulier pour la tâche de segmentation [Kaiser, 1999]. Dans notre approche, nos motivations concernent essentiellement l'efficacité de l'analyse, l'homogénéisation des formalismes utilisés pour l'étiquetage et la segmentation, ainsi que la simplicité de description des grammaires considérées.

L'analyse est déterministe. La gestion de l'ambiguïté est réalisée en favorisant les groupes syntagmatiques les plus longs. Ainsi le syntagme *aimerais pouvoir faire* sera considéré comme un seul syntagme<sup>37</sup> :

$$[\text{aimerais pouvoir (faire)GV}_{inf}] \text{GV}_{mod}$$

et non :

$$[\text{aimerais (pouvoir)GV}_{inf}] \text{GV}_{mod} [\text{faire}] \text{GV}_{inf}$$

Les groupes grammaticaux considérés sont minimaux et non récursifs, cette identification peut ainsi être réalisée par l'application en cascade des transducteurs (les groupes susceptibles d'être inclus dans d'autres groupes étant identifiés en premier). La gestion du parenthésage interne au syntagme est gérée par un transducteur particulier noté  $cat(A)$  permettant de remplacer toute séquence de la forme  $[B]A$  en  $(B)A$ <sup>38</sup>.

Les règles utilisées pour les groupes grammaticaux suivent les définitions que nous avons données dans le tableau 5.1 et reposent essentiellement sur les catégories gram-

<sup>36</sup>Cet utilitaire a été conçu pour le traitement automatique des langues [Noord et Gederman, 1999].

<sup>37</sup>La tête lexicale de ce groupe est le verbe infinitif le plus à droite.

<sup>38</sup> $B$  dénote toute séquence de caractères de l'alphabet du transducteur à l'exception des crochets utilisés pour la délimitation des syntagmes.

$$cat(A) \stackrel{def}{=} [ ' [ ' : ' ( ' , ( ? - \{ [ ] ; ' \} ) * , ' ] : ' ) ' , A ]$$

Un autre opérateur  $catm(A, M)$  nous permet de réaliser la même opération en spécifiant cette fois le mot interne au syntagme (le remplacement de  $[B]A$  n'est réalisé que si  $B$  contient le mot  $M$ ).

maticales. Ainsi, la règle suivante est l'une des règles permettant d'identifier les groupes nominaux<sup>39</sup> :

$$[ @DET^{\wedge}, num^{\wedge}, cat(GADJ)^*, nom ]$$

Les règles exprimant les heures, les dates, les adresses, etc. ne font intervenir essentiellement que des informations lexicales. Contrairement aux groupes grammaticaux minimaux, nous utilisons pour ces dernières un opérateur spécifique de remplacement pour assurer l'identification des groupes maximaux en une seule passe. Cet opérateur est également utilisé pour identifier la couverture maximale des marques des hésitations, répétitions et des répétitions<sup>40</sup>. L'utilisation de cet opérateur pour optimiser une analyse en *chunk* a été expérimentée dans plusieurs travaux [Roche, 1997; Karttunen et al., 1996; Grefenstette, 1996]. Son implémentation dans l'utilitaire FSA [Noord et Gederman, 1999] suit la définition donnée dans [Karttunen, 1996]. Il est dénoté par  $A@ \rightarrow B$  ( $A$  est la catégorie du syntagme à identifier ;  $B$  est le langage régulier dénotant ce syntagme). L'opérateur introduit autour de l'instance maximale de  $B$  les marqueurs de délimitation et la catégorie :  $[ B ]A$ . La règle suivante (simplifiée<sup>41</sup>) est utilisée pour l'identification des expressions correspondant aux dates<sup>42</sup> :

$$\begin{aligned} \text{DATE } @ \rightarrow \{ & @JOUR_{adv}, \\ & [ @JOUR, prochain/adj ], \\ & [ (le/art_{defS})^{\wedge}, \{ @NUM_{mois}, [ (sur)^{\wedge}lendemain/nom ], \}, \\ & [ (le/art_{defS})^{\wedge}, @JOUR, @NUM_{mois}^{\wedge} ], \\ & [ [ (le/art_{defS})^{\wedge}, @JOUR^{\wedge}, @NUM_{mois} ]^{\wedge}, @MOIS ] \} \end{aligned}$$

Notre analyse est réalisée en 4 étapes. Chacune de ces étapes applique simultanément plusieurs transducteurs. La séquence ordonnée des traitements est présentée figure 5.3.

Pour un même énoncé, nous ne conservons pas, à l'issue de l'étape de segmentation, toutes les séquences produites par l'étiquetage. Nous ne conservons pour l'étape supérieure, que les séquences ayant conduit à la meilleure couverture locale en syntagme. Seuls les mots outils (c'est-à-dire les déterminants et les prépositions, qui ne peuvent

<sup>39</sup>Dans cet exemple, la signification des opérateurs réguliers est la suivante :  $[A,B]$  dénote la concaténation des langages  $A$  et  $B$  et  $\{A,B\}$  l'union de  $A$  et  $B$ .  $A^{\wedge}$  indique que élément  $A$  est optionnel et  $A^*$  indique 0 ou plusieurs éléments  $A$ .  $@DET$  regroupe l'ensemble de nos étiquettes jouant le rôle de déterminant.

<sup>40</sup>Notons que l'heuristique favorisant les groupements les plus longs aurait pu être gérée sans cet opérateur en modifiant l'algorithme général de l'application des transducteurs sur la chaîne d'entrée. L'incorporation dans l'algorithme général du nombre de cas à traiter rend toutefois cette solution assez coûteuse pour l'identification des dates et des heures. L'utilisation de l'opérateur de [Karttunen, 1996] n'est qu'un exemple des traitements optimisés que l'on peut envisager pour le traitement automatique des langues reposant sur les automates à états finis [Roche et Schabes, 1997b].

<sup>41</sup>Sont aussi considérées des expressions du type *mi décembre, fin décembre, début décembre*. En revanche, un énoncé du type *pour le début du mois de décembre* sera segmenté en 3 syntagmes :

$$[pour (le début)GN]GP [du (mois)GN]GP [de (décembre)DATE]GPDATE$$

dont seul le dernier porte l'indentifiant de date.

<sup>42</sup>Dans cet exemple,  $@JOUR$  représente les jours de la semaine (*lundi...dimanche*) ;  $@MOIS$  les mois de l'année ;  $@NUM_{mois}$  les nombres de *deux* à *trente-et-un* et  $@JOUR_{adv}$  les adverbes *aujourd'hui, demain, après-demain*.

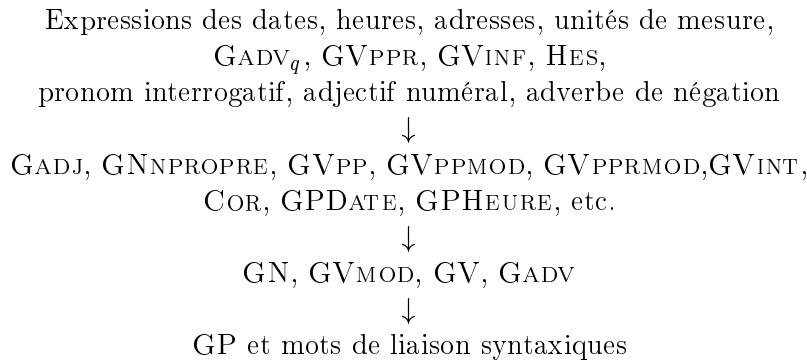


FIG. 5.3 – Séquence ordonnée de l'application des transducteurs

constituer un syntagme à eux seuls) sont concernés par cette élimination d'hypothèses. Si un de ces éléments n'est pas intégré dans un segment dans une des séquences d'étiquetage de l'énoncé, alors qu'il est intégré dans un segment situé à sa droite dans une autre analyse, nous ne conservons que cette dernière hypothèse. Nous illustrons ce cas de figure sur les fragments d'énoncés suivants : *dans la gare, le premier, un des hôtels*. Chacun de ces fragments est ambigu à l'issue de l'étiquetage<sup>43</sup> :

dans/prep la/pr<sub>persS</sub> gare/v le/art<sub>defS</sub> premier/adj  
 dans/prep la/art<sub>defS</sub> gare/nom le/art<sub>defS</sub> premier/nom  
 un/art<sub>indS</sub> des/prep hôtels/nom  
 un/nom des/prep hôtels/nom

Chacune de ces séquences étiquetées donne lieu à une unique segmentation :

dans/prep [la/pr<sub>persS</sub>]GN [gare/v]GV le/art<sub>defS</sub> [premier/adj]GADJ  
 [dans/prep (la/art<sub>defS</sub> gare/nom)GN]GP [le/art<sub>defS</sub> premier/nom]GN  
 un/art<sub>indS</sub> [des/prep (hôtels/nom)GN]GP  
 [un/nom]GN [des/prep (hôtels/nom)GN]GP

Pour le cas du premier fragment, la préposition *dans* ne fait partie d'aucun syntagme dans la première analyse, alors qu'elle est intégrée dans un syntagme prépositionnel dans la seconde. Seule la seconde hypothèse sera conservée. Ce choix est identique pour le second fragment. En revanche, pour le troisième fragment, même si la forme *un* n'est pas intégrée dans un segment dans la première séquence, elle n'est pas non plus intégrée à un syntagme comportant le mot situé à sa droite (puisqu'elle constitue un syntagme à elle seule) dans la seconde. Dans ce cas de figure, toutes les hypothèses seront conservées pour l'étape supérieure. Ce choix est légitime car ce fragment peut, en dehors de tout contexte global, soit correspondre à une hésitation provoquant la répétition du début d'un syntagme, soit être interprété par *un des hôtels* (parmi ceux précédemment évoqués).

<sup>43</sup>Le dernier fragment possède 4 séquences d'étiquetage différentes. Nous ne représentons pas l'ambiguïté sur la forme *des* entre préposition contractée et article indéfini pluriel dans un souci de lisibilité.



L'étape de segmentation permet ainsi de réduire de plus de moitié l'ambiguïté résultante de l'étape d'étiquetage.

Remarquons enfin que les mots outils non intégrés dans un syntagme dans les hypothèses d'étiquetage retenues, sont éliminés de l'énoncé. Ainsi, dans la première hypothèse correspondant au dernier fragment (hypothèse correspondant à une répétition), l'article *un* sera ignoré dans les traitements supérieurs. Ce premier filtrage est similaire à celui qu'auraient pu donner des méthodes par *patterns* mais résulte d'un processus moins aveugle que ces dernières approches.

### 5.3.3.2 Évaluation

Les performances dépendent de celles du processus d'étiquetage. Les erreurs attestées sur la segmentation d'une séquence étiquetée donnée, par rapport à une segmentation manuelle sur cet énoncé brut (c'est-à-dire non étiqueté), ne sont liées qu'à la précision de l'étiquetage pour cette séquence. L'ambiguïté résiduelle à l'issue de l'étape de segmentation est préservée. Ceci nous garantit d'avoir conservé la segmentation "correcte" de l'énoncé sous réserve qu'une séquence d'étiquetage acceptable soit présente parmi ces hypothèses. Le caractère minimal des syntagmes que nous avons retenus ainsi que l'analyse incrémentale réalisée nous interdisent par ailleurs des regroupements de syntagmes erronés.

On notera cependant quelques cas d'élimination abusive de prépositions alors que dans l'optique de l'approche proposée, ces prépositions auraient dû faire partie d'un syntagme. C'est le cas par exemple du fragment d'énoncé *pour euh voir un film* dans lequel l'interjection *euh* ne nous permet pas la caractérisation du groupe prépositionnel *pour voir*. Ceci n'est cependant pas lié au processus de segmentation en lui-même mais à la définition des syntagmes que nous avons considérés. Ces cas, non conformes aux hypothèses adoptées des régularités de l'oral, semblent cependant très rares. Si la non répétition de la préposition dans l'exemple précédent semble légitime après une courte pause remplie<sup>44</sup>, elle semble moins vraisemblable dans un cas tel que *les horaires des euh films*. D'après nos études sur le corpus Murol, ces cas ne semblent se produire que pour certains types de groupes prépositionnels (*pour, sans* suivi d'un groupe verbal par exemple). Nous envisageons à terme de mener une étude de corpus plus détaillée sur ce point précis.

Nous évaluerons enfin l'influence des erreurs de reconnaissance sur cette analyse au chapitre 7.

Le segmenteur (et l'étiqueteur sur lequel il repose) seront évalués précisément dans le cadre du projet *Technolangue* EASY auquel nous participons. Ce projet concerne l'évaluation des analyseurs syntaxiques. Notre système sera uniquement évalué sur l'étape de segmentation. Les objectifs de ce projet sont intéressants car ils visent notamment à évaluer les différents systèmes sur des genres différents : corpus journalistiques, énoncés

---

<sup>44</sup>En supposant d'ailleurs qu'elle soit détectée par le module de reconnaissance. Nous reviendrons sur ce point au chapitre 7 §7.2.3.

oraux transcrits et courrier électronique. Il sera intéressant d'évaluer notre segmenteur et de comparer ses performances avec les autres systèmes participants sur des données proches de celles pour lesquelles notre segmenteur a été conçu.

## 5.4 Conclusions

Cette première étape d'analyse permet un découpage de l'énoncé. Ce découpage, linguistiquement motivé, reflète la façon dont les mots se regroupent en surface et caractérise les unités susceptibles d'être impliquées dans les principaux phénomènes de l'oral : la production des inattendus structuraux et les mouvements dans l'énoncé.

Ces unités se rapprochent des *segments conceptuels* envisagés dans les méthodes sélectives. À la différence de ces dernières approches, nous ne préjugeons pas, à ce niveau de l'analyse, de leur existence *a priori* en fonction du domaine d'application envisagé ni de leur position en un point donné de l'énoncé.

On peut résumer l'intérêt de ce traitement par les points suivants qui soulignent l'importance du rôle de la syntaxe au niveau local des syntagmes :

- **Généricité.** En faisant reposer la caractérisation de ces syntagmes sur des critères grammaticaux minimaux, nous garantissons la portabilité et la réutilisabilité de cette partie du système à d'autres domaines applicatifs. La séparation des traitements relevant de critères liés à la langue de ceux liés au domaine d'application traité nous semble être, à l'image de [Seneff, 1998], un critère essentiel pour le développement de systèmes de compréhension portant sur des domaines plus riches. La capitalisation des traitements permettant la caractérisation des structures communes à plusieurs domaines est de ce point de vue également importante, ne serait-ce que pour la réduction des modèles à mettre en œuvre en compréhension sélective. L'approche adoptée par cascades de transducteurs souligne que ces traitements peuvent efficacement et relativement simplement être mis en œuvre.
- **Couverture et analyse de la structure interne des syntagmes.** Cette analyse ne se limite pas à la seule identification des segments portant le sens dit "utile" de l'énoncé. La segmentation couvre la totalité de l'énoncé et permet la conservation de chacun des syntagmes identifiés (marques des hésitations et des corrections, mots de liaison syntaxiques) qui constituent autant d'indices nécessaires à une analyse détaillée de l'énoncé. La caractérisation de la structure interne des syntagmes grammaticaux dès ce niveau de l'analyse permet en outre l'extraction précise d'informations – propriétés morphologiques retenues et dépendances locales (adjectifs qualifiant le nom pour les groupes nominaux par exemple) – qui n'auront plus à être analysées à l'étape suivante.
- **Absence de prétraitements particuliers.** L'exploitation, par le découpage syntagmatique, des régularités observées dans la production orale permet enfin d'accepter les phénomènes "perturbateurs" sans devoir recourir aux étapes délicates de prétraitement parfois envisagées. Elle permet en outre un premier filtrage des mots les plus fréquemment répétés (ceux débutant un syntagme).

Enfin cette première étape est surtout l'amorce d'un traitement robuste des énoncés. Comme le note [Abney, 1996] :

(Cit.5.4) *“By reducing the sentence to chunks, there are fewer units whose associations must be considered, and we can have more confidence that the pairs being considered actually stands in the syntactic relation of interest, rather than being random pairs of words that happen to appear near each other”*[Abney, 1996]

Le chapitre suivant explique la façon dont nous rattachons les syntagmes identifiés pour la construction de la représentation sémantique.



## Chapitre 6

# Analyse sémantico-pragmatique détaillée

### 6.1 Introduction à l'analyse par dépendance

Comme nous l'avons souligné au chapitre précédent (§ 5.2.1), notre système s'inscrit dans la lignée des analyseurs syntaxiques hybrides développés pour le TAL robuste [Grefenstette, 1992; Giguët, 1998; Aït-Mokhtar et Chanod, 1997; Basili et al., 1999], associant analyse syntaxique de surface en *chunks* puis mise en relation de ces syntagmes fondée sur un formalisme inspiré de la notion de *dépendance*.

L'une des premières théories linguistiques fondées sur la dépendance est celle de [Tesnière, 1959] :

(Cit.6.1) *La phrase est un ensemble organisé dont les éléments sont les mots. Tout mot qui fait partie d'une phrase cesse par lui-même d'être isolé comme dans le dictionnaire. Entre lui et ses voisins, l'esprit aperçoit des connexions, dont l'ensemble forme la charpente de la phrase. (...) Les connexions structurales établissent entre les mots des rapports de dépendance.* [Tesnière, 1959, page 11]

Cette vision de la structure de la phrase explique ainsi la façon dont les mots, par leur présence, dépendent les uns des autres. Étant donnés deux mots du langage, on établit entre eux une relation de dominé (*dépendant*) à dominant (*gouverneur*) que l'on peut représenter par un arc entre deux nœuds, chaque nœud étant étiqueté par un mot. Cette représentation de l'organisation syntaxique de la phrase diffère de celle adoptée par les grammaires syntagmatiques. Dans ces dernières, elle correspond à une vision hiérarchique de la phrase décrite par la façon dont les mots peuvent être groupés selon la notion grammaticale abstraite de constituants<sup>1</sup>.

---

<sup>1</sup>Il existe, entre ces deux paradigmes, des oppositions plus fondamentales que la représentation adoptée de l'organisation syntaxique de la phrase. Nous ne les aborderons pas dans cette présentation succincte des grammaires de dépendance.

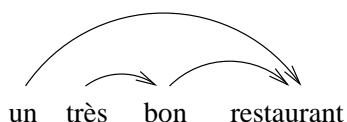
Il n'existe pas de formalisme de référence pour les grammaires de dépendance, comme peuvent l'être les grammaires de réécriture hors-contexte de [Chomsky, 1957] pour les grammaires syntagmatiques. Ainsi les différents formalismes et théories fondés sur cette notion (théorie Sens-Texte de [Mel'čuk, 1988], *Word Grammar* [Hudson, 1984], *Functional Grammar* [Halliday, 1994], *Constraint Grammar* [Karlsson et al., 1995] pour n'en citer que quelques unes) n'ont parfois comme seul point commun que de considérer des relations orientées entre mots. On notera par ailleurs que la plupart des modèles linguistiques issus de la mouvance chomskienne (GB, LFG, HSPG) ont introduit l'usage de la dépendance syntaxique sous des formes plus ou moins cachées (fonctions syntaxiques, constituants avec tête, cadre de sous-catégorisation) [Kahane, 2001].

Courant marginal jusqu'à la fin des années 80, la dépendance est revenue au premier plan en raison de deux évolutions des recherches en linguistique et en linguistique informatique : la tendance à la lexicalisation d'une part et la réintégration de la composante sémantique dans les systèmes de TAL d'autre part. Ces deux évolutions sont liées, l'intégration de connaissances sémantiques relevant naturellement du lexique : la sémantique est soit propre au mot soit décrite par les relations que ce mot peut entretenir avec les autres. On notera que les grammaires complètement lexicalisées (comme TAG ou les grammaires catégorielles), qui construisent des phrases en combinant des structures élémentaires associées aux mots, produisent naturellement des liens entre les mots et donc des structures de dépendance<sup>2</sup>. En ce qui concerne la sémantique, les structures de dépendance, qui distinguent l'ordre des mots et la structure syntaxique proprement dite, se rapprochent davantage d'une structure sémantique que ne le fait une structure syntagmatique. Les relations sémantiques prédicat-argument (parfois appelées dépendances sémantiques), que nous cherchons à caractériser dans le cadre qui nous s'intéresse, coïncident d'ailleurs en partie avec les dépendances syntaxiques [Kahane et Mel'čuk, 1999].

Alors que Tesnière définissait la connexion entre les mots, nous appliquons cette notion de sorte qu'elle relie non plus tous les mots mais seulement les têtes lexicales des syntagmes minimaux caractérisés par l'analyse syntaxique de surface. La structure de ces groupes élémentaires (que nous qualifierons dans la suite de ce document de *segments minimaux*) peut en effet être assimilée à un arbre de dépendance local articulé autour de la tête lexicale :

[*un* ((*très*)GADV<sub>q</sub> *bon*)GADJ *restaurant*]GN

≡



<sup>2</sup>La notion de hiérarchie reste toutefois présente.

Cette caractéristique différencie notre approche des principaux travaux en analyse syntaxique automatique fondés sur le modèle de la dépendance. Dans les paragraphes suivants, nous rappelons les principaux objectifs de cette mise en relation des segments puis présentons le formalisme que nous avons retenu pour effectuer cette analyse. Nous explicitons ensuite les informations liées aux segments minimaux que nous exploitons et la nature des relations impliquées. Concernant les formalismes et les travaux fondés sur la notion de dépendance, on pourra se reporter à [Kahane, 2000b, 2001] et au portail officiel des grammaires de dépendance<sup>3</sup>.

## 6.2 Mise en relation des segments dans le système ROMUS

### 6.2.1 Objectifs

Notre analyse consiste à combiner les unités lexicales considérées selon les relations sémantico-pragmatiques<sup>4</sup> qui existent entre elles afin d'en déduire la représentation sémantique de l'énoncé. La combinaison des unités ne peut pas être uniquement fondée sur les contraintes liées à ces unités (catégorie grammaticale ou propriétés morphologiques). Un processus d'analyse fondé sur la dépendance doit avoir recours à d'autres connaissances pour devenir maîtrisable et contrôler la combinatoire de l'instanciation des relations entre les unités. Plusieurs solutions sont envisageables comme par exemple :

- laisser la possibilité d'introduire des obligations ou interdictions de relations entre deux unités,
- permettre la définition de la position relative des unités devant être reliées.

Quels que soient les choix adoptés, l'intégration de ces contraintes ne doit pas faire oublier les objectifs de l'analyse des énoncés oraux. D'une part, notre analyse doit pouvoir rester partielle (en cas d'inachèvements ou d'erreurs de reconnaissance<sup>5</sup> par exemple). D'autre part, l'analyse des énoncés oraux doit prévoir le traitement de la variabilité faible de l'ordre des mots.

D'autres contraintes permettent toutefois de limiter la combinatoire des relations. C'est le cas de la contrainte de *projectivité* qui caractérise la structure relationnelle de l'ensemble de la phrase dans le cadre de la théorie des grammaires de dépendance. L'arbre de dépendance est dit *projectif* s'il est planaire dans le demi-plan délimité par les mots de la phrase [Mel'čuk, 1988]. En d'autres termes, les relations ne peuvent pas se croiser. Nous avons vu que les structures discontinues non projectives sont rares en français parlé. L'analyse peut ainsi reposer sur un formalisme projectif.

---

<sup>3</sup><http://ulaf.mff.cuni.cz/dg.html>

<sup>4</sup>Le terme pragmatique fait ici référence au domaine de l'application et non à la prise en compte du dialogue.

<sup>5</sup>Comme nous le verrons au chapitre 7 §7.2, les erreurs de reconnaissance peuvent se caractériser par une substitution, une insertion ou une élimination d'un mot dans l'énoncé. Chacune de ces erreurs peut provoquer une caractérisation erronée des segments de l'énoncé.

Le formalisme que nous avons expérimenté pour l'analyse sémantico-pragmatique, les grammaires de lien (*Link Grammars*), proches des grammaires catégorielles, a été conçu à l'université de Carnegie-Mellon [Sleator et Temperley, 1991]. Il a montré son efficacité pour l'analyse syntaxique robuste de textes libres et de corpus oraux [Grinberg, Lafferty, et Sleator, 1995]. Il a par ailleurs été utilisé avec succès pour la construction de représentations sémantiques (à partir des structures syntaxiques délivrées par l'analyseur) dans un système d'extraction automatique de réponses (EAR)<sup>6</sup> [Mollá, Schneider, Schwitter, et Hess, 2000].

### 6.2.2 Grammaires de lien

Les grammaires de lien [Sleator et Temperley, 1991; Grinberg et al., 1995] ne sont pas à proprement parler des grammaires de dépendance, dans la mesure où, contrairement aux dépendances, les "liens" définis par ce formalisme ne sont pas orientés<sup>7</sup>. En conséquence, des structures cycliques sont autorisées<sup>8</sup>. On retrouve par contre beaucoup de notions définies dans le cadre des grammaires de dépendance telle que la projectivité. Comme dans la plupart des formalismes fondés sur la dépendance, les liens sont étiquetés par des fonctions syntaxiques<sup>9</sup>.

La grammaire est lexicalisée : elle associe à tout mot une structure représentant les liens que le mot peut établir avec d'autres mots de la phrase. Les liens sont regroupés en deux listes : une liste droite indiquant les liens que le mot peut entretenir avec d'autres mots se trouvant à sa droite et, symétriquement, une liste gauche. Les liens sont ainsi systématiquement dupliqués. En prenant l'exemple de l'utilisation de ce formalisme pour l'analyse syntaxique, un verbe, par exemple, aura un lien gauche étiqueté  *sujet*  et tout substantif un lien droit du même type. Nous avons représenté dans la figure 6.1 un exemple des structures pouvant être associées aux mots. Dans cet exemple, nous avons représenté l'extrémité des liens par une forme géométrique, appelée connecteur. Pour qu'un lien soit établi entre deux mots, un connecteur droit (dirigé vers la droite dans cette illustration) de l'un doit pouvoir s'apparier avec un connecteur gauche de l'autre mot.

Les attentes de connexion de chaque entrée sont représentées par un ensemble disjonctif de règles. Chaque règle exprime un comportement particulier de l'unité dans l'énoncé c'est-à-dire une des manières dont elle peut être mise en relation avec les

---

<sup>6</sup>Ce domaine applicatif particulier a pour but de trouver les passages d'un document (ou d'un ensemble de documents) qui répondent directement à une requête utilisateur exprimée en langage naturel. L'EAR est ainsi plus ambitieuse que la recherche et l'extraction d'informations (la requête est exprimée librement et les résultats de recherche ne sont pas des documents entiers mais des phrases). Cette tâche se rapproche de celle des systèmes de question-réponse.

<sup>7</sup>Les notions de *dépendants* et de *gouverneurs* sont absentes dans ce formalisme.

<sup>8</sup>L'absence d'orientation des liens considérés autorise ainsi pour un élément de dépendre de plusieurs gouverneurs.

<sup>9</sup>On pourra se reporter à [Schneider, 1998] pour une étude complète sur les différences et les similitudes entre grammaires de dépendance et grammaires de lien.



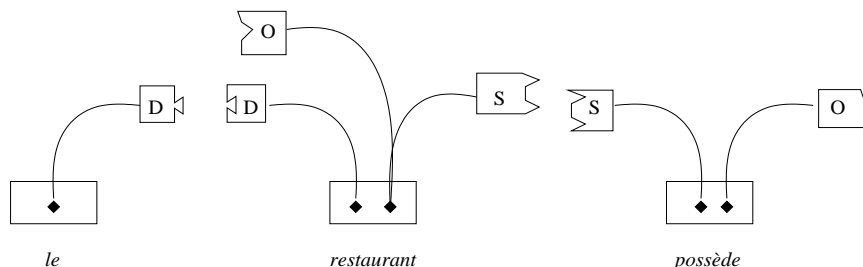


FIG. 6.1 – Exemples de structures élémentaires d’une grammaire de lien

autres. Ainsi, si on veut exprimer que le mot *restaurant* fait partie d’un groupe nominal minimal (précédé d’un déterminant  $D$  et éventuellement d’un adjectif  $A$ ) et peut être sujet  $S$  ou complément d’objet  $O$  d’une phrase, un ensemble disjonctif de 4 règles est nécessaire (la position du mot *restaurant* est indiquée par  $\diamond$ ) :

$$\begin{aligned} & ((D, A) \diamond (S)) \\ & ((D) \diamond (S)) \\ & ((D, A, O) \diamond ()) \\ & ((D, O) \diamond ()) \end{aligned}$$

Cet ensemble de règles peut être exprimé de manière compacte par la formule :

$$D^- \text{ and } (A^- \text{ or } ()) \text{ and } (S^+ \text{ or } O^-)$$

On notera que les listes des connecteurs droite et gauche sont ordonnées suivant la position relative des liens à établir par rapport au mot. Ainsi la formule  $(D^- \text{ and } O^-)$  indique que le déterminant est plus proche que le verbe par rapport à l’entrée lexicale considérée. Cette indication de la structure d’ordre rend délicate la conception de la grammaire de lien qui doit exprimer l’ensemble des règles envisageables pour une entrée lexicale donnée. Elle garantit toutefois l’efficacité de l’algorithme d’analyse en permettant de maîtriser la combinatoire de recherche<sup>10</sup>.

En dépit de cette description grammaticale assez coûteuse, ce formalisme offre une certaine lisibilité des liens devant être introduits pour chaque unité lexicale ; la description dans l’ordre de tous les liens droits et gauches pouvant être générée semi-automatiquement à partir d’une première description des relations impliquées. Par ailleurs, il offre un ensemble d’opérateurs permettant la formulation simple des descriptions lexicales (expression du caractère optionnel d’un lien ou la possibilité de définir

<sup>10</sup>On retrouve, pour les mêmes raisons, la prise en compte des contraintes topologiques d’ordre dans les approches de l’analyse des énoncés oraux fondées sur les grammaires d’arbres adjoints lexicalisées. La prise en compte de ce facteur structurel s’appuie sur la notion de “type linéaire d’arbre” associé à chaque arbre élémentaire. Cette “projection à plat” de la structure arborescente est représentée de la même manière par des listes (liste droite et liste gauche) ordonnées de connecteurs (dénnotant dans ce formalisme les opérations de substitution et d’adjonction) [Halber, 1999, pages 117–134]. L’extraction de ce type linéaire d’arbre peut être réalisée par parcours connexe des arbres élémentaires [Lopez, 1999].

plusieurs liens identiques pour une même unité<sup>11</sup>). Il faut noter enfin l'existence de modèles probabilistes fondés sur ce formalisme [Lafferty et al., 1992]<sup>12</sup>.

L'analyseur développé a été adapté à partir de l'analyseur syntaxique fondé sur les grammaires de lien conçu par [Sleator et Temperley, 1991]. Nous décrivons dans la suite de ce chapitre le principe de l'algorithme et les quelques adaptations que nous y avons apportées. Nous insistons également sur le processus de conception de la grammaire de lien qui constitue l'étape la plus délicate pour le développement de l'analyseur. Nous présentons tout d'abord les contraintes utilisées pour caractériser les liens et effectuer les rattachements entre les unités de l'énoncé.

### 6.2.3 Contraintes exploitées

#### 6.2.3.1 Contraintes sur les unités à relier

En analyse syntaxique, les formalismes fondés sur les grammaires de dépendance autorisent généralement la spécification des relations en définissant des contraintes sur la catégorie grammaticale des deux unités impliquées et des contraintes morphologiques entre ces deux unités.

L'analyse syntaxique de surface nous a permis d'obtenir une caractérisation des groupes grammaticaux élémentaires. Cette caractérisation repose à la fois sur la catégorie du segment et sur sa structure interne. La structure interne est caractérisée par les dépendances locales du segment et peut être précisée par les informations portées par les étiquettes grammaticales associées à chaque mot (notamment par les quelques propriétés morphologiques que nous avons choisies de retenir).

La mise en relation des segments reposera ainsi sur un triplet comportant trois types d'information différente associés aux segments :

- la catégorie du segment (GN, GADJ, GP, DATE, GPHEURE, etc.),
- sa tête lexicale,
- et une partie des informations caractérisant sa structure interne.

On notera que la catégorie du segment est suffisante pour la caractérisation des groupes élémentaires exprimant les expressions langagières indépendantes du domaine d'application. Nous précisons ci-après la nature des informations retenues en ce qui concerne les groupes grammaticaux élémentaires :

**tête lexicale.** La construction des segments effectuée dans l'étape précédente nous assure que la tête lexicale de chaque groupe grammatical soit l'élément le plus à droite du segment. Ainsi, la tête lexicale d'un verbe composé (*ai vu*) est le participe. En analyse syntaxique, plusieurs travaux fondés sur la notion de dépendance considèrent au

<sup>11</sup>Il est par exemple possible d'exprimer qu'un verbe ait plusieurs compléments d'objets situés à sa droite en l'absence de coordination. Cette caractéristique nous sera utile pour le traitement des inattendus structuraux.

<sup>12</sup>Nous n'avons pas utilisé de modèle probabiliste dans notre expérimentation. Cette caractéristique est toutefois intéressante pour prendre en compte à terme, conjointement aux probabilités associées aux liens, le score de reconnaissance associé à chaque énoncé.

contraire que l’auxiliaire est la tête et le participe son dépendant, car c’est l’auxiliaire qui hérite des marques grammaticales (*il pense avoir vu*) et porte la négation<sup>13</sup>. Le choix que nous faisons est cependant légitime étant donné notre objectif visant à construire une représentation sémantique de l’énoncé : les relations sémantiques que ce groupe peut entretenir avec les autres dépendent principalement du participe. On notera en outre que cette information (verbe composé) est présente dans la catégorie grammaticale du segment.

En ce qui concerne les groupes prépositionnels, nous retenons comme tête de segment la préposition. Notre domaine d’application nous permet en effet, dans certains cas, de définir entièrement le rôle sémantique du segment en fonction de la préposition. Par exemple, la locution prépositionnelle *proche-de* n’a pas d’autre rôle dans notre domaine que d’indiquer une propriété de localisation d’un objet par rapport à un autre<sup>14</sup>. La préposition ne peut pas toutefois suffire à elle seule pour caractériser le groupe prépositionnel dans tous les cas de figure. Les informations supplémentaires nécessaires sont alors conservées dans le troisième élément du triplet retenu.

**Éléments de la structure interne retenus.** Les éléments que nous retenons dépendent du type du segment considéré :

- **Groupes verbaux.** La catégorie du segment porte l’information sur le type du groupe verbal considéré (participes, infinitif ou présence d’un modal). Nous ne conservons donc ici que le modal éventuel impliqué. Il s’agit de différencier les actes dialogiques des énoncés comme *je veux réserver pour demain* et *je peux réserver pour demain* par exemple.
- **Groupes nominaux.** Les informations retenues concernent essentiellement les propriétés portées par les étiquettes grammaticales des éléments qualifiant le nom : le caractère *défini* ou *indéfini* des articles, les *possessifs* et *démonstratifs* ainsi que la présence d’un adjectif *interrogatif* (*quel musée*).
- **Groupes prépositionnels.** Lorsque la préposition ne suffit pas à elle seule pour caractériser précisément le rôle sémantique du groupe dans le domaine, nous conservons la tête lexicale du groupe dépendant ainsi que les éléments correspondant à sa structure interne<sup>15</sup>.

Les autres dépendances locales du segment (par exemple un adjectif dans le cas des groupes nominaux) n’interviennent pas dans le processus de mise en relation.

<sup>13</sup>Il n’existe pas de réel consensus sur ces cas litigieux [Kahane, 2001].

<sup>14</sup>Dans le corpus qui nous a servi à caractériser notre domaine applicatif, nous n’avons rencontré qu’un seul autre rôle de cette forme, assimilée à une locution prépositionnelle : l’expression d’un ordre de grandeur sur les prix (*proche de dix euros*). Or, dans ce cas, le segment a été étiqueté GP<sub>prix</sub> par l’étape de segmentation. Ceci nous permet de lever l’ambiguïté sur les autres groupes prépositionnels introduits par *proche-de*.

<sup>15</sup>Si le dépendant est un nom propre seule l’étiquette NPROPRE est suffisante

### 6.2.3.2 Contraintes sur les relations

La définition des liens est fondée sur l'utilisation de deux types d'information : des connaissances sémantiques liées au domaine d'application d'une part et des informations liées à la fonction syntaxique des éléments nécessaires à la structuration de l'énoncé (co-ordinations, relatifs, conjonctions) d'autre part. La caractérisation des liens sémantico-pragmatiques envisagés repose sur l'expertise du domaine présentée au chapitre 4. Nous avons retenu 144 étiquettes sémantico-pragmatiques différentes qui correspondent :

- aux concepts propriétés portant sur les objets du domaine applicatif (**Catégorie** d'un hôtel, **Spécialité** d'un restaurant, **Version** d'un film, etc.),
- aux concepts arguments impliqués dans les actes dialogiques considérés (**Objet-tarif**, **Objet-horaire**, **Etablissement-référent**, etc.),
- et aux schémas retenus.

On trouvera en annexe C.2 la liste complète des étiquettes retenues.

La notion de schéma est par nature limitative. En fondant la définition des liens sur les représentations attendues, nous restons sur ce point assez proches des méthodes sélectives. On notera toutefois que nous chercherons à effectuer le plus de rattachements possibles entre les segments de l'énoncé. Par ailleurs, nous n'imposons pas la présence dans l'énoncé de l'ensemble des concepts attendus par rapport à un schéma donné. Ces derniers ne servent qu'à la définition des attentes conceptuelles potentielles liées à chaque segment. Ainsi, la présence du lien permettant d'identifier l'objet sur lequel porte une requête, par exemple, est facultative<sup>16</sup>. Ceci nous permet de traiter des énoncés du type *quels sont les horaires* sans imposer que le segment nominal *les horaires* soit obligatoirement relié par un lien **Objet-horaire** à un autre segment :

```

+-----InfoHORAIRE-----+
|                               |
////  GVint(etre)                GN(horaire.defini)

```

Ce n'est que lors de la conversion de la structure de lien vers la représentation sémantique attendue (cf. § 6.2.6) que la valeur du concept **Objet-horaire** sera instanciée comme faisant potentiellement partie du contexte de l'interaction.

Nous imposons toutefois quelques restrictions sur certaines combinaisons de liens. Ces restrictions ne portent principalement que sur les informations grammaticales attachées à la structure interne du segment, notamment le caractère défini et indéfini<sup>17</sup>, essentiel pour traiter la référence.

<sup>16</sup>Il en va de même pour l'ensemble des propriétés pouvant se rattacher à un objet particulier.

<sup>17</sup>À titre d'exemple, un segment caractérisant un acte de demande d'information de **Localisation** (*l'adresse*) ne peut être relié qu'avec un groupe prépositionnel défini (*de l'hôtel, de cet hôtel*). Un lien caractérisant l'objet d'une requête de **Sélection** sera en revanche préféré dans le cas d'un groupe prépositionnel indéfini (*d'un hôtel*) ou en cas de la présence conjointe du groupe prépositionnel défini et d'un segment du type *le plus proche* ou *le moins cher*. Dans ce cas de figure, la présence du groupe nominal *l'adresse* est conservée dans la représentation. Cette information est utile pour pouvoir préciser la réponse à fournir (l'adresse une fois la sélection effectuée et éventuellement validée par l'utilisateur).

En ce qui concerne les liens sémantico-pragmatiques, la combinatoire de recherche sera ainsi limitée principalement par les attentes conceptuelles (même si elles restent facultatives) de chaque segment. La contrainte de projectivité permet quant à elle de limiter la portée de la recherche de ces attentes dans l'énoncé. Des contraintes plus fortes sont cependant imposées pour la construction des liens impliquant les segments de liaison syntaxique (coordinations, relatifs, conjonction). Nous donnons dans la suite de ce chapitre l'exemple du traitement de la coordination.

#### 6.2.4 Construction de la grammaire de lien

La grammaire de lien doit encoder la manière dont chaque unité peut participer à la structure globale de l'énoncé. Il ne s'agit pas seulement de représenter le comportement sémantique de chacune de ces unités dans le domaine d'application (attentes conceptuelles). Le formalisme utilisé et l'algorithme d'analyse imposent également que toutes les séquences ordonnées de liens soient représentées. La qualité de l'analyse repose ainsi principalement sur la complétude et la représentativité de la grammaire à concevoir. Le processus de conception doit par ailleurs être suffisamment souple pour pouvoir permettre une intégration relativement aisée d'une nouvelle entrée ou d'un nouveau type de lien.

Nous avons adopté une démarche incrémentale et semi-automatique pour la conception de la grammaire de lien à partir de l'expertise du domaine :

1. **Création d'un dictionnaire sémantico-pragmatique.** Nous avons développé manuellement un dictionnaire représentant, sans considération d'ordre, l'ensemble des liens pouvant être attachés à une unité donnée. Nous nous sommes appuyés sur la catégorisation réalisée à partir de l'étude du corpus PARISCORP<sup>18</sup>. Nous avons considéré dans un premier temps comme entrées de ce dictionnaire les unités issues de la segmentation de la moitié du corpus PARISCORP (700 énoncés). Afin de limiter les entrées, nous avons regroupé les segments de têtes lexicales différentes mais ayant le même comportement sémantique dans le domaine<sup>19</sup>. Par ailleurs, il n'est pas nécessaire de créer une entrée pour toutes les combinaisons de catégorie, tête lexicale et éléments de la structure interne. Des règles additionnelles permettant de compléter le comportement général de l'unité en fonction des éléments de sa structure interne<sup>20</sup> sont ajoutées suivant les besoins. Les liens pouvant être attachés à un même segment sont classés en 5 types distincts : concepts de propriété (CP), concepts arguments (CA), schémas (SC), Modificateurs (Mod) et liens syntaxiques (S). Nous donnons un exemple illustratif du dictionnaire sémantique

<sup>18</sup>Cette étude nous avait permis, par exemple, d'organiser en ensembles structurés les différents objets de notre domaine pouvant faire l'objet d'un même type de requêtes (cf. Annexe C.1).

<sup>19</sup>Par exemple, l'entrée GN(\$RESTAURANT) représente dans le dictionnaire tous les groupes nominaux de tête *restaurant*, *pizzeria*, *snack*, etc.

<sup>20</sup>Principalement le caractère défini/indéfini.

en annexe C.3.

2. **Compilation du dictionnaire sémantico-pragmatique en grammaire de lien.** Les règles exprimant chacune des séquences ordonnées de liens potentielles sont générées automatiquement à partir du dictionnaire précédent. Des processus de compilation différents sont appliqués en fonction de la nature du segment. Par exemple, dans le cas d'un segment correspondant à un objet du domaine, nous n'acceptons la présence que d'un concept argument par règle<sup>21</sup> alors que chacune des règles intègre une distribution particulière de l'ensemble des concepts propriétés. Plusieurs concepts arguments, éventuellement du même type<sup>22</sup>, sont en revanche acceptés au sein d'une règle pour les segments verbaux. La compilation est réalisée en plusieurs étapes construisant progressivement des règles de plus en plus complètes.

L'évaluation de la couverture de la grammaire de lien obtenue a été effectuée par l'analyse des 700 premiers énoncés du corpus PARISCORP. L'examen des structures de liens établies a permis la vérification de la consistance de l'ensemble des étiquettes retenues et l'enrichissement du dictionnaire sémantico-pragmatique. Le nouveau dictionnaire enrichi a ensuite été compilé de façon à compléter la grammaire de lien. Cette procédure a été réitérée sur la seconde partie du corpus PARISCORP jusqu'à obtenir une couverture satisfaisante de la grammaire de lien.

La grammaire de lien à laquelle nous avons ainsi progressivement abouti comporte environ 1000 entrées. Nous donnons en annexe C.4 un sous-ensemble de cette grammaire. Le nombre de règles pour une entrée est très variable, allant d'une à plusieurs dizaines. Ce processus aveugle de constitution de toutes les combinaisons ordonnées de liens induit en effet une certaine redondance<sup>23</sup>. En contrepartie, cette solution permet d'offrir une interface, via le dictionnaire sémantico-pragmatique, pour la maintenance de la grammaire de lien<sup>24</sup>. Ainsi l'ajout d'une nouvelle entrée ou d'un nouveau lien pour une entrée existante se limite à la définition ou la modification des attentes conceptuelles liées à cette entrée dans le dictionnaire puis à la compilation de la règle correspondante. Notons enfin qu'une même règle n'implique pas plus de 5 liens (dont au maximum 2 sont obligatoires). L'analyse établit quant à elle rarement plus de 3 liens à partir d'un même segment.

<sup>21</sup>Nous évoquons dans la suite le traitement envisagé en ce qui concerne les coordinations.

<sup>22</sup>Nous utilisons les opérateurs définis dans [Sleator et Temperley, 1991] pour l'expression des règles. La possibilité de définir un nombre indéterminé de liens de même type à la droite ou à la gauche du segment (ce qui correspond au phénomène d'entassement paradigmatique) est en particulier définie au moyen d'un opérateur spécifique.

<sup>23</sup>Un nombre élevé de règles est en particulier attesté sur les objets du domaine comptant plus de 3 concepts de propriété. Ces liens sont effet optionnels pour le segment correspondant à l'objet considéré. Ainsi les règles  $(() \diamond (A C B))$  et  $(() \diamond (A B C))$  où A, B et C dénotent des liens de propriété sont équivalentes en l'absence de C dans l'énoncé.

<sup>24</sup>La modification directe de la grammaire de lien est hasardeuse. Il est en effet difficile de se rendre compte des structures qu'autorise une combinaison de liens particulière.

### 6.2.5 Algorithme d'analyse et gestion de l'ambiguïté

L'algorithme d'analyse utilisé a été adapté à partir de celui décrit dans [Grinberg et al., 1995]. Son principe général est assez simple : toute unité tente d'établir des liens avec les autres unités conformément à la description des liens de la grammaire. L'algorithme est récursif : lorsqu'un lien est établi entre deux unités, l'analyse est relancée sur la portion de l'énoncé située entre ces deux unités. La combinaison des liens est maîtrisée par la contrainte de projectivité. À la différence de l'analyseur par grammaire de lien décrit dans [Sleator et Temperley, 1991], les analyses partielles sont autorisées. L'analyse est ainsi réalisée en trois étapes<sup>25</sup> :

- la première étape cherche à satisfaire l'ensemble des liens obligatoires de chaque unité et à relier chacune d'entre elles à la structure de lien globale. L'algorithme, par programmation dynamique, utilise la technique dite de *mémoïsation* qui consiste, lors d'un retour arrière, à conserver en mémoire les analyses déjà faites pour ne pas avoir à les reconstruire,
- si aucune structure ne peut être établie entre l'ensemble des unités en présence, une seconde analyse est effectuée. Cette étape cherche cette fois à construire une structure complète en ajoutant des liens fictifs entre unités adjacentes. Ces liens fictifs servent soit à relier une unité qui ne peut pas participer à la structure de liens globale soit plusieurs structures complètes sur une portion d'énoncé. Le nombre de liens fictifs devant être introduits sert d'heuristique pour élaguer l'espace de recherche<sup>26</sup> [Grinberg et al., 1995, pages 5–9],
- la dernière étape classe les structures de liens concurrentes en appliquant un système d'heuristiques. Le nombre de liens fictifs devant être introduits pour les analyses partielles est un premier critère de classement. Nous classons ensuite les structures résultantes selon deux critères :
  - le nombre d'unités ignorées par l'analyse. La priorité est ainsi donnée aux analyses en îlots (plus de deux structures de dépendance complètes mais non reliées entre elles, c'est-à-dire sans unités ignorées) par rapport aux analyses ignorant une ou plusieurs unités.
  - en cas d'analyses partielles de scores identiques (même nombre d'unités ignorées), la priorité est donnée aux analyses comportant le moins de dépendances longue-distance<sup>27</sup>.

---

<sup>25</sup>On trouvera dans [Sleator et Temperley, 1991; Grinberg et al., 1995] une description des techniques utilisées pour optimiser l'accès à la grammaire et faire une première sélection des règles à prendre en compte (Si un lien  $A^+$  fait partie d'une des règles d'un élément de l'énoncé, alors que le lien correspondant  $A^-$  n'appartient à aucune des règles associées aux autres unités présentes, cette règle est éliminée avant l'analyse).

<sup>26</sup>Si un chemin conduit à une solution nécessitant plus de liens fictifs qu'une solution précédemment trouvée, ce chemin est abandonné.

<sup>27</sup>La portée d'une dépendance est évaluée par le nombre d'unités considérées dans l'énoncé entre les deux unités reliées par la dépendance. Deux unités adjacentes reliées ont une "longueur" nulle. Cet heuristique est utile pour des énoncés comportant plusieurs actes dialogiques différents, par exemple un assertif suivi d'une requête (cf. 6.3.3.1).

La *mémoïsation* des structures partielles analysées au cours de la première étape permet d’assurer une complexité cubique par rapport au nombre de segments dans l’énoncé<sup>28</sup>.

Les différentes segmentations d’un même énoncé sont analysées en parallèle. Pour chaque énoncé, nous retenons les analyses de plus faible coût selon les heuristiques que nous venons de décrire. Une procédure de classification supplémentaire est effectuée selon ces mêmes critères pour ordonner les analyses concurrentes issues des différentes segmentations de l’énoncé. Toutes les analyses de scores identiques sont conservées. Dans le cas de scores identiques sur des analyses issues de segmentations différentes d’un même énoncé, nous conservons celles comportant le plus grand nombre d’unités<sup>29</sup>. Nous donnons dans la suite de ce chapitre quelques exemples des ambiguïtés restantes dues à la stratégie adoptée ainsi que quelques exemples de structures de liens obtenues.

### 6.2.6 Génération des représentations sémantiques finales

La structure de liens obtenue est un graphe non orienté de relations entre les unités considérées. L’étape de génération de la représentation sémantique finale est plus qu’une simple conversion de la structure de liens en une représentation sous forme de structure de traits. C’est également lors de cette étape que sont traitées, en fonction de la caractérisation des unités mises en jeu, l’identification potentielle des références au contexte et les ambiguïtés résiduelles. Nous donnons dans le paragraphe suivant quelques exemples des traitements spécifiques dédiés à cette étape en rapport avec les phénomènes oraux ou linguistiques identifiés. Nous illustrons le processus de conversion sur un exemple simple correspondant à l’analyse de l’énoncé “*je voudrais la liste des hôtels trois étoiles à proximité de la gare*” qui donne lieu à la structure de lien suivante :

```

+-----CRITERESELECTLoc-----+
+--AGENT--+SELECTION--+OBJSELECT--+CATEGORIE-+
|           |           |           |           |
//// GN(je) GV(vouloir) GN(liste.      GP(de,hôtel) GN(étoile) GP(proche-de,
                             defini)                                gare.defini)

```

La conversion est réalisée en deux étapes. Les liens n’étant pas orientés, la structure de lien est “mise à plat” en indiquant les relations établies entre les différentes unités :

<sup>28</sup>Sur l’ensemble des tests utilisés pendant la campagne DEFI, (cf. chapitre suivant), le nombre moyen de mots par énoncé est de l’ordre de 13. Sur ce même ensemble, il y a en moyenne entre 7 et 8 *chunks* par énoncé.

<sup>29</sup>Ainsi que nous l’avons vu au § 5.3.3 l’énoncé “*donnez moi l’adresse d’un de ces hôtels*” donne lieu à au moins deux segmentations : l’une éliminant *d’un* l’autre conservant cette forme comme groupe prépositionnel. Chacune de ces segmentations donne lieu à une analyse sous forme de structures de lien présentant un score identique (analyse complète, aucune unité ignorée). La seconde est retenue car elle comporte une unité de plus.



```
1:GN(je) "AGENT" 2:GV(vouloir);
2:GV(vouloir) "SELECTION" 3:GN(liste.defini);
3:GN(liste.defini) "OBJSELECT" 4:GP(de,hôtel);
3:GN(liste.defini) "CRITERESELECTloc" 6:GP(proche-de,gare.defini);
4:GP(de,hôtel) "CATEGORIE" 5:GN(étoile);
```

Le parcours de cette structure est ensuite effectué à partir du lien "dominant" c'est-à-dire celui qui correspond à l'acte de dialogue impliqué<sup>30</sup> (dans cet exemple, il s'agit du lien SELECTION). En l'absence de tels liens, établis généralement à partir d'un groupe verbal (et donc absents des énoncés elliptiques comme *les hôtels 3 étoiles* ou *pas trop cher* par exemple), on recherche les concepts arguments présents puis par défaut les concepts de propriété. Les valeurs des concepts sont ensuite instanciées au cours du parcours. Sur l'exemple précédent, on obtient la représentation sémantique suivante :

<pre>&lt;ignoré&gt; : 0 &lt;acte&gt; : demande-information   Type-information : sélection   Objet-Select : hôtel   Quantité : plusieurs   Catégorie : 3   Critère-Selectloc : proche-de   Référentloc : gare   Référence : <i>ctxt</i></pre>
--

On notera la présence de concepts absents de la structure de liens. Les concepts de propriété **Quantité**, **Référentloc** et la valeur du concept **Catégorie** sont issus des dépendances locales des segments *les hôtels*, *de la gare* et *3 étoiles*. Le groupe prépositionnel *de la gare* est défini mais non caractérisé dans l'énoncé, la valeur *ctxt* est donc instanciée pour le concept **Référence**, indiquant que cette information est éventuellement à rechercher dans le contexte dialogique. Nous intégrons également dans la représentation le degré de complétude de l'analyse, indiqué par le nombre de segments ignorés<sup>31</sup>.

### 6.3 Quelques exemples de traitement de phénomènes oraux et linguistiques

Dans ce paragraphe, nous présentons brièvement sur quelques exemples les traitements envisagés pour les principaux phénomènes linguistiques rencontrés ainsi que pour la prise en compte des actes dialogiques multiples et le relâchement de contraintes caractéristiques de la richesse du domaine applicatif.

<sup>30</sup>Nous évoquons dans la suite les énoncés impliquant plusieurs actes dialogiques.

<sup>31</sup>Le cas échéant, les segments ignorés sont également conservés dans la représentation.

### 6.3.1 Traitement des coordinations et des inattendus structuraux

#### 6.3.1.1 Coordinations

Le traitement de la coordination est un problème général pour un analyseur [Abeillé et Blache, 2000]. Nous nous limiterons ici aux coordinations logiques. Nous avons choisi de faire porter le lien marquant la coordination par la conjonction de coordination elle-même :

```

+-----OBJHORAIRE-----+
+Info-HORAIR+           +--COOhor---+---COOhor---+
|           |           |           |           |
//// GN(horaire.defini) GP(de,musée) Coo(et) GP(de,cinéma.defini)

```

Pour les segments impliqués dans cet exemple, le lien `COOhor` se substitue au lien `ObjetHoraire` que chacun d'entre eux aurait pu établir avec le segment nominal *horaire*. La position de la coordination par rapport aux segments qu'elle relie est fixe<sup>32</sup>. Cette contrainte est utilisée pour la génération automatique de la grammaire de lien. Le calcul de la portée de la coordination est effectué sur des critères sémantico-pragmatiques. Pour que deux segments soient reliés à une coordination par des liens de type  $COO_X$  il faut que ces deux segments soient *compatibles* selon la relation  $X$ . En d'autres termes, ces segments doivent partager le même comportement sémantique dans l'énoncé vis-à-vis de l'élément avec lequel la coordination logique va être reliée (le segment *horaire* dans l'exemple précédent). Cette information est déduite automatiquement de la grammaire modulaire et portée par l'étiquette du lien de coordination<sup>33</sup>. Ainsi, dans l'énoncé "*donnez moi les horaires du musée et les tarifs des visites*", le lien de coordination ne sera pas instancié entre *du musée* et *les tarifs* mais entre *les horaires* et *les tarifs*, ces deux segments partageant la même fonction d'identification d'acte de demande d'information. Dans le cas d'une énumération, *les horaires de A de B et de C*, nous obtenons la structure suivante :

```

+-----OBJHORAIRE-----+
+Info-HORAIR+---OBJHORAIRE---+           +--COOhor---+---COOhor+
|           |           |           |           |           |
//// GN(horaire.defini) GP(de,A) GP(de,B)   Coo(et) GP(de,C)

```

étant donnée la possibilité pour le segment *horaire* d'accepter plusieurs concepts arguments de même type successifs. La présence de la coordination entre les deux derniers éléments est exploitée lors du parcours de la structure de lien pour identifier la coordination entre les trois éléments. Lorsqu'une propriété qualifie le deuxième élément et que

<sup>32</sup>Elle précède immédiatement le deuxième élément coordonné. Seuls des liens exprimant des concepts de propriété peuvent éventuellement être établis entre des segments séparant le premier élément coordonné de la coordination.

<sup>33</sup>Pour la coordination logique *et* entre les segments  $S_1$  et  $S_2$ , la compatibilité est stricte dans le cas des concepts arguments (dans la grammaire modulaire  $S_1$  et  $S_2$  doivent avoir en commun le concept argument considéré) ou large dans le cas des concepts de propriété ou des schémas ( $S_1$  et  $S_2$  sont porteurs de liens de type propriété mais leur nature est différente). Cette définition diffère dans le cas de la coordination logique *ou*, qui peut être stricte sur les concepts de propriété.

sa portée est ambiguë, nous intégrons l’ambiguïté dans la représentation sémantique finale. Ainsi, la structure de lien sur l’exemple “(...) *des restaurants et des hôtels pas chers*”,

```

+-----OBJET-----+
|           +-----COO---+---COO---+---PTETARIF---+
|           |           |           |           |
//// GP(de,restaurant) Coo(et) GP(de,hotel) Gadj(cher)

```

est convertie en :

Objet : <i>et</i>
Objet : hôtel
PteTarif : <i>pas cher</i>
Objet : restaurant
PteTarif : ? : <i>pas cher</i>

### 6.3.1.2 Inattendus structuraux

Nous envisageons le traitement des inattendus structuraux de la même manière que les coordinations. Nous nous appuyons en effet, quand elles existent, sur les marques identifiant ces phénomènes. Ainsi, la structure de lien associée à l’énoncé *les horaires du musée non du cinéma* est analogue à celle présentée dans le paragraphe précédent<sup>34</sup>. La marque *non* sert de connecteur entre les deux éléments repris<sup>35</sup>. Quand elle est explicite (*non, pardon*), seul le second élément est intégré dans la représentation. Dans le cas contraire (présence d’une interjection comme *eh* par exemple), l’ambiguïté entre coordination et correction est indiquée dans la représentation<sup>36</sup>.

Nous nous appuyons, pour les répétitions et les corrections non marquées, sur la possibilité d’instancier plusieurs liens de même type entre unités différentes. Le traitement de ces phénomènes s’apparente ainsi à celui des énumérations décrit dans le paragraphe précédent.

```

+-----OBJHORAIRE-----+
+----OBJHORAIRE-----+ |
+Info-HORAIR+-OBJHORAIRE-+ | |
|           |           |           |           |
//// GN(horaire.defini) GP(de,A) GP(de,B) GP(de,C)

```

Dans cet exemple, en l’absence de toute marque (coordination logique ou marque d’hésitation ou de correction), la représentation sémantique intègre l’ambiguïté entre énu-

<sup>34</sup>Le lien COOhor est remplacé par le lien CORhor.

<sup>35</sup>À la différence de la coordination logique, la distribution des liens associés à chaque marque dans la grammaire de lien est variable selon les usages de ces marques. *pardon* par exemple ne se situe pas forcément entre les éléments répétés.

<sup>36</sup>Le coordinateur *et* peut ne pas avoir été reconnu par la reconnaissance de la parole. Nous faisons porter l’ambiguïté sur le premier élément coordonné.

mération et correction pour les deux premiers concepts-arguments<sup>37</sup>.

Dans le cas des enrichissements lexicaux, c'est-à-dire lorsque les segments sont identiques (catégorie grammaticale, tête lexicale et éléments de la structure interne), nous faisons porter cette ambiguïté sur les propriétés qui différencient les segments impliqués.

### 6.3.2 Procédés réguliers

#### 6.3.2.1 Mouvements des constituants

Le formalisme utilisé, qui encode toutes les combinaisons de concepts possibles, permet d'accepter les principaux phénomènes de mouvements des segments dans l'énoncé n'entraînant pas de structures discontinues. Il convient toutefois de s'interroger sur la surgénérativité de l'analyse pouvant être induite par le processus aveugle de génération automatique des séquences autorisées. Nous évaluons la robustesse de l'analyseur par rapport à ces phénomènes au chapitre suivant. Précisons que nous avons choisi de conserver dans la représentation sémantique certains indices permettant de caractériser un phénomène d'extraction. Ainsi dans le cas de présence conjointe d'un pronom et d'un argument reliés tous deux à un même groupe verbal, nous indiquons, lors de la conversion de la structure de lien, la présence d'un double marquage<sup>38</sup>. De la même manière, nous réalisons l'analyse des procédés d'extraction à présentatifs de façon à pouvoir indiquer la présence de ce procédé dans l'énoncé<sup>39</sup>. Nous ne faisons aucune interprétation à ce sujet lors de la génération de la représentation sémantique finale qui se contente d'indiquer qu'un des éléments a été extrait. On sait que ces procédés sont le plus souvent utilisés par les locuteurs pour mettre en avant une information particulière de leur énoncé. Le choix de conserver cet indice peut éventuellement être utile notamment pour identifier la partie importante des énoncés correctifs des locuteurs.

#### 6.3.2.2 Ellipses

Deux traitements sont envisagés en ce qui concerne les ellipses. Le premier, que nous avons déjà évoqué, résulte de notre choix de préserver les segmentations d'un même énoncé issues des différentes séquences d'étiquetage obtenues. Les formes nominales ou nominales des mots comme *un*, *une*, *premier*, etc. nous permettent de caractériser des groupes nominaux dans les énoncés du type "*je voudrais le premier*" et dans au moins une des segmentations retenues pour des énoncés tel que "*j'en voudrais une double*". En ce qui concerne l'absence d'une structure prédicative verbale et les énoncés ne comportant qu'un segment (*les hôtels*, *les horaires*, *double*, etc.) nous

<sup>37</sup>Qu'il s'agisse d'une énumération ou d'une correction, le dernier élément est le seul qui puisse être considéré comme valide.

<sup>38</sup>Si le pronom est seul, la valeur *ctxt* est instanciée pour le concept argument concerné.

<sup>39</sup>Dans le cas d'un procédé d'extraction à présentatif (*c'est pour demain que je veux réserver* par exemple), des liens de type **Extraction** relient de part et d'autre l'élément extrait à l'élément initial (*c'est* ou les introductifs construits à partir du verbe avoir) et l'élément introduisant la (fausse) subordonnée. Ces derniers (*que* ou *qui*) ne sont pas reliés au reste de l'énoncé à la différence de l'élément extrait.

utilisons une unité fictive située à la droite de l'énoncé pour établir les liens manquants à partir de ces unités. L'énoncé comprenant le seul segment *le film* par exemple donnera lieu à plusieurs structures de lien de la forme :

```

+--OBJETHORAIRE--+          +---OBJETDUREE---+
|                          |                    |
/////          GN(film.defini)  /////          GN(film.defini)
    
```

L'établissement des liens à partir de l'unité fictive est pénalisé par un coût comptabilisé par le système d'heuristique pour le classement des analyses concurrentes. Ainsi l'analyse établissant un lien de même type avec un autre segment de l'énoncé sera préférée à celle utilisant l'unité fictive. Lors de la conversion de la structure de lien l'ambiguïté entre les différentes requêtes (dans le cas de l'exemple précédent une demande d'information portant sur les horaires ou la durée) est préservée.

### 6.3.2.3 Anaphores

Il n'y a pas à proprement parlé de traitement spécifique concernant les anaphores. La conservation du critère indéfini/défini et des adjectifs démonstratifs dans la structure interne des groupes nominaux et prépositionnels permet simplement d'identifier, pour certains concepts<sup>40</sup> une référence éventuelle au contexte lors de la conversion de la structure de lien. Les pronoms démonstratifs portent quant à eux des liens de type générique (schéma, arguments ou propriété) leur permettant d'être reliés aux autres segments de l'énoncé indépendamment du rôle sémantique de l'élément auquel ils réfèrent. Si cet élément est présent dans l'énoncé, comme dans l'exemple suivant où *ceux* est relié au groupe verbal lui-même relié au segment *les horaires*, il peut être identifié.

```

+-----OBJETHORAIRE-----+
+-----Info-HORAIRE-----+
|                               +--AGENT--+--SCHEMA--+--OBJET--+
|                               |           |           |           |
///// GN(horaire.defini)  GN(je)  GV(vouloir)  GN(ceux)  GP(de,musée)
    
```

Dans le cas contraire, où si le lien de même type ne peut être établi avec l'élément référent (par exemple dans le cas d'un groupe nominal avec adjectif démonstratif *cet hôtel*), nous nous contentons d'intégrer dans la représentation sémantique l'ambiguïté sur l'élément référent (présence dans le contexte dialogique ou parmi les segments de l'énoncé).

### 6.3.3 Richesse du domaine applicatif

#### 6.3.3.1 Actes dialogiques multiples

Le traitement présenté précédemment concernant les coordinations nous permet d'identifier la présence de requêtes multiples dans l'énoncé ou d'une même requête portant sur plusieurs objets de l'application. Nous précisons dans ce paragraphe la façon

<sup>40</sup>Voir l'exemple donné au § 6.2.6 pour le concept argument CRITERESELECTLoc.

dont sont appréhendés les énoncés intégrant des actes dialogiques différents.

Pour l'identification des actes de fermeture et d'ouverture du dialogue, nous nous appuyons sur l'unité fictive introduite pour le traitement des ellipses. Cette unité est utilisée pour porter les relations caractérisant ces actes ainsi que la présence des marques de politesse :

```

+-----POLITESSE-----+
+-OUVERTU-+          +-OBJSELECT-+-CATEGORIE-+          |
|           |          |           |           |           |
//// Gadv(bonjour) GN(liste.   GP(de,   GN(étoile) GAdv(s'il-vous-plait)
                    defini)   hôtel)

```

Ce choix nous permet d'éviter d'avoir à relier les segments caractéristiques de ces actes dialogiques avec un élément particulier de l'énoncé et facilite leur identification lors de la conversion de la structure de lien. Ce procédé introduit en revanche des discontinuités pour des énoncés du type "*donnez moi s'il vous plaît les horaires du musée*" dans lequel la locution *s'il vous plaît* ne peut être reliée à l'unité fictive en raison de la contrainte de projectivité<sup>41</sup>. Comme pour son utilisation pour les structures elliptiques, l'instanciation des liens établis à partir de l'unité fictive est pénalisée. Il s'agit ainsi de préférer une analyse qui ignore ces marques, lorsqu'elles introduisent une discontinuité, à une analyse en îlots cassant la structure globale de l'énoncé (le lien entre *donnez* et *les horaires* dans notre exemple). Dans ce cas de figure, notons que le segment ignoré est intégré dans la représentation. Son rôle n'étant pas ambigu, l'information peut être facilement récupérée.

Les actes de demande d'information et les assertifs sont quant à eux traités au même niveau :

```

+-ASSER-+---LOC---+          +---TRAJET---+---DEST---+
|           |           |           |           |           |
//// GN(je) GV(etre) GP(à,gare  GVint(faire, GP(pour,   GP(à,
                    .defini)   comment)   aller)   NPropre)

```

Le parcours de la structure de liens permet de caractériser les deux actes impliqués (assertif sur une localisation et demande d'information sur un trajet). On notera toutefois que notre analyse est imparfaite sur ce type d'énoncé et nécessite une évaluation précise sur ces phénomènes<sup>42</sup>. Les heuristiques utilisées nous permettent par exemple de ne pas retenir, dans l'énoncé "*le musée du Louvre est fermé à quelle heure ouvre le musée Grévin*", l'analyse instanciant le musée du Louvre comme objet de la requête sur les horaires. Cette analyse aura en effet un coût plus élevé étant donné que la dépendance

<sup>41</sup>Le lien devant être établi entre l'unité fictive et la locution croise celui entre le groupe verbal et le segment *les horaires*.

<sup>42</sup>Ces phénomènes étaient relativement peu présents de l'évaluation réalisée dans le cadre de la campagne DEFI que nous présentons dans le chapitre suivant.

entre *ouvre* et *le musée (du Louvre)* est plus “longue” que celle entre *ouvre* et *le musée (Grévin)*<sup>43</sup>. En revanche, la validité de cette heuristique est plus incertaine en cas d’actes dialogiques successifs présentant des phénomènes de mouvements de segments<sup>44</sup>.

### 6.3.3.2 Relâchement de contraintes

Évoquons enfin le cas des locutions adverbiales venant relâcher certaines contraintes du locuteur (*à la rigueur, de préférence, si possible*), rarement prises en compte dans les systèmes sélectifs. Chacune de ces formes peut établir un lien de type **Modifieur** avec un segment qui lui est adjacent. Ces locutions adverbiales peuvent toutefois être présentes à la droite ou à la gauche des contraintes qu’elles modifient. Ainsi si le coordinateur logique *ou* n’est pas reconnu par la reconnaissance vocale dans le cas de l’énoncé “*proche de la gare si possible ou dans le centre*”, nous obtenons deux analyses ambiguës de coût identique, la locution *si-possible* pouvant se rattacher à l’un ou à l’autre des groupes prépositionnels. Dans ce cas de figure, l’ambiguïté est intégrée dans la représentation sémantique.

## 6.4 Bilan

Nous avons présenté dans ce chapitre les principes fondamentaux de l’analyse permettant la mise en relation des segments identifiés lors de la première étape. Les intérêts de l’analyse par dépendance réalisée peuvent être résumés par les points suivants :

- **Robustesse** La robustesse de l’analyse est due à deux caractéristiques principales : le caractère lexicalisé du formalisme utilisé d’une part et l’utilisation de critères sémantico-pragmatiques pour l’établissement des relations envisagées (attentes conceptuelles mais aussi traitement des phénomènes linguistiques telles les coordinations, les répétitions et les corrections) d’autre part. L’étape de segmentation préliminaire est également un facteur important permettant d’y contribuer. Cette dernière permet non seulement de réduire la combinatoire de recherche des dépendances au sein de l’énoncé (en caractérisant les dépendances internes aux segments) mais offre également la possibilité d’envisager le traitement des phénomènes de mouvement et des inattendus structuraux directement sur les unités impliquées dans ces phénomènes.
- **Niveau de détail des représentations sémantiques générées** En exploitant un grand nombre des indices retenus (catégorie grammaticale des segments, éléments de leur structure interne, marques des corrections), nous avons cherché

<sup>43</sup>On remarquera que si nous autorisons l’établissement de plusieurs liens de même type entre une structure verbale et un objet du domaine pour le traitement des énumérations et des inattendus structuraux non marqués (cf. § 6.3.1.2), nous imposons que les objets du domaine soient situés du même côté de la structure verbale.

<sup>44</sup>On pourrait imaginer la production d’un énoncé tel que “*il est fermé le musée du Louvre il ouvre à quelle heure le musée Grévin*”, sur lequel nous obtiendrions deux structures de liens de coût identique et une ambiguïté sur l’objet de la requête horaire. Ce problème se retrouve en cas d’incises longues dans l’énoncé.

à produire une représentation sémantique fine des énoncés. Nous ne cherchons toutefois pas à interpréter les énoncés, nous nous contentons d'intégrer dans la représentation sémantique finale les différents indices qui peuvent être utiles pour leur interprétation. Au travers des exemples que nous avons présentés, nous avons en particulier insisté sur l'intégration au sein de la représentation des ambiguïtés résiduelles de l'analyse ainsi que de son degré de complétude et les éventuels éléments ignorés. Le choix de conserver les ambiguïtés permet de ne pas être trop dépendant des éventuelles erreurs de reconnaissance comme l'absence d'un coordonnateur logique par exemple. On notera par ailleurs que le formalisme lexicalisé nous permet assez facilement l'analyse d'éléments permettant de mieux prendre en compte les attentes des utilisateurs (comme les relâchements de contraintes).

Nous évaluons dans le chapitre suivant la robustesse de l'analyse face aux principaux phénomènes rencontrés : mouvement, inattendus structuraux et erreurs de reconnaissance. Nous dressons dans le dernier chapitre un bilan plus complet et critique de l'approche adoptée à partir de l'analyse qualitative des résultats de l'évaluation.



Troisième partie

Bilan et perspectives



## Chapitre 7

# Évaluation

Depuis une dizaine d'années, plusieurs campagnes d'évaluation internationales ont été menées portant sur les domaines traditionnels du T.A.L. : étiquetage, analyse syntaxique, calcul sémantique, etc. En dialogue oral, chacun des modules composant le système nécessite un paradigme d'évaluation différent. En ce qui concerne plus précisément la compréhension hors ou en contexte, il n'existe actuellement pas de méthodologie standard, ni même de pratique communément admise dans la communauté scientifique pour évaluer et comparer chaque système. Les protocoles d'évaluation sont généralement fondés sur des métriques quantitatives qui ont pour intérêt de permettre une étude objective<sup>1</sup> et reproductible des systèmes. L'évaluation d'un module de compréhension peut se faire en comparant la représentation sémantique fournie par le module de compréhension testé par rapport à une représentation prédéfinie servant de référence [Mariani, 1998]. Une autre solution est d'évaluer la réponse du système délivrée par l'accès à la base de données en la comparant à une réponse type [MADCOW, 1992; Antoine et al., 2000]. La référence est dans ce cas plus difficile à définir. Ces deux paradigmes ne peuvent être envisagés que dans le cadre d'une campagne d'évaluation d'envergure impliquant une étape lourde de définition des références prises en compte (voire d'homogénéisation des domaines d'application) entre les différents partenaires.

De plus, en se limitant à un champ applicatif précis ainsi qu'au calcul de taux de robustesse globaux sur des corpus supposés représentatifs, ces programmes ne fournissent qu'un diagnostic très grossier, et spécifique à la tâche considérée, du comportement des systèmes étudiés. En dehors du problème de généralité vis-à-vis du domaine applicatif, elles se caractérisent avant tout par un très faible pouvoir diagnostic, les performances globales du système étant très difficilement interprétables. Elles sont par conséquent dans l'incapacité de nous guider sur les évolutions futures du domaine [Antoine et Caelen, 1999]. Plusieurs évaluations objectives ont toutefois été proposées avec l'objectif de fournir une analyse qualitative en fonction des principales difficultés liées au trai-

---

<sup>1</sup>On distingue en dialogue oral les évaluations *objectives* qui fournissent une mesure sur les performances attendues du système (mesures le plus souvent quantitatives et plus rarement qualitatives) des évaluations *subjectives* fondées sur le jugement des utilisateurs du système en terme de convivialité, fiabilité et facilité d'utilisation du système.

tement de l'oral et des phénomènes linguistiques attestés. Ces propositions [FRACAS, 1996; Antoine et al., 2000] reposent sur un schéma d'évaluation détaillé mettant en jeu plusieurs séries de tests exhaustifs dédiés à des phénomènes bien spécifiés dont la conception est également très coûteuse.

La campagne d'évaluation DEF1 [Antoine et al., 2002] à laquelle nous avons participé dans le cadre du GDR-I3 *Compréhension robuste* du CNRS a précisément pour objectif de concilier précision d'analyse et légèreté d'évaluation. Nous présentons brièvement la méthodologie adoptée en précisant les limitations de cette première campagne puis détaillons les résultats obtenus pour notre système<sup>2</sup>.

## 7.1 Campagne DEF1

### 7.1.1 Méthodologie d'évaluation

L'un des objectifs de la méthodologie d'évaluation DEF1 est de permettre un diagnostic fin du comportement de chaque système par rapport à plusieurs phénomènes spécifiques de l'oral. Afin de s'affranchir de la définition d'une application ou d'une représentation sémantique commune, l'évaluation a porté sur un jeu de tests spécifiques à chaque système. Ce jeu de tests a été élaboré à partir d'une série de 20 énoncés (dits *initiaux*) fournis par le concepteur du système testé comme représentatifs de la tâche concernée. Chacun des autres participants a ensuite proposé 15 réécritures ou complexifications des énoncés initiaux (dits énoncés *dérivés*) en y intégrant des phénomènes spécifiques (hésitations, corrections, antépositions, incisives, etc.). Un participant propose ainsi une série de 300 tests pour chaque système à évaluer. Les phénomènes testés sont le reflet des centres d'intérêts et des approches de la compréhension adoptées par chacun des participants<sup>3</sup>. Une typologie des phénomènes testés a pu toutefois être dégagée. Nous en retenons 4 pour l'évaluation de notre système :

- tolérance aux erreurs de la reconnaissance de la parole,
- modélisation de la complexité structurale (requêtes multiples, reconnaissance des objets complexes),
- traitement robuste des incisives et des inattendus structuraux,
- traitement robuste des dislocations et altérations de l'ordre canonique de l'énoncé.

5 systèmes ont participé à l'évaluation, chacun des systèmes a ainsi été évalué sur 1200 énoncés dérivés<sup>4</sup>.

---

<sup>2</sup>Pour une présentation plus détaillée des objectifs et de la méthodologie de la campagne DEF1, on pourra se reporter à [Antoine et al., 2002] ou au site présentant les travaux du GDR <http://www.univ-ubs.fr/valoria/antoine/gdri3/index.html>

<sup>3</sup>Afin de se rendre compte de la diversité des phénomènes introduits, on trouvera 3 séries d'énoncés dérivés en annexe D, chacune ayant été produite par un participant différent à partir du même énoncé initial.

<sup>4</sup>Ont participé à cette campagne d'évaluation, le CLIPS (réservation hôtelière), l'IRIT (renseignement ferroviaire, tâche portant sur les horaires), le LIMSI (renseignement ferroviaire) et le VALORIA (notre laboratoire présentant deux systèmes de compréhension sur le domaine du renseignement tou-

### 7.1.2 Limites de la première campagne DEFI

Comme nous l'avons signalé, le but principal de cette campagne d'évaluation est de fournir un diagnostic objectif du comportement de chacun des systèmes. Le fait d'utiliser un jeu de tests spécifiques à chaque système ne peut donner lieu à comparaison. Tous les systèmes n'ont par ailleurs pas pu être évalués sur toutes les classes de problèmes (par exemple parce que les tests sortaient du périmètre de la tâche). Lors de cette première campagne, une liberté totale a été donnée à chaque concepteur des énoncés dérivés. Si ce choix a permis des échanges intéressants entre les participants à ce groupe de travail, il en résulte toutefois deux limitations principales<sup>5</sup> :

- Il s'agit d'énoncés simulés. En conséquence, on peut parfois s'interroger sur le réalisme et la pertinence réelle de certains de ces énoncés. Cette appréciation est évidemment très subjective et est révélatrice des stratégies de compréhension adoptées par chacun des participants. Nous donnons dans la suite de ce chapitre quelques exemples d'énoncés qui nous semblent, de notre point de vue, sujets à caution.
- Certains tests comportent plusieurs phénomènes distincts. Ces tests sont intéressants car certaines erreurs peuvent justement provenir de la présence conjointe de plusieurs phénomènes. Toutefois, le diagnostic détaillé de l'origine des erreurs attestées est, dans ce cas de figure, parfois difficile à effectuer.

### 7.1.3 Résultats et discussion

#### 7.1.3.1 Présentation de l'évaluation du système ROMUS

Nous avons effectué une répartition des 1200 énoncés testés en fonction de la présence, dans chacun d'entre eux, des phénomènes que nous cherchons à évaluer. Cette classification reste assez grossière étant donnée la difficulté de cette tâche de répartition. Elle nous a semblé toutefois acceptable pour fournir un premier diagnostic du comportement de notre système<sup>6</sup>. La plupart des énoncés comportent plus de deux phénomènes. Nous les avons classés en fonction du phénomène impliquant le plus de variations par rapport à l'énoncé initial<sup>7</sup>. Nous avons cependant introduit deux classes supplémentaires. La première regroupe les énoncés intégrant simultanément plusieurs phénomènes

---

ristique). On pourra trouver la description des approches adoptées et quelques entrées bibliographiques concernant ces systèmes dans [Antoine et al., 2002].

<sup>5</sup>Ces insuffisances ne remettent pas en cause les principes fondateurs de la méthodologie d'évaluation par défi. Elles dénotent simplement la nécessité de la définition d'un processus plus rigide des énoncés dérivés.

<sup>6</sup>Une autre présentation des résultats aurait pu être effectuée par série de tests. Chaque série est en effet d'un niveau de difficulté différent et il est plus facile d'identifier, au sein de chaque série, les phénomènes ayant été intégrés. Ceux-ci répondent en effet aux motivations de chacun des concepteurs et on trouve une certaine régularité dans leur manifestation qui aurait permis une meilleure classification. Toutefois, une telle présentation rend difficile l'interprétation globale des performances du système par type de phénomènes.

<sup>7</sup>Par exemple, nous avons considéré qu'une dislocation présente une plus forte variation que la présence d'une seule interjection comme *eah* entre deux segments.

imbriqués, pour lesquels il était difficile de décider quel était celui le plus perturbateur. La seconde regroupe les énoncés ne comportant aucun des phénomènes que l'on cherche à évaluer (classe notée DIVERS). Ces énoncés correspondent à des reformulations “standards” de l'énoncé initial<sup>8</sup>. Le nombre d'énoncés testés par classe de phénomènes est représenté dans le tableau 7.2. Chaque énoncé comporte en moyenne 13 mots.

Nous avons utilisé trois scores pour mesurer les performances de notre système :

- B (Bon) dénote le nombre d'énoncés donnant lieu à une représentation sémantique complète et correcte de l'énoncé,
- K (oK) dénote le nombre d'énoncés donnant lieu à une représentation sémantique incomplète mais intégrant le sens global de l'énoncé (omission d'une propriété ou d'une contrainte, ambiguïté non représentée par exemple),
- M (Mauvais) dénote le nombre d'énoncés dont la représentation est incorrecte (mauvaise identification des actes dialogiques, objet erroné d'une requête par exemple).

On notera que l'ensemble  $B + K$  dénote le nombre d'énoncé donnant lieu à une représentation sémantique *acceptable*. L'ensemble  $K + M$  peut être quant à lui considéré comme le nombre d'erreurs par rapport à nos objectifs de compréhension détaillée.

Deux versions de notre système ( $R_1$  et  $R_2$ ) ont été évaluées. Ces deux versions se distinguent principalement par la couverture de la grammaire de lien d'une part (environ 800 entrées pour  $R_1$  contre 1000 pour la version actuelle  $R_2$ ) et le processus d'étiquetage mis en œuvre d'autre part. Dans la première version du système, la stratégie envisagée pour l'étiquetage était de chercher à désambigüiser le plus possible les séquences étiquetées. Le taux de décision de cet étiqueteur était de 97% pour un taux de précision de l'ordre de 92% contre des taux respectifs de 80.4% et 97.5% pour l'étiqueteur actuel<sup>9</sup>.

### 7.1.3.2 Résultats

Nous présentons dans ce paragraphe les résultats de l'évaluation. Le tableau 7.1 indique les performances globales sur l'ensemble des tests pour chacune des versions du système. Le tableau 7.2 donne les taux de performance par classe de phénomènes testés. Enfin, le tableau 7.3 classe les principales erreurs attestées en fonction de l'origine supposée de l'erreur (phénomène impliqué) pour le système actuel.

D'une manière générale, les performances globales obtenues par notre système sur ces énoncés complexifiés sont acceptables. Elles rejoignent celles attestées par le sys-

---

<sup>8</sup>C'est le cas par exemple de l'énoncé “*est ce que le musée est ouvert demain matin*” dérivé à partir de “*donnez moi les horaires de fermeture du musée Grévin et du musée du Louvre*” cf. annexe D. Ces énoncés comportent des phénomènes divers (comme par exemple la référence implicite au contexte) pourtant importants, mais qui ne pas suffisamment représentés dans le jeu de tests car ne faisant pas partie des objectifs initiaux de l'évaluation.

<sup>9</sup>Les deux versions des étiqueteurs ont été évaluées sur le même ensemble constitué des 1200 énoncés de la campagne DEFI.

<i>Système</i>	B	K	M	% d'erreurs (K+M)	% acceptables (B+K)
$R_1$	952	59	189	20.6	84.2
$R_2$	1021	85	94	14.9	92.1

TAB. 7.1 – Performances globales des systèmes  $R_1$  et  $R_2$ 

tème LOGUS [Villaneau, Antoine, et Ridoux, 2002] sur le domaine du renseignement touristique<sup>10</sup>. En raison de la différence importante des performances attestées entre le premier prototype de notre système et sa version actuelle, nous avons examiné la différence de répartition de chacun de nos énoncés dans les trois catégories B, K et M considérées. Cet examen nous permet de souligner deux points importants :

- le taux de décision élevé de l'étiqueteur utilisé dans le premier système est la raison principale des erreurs attestées. Dans notre approche, une erreur dès la première étape de l'analyse est très pénalisante car elle est répercutée sur les niveaux supérieurs. L'amélioration constatée, notamment pour les erreurs les plus graves (c'est-à-dire celles sur le sens global de l'énoncé), souligne la pertinence de la diminution opérée du taux de décision de l'étiqueteur et la validité *a priori* des heuristiques successives utilisées pour la gestion de la multiplicité des analyses à prendre en compte.
- l'augmentation de la couverture de la grammaire de lien réalisée entre les deux prototypes<sup>11</sup>, contribue également, mais plus faiblement, à l'amélioration constatée. Il est important de noter que cette augmentation n'introduit pas d'ambiguïtés supplémentaires dans les représentations sémantiques correctement générées par le premier prototype.

Le tableau 7.2 indique les taux de performance par classe de phénomènes envisagés. Nous n'indiquons pas de pourcentage d'erreurs sur les énoncés de la catégorie DIVERS<sup>12</sup>. Les erreurs sur ce type d'énoncés sont pourtant les plus fréquentes. Elles correspondent toutefois à une diversité de phénomènes qui ne sont pas suffisamment représentés dans les tests pour que nous puissions en tirer des conclusions pertinentes sur chacun d'entre eux<sup>13</sup>.

<sup>10</sup>Le système LOGUS est développé au laboratoire VALORIA et repose sur une stratégie identique à la nôtre de segmentation en *chunks* suivi de leur mise en relation. Ce système propose une approche logique et lexicalisée, les représentations sémantiques étant obtenues par composition de  $\lambda$ -termes suivant le principe du calcul de Lambeck [Moorgat, 1997].

<sup>11</sup>Cette augmentation a principalement concerné l'intégration de nouvelles unités dans la grammaire. Seuls une dizaine de nouveaux concepts ont été ajoutés.

<sup>12</sup>'n.s.' signifie "non significatif".

<sup>13</sup>Ces phénomènes correspondent principalement aux ellipses, à certaines anaphores ou à la juxtaposition d'actes dialogiques différents. Chacun de ces phénomènes nécessiterait une évaluation sur un nombre d'énoncés plus important. L'évaluation de notre système dans le cadre de la campagne d'évaluation MEDIA devrait permettre de pouvoir apprécier leur influence sur la stratégie adoptée. On notera également qu'un nombre non négligeable de ces énoncés n'ont pu être traités car ils correspondent à des actes dialogiques de demande d'information non prévus par la grammaire.

Classes de phénomènes	Nombre de tests	$R_1$			$R_2$		
		K	M	% erreur	K	M	% erreur
Erreurs de reconnaissance	93	18	32	53.7	23	10	35.4
Inattendus structuraux	390	12	30	10.7	22	7	7.4
Mouvements	293	5	10	5.1	8	7	5.1
Complexité structurale	196	9	11	10.2	4	6	5.1
Phénomènes mul. imbriqués	96	10	36	47.9	18	20	39.5
DIVERS	132	5	70	n.s.	10	44	n.s.

TAB. 7.2 – Taux de performance attestés sur les différentes classes de phénomènes envisagées

Ces résultats soulignent une robustesse appréciable de notre système en présence des inattendus structuraux et sur les énoncés complexes (requêtes multiples, objets de requête de multiples, etc.). Les taux constatés d'erreur sur les énoncés comportant ces phénomènes sont proches (10% pour le premier prototype). Ceci n'est pas étonnant étant donnée la proximité des traitements mis en œuvre pour y faire face. On notera que la plupart des erreurs concernant ces phénomènes sont attestées sur des énoncés qui ne respectent pas les hypothèses sur lesquelles nous avons fondé notre stratégie d'analyse. C'est le cas par exemple des deux énoncés dérivés suivants :

- *le musée Grévin ses horaires de fermeture c'est quoi et le musée du Louvre*<sup>14</sup>
- *bonjour je voudrais la liste des hôtels trois étoiles à proximité de la gare s'il vous plaît*

Dans le premier cas, les contraintes imposées pour le traitement des coordinations ne nous permettent pas d'identifier directement le musée du Louvre comme objet de la requête horaire. On peut toutefois s'interroger sur la pertinence réelle de cet énoncé. Dans le second cas, l'hésitation simulée ne provoque pas la répétition entière du groupe 'à proximité'. L'étape de segmentation ne permet pas d'identifier le groupe prépositionnel 'à proximité de la gare', cette propriété est ainsi perdue<sup>15</sup>. Les améliorations constatées sur le traitement de ces phénomènes sont dues à l'étiquetage réalisé dans la version actuelle. Remarquons enfin que la classe des inattendus structuraux regroupe à la fois les phénomènes de répétition ou de correction ainsi que les incises et les faux-départs. Ces derniers phénomènes sont les causes principales des erreurs sur cette classe n'ayant pas pu être résolues avec la seconde version du système<sup>16</sup>.

Notre système montre de bonnes performances sur les phénomènes de mouvement des segments dans l'énoncé, phénomène qui a motivé notre approche. Nous discutons de la généralité de cette conclusion au chapitre suivant. Ces résultats font en revanche

<sup>14</sup>Cet énoncé a été classé dans la catégorie "Phénomènes multiples".

<sup>15</sup>L'erreur provient également de notre choix d'identifier la locution à-proximité plutôt que de construire des liens entre 'à', 'proximité' et 'de la gare' par exemple. Les raisons de ce choix reposaient toutefois principalement sur les régularités de production de parole.

<sup>16</sup>On notera en particulier la présence d'incises à l'intérieur d'un segment parmi les énoncés évalués.



nettement apparaître une insuffisance des deux prototypes de notre système face aux erreurs de reconnaissance. Ce résultat est toutefois à relativiser en raison de la sous-représentativité de l'ensemble de tests concernant ce phénomène (93 énoncés contre 390 pour les inattendus structuraux par exemple). On note par ailleurs que les erreurs de reconnaissance auxquelles notre système n'apporte pas une tolérance suffisante se distribuent selon des caractéristiques particulières. L'omission de certaines prépositions<sup>17</sup> a par exemple un impact plus important que la substitution d'un verbe par exemple<sup>18</sup>. D'autre part, soulignons que ces erreurs ont été simulées par les concepteurs des tests avec les réserves que l'on pourrait formuler sur certaines d'entre elles.

Nous résumons dans le tableau 7.3 les pourcentages des erreurs ayant été occasionnées par chacun des phénomènes attestés. Quand cela a été possible, nous avons reporté les énoncés comportant simultanément plusieurs phénomènes (classe "phénomènes multiples" dans les tableaux précédents) dans la catégorie du phénomène ayant vraisemblablement été à l'origine de l'erreur. La catégorie "combinaison" ne comptabilise ainsi que les erreurs occasionnées par la combinaison d'au moins deux phénomènes au sein du même énoncé.

<i>Caractéristique de l'erreur</i>	Nombre d'erreurs	ORIGINE de l'erreur (%)				
		Reco.	Inatt.	Mouv.	Compl.	combinaison
Compréhension erronée (M)	94	32.9	29.4	9.4	5.8	10.5
Compréhension incomplète (K)	85	13.8	10.6	10.6	8.5	10.6
Bilan (M+K)	179	22.9	19.5	10	7.2	10.6

TAB. 7.3 – Classification des sources d'erreurs – système  $R_2$

On retrouve sur les erreurs observées l'influence principale des inattendus structuraux et des erreurs de reconnaissance. Comme nous l'avons déjà souligné, les erreurs de la classe des inattendus structuraux concernent principalement les incises, notamment lorsque la simulation de ces phénomènes fait intervenir l'incise à l'intérieur d'un segment sans que celui-ci soit entièrement répété. L'influence des erreurs de reconnaissance est plus problématique étant donné que celles-ci sont les plus imprévisibles et les plus susceptibles de remettre en cause les régularités de la production orale sur lesquelles nous nous appuyons. On notera par ailleurs que ces erreurs sont celles qui conduisent le plus souvent à une compréhension erronée de l'énoncé. Ce type d'erreur implique que le locuteur engage un sous dialogue de clarification ou d'explicitation (dont l'ana-

<sup>17</sup>Les conséquences de l'omission d'une préposition varient en fonction du rôle de cette préposition dans l'énoncé et de son usage dans le domaine considéré. Par exemple, l'omission de la préposition "pour" dans l'énoncé "*y a t il des réductions pour les étudiants*" est sans conséquences étant donné que le rôle "propriété" de la réduction peut également être portée par le groupe "*étudiant*" seul.

<sup>18</sup>Dans ce dernier cas l'analyse se ramène à celle d'une structure elliptique.

lyse ne sera pas forcément très simple) voire une réinitialisation complète de l'échange utilisateur-système. Afin de nous faire une meilleure idée de l'influence de ces erreurs sur notre analyse, nous présentons dans le paragraphe suivant une évaluation plus précise dédiée à ce phénomène.

## 7.2 Influence des erreurs de reconnaissance

### 7.2.1 Tests à partir d'une dictée vocale

En l'absence de système de reconnaissance continue disponible au laboratoire, nous avons choisi de faire une première évaluation de l'influence des erreurs de reconnaissance sur la stratégie de compréhension adoptée à partir d'un système de dictée vocale. Notre choix s'est porté sur la dictée vocale d'IBM ViaVoice qui offre la possibilité d'être intégrée à terme à notre système pour la réalisation de tests plus ambitieux.

Le système de dictée vocale a été entraîné sur deux ensembles d'énoncés :

- dans un premier temps sur une lecture des énoncés du corpus PARISCORP. Il s'agissait avant tout d'entraîner le système sur le vocabulaire de base de notre application,
- dans un deuxième temps, nous avons utilisé deux séries d'énoncés dérivés de la campagne DEFI (600 énoncés). Nous avons cette fois cherché à "interpréter"<sup>19</sup> ces énoncés le plus naturellement possible en faisant varier la vitesse d'élocution et en forçant la reconnaissance des interjections telles que *eah*.

Une fois une certaine stabilité des performances atteinte, nous avons réalisé nos tests avec les 600 autres énoncés de la campagne DEFI<sup>20</sup>. Ces énoncés ont été de la même manière "interprétés" sans la référence à l'énoncé testé sous les yeux au moment de l'enregistrement. Nous avons ensuite retranscrit les énoncés réellement produits à partir de leur enregistrement. Cette retranscription nous a servi de référence pour l'identification des erreurs de reconnaissance éventuelles dans l'hypothèse de reconnaissance délivrée par le système. La dictée vocale délivre la séquence de mots la plus probable et non un graphe de mots.

### 7.2.2 Typologies des problèmes évalués

Classiquement, l'évaluation d'un module de reconnaissance est effectuée en comparant la séquence de mots reconnus (l'hypothèse) avec la séquence de mots attendus (la référence, dans notre cas la transcription des enregistrements réellement produits).

<sup>19</sup>Les énoncés n'ont pas été lus. Nous avons cherché à retranscrire leur structuration et les phénomènes présents sans avoir sous les yeux l'énoncé au moment de l'enregistrement. Cette solution a été suffisante pour adapter la dictée vocale à l'oralité.

<sup>20</sup>Ce choix introduit un certain nombre de biais dans cette expérience. En particulier, le modèle de langage a été particulièrement bien adapté au vocabulaire utilisé et à l'ensemble des dislocations correspondant aux énoncés testés. Ces biais sont inévitables dès lors que l'évaluation est réalisée à partir d'une dictée vocale. Ils ne sont toutefois pas trop préjudiciables par rapport à nos objectifs d'évaluation de l'influence des erreurs de reconnaissance vocale.

Cette comparaison donne une mesure en terme de substitutions, d'insertions et d'élisions de mots. Il est important de distinguer ces trois types d'erreurs car ils n'ont pas la même importance par rapport à l'analyse effectuée en aval. Nous avons également annoté à part les erreurs d'accord en genre et en nombre. Sur les 600 énoncés testés,

- 102 ne comportent aucune erreur de reconnaissance,
- 358 présentent au moins une erreur de reconnaissance "majeure" (c'est-à-dire au moins une insertion, une substitution ou une élision avec éventuellement des erreurs d'accord en genre et en nombre). Environ 60% de ces erreurs concernent l'élision ou la substitution de mots outils (prépositions, déterminants). Sur l'ensemble des erreurs "majeures" attestées on compte 57% de substitutions, 27% d'élisions et 14% d'insertions.
- 140 comportent au moins une erreur d'accord en genre ou en nombre, à l'exclusion de toute autre erreur de reconnaissance.

On notera que le système de dictée vocale présente d'une manière générale des performances meilleures que ne le laissent présager ces chiffres. Certaines erreurs ont été délibérément provoquées par la vitesse (relative) et le style d'élocution utilisés au moment de l'enregistrement. Il s'agissait d'obtenir un ensemble d'erreurs assez large afin de tester la robustesse de notre système sur une situation extrême.

### 7.2.3 Évaluation

Chacun des deux ensembles d'énoncés présentant des erreurs (erreurs "majeures" et erreurs d'accord en genre ou en nombre) a été analysé par notre système de compréhension. Les représentations sémantiques ont été comparées à celles obtenues à partir des énoncés de référence (transcription de l'enregistrement produit)<sup>21</sup>.

**Erreurs d'accord en genre et en nombre.** Ces erreurs sont globalement sans conséquences sur la qualité de l'analyse produite. Ce résultat était attendu étant donné que nous n'utilisons que peu ces informations morphologiques pour l'analyse<sup>22</sup>. On notera toutefois l'influence de deux types d'erreur particuliers : l'erreur entre le participe passé et le verbe à l'infinitif (é/er) et la confusion entre *ou* et *où*. La première erreur a été présente dans 38 énoncés. Elle a eu des répercussions sur 4 d'entre eux, dans lesquels la catégorie  $GV_{inf}$  était indispensable pour caractériser le rôle du groupe considéré. La confusion entre *ou* et *où* induit une erreur dans le type de demande d'information caractérisé (**Sélection** à la place de **Localisation**)<sup>23</sup>.

<sup>21</sup>Nous n'avons pas pu utiliser directement les représentations sémantiques déjà analysées dans le cadre de DEFI étant donné que les productions orales correspondant à ces énoncés sont parfois assez différentes des énoncés dérivés.

<sup>22</sup>Nous avons par ailleurs considéré dans l'ensemble suivant les substitutions portant sur les déterminants : *ses vs. ces*, *le vs. la*, ...

<sup>23</sup>Des processus de correction *ad hoc* pourraient être utilisés pour limiter ces erreurs, par exemple en faisant porter sur la coordination logique *ou* les rôles sémantiques de la forme *où* (reliée dans notre domaine applicatif à un nombre limité de verbes (être, se trouver, ...))

**Substitutions, insertions et élisions.** Ces erreurs ont eu une influence sur la représentation sémantique produite pour 60 énoncés sur les 373 (environ 16%). En revanche, elles n'ont jamais conduit à une compréhension erronée de l'énoncé mais uniquement à des représentations sémantiques incomplètes de celui-ci. Cette relativement faible influence s'explique par le fait que la plupart des substitutions ont été observées sur des structures verbales (*séquelles* à la place de *c'est quel, est ce qui vient pour est ce qu'il y a un, ...*). Notre analyse permet de traiter des structures elliptiques ne comportant pas ces verbes et ainsi limiter l'influence de ces erreurs.

Le système s'est montré très tolérant aux insertions de mots. Les segments erronés introduits ont été ignorés de la structure de liens produite. Nous n'avons toutefois observé que peu d'insertions de mots "pleins" ayant un rôle sémantique dans notre domaine d'application (20 énoncés). La plupart des autres insertions concernent des mots outils qui n'ont pas introduit de segmentation erronée de l'énoncé.

Pour les élisions, la robustesse est également appréciable. 20 énoncés (sur les 100 comportant une élision) ont été incomplètement compris. Les erreurs ont essentiellement concerné les omissions de certaines prépositions. L'omission des déterminants n'entraîne pas d'erreur de compréhension mais génère des ambiguïtés dans la représentation (notamment en raison de la perte du caractère défini / indéfini des articles).

Les erreurs de substitution sont les plus pénalisantes. En dehors de quelques cas marginaux sur des objets et propriétés du domaine<sup>24</sup>, les principales sources d'erreurs concernent essentiellement les substitutions de prépositions (perte du rôle sémantique correspondant à un segment) et de certains pronoms (*le quel* pour *lequel* par exemple).

#### 7.2.4 Conclusions

Les résultats obtenus sur cette évaluation permettent de relativiser l'influence de la reconnaissance observée dans le cadre de l'évaluation DEFI. Sur un ensemble plus conséquent de tests concernant ce phénomène (373 contre 93 énoncés) nous n'obtenons en effet que 16% d'erreurs ayant des répercussions (partielles) sur la qualité de la représentation sémantique produite. Ceci s'explique principalement par deux raisons :

- le caractère partiel de la mise en relation des segments et l'analyse possible des structures elliptiques d'une part,
- et du fait de l'intégration des ambiguïtés au sein de la représentation sémantique d'autre part. Ainsi une substitution sur un mot ayant le rôle de déterminant par exemple ne conduit qu'à l'intégration d'une ambiguïté supplémentaire dans la représentation.

Nous sommes conscients qu'il est également difficile de tirer des enseignements généraux à partir de cette évaluation. En particulier, nous n'avons pas pu évaluer les conséquences de l'insertion d'un mot ayant un rôle sémantique important dans le domaine

<sup>24</sup>pour cinq remplacé par '%' par exemple.

dans l'énoncé. Elle nous a seulement semblé plus pertinente que l'interprétation des résultats de la campagne DEFI sur ce point.

D'une manière générale, il semble difficile de faire face aux mots mal reconnus en ne considérant que la meilleure hypothèse de reconnaissance. Une solution viable ne peut être trouvée que par la prise en considération de l'ensemble des hypothèses. Nous abordons brièvement ce point dans le chapitre suivant.



## Chapitre 8

# Conclusions et perspectives

### 8.1 Conclusions

Nous pouvons résumer les principales caractéristiques de la stratégie de compréhension proposée par les aspects suivants :

1. Si on souhaite permettre la mise en œuvre de dialogues homme-machine naturels et accessibles au plus grand nombre, il faut, de notre point de vue, appréhender toute la variabilité et la complexité des phénomènes langagiers observés sur les corpus oraux. Notre approche est ainsi fondée sur la prise en compte explicite des régularités et des phénomènes observés dans les productions orales. Du modèle descriptif issu des travaux de Blanche-Benveniste sur le français parlé, nous avons déduit une stratégie d'analyse des énoncés oraux en deux étapes. Le premier niveau, une segmentation des énoncés, correspond à l'identification des unités minimales intervenant dans le processus de production de la parole. Il s'agit en fait directement de l'analyse des développements selon l'axe syntagmatique. Le second niveau couvre les variations paradigmatiques en utilisant des traitements spécifiques reposant sur des critères sémantico-pragmatiques. Des quelques études linguistiques de corpus réalisées, nous avons identifié les principaux phénomènes à prendre en compte et les conséquences qu'ils imposent en matière de traitement automatique. De là découle en particulier notre choix de l'utilisation d'un formalisme projectif. Contrairement aux méthodes sélectives stochastiques, notre approche permet une analyse qualitative à pouvoir diagnostique par rapport aux différents phénomènes caractérisés.
2. Nous avons souhaité développer une stratégie la plus générique possible. Ainsi, l'étape de segmentation des énoncés oraux est indépendante de l'application considérée et nous avons essayé de capitaliser dès cette étape l'analyse des concepts dont la réalisation est commune à plusieurs domaines applicatifs. On soulignera que la prise en compte des phénomènes oraux ne nécessite pas, dans cette étape, de filtrage particulier des énoncés. Les éléments traditionnellement éliminés avant toute analyse par un filtrage par *patterns* par exemple peuvent par ailleurs être des indices utiles pour la suite de l'analyse.

3. Enfin, nous espérons avoir montré qu'une analyse détaillée de l'énoncé était possible tout en respectant la contrainte de robustesse imposée. Notre approche permet d'aborder le traitement des énoncés dont l'analyse est essentielle dans des domaines relativement complexes : requêtes multiples, objets de requêtes multiples, relâchement de contraintes. Nous avons également insisté sur le fait qu'il était possible d'extraire, dès l'étape de compréhension hors-contexte, un certain nombre d'indices dont nous pensons qu'ils peuvent être utiles à la caractérisation plus précise des attentes de l'utilisateur : type du déterminant caractérisant les objets du domaine pour le traitement de la référence et présence des modaux par exemple. Par ailleurs, nous avons fait le choix, pour les mêmes raisons, d'intégrer au sein de la représentation sémantique les ambiguïtés identifiées de l'analyse (énumérations/corrections, portée d'une propriété, référence au contexte).

L'évaluation réalisée dans le cadre de la campagne d'évaluation DEFI a permis de souligner la robustesse de notre analyse face aux inattendus structuraux d'une part (principalement les répétitions et les corrections) mais surtout face aux phénomènes linguistiques complexes (mouvements –extractions, dislocations–, objets ou requêtes complexes) dont les méthodes sélectives stochastiques ont davantage de difficultés à rendre compte. Étant donnée l'extrême variabilité de leurs manifestations, chacune d'entre elles ne peut être significativement représentée dans les corpus qui servent d'ensembles d'apprentissage à ces systèmes. L'évaluation a notamment mis en évidence la pertinence de la première étape de l'analyse pour le traitement des énoncés oraux. Sans doute serait-il très excessif d'en conclure qu'une analyse fine des énoncés oraux doit être fondée sur un tel modèle. On peut cependant affirmer qu'une telle analyse ne nuit pas à la compréhension et que, au contraire, elle peut apporter, tout en restant efficace, des informations pertinentes, principalement pour le dialogue correctif.

Les limites de notre approche restent toutefois nombreuses et concernent principalement le second niveau de l'analyse. Par rapport aux objectifs que nous nous étions fixés au chapitre 2, soulignons en particulier que notre objectif de généralité n'est que partiellement atteint. Plusieurs raisons en sont la cause :

- Notre analyse sémantique est fondée sur les attentes conceptuelles de chaque segment dans l'énoncé. En dépit du fait que nous ne préjugeons pas de la présence *a priori* de l'ensemble de ces segments, la définition des attentes conceptuelles découle d'une caractérisation du domaine principalement déduite à partir d'une liste de schémas attendus. Nous restons ainsi sur ce point assez proche des méthodes sélectives.
- La qualité de l'analyse dépend principalement de la couverture de la grammaire de lien produite. Ce formalisme a principalement été choisi en raison de l'efficacité des algorithmes d'analyses sous-jacents. On regrettera toutefois qu'il soit difficile d'apprécier *a priori* la couverture de la grammaire de lien générée et les conséquences de l'intégration de nouveaux concepts. Le processus de génération automatique de la grammaire nous permet néanmoins une certaine souplesse pour sa maintenance. En ce qui concerne sa conception, on notera que nous avons veillé à ce que soient



séparées les connaissances liées au domaine (encodées dans la grammaire modulaire) et celles nécessaires au traitement des phénomènes linguistiques (ordre fixe des éléments coordonnés par exemple) encodées dans le processus de compilation spécifique aux unités les caractérisant. Ainsi, la conception d'une grammaire de lien sur un autre domaine est relativement aisée sous réserve d'une caractérisation cohérente du domaine concerné. On peut d'ailleurs s'interroger sur la réalité de la conception de systèmes génériques. Une autre solution est en effet de s'orienter vers des systèmes *adaptatifs* (ainsi que le suggèrent les travaux du MIT). Dans ce contexte, notre stratégie est intéressante étant donné le caractère générique de la première étape.

- Enfin, nous avons choisi le domaine du renseignement touristique en raison de l'ambiguïté sémantique qu'il était susceptible de présenter. Si le nombre de concepts à prendre en compte est sensiblement plus important que dans un domaine tel que *ARISE* par exemple, l'ambiguïté sémantique des énoncés oraux du domaine que nous avons étudiés reste toutefois relativement limitée. Ainsi, il est difficile de conclure sur l'adéquation de l'analyse en structures de lien et du système d'heuristiques utilisé pour classer les analyses concurrentes.

## 8.2 Perspectives

Parallèlement à notre participation à la campagne d'évaluation des systèmes de compréhension dans le cadre du projet *Technolangue MEDIA* et l'évaluation nécessaire de notre système sur les phénomènes peu représentés dans le cadre de la campagne *DEFI*, nous envisageons trois développements principaux de nos travaux :

### Linguistique de corpus

- Les hypothèses sur lesquelles sont fondées notre approche nécessitent d'être validées par des études linguistiques de corpus oraux de plus grande envergure. Nous souhaitons en particulier étudier les régularités de production des inattendus structuraux en ce qui concerne les groupes prépositionnels. Les études que nous avons menées jusqu'ici semblent suggérer que la préposition est reprise en cas d'interruptions. Nous avons ainsi choisi de caractériser les groupes prépositionnels dès l'étape de segmentation au détriment d'un rattachement de la préposition au cours de l'analyse des dépendances sémantiques, choix pourtant plus cohérent d'un point de vue théorique par rapport à l'approche adoptée. Cette régularité semble toutefois être remise en cause dans certains cas de figure.
- Nous souhaitons également approfondir l'analyse des usages sur le domaine du renseignement touristique. Cette analyse est importante pour la définition de la grammaire de lien, notamment en ce qui concerne les collocations.
- Enfin, en dehors des marques caractérisant les répétitions et les corrections, nous avons peu exploité les indices véhiculés par les appuis du discours et les éléments phatiques, alors même que nous les identifions. Nous comptons à terme exploiter les études linguistiques existantes à ce sujet.

Ces travaux pourront être menés sur le corpus OTG (*Office du Tourisme de Grenoble*) recolté dans la cadre de l'Action de Recherche Concertée "Dialogue Oral" de l'AUF, dont notre laboratoire vient d'assurer la transcription.

### Traitement des mots inconnus et hypothèses de reconnaissance

- Nous n'avons pas proposé de traitements spécifiques concernant les mots inconnus (c'est-à-dire absents du lexique utilisé par la première étape de l'analyse). Notre analyse se contente d'identifier leur présence dans l'énoncé. Cette information est conservée dans la représentation sémantique finale. Nous envisageons de compléter cette information en introduisant deux types de probabilités : un score de prédiction de la catégorie grammaticale (fondée sur les étiquettes des mots et des catégories des segments adjacents dans l'énoncé) et un score de prédiction du rôle sémantique de ce mot dans l'énoncé. Cette dernière information peut être dérivée des attentes conceptuelles non remplies des segments adjacents ainsi que des concepts obligatoires (par rapport au schéma identifié) non instanciés dans l'énoncé. Il serait toutefois hasardeux de chercher à instancier le rôle sémantique sur la base de ces informations étant donnée la richesse relative du domaine. Il semble préférable, de notre point de vue, de se contenter d'enrichir la représentation sémantique d'indices permettant au gestionnaire de dialogue de générer un dialogue pertinent. Une même stratégie est envisageable pour la prédiction des rôles sémantiques des segments ignorés, également conservés dans la représentation. Un segment peut en effet avoir été ignoré soit en raison d'une erreur de reconnaissance (sur une préposition par exemple) soit en raison d'une insuffisance de la couverture sémantique de la grammaire de lien.
- Comme nous l'avons souligné au chapitre précédent, le problème des mots mal reconnus nécessite la prise en compte de l'ensemble (ou d'un sous-ensemble) des hypothèses de reconnaissance. Une stratégie intéressante, compte tenu de notre approche, pourrait être d'envisager le processus d'étiquetage (voire de segmentation) sur le graphe de mots issus de la reconnaissance. Il a été montré qu'une telle analyse pouvait être efficacement menée pour le filtrage des hypothèses de reconnaissance au moyen d'automates à états finis [Noord, 2001].

**Compréhension contextuelle et dialogue** Notre principale priorité à moyen terme concerne principalement le traitement de la compréhension contextuelle. On peut imaginer que l'interprétation contextuelle doit être séparée de la compréhension hors-contexte et confiée au gestionnaire du dialogue. Nous pensons au contraire qu'elle doit être incorporée à notre système. Nous espérons avoir montré qu'une analyse plus raffinée et linguistiquement motivée des énoncés pouvait être menée dès l'étape de compréhension littérale, sans pour autant nuire à la robustesse de cette analyse. Il s'agit maintenant d'évaluer la pertinence de cette stratégie d'analyse et des indices que nous avons cherché à intégrer dans la représentation sémantique que ce soit pour la compréhension en contexte ou comme aide à la gestion du dialogue. Sur ce dernier point, nous pensons que notre stratégie de compréhension mériterait d'être évaluée sur le problème du dialogue correctif.

# Bibliographie

- Abeillé, A. [1991]. *Une grammaire lexicalisée d'Arbres Adjoints pour le français*. Thèse de Doctorat, Université de Paris 7.
- Abeillé, A., et Blache, P. [2000]. Grammaires et analyseurs syntaxiques. In J.-M. Pierrel (Ed.), *Ingénierie des langues* (p. 51–76). Hermès.
- Abeillé, A., Clément, L., et Kinyon, A. [2000]. Building a treebank for French. In *Proceedings of the Second International Conference on Language Resources and Evaluation, LREC 2000* (p. 87–94). Athènes, Grèce.
- Abney, S. [1991]. Parsing by chunks. In R. Berwick, S. Abney, et C. Tenny (Eds.), *Principle based parsing*. Kluwer Academic.
- Abney, S. [1994]. Partial parsing. In *Tutorials of ANCL'94*. Stuttgart. (<http://www.sfs.nphil.uni-tuebingen.de/~abney/Papers.html>)
- Abney, S. [1996]. Partial parsing via finite-state cascades. In J. Carroll (Ed.), *Workshop on Robust Parsing, Proceedings of ESSLLI'96* (p. 8–15).
- Adda, G., Mariani, J., Paroubek, P., Rajman, M., et Lecomte, J. [1999]. Métrique et premier résultat de l'évaluation GRACE des étiqueteurs morphosyntaxiques pour le français. In *Actes de TALN'99* (p. 15–24). Cargèse, France.
- Alessandro, C. d, et Tzoukermann, E. [2001]. La synthèse de la parole à partir du texte et le traitement automatique des langues. *Traitement Automatique des Langues, T.A.L., "La synthèse de la parole à partir du texte"*, 42(1), 7–15.
- Alexandersson, J., Reithinger, N., et Maier, E. [1997]. Insights into the dialogue processing of VERMOBIL. In *Proceedings of the 5<sup>th</sup> Conference on Applied Natural Language ANLP'97* (p. 33–40). Washington, États-Unis.
- Allen, J. [1998]. Natural language understanding. In B. Cummings (Ed.), (2 ed.). Benjamins Cummings.
- Allen, J., Miller, B., Ringger, E., et Sikorski, T. [1996]. A Robust System for natural Spoken Dialogue. In *Proceedings of ACL'96*.
- Antoine, J.-Y. [1994]. *Coopération syntaxe-sémantique pour la compréhension automatique de la parole spontanée*. Thèse de Doctorat, Institut National Polytechnique de Grenoble.
- Antoine, J.-Y., Bousquet-Vernhettes, C., Goulian, J., Kurdi, M., Rosset, S., Vigouroux, N., et Villaneau, J. [2002]. Predictive and objective evaluation of speech understanding : the challenge evaluation campaign of the I3 speech workgroup of the

- French CNRS. In *Proceedings of the Third International Conference on Language Resources and Evaluation, LREC 2002*.
- Antoine, J.-Y., et Caelen, J. [1999]. Pour une évaluation objective, prédictive et générique de la compréhension en CHM orale. *Langues*, 2(2), 130–139.
- Antoine, J.-Y., et Genthial, D. [1999]. Méthodes hybrides issues du TALN et du TAL Parlé : état des lieux et perspectives. In *Actes de TALN'99* (p. 1–17). Cargèse, France.
- Antoine, J.-Y., et Goulian, J. [1999]. Le français parlé spontané est-il un langage à ordre variable ? In *Actes des journées internationales de linguistique appliquée, JILA '99*. Nice, France.
- Antoine, J.-Y., et Goulian, J. [2001a]. Etude des phénomènes d'extraction en français parlé sur deux corpus de dialogue oral finalisé, application à la CHM orale. *T.A.L.*, 42(2), 413–439.
- Antoine, J.-Y., et Goulian, J. [2001b]. Word order variations and spoken man-machine dialogue in French : a corpus analysis on the ATIS domain. In *Proceedings of Corpus Linguistics 2001* (p. 22–30). Lancaster.
- Antoine, J.-Y., Siroux, J., Caelen, J., Villaneau, J., Goulian, J., et Ahafhaf, M. [2000]. Obtaining Predictive Results with an Objective Evaluation of Spoken Dialogue Systems : Experiments with the DCR Assessment Paradigm. In *Proceedings of the Second International Conference on Language Resources and Evaluation, LREC 2000*. Athènes, Grèce.
- ARPA (Ed.). [1995]. *Proceedings of 1995 ARPA spoken language systems technology workshop*. Austin, Texas.
- Aït-Mokhtar, S., et Chanod, J.-P. [1997]. Incremental finite-state parsing. In *Proceedings of ANLP'97* (p. 72–79). Washington.
- Aust, H., Oerder, M., Seide, F., et Steinbiss, V. [1995]. The Philips automatic train timetable information system. *Speech Communication*, 17, 249–262.
- Autebert, J. [1994]. *Théorie des langages et des automates*. Paris : Masson.
- Bangalore, S., et Joshi, A. [1999]. Supertagging : an approach to almost parsing. *Computational Linguistics*, 25(2), 237–265.
- Basili, R., Pazienza, M., et Zanzotto, F. [1999]. Lexicalizing a shallow parser. In *Actes de TALN'99* (p. 25–34). Cargèse, France.
- Bear, J., Dowding, J., et Shriberg, E. [1992]. Integrating multiple knowledge sources for detection and correction of repairs in Human-Computer dialogue. In *Proceedings of ACL'92* (p. 56–63). Newark, Danemark.
- Bear, J., Dowding, J., Shriberg, E., et Price, P. [1993]. *A system for labeling self-repairs in speech* (Rap. Tech. No. 552). SRI International.
- Bear, J., et Price, P. [1990]. Prosody, syntax and parsing. In *Proceedings of ACL'90* (p. 17–22). Pittsburgh.

- Bellegarda, J. [2001]. Robustness in statistical language modelling : review and perspectives. In J.-C. Junqua et G. van Noord (Eds.), *Robustness in language and speech technology* (p. 101–121). Dordrecht, Pays-Bas : Kluwer Academic Publishers.
- Bellenger, L. [1979]. *L'expression orale*. Presses Universitaires de France. (Collection "que sais-je")
- Bennacef, S., Bonneau-Maynard, H., Gauvain, J.-L., Lamel, L., et Minker, W. [1994]. A spoken language system for information retrieval. In *Proceedings of ICSLP'94* (p. 1271–1274). Yokohama, Japon.
- Bennacef, S., Devillers, L., Rosset, S., et Lamel, L. [1996]. Dialog in the RAILTEL Telephone-based System. In *Proceedings of ICSLP'96* (p. 550–553). Philadelphie, Pensylvanie.
- Bernsen, N., Dybkjaer, L., et Dybkjaer, H. [1996]. Cooperativity in human-machine and human-human spoken dialogue. *Discourse Processes*, 21(2), 213–236.
- Bessac, M., et Caelen, J. [1995]. Analyses pragmatiques, prosodiques et lexicales d'un corpus de dialogue oral homme-machine. In *Actes de JADT'95* (p. 363–370). Rome, Italie.
- Biber, D. [1986]. Spoken and written textual dimensions in English resolving the contradictory findings. *Language*, 62(2).
- Biber, D. [1993]. Using Register-Diversified Corpora for General Language Studies. *Computational Linguistics*, 19(2), 219–241.
- Biber, D., Johansson, S., Leech, G., Conrad, S., et Finegan, E. [1999]. *Longman grammar for spoken and written English*. Longman.
- Bilange, E. [1992]. *Dialogue personne-machine. Modélisation et réalisation informatique*. Paris : Hermès.
- Blanche-Benveniste, C. [1990]. *Le français parlé ; études grammaticales*. Paris : CNRS Editions.
- Blanche-Benveniste, C., et Bilger, M. [1999]. Français parlé – oral spontané, quelques réflexions. *Revue française de linguistique appliquée*, IV(2), 21–30. (dossier : l'oral spontané)
- Block, H.-U., et Schachtl, S. [1995]. What a grammar of spoken dialogues has to deal with. In G. Heyer et H. Haugeneder (Eds.), (p. 101–126).
- Bloomfield, L. [1933]. *Language*. New York.
- Bod, R. [1998]. Spoken Dialogue Interpretation with the DOP Model. In *Proceedings of COLING'98* (p. 138–144). Montréal, Canada.
- Bonneau-Maynard, H., et Devillers, L. [1998]. *Acquisition, Transcription et Annotation du Corpus Pariscorp* (Rap. Tech.). Action de Recherche Concertée "Dialogue Oral" (Arc B2) de l'AUF.
- Booth, T., et Thompson, R. [1973]. Applying probability measures to abstract languages. *IEEE Transactions on Computers*, 22(5), 442–450.

- Boros, M., Aretoulaki, M., Gallwitz, F., Nöth, E., et Niemann, H. [1997]. Semantic processing of out-of-vocabulary words in a spoken dialog system. In *Proceedings of Eurospeech'97* (p. 1887-1890).
- Boufaden, N., Delisle, S., et Moulin, B. [1998]. Analyse syntaxique robuste de dialogues transcrits : peut-on vraiment traiter l'oral à partir de l'écrit ? In *Actes de TALN'98* (p. 112-121). Paris, France.
- Bousquet-Vernhettes, C. [2002]. *Compréhension robuste de la parole spontanée dans le dialogue oral homme-machine - décodage conceptuel stochastique*. Thèse de Doctorat, Université Paul Sabatier, Toulouse.
- Brants, T. [1999]. Cascaded markov models. In *Proceedings of the 9<sup>th</sup> Conference of the European Chapter of the association for Computational Linguistics, EAACL'99*. Bergen, Norvège.
- Brill, E., Florian, R., Henderson, J., et Mangu, L. [1998]. Beyond N-Grams : can Linguistic Sophistication Improve Language Modeling. In *Proceedings of COLING'98* (p. 186-190). Montréal, Canada.
- Bruce, B. [1975]. Case systems for natural language. *Artificial Intelligence*, 6, 327-360.
- Bub, W., et Schwimm, J. [1996]. Vermobil : the Evolution of a Complex Large Speech-to-Speech Translation System. In *Proceedings of ICSLP'96* (p. 2371-2374).
- Butzberger, J., Murveit, H., et Weintraub, M. [1992]. Spontaneous Speech Effects in Large Vocabulary Speech Recognition Applications. In *Proceedings ARPA Workshop on Speech and Natural Language* (p. 339-344).
- Caelen, J. [1997]. Les corpus pour l'évaluation du dialogue homme-machine. In *Actes de JST'97 FRANCIL* (p. 215-222). Avignon, France.
- Calder, J., Klein, E., et Zeevat, H. [1988]. Unification Categorical Grammar : A Concise, Extendable Grammar for Natural Language Processing. In *Proceedings of the 12<sup>th</sup> International Conference on Computational Linguistics (COLING'88) and the 24<sup>th</sup> Annual Meeting of the Association for Computational Linguistics* (Vol. I, p. 83-86). Budapest.
- Candéa, M. [2000]. *Contribution à l'étude des pauses silencieuses et des phénomènes dits "d'hésitation" en français oral spontané. étude sur un corpus de récits en classe de français*. Thèse d'État. Université de la Sorbonne Nouvelle, Paris III.
- Carlson, R., et Hunnicutt, S. [1996]. Generic and Domain-Specific Aspects of the WAXHOLM NLP and Dialog Modules. In *Proceedings of ICSLP'96* (p. 677-680). Philadelphie, États-Unis.
- Carroll, J., et Briscoe, T. (Eds.). [1996]. *Proceedings of the Workshop On Robust Parsing, ESSLLI'96*. Prague, République Tchèque.
- Castellanos, A., Vidal, E., Varó, M., et Oncina, J. [1998]. Language understanding and subsequential transducer learning. *Computer Speech and Language*, 12, 193-228.
- Chanet, C. [2001]. 1700 occurrences de la particule quoi en français parlé contemporain : approche de la "distribution" et des fonctions du discours. *Marges Linguistiques*.

- Chanod, J. [2001]. Robust Parsing and Beyond. In J.-C. Junqua et G. van Noord (Eds.), *Robustness in Language and Speech Technology* (p. 187–204). Dordrecht, Pays-Bas : Kluwer Academic Publishers.
- Chanod, J.-P. [1993]. Problèmes de robustesse en analyse syntaxique. In *Conférence Informatique et Langue Naturelle* (p. 223–244).
- Chanod, J.-P., et Tapanainen, P. [1996]. A Robust Finite-State Parser for French. In *Proceedings of ESSLI'96*. Prague. (Robust Parsing Workshop)
- Charniak, E. [1993]. *Statistical Language Learning*. Cambridge : MIT Press.
- Chelba, C. [2000]. *Exploiting syntactic structure for language modeling*. Thèse de Doctorat, Johns Hopkins University.
- Chelba, C., et Jelinek, F. [1998]. Exploiting syntactic structure for language modeling. In *Proceedings of COLING'98* (p. 225–231). Montreal, Canada.
- Chomsky, N. [1957]. *Syntactic structure*. Cambridge : MIT Press.
- Chomsky, N. [1981]. *Lectures on government and binding*. Dordrecht : Foris. (Collection : Studies in generative grammar)
- Chung, G. [2000]. A three-stage solution for flexible vocabulary speech understanding. In *Proceedings of the International Conference on Spoken Language Processing ICSLP'2000*.
- Church, K. [1988]. A stochastic parts program and noun phrase parser for unrestricted text. In *Proceedings of the Second Conference on Applied NLP, ACL* (p. 136–143). Austin.
- Colineau, N. [1997]. *étude des marqueurs discursifs dans le dialogue finalisé*. Thèse de Doctorat, Université Joseph Fournier, Grenoble I.
- Covington, M. [1994]. *Discontinuous Dependency Parsing of Free and Fixed Word Order* (Artificial Intelligence Programs No. AI-1994-02). University of Georgia.
- Cunningham, H. [1999]. A definition and short history of Language Engineering. *Natural Language Engineering*, 5(1), 1–16.
- Dacuik, J. [1998]. *Incremental construction of Finite-State Automata and Transducers and their use in Natural Language Processing*. Thèse de Doctorat, Université de Gdansk.
- Daelemans, W., Buchholz, S., et Veenstra, J. [1999]. Memory-based shallow parsing. In *Proceedings of CoNLL*. Bergen.
- Delisle, S., Boufaden, N., et Moulin, B. [1998]. L'analyse syntaxique robuste de dialogues transcrits : de la parole aux actes en passant par l'écrit. In *Actes des XXIIèmes Journées d'Étude sur la parole, JEP'98* (p. 359–362). Martigny, Suisse.
- De Mori, R. [1994]. Apprentissage automatique pour l'interprétation sémantique. In *Proceedings of JEP'94* (p. 11–19). Tregastel, France.
- De Mori, R. [1998]. *Spoken Dialogues with computers*. Academic Press.
- Déjean, H. [1998]. *Concepts et algorithmes pour la découverte des structures formelles des langues*. Thèse de Doctorat, Université de Caen.

- Dontchev, D. [2000]. *Dictionnaire du français argotique, populaire et familier*. Éditions du Rocher.
- Dowding, J., Gawron, J., Applet, D., Bear, J., Cherny, L., Moore, R., et Moran, D. [1993]. GEMINI : a natural language system for spoken-language understanding. In *Proceedings of ACL'93* (p. 54–61).
- Dudewicz, E., et Mishra, S. [1988]. *Modern mathematical statistics*. New York, États-Unis. (Wiley series in probability and mathematical statistics)
- Dybkjaer, L., et Bernsen, N. [2000]. Usability issues in spoken dialogue systems. *Natural Language Engineering*, 6 (3–4), 243–272.
- Ehrlich, U., et Hanrieder, G. [1996]. Robust speech parsing. In J. Carroll (Ed.), *Proceedings of Workshop on Robust Parsing* (p. 26–34). Prague.
- Ehrlich, U., Hanrieder, G., Hitzemberger, L., Heisterkamp, P., Mecklenburg, K., et Regel-Brietzmann, P. [1997]. ACCeSS – Automated Call Center through Speech understanding System. In *Proceedings of Eurospeech'97* (p. 1823–1826). Rhodes, Grèce.
- Ejerhed, E. [1993]. Nouveaux courants en analyse syntaxique. *T.A.L.*, 34(1), 61–82.
- Ejerhed, E. [1996]. Finite State Segmentation of Discourse into Clauses. In A. Kornai (Ed.), *Proceedings of ECAI Workshop on Extended Finite State Model of Language* (p. 24–33).
- Engel, R. [2000]. CHUNKY : an example based machine translation system for spoken dialog. In *Proceedings of ICSLP'00*.
- Fillmore, C. [1968]. The case for case. In E. Bach et R. Harms (Eds.), *Universals in Linguistics Theory* (p. 1–90). Holt and Rinehart and Winston Inc.
- FRACAS. [1996]. *Using the framework* (Rap. Tech. No. D 16). FRACAS project, LRE 62–051.
- Fraser, N., et Gilbert, G. [1991]. Effects of system voice quality on user utterances in speech dialogue systems. In *Proceedings of Eurospeech'91* (p. 57–60). Gènes, Italie.
- Furui, S. [2000]. Steps towards natural Human Machine Communication in the 21th century. In *Proceedings ISCA workshop on voice operated telecom services* (p. 17–24). Belgique.
- Gadet, F. [1989]. *Le Français ordinaire*. Paris : Colin.
- Gadet, F. [1992]. *Le Français populaire*. Paris : PUF.
- Gala Pavia, N. [2001]. A two-tier corpus-based approach to robust syntactic annotation of unrestricted corpora. *Traitement Automatique des Langues, TAL*, 42(2), 381–411.
- Gallwitz, F., Aretoulaki, M., Boros, M., Haas, J., Harbeck, S., Huber, R., Niemann, H., et Nöth, E. [1998]. The Erlangen Spoken Dialogue System EVAR : A State-of-the-Art Information Retrieval System. In *Proceedings of the International Symposium on Spoken Dialogue* (p. 19–26). Sydney, Australie.



- Gauvain, J., De Mori, R., et Lamel, L. [2002]. Advances in Large Vocabulary Speech Recognition. *Computer Speech and Language*, 16. (Special Issue)
- Gee, J., et Gorsjean, F. [1983]. Performances structures : psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15, 411–458.
- Giguet, E. [1998]. *Méthode pour l'analyse automatique de structures formelles sur documents multilingues*. Thèse de Doctorat, Université de Caen.
- Gillet, J., et Ward, W. [1998]. A language model combining trigrams and stochastic context-free grammars. In *Proceedings of ICSLP'98* (Vol. 6, p. 2319–2322).
- Glass, J., Flammia, G., Goodine, D., Phillips, M., Polifroni, M., Seneff, S., et Zue, V. [1995]. Multilingual spoken language understanding in the MIT Voyager system. *Speech Communication*, 17, 1–18.
- Goddeau, D., Brill, E., Glass, J., Pao, C., Phillips, M., Polifroni, M., Seneff, S., et Zue, V. [1994]. GALAXY : A human-language interface to on-line travel information. In *Proceedings of ICSLP'94* (p. 707–710).
- Goddeau, D., et Zue, V. [1992]. Integrating Probabilistic LR-parsing into speech understanding systems. In *Proceedings of ICASSP'92* (Vol. 1, p. 181–184).
- Gorin, A., Riccardi, G., et Wright, J. [1997]. How may I help you? *Speech Communication*, 23, 113–127.
- Goulian, J. [1998]. *Analyse robuste du français parlé*. Mémoire de DEA, I.N.P.G, Université Joseph Fourier—Grenoble I.
- Grefenstette, G. [1992]. SEXTANT : Exploring Unexplored Contexts for Semantic Extraction from Syntactic Analysis. In *Proceedings of ACL'92*. Newark, Delaware.
- Grefenstette, G. [1996]. Light Parsing as Finite-state Filtering. In *Proceedings of ECAI'96, Workshop on Extended Finite State Models of Language*. Budapest, Hongrie.
- Grinberg, D., Lafferty, J., et Sleator, D. [1995]. *A robust parsing algorithm for Link Grammars* (Rap. Tech. No. CMU-CS-TR-95-125). États-Unis : CMU.
- Habert, B. [2000]. Des corpus représentatifs : de quoi, pour quoi, comment ? In M. Bilger (Ed.), *Linguistique sur corpus. études et réflexions* (p. 11–58). Perpignan : Champion & Presses Universitaires de Perpignan.
- Habert, B., Nazarenko, A., et Salem, A. [1997]. *Les linguistiques de corpus*. Paris : Armand Colin/Masson.
- Halber, A. [1999]. *Stratégie d'analyse pour la compréhension de la parole : vers une approche à base de Grammaires d'Arbres Adjoints Lexicalisées*. Thèse de Doctorat, École Nationale Supérieure des Télécommunications.
- Halliday, M. [1985]. *Spoken and written language*. Oxford : Oxford University Press.
- Halliday, M. [1994]. *An introduction to functional grammar* (deuxième ed.). Londres : Arnold.
- Haton, J.-P., Pierrel, J.-M., Perennou, G., et Gauvain, J.-L. [1991]. *Reconnaissance automatique de la parole*. Dunod.

- Heeman, P., et Allen, J. [1995]. The TRAINS spoken dialog corpus. *CD-ROM*.
- Heeman, P., et Allen, J. [1999]. Speech repairs, intentional phrases and discourse markers : modeling speakers' utterances in spoken dialogue. *Computational Linguistics*, 25(4), 527–573.
- Heeman, P., et Allen, J. [2001]. Improving robustness by modeling spontaneous speech events. In J.-C. Junqua et G. van Noord (Eds.), *Robustness in language and speech technology* (p. 123–152). Dordrecht, Pays-Bas : Kluwer Academic Publishers.
- Henry, S. [2002]. étude des répétitions en français parlé spontané pour les technologies de la parole. In *Actes de RECITAL-TALN'2002* (p. 467–476).
- Hetherington, I., et Zue, V. [1991]. New Words : Implications for Continuous Speech Recognition. In *Proceedings of Eurospeech'91*.
- Hindle, D. [1983]. Deterministic parsing of syntactic nonfluencies. In *Proceedings of the 21<sup>th</sup> Annual Meeting of the Association for Computational Linguistics, ACL'83* (p. 123–128).
- Hirose, K., Minematsu, N., Hashimoto, Y., et Iwano, K. [2001]. Continuous Speech Recognition of Japanese Using Prosodic Word Boundaries Detected by Mora Transduction Modeling of Fundamental Frequency Contours. In *Proceedings ISCA Tutorial and Research Workshop on Prosody in Speech Recognition and Understanding* (p. 61–66). Red Bank.
- Hirschman, L. [1998]. Language understanding evaluation : lessons learned from MUC and ATIS. In *Proceedings of the First International Conference on Language Resources and Evaluation, LREC'98* (p. 117–122). Grenade, Espagne.
- Hobbs, J., Applet, D., Bear, J., Israel, D., Kameyama, M., Stickel, M., et Tyson, M. [1997]. FASTUS : A cascaded finite-state transducer for extracting information from natural-language text. In Roche et Schabes (Eds.), *Finite State Devices for Natural Language Processing*. MIT Press.
- Holan, T., Kubon, V., Oliva, K., et Plátek, M. [2000]. On complexity of word order. *Traitement Automatique des Langues, T.A.L.*, 41(1), 273–300.
- Huckvale, M. [1996]. *Learning from the experience of building automatic speech recognition systems* (Rap. Tech.). University College London, Department of phonetics and linguistics. (<http://www.phon.ucl.ac.uk/research/speechrec.htm>)
- Hudson, R. [1984]. *Word Grammar*. Oxford : Blackwell.
- Hudson, R. [2000]. Discontinuity. *T.A.L.*, 41(1), 15–56.
- Hunt, A. [1994]. Improving speech understanding through integration of prosody and syntax. In *Proceedings of the 7<sup>th</sup> Aust. Joint Conference on Artificial Intelligence* (p. 442–449). Armidale, Australie.
- Hunt, M. [1998]. Les systèmes de dictée automatique dans la pratique. *La lettre d'information d'ELRA*, 4–7.
- Issar, S., et Ward, W. [1993]. CMU's Robust Spoken Language Understanding System. In *Proceedings of Eurospeech'93* (p. 2147–2150).

- Iyer, R., et Ostendorf, M. [1999]. Modeling long distance dependencies in language : Topic mixtures versus dynamic cache models. *IEEE Transactions on Speech and Audio Processing*, 7(1), 30–39.
- Jardino, M., et Adda, G. [1993]. Automatic word classification using simulated annealing. In *Proceedings of the 1993 IEEE international conference on Acoustics, Speech and Signal Processing* (Vol. 1, p. 41–44). Minneapolis.
- Jardino, M., et Beaujard, C. [2000]. Rôle du contexte dans les modèles "n-classes". Application et évaluation sur MASK et RAILTEL. In K. Chibout, J. Mariani, N. Masson, et F. Néel (Eds.), *Ressources et évaluation en ingénierie des langues* (p. 379–389). De Boeck Université, Duculot.
- Jelinek, F., et Chelba, C. [1999]. Putting language into language modeling. In *Proceedings of Eurospeech'99* (Vol. 1, p. KN1–KN5). Budapest, Hongrie.
- Jelinek, F., et Lafferty, J. [1991]. Computation of the probability of initial substring generation by stochastic context-free grammars. *Computational Linguistics*, 17, 315–323.
- Jones, B. [1994]. Exploring the role of punctuation in parsing natural text. In *Proceedings of COLING'94*. Budapest, Hongrie.
- Junqua, J.-C., et Haton, J.-P. [1996]. *Robustness in Automatic Speech Recognition*. Kluwer Academic Publishers.
- Kahane, S. [2000a]. Extractions dans une grammaire de dépendance lexicalisée à bulles. *T.A.L.*, 41(1), 211–243.
- Kahane, S. [2000b]. Les grammaires de dépendance. *T.A.L.*, 41(1), 7–13.
- Kahane, S. [2001]. Grammaires de dépendance formelles et théorie Sens-Texte. In *Tutoriel TALN'2001* (Vol. 2, p. 17–76). Tours.
- Kahane, S. [2002]. *Grammaires d'Unification Sens-Texte. Vers un modèle mathématique articulé de la langue*. Document de Synthèse d'Habilitation à Diriger des Recherches. Université Paris 7.
- Kahane, S., et Mel'čuk, I. [1999]. La synthèse sémantique ou la correspondance entre graphes sémantiques et arbres syntaxiques—le cas des phrases à extraction en français contemporain. *T.A.L.*, 40(2), 25–85.
- Kaiser, E. [1999]. Robust, Finite-State Parsing for Spoken Language Understanding. In *Proceedings of ACL'99*.
- Karlsson, F., Voutilainen, A., Heikkilä, J., et Anttila, A. [1995]. *Constraint grammar : A language-independent system for parsing unrestricted text*. Berlin/New York : Mouton de Gruyter.
- Karsenty, L. [2001]. *Quelles stratégies de transparence pour la gestion des erreurs de compréhension dans le dialogue vocal en langage naturel* (Rap. Tech.). IRIT.
- Karttunen, L. [1996]. Directed replacement. In *Proceedings of 34<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*. Santa Cruz.

- Karttunen, L., Chanod, J.-P., Grefenstette, G., et Schiller, A. [1996]. Regular expressions for language engineering. *Natural Language Engineering*, 2(4), 305–328.
- Kerbrat-Orecchioni, C. [1999]. L'oral dans l'interaction : une liberté surveillée. *Revue française de linguistique appliquée*, IV(2), 41–55. (dossier : l'oral spontané)
- Kerleroux, F., et Gadet, F. [1998]. Grammaire et données orales. *LINX Linguistique Institut Nanterre Paris X*, 18. (Etudes grammaticales de corpus oraux)
- Kipp, M., Alexandersson, J., et Reithinger, N. [1999]. Understanding Spontaneous Negotiation Dialogue. In *Proceedings Workshop on Knowledge and Reasoning in Practical Dialogue, Proceedings of ISCAI'99* (p. 57–64).
- Kita, K., Morimoto, T., Ohkura, K., Sagayama, S., et Yano, Y. [1994]. Spoken sentence recognition based on HMM-LR with hybrid language modeling. *IEICE Transactions of Inf. and Syst.*, E77-D(2), 258–265.
- Kuhn, R., et De Mori, R. [1990]. A cache-based natural language method for speech recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI*, 12(6), 570–582.
- Kuhn, R., et De Mori, R. [1995]. The application of semantic classification trees to natural language understanding. *IEEE Trans. Pattern Anal. Machine Intell.*, 17(5), 449–460.
- Kuhn, R., et De Mori, R. [1997]. Sentence Interpretation. In R. De Mori (Ed.), *Spoken Dialogues with Computers* (p. 485–520). Academic Press.
- Lafferty, J., Sleator, D., et Temperley, D. [1992]. Grammatical trigrams : a probabilistic model of link grammar. In R. Goldman (Ed.), *AAAI Fall symposium on Probabilistic Approaches to Natural Language Processing*. Cambridge : AAAI Press.
- Lamel, L., Bennacef, S., Gauvain, J., Dartigues, H., et Temem, J. [2002]. User evaluation of the MASK kiosk. *Speech Communication*. (à paraître)
- Lamel, L., Rosset, S., Gauvain, J., Bennacef, S., Garnier-Rizet, M., et Prouts, B. [2000]. The LIMSI ARISE system. *Speech Communication*, 31, 339–353.
- Lavie, A. [1996]. *GLR\* : A Robust Grammar-Focused Parser for Spontaneously Spoken Language*. Thèse de Doctorat, Carnegie Mellon University, Pittsburgh, PA.
- Letellier-Zarshenas, S., Nicolas, P., Goulian, J., et Antoine, J.-Y. [1999]. Inattendus structurels et communication orale finalisée : influence de la tâche et du contexte interactif. In *Actes des Journées Internationales de Linguistique Appliquée, JI-LA'99*. Nice, France.
- Levelt, W. [1983]. Monitoring and self-repair in speech. *Cognition*, 14, 41–104.
- Levin, E., et Pieraccini, R. [1995]. Concept-based spontaneous speech understanding system. In *Proceedings of Eurospeech'95* (p. 555–558). Madrid, Spain.
- Levin, L., Lavie, A., Woszczyna, M., Gates, D., Galvadà, M., Koll, D., et Waibel, A. [2000]. The JANUS-III Translation System : speech-to-speech translation in multiple domains. *Machine Translation*, 15(1), 3–25.
- Lokbani, M., et White, S. [1999]. La reconnaissance de la parole. *La recherche*, 319, 82.

- Lopez, P. [1999]. *Analyse d'énoncés oraux pour le dialogue homme-machine à l'aide de grammaires lexicalisées d'arbres*. Thèse de Doctorat, UHP—Nancy I.
- Luzzati, D. [1989]. *Recherches sur le dialogue homme-machine : modèles linguistiques et traitements linguistiques*. Doctorat d'État. Université Paris III.
- MADCOW. [1992]. Multi-Site Data Collection for a spoken language Corpus. In *Proceedings of DARPA Speech and Natural Language Workshop* (p. 7–14).
- Mariani, J. [1998]. The Aupelf-Uref Evaluation-based Language Engineering Actions and Related Projects. In *Proceedings of the First International Conference on Language Resources and Evaluation, LREC'98* (p. 123–128). Grenade, Espagne.
- Martínez, F., Tapias, D., Álvarez, J., et León, P. [1997]. Characteristics of slow, average and fast speech and their effects in large vocabulary continuous speech recognition. In *Proceedings of Eurospeech'97* (Vol. 1, p. 469–472). Rhodes, Grèce.
- Maynard, H., et Lefèvre, F. [2002]. Apprentissage d'un module stochastique de compréhension de la parole. In *Actes de JEP-TALN'02*. Nancy, France.
- Mecklenburg, K., Heisterkamp, P., et Hanrieder, G. [1995]. A robust parser for continuous spoken language using Prolog. In *Proceedings of the Workshop on Natural Language Understanding and Logic Programming, NLULP'95* (p. 127–141). Lisbonne, Portugal.
- Megyesi, B. [2002]. Shallow Parsing with PoS Taggers and Linguistic Features. *Journal of Machine Learning Research*, 2, 639–668.
- Mel'čuk, I. [1988]. *Dependency Syntax, Theory and Practice*. State University of New York Press, Alb.
- Mercier, G., Bigorgne, D., Miclet, L., Le Guennec, L., et Querré, M. [1990]. Recognition of Speaker Dependent Continuous Speech with KEAL. In Waibel et Lee (Eds.), *Readings in Speech Recognition* (p. 225–234).
- Mertens, P. [2002]. Le corpus de français ELICOP : consultation et exploitation. In J. Binon, P. Desmet, J. Elen, P. Mertens, et L. Sercu (Eds.), *Tableaux vivants : Opstellen aer-taal-en-onderwijs aangeboden aan Mark Debrock*. Leuven, Belgique : Presses Universitaires.
- Miller, S., Stallard, D., Bobrow, R., et Schwartz, R. [1996]. A Fully Statistical Approach to Natural Language Interfaces. In *Proceedings of the 34<sup>th</sup> Annual Meeting of the ACL* (p. 55–61).
- Minami, Y., Shikano, K., Takahashi, S., Yamada, T., Yoshioka, O., et Furui, S. [1995]. Large-vocabulary continuous speech recognition algorithm applied to a multi-modal telephone directory assistance system. *Speech Communication*, 15(301–310).
- Minker, W. [1996]. Compréhension dans le domaine ATIS. In *Actes des 21<sup>èmes</sup> Journées d'Étude sur la parole, JEP'96* (p. 417–420). Avignon, France : JEP'96.
- Minker, W. [1999]. *Compréhension automatique de la parole spontanée*. L'Harmattan.
- Méloni, H. [1982]. *étude et réalisation d'un système de reconnaissance de la parole continue*. Doctorat d'État. Université d'Aix-Marseille II.

- Moeschler, J., et Auchlin, A. [1997]. *Introduction à la linguistique contemporaine*. Masson. (Collection "Cursus", série "Lettres")
- Mollá, D., Schneider, G., Schwitter, R., et Hess, M. [2000]. Answer extraction using dependency grammar in ExtrAns. *Traitement Automatique des Langues, T.A.L.*, 41(1), 145–178.
- Monceaux, L., et Robba, I. [2002]. Les analyseurs syntaxiques : atouts pour une analyse des questions dans un système de question-réponse ? In *Actes de TALN 2002* (p. 195–204). Nancy, France.
- Moorgat, M. [1997]. Categorical Type Logics. In J. van Benthem et A. ter Meulen (Eds.), *Handbook of Logic and Language* (p. 93–177). Elsevier Science B.V.
- Morel, M.-A. [1988]. Dialogue oral homme-machine : prévisibilité et données d'observation. *Revue LINX*, 18, 137–152.
- Morel, M.-A. [1989]. *Analyse linguistique de corpus*. Paris : Publications de la Sorbonne Nouvelle.
- Murveit, H., Butzberger, J., Digalakis, V., et Weintraub, M. [1993]. Large vocabulary dictation using SRI's DECIPHER speech recognition system : Progressive search techniques. In *Proceedings IEEE ICASSP'93* (Vol. 2, p. 319–322).
- Nakatani, C., et Hirschberg, J. [1994]. A corpus-based study of repair cues in spontaneous speech. *Journal of the Acoustical Society of America*, 95(3), 1603–1616.
- Nederhof, M., Bouma, G., Koeling, R., et Noord, G. van. [1997]. Grammatical analysis in the OVIS spoken-dialogue system. In *Proceedings of ACL/EACL workshop on Spoken Dialog Systems* (p. 66–73). Madrid, Espagne.
- Néel, F., Chollet, G., Lamel, L., Minker, W., et Constantinescu, A. [1996]. Reconnaissance et compréhension de la parole : Évaluation et applications. In H. Méloni (Ed.), *Fondements et perspectives en Traitement Automatique de la Parole* (p. 331–367). AUPELF-UREF.
- Niesler, T., et Woodland, P. [1996]. A variable-length category-based n-gram language model. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing* (Vol. 1, p. 164–167). Atlanta.
- Noord, G. van. [1997a]. An efficient implementation of the head corner parser. *Computational Linguistics*, 23(3), 425–456.
- Noord, G. van. [1997b]. FSA Utilities : A toolbox to manipulate finite-state automata. In D. Raymond, D. Wood, et S. Yu (Eds.), *Automata Implementaton* (p. 87–108). Springer Verlag.
- Noord, G. van. [2001]. Robust Parsing of Word Graphs. In J.-C. Junqua et G. van Noord (Eds.), *Robustness in language and speech technology* (p. 205–238). Dordrecht, Pays-Bas : Kluwer Academic Publishers.
- Noord, G. van, Bouma, G., Koeling, R., et Nederhof, M.-J. [1999]. Robust Grammatical Analysis for Spoken Dialogue Systems. *Natural Language Engineering*, 5(1).

- Noord, G. van, et Gederman, D. [1999]. An extensible Regular Expression Compiler for Finite-state approaches in Natural Language Processing. In O. Boldt et H. Juer-gensen (Eds.), *Automata implementation, Proceedings of the 4<sup>th</sup> International Workshop on Implementation Automata, WIA'99*. Postdam, Allemagne.
- Okada, M., et Otsuka, H. [1993]. Incremental elaboration in generating spontaneous speech. In *Proceedings of International Symposium on Spoken Dialogue, ISSD'93* (p. 53–56). Tokyo, Japon.
- Olive, J., Greenwood, A., et Coleman, J. [1993]. *Acoustics of American English Speech*. Springer Verlag.
- Os, E. den, Boves, L., Lamel, L., et Baggia, P. [1999]. Overview of the ARISE project. In *Proceedings of Eurospeech'99* (p. 1527–1530). Budapest, Hongrie.
- Oviatt, S. [1996]. Multimodal interfaces for dynamic interactive maps. In *Proceedings of the Conference on Human Factors in Computing Systems CHI'96* (p. 95–102).
- Ozkan, N. [1994]. *Vers un modèle dynamique du dialogue : analyses de dialogues finalisés dans une perspective communicationnelle*. Thèse de Doctorat, Institut National Polytechnique de Grenoble.
- Pallett, D., Fiscus, J., Fisher, W., Garofolo, J., Lund, B., et Pryzbocki, M. [1994]. 1993 benchmark tests for the ARPA spoken language program. In *Proceedings of DARPA Speech and Natural Language Workshop* (p. 14–27).
- Paroubek, P., et Rajman, M. [2000]. étiquetage morpho-syntaxique. In J.-M. Pierrel (Ed.), *Ingénierie des langues* (p. 131–150). Hermès.
- Peckham, J. [1993]. A New Generation of Spoken Dialogue Systems : Results and Lessons from the SUNDIAL Project. In *Eurospeech'93* (p. 33–40).
- Pereira, F., et Warren, D. [1980]. Definite Clause Grammars for language analysis – a survey of the formalism and a comparaison with Augmented Transitions Networks. *Artificial Intelligence*, 13, 231–278.
- Pieraccini, R., Levin, E., et Vidal, E. [1993]. Learning how to understand language. In *Proceedings of Eurospeech'93* (p. 1407–1412).
- Pieraccini, R., Tzoukermann, E., Garelov, Z., Gauvain, L., Levin, E., Lee, C.-H., et Wilpon, J. [1992]. A Speech Understanding System Based on Statistical Representation of Semantics. In *International Conference on Acoustics, Speech and Signal Processing, ICASSP'92* (p. 193–196). San-Francisco, États-Unis.
- Pierrel, J.-M. [1988]. Dialogue homme-machine en langage naturel écrit ou oral. In EC2 (Ed.), *Actes des 1<sup>ères</sup> journées du PRC Communication Homme-Machine* (p. 151–182). Paris.
- Pierrel, J.-M., et Romary, L. [2000]. Dialogue Homme-Machine. In J.-M. Pierrel (Ed.), *Ingénierie des langues* (p. 331–349). Hermès.
- Pollard, C., et Sag, I. [1994]. *Head-driven Phrase Structure Grammar*. Chicago : University of Chicago Press.
- Price, P. [1990]. Evaluation of spoken language systems : the ATIS domain. In *Proceedings DARPA Speech and Natural Language Workshop* (p. 91–95).

- Price, P. [1996]. In R. Cole (Ed.), *Survey of the state of the art in Human Language Technology*. Cambridge University Press.
- Pérénou, G., Calmès, M. de, Lavelle, A., et Tronel, R. [1998]. DEMON : un système de dialogue oral spontané pour l'accès téléphonique aux informations d'horaire de train—Problèmes de robustesse. *Lettre de l'IA*(134-135-136).
- Rabiner, L. [1989]. A tutorial on hidden Markov models and selected applications in speech recognition. In *Proceedings of IEEE* (Vol. 77, p. 257–286).
- Rabiner, L., et Juang, B. [1986]. An introduction to Hidden Markov Models. *IEEE Transactions on acoustics, Speech and Signal processing*, 3(1), 4–16.
- Rabiner, L., et Juang, B. [1993]. *Fundamentals of Speech Recognition*. Prentice Hall.
- Rambow, O., et Joshi, A. [1994]. A formal look at dependency grammars and phrase-structure grammars with special considerations of word-order phenomena. In W. L. (Ed.), *Current issues in Meaning Text* (p. 1–20). Londres : Pinter.
- Roark, B. [2002]. Probabilistic Top-Down Parsing and Language Modeling. *Computational Linguistics*, 27(2), 249–276.
- Roche, E. [1997]. Parsing with finite-state transducers. In E. Roche et Y. Schabes (Eds.), *Finite-state language processing* (p. 241–281). MIT press.
- Roche, E., et Schabes, Y. [1997a]. Deterministic Part-of-Speech Tagging with Finite State Transducers. In E. Roche et Y. Schabes (Eds.), *Finite state Language Processing* (p. 205–239). MIT Press.
- Roche, E., et Schabes, Y. (Eds.). [1997b]. *Finite state Language Processing*. MIT Press.
- Rosé, C., et Lavie, A. [1997]. An efficient distribution of labor in two stage robust interpretation process. In *Proceedings of Empirical Methods in Natural Language Processing EMNLP'97*. États-Unis.
- Rosé, C., et Lavie, A. [2001]. Balancing robustness and efficiency in unification-augmented context-free parsers for large practical applications. In J.-C. Junqua et G. van Noord (Eds.), *Robustness in language and speech technology* (p. 239–269). Dordrecht, Pays-Bas : Kluwer Academic Publishers.
- Rose, R., Yao, H., Riccardi, G., et Wright, J. [2001]. Integration of utterance verification with statistical language modeling and spoken language understanding. *Speech Communication*, 34, 321–331.
- Rossato, S., Blanchon, H., et Besacier, L. [2002]. Évaluation du premier démonstrateur de traduction de parole dans le cadre du projet nespole! In *Actes de TALN'02, Conférence associée "Couplage de l'écrit avec l'oral"* (Vol. 2, p. 149–161). Nancy, France.
- Rosset, S. [2000]. *Stratégies et gestionnaire de dialogue pour des systèmes d'interrogation de bases de données à reconnaissance vocale*. Thèse de Doctorat, Université Paris XI—Orsay.
- Rosset, S., Lamel, L., Bennacef, S., Devillers, L., et Gauvain, J.-L. [2000]. Corpus oral de renseignement touristique. In K. Chibout, J. Mariani, N. Masson, et F. Néel



- (Eds.), *Ressources et évaluation en ingénierie des langues* (p. 483–489). De Boeck Université, Duculot.
- Roubaud, M. [2000]. *Les constructions pseudo-clivées en français contemporain*. Honoré Champion.
- Roukos, S. [1996]. In R. Cole (Ed.), *Survey of the state of the art in Human Language Technology* (p. 35–42). Cambridge University Press.
- Ruland, T., Rupp, C., Spilker, J., Weber, H., et Worm, K. [1998]. Making the most of multiplicity : a multi-parser multi-strategy architecture for the robust processing of spoken language. In *Proceedings of the International Conference on Spoken Language Processing, ICSLP'98* (p. 570–573). Sydney, Australie.
- Sabah, G. [1994]. *Projet DALI*. Rapport d'activité GDR-PRC Communication Homme-Machine.
- Sabah, G. [1997]. *Rapport final du projet DALI (dialogue Adaptatif : Langue et Interaction* (Rap. Tech.). LIMSI. ([http://www-geod.imag.fr/pages\\_html/projets/DALI.html](http://www-geod.imag.fr/pages_html/projets/DALI.html))
- Sabah, G., Vivier, J., Vilnat, A., Pierrel, J.-M., Romary, L., et Nicolle, A. [1997]. *Machine, langue et dialogue*. L'Harmattan.
- Sadek, D. [1996]. Le dialogue homme-machine : de l'ergonomie des interfaces à l'agent intelligent dialoguant. In Lavoisier (Ed.), *Nouvelles Interfaces Homme-Machine* (Vol. 18, p. 277–321). (Série ARAGO)
- Schabes, Y. [1992]. Stochastic lexicalized tree-agjoining grammars. In *Proceedings of the 14<sup>th</sup> International Conference on Computational Linguistics, COLING'92*. Nantes, France.
- Schneider, G. [1998]. *A linguistic comparison of constituency, dependency and link grammar* (Rap. Tech.). ExtrAns Research Report, Institut d'informatique, Université de Zürich.
- Seneff, S. [1989]. TINA : A Probabilistic Syntactic Parser for Speech Understanding Systems. In *Proceedings DARPA Speech and Natural Language Workshop* (p. 168–178). Philadelphie, États-Unis.
- Seneff, S. [1992a]. Robust parsing for spoken language systems. In *Proceedings of ICASPP'92* (p. 189–192).
- Seneff, S. [1992b]. TINA : a natural language system for spoken language applications. *Computational Linguistics*, 18(1), 61–86.
- Seneff, S. [1998]. The use of linguistic hierarchies in speech understanding. In *Proceedings of the International Conference on Spoken Language Processing, ICSLP'98* (p. 3321–3333). Sydney, Australie.
- Seneff, S., Lau, R., et Poliforni, J. [1999]. Organization, Communication, and Control in the GALAXY-II Conversational System. In *Proceedings of Eurospeech'99* (p. 1271–1274). Budapest, Hongrie.

- Shao, J., Tazine, N., Lamel, L., Prouts, B., et Schröter, S. [1998]. An Open System Architecture for a Multimedia and Multimodal User Interface. In *Proceedings of the 3rd Technology for Inclusive Design and Equality (TIDE)*. Helsinki, Finlande.
- Shriberg, E. [1994]. *Preliminaries to a theory of speech disfluencies*. Thèse de Doctorat, Département de Psychologie, Université de Californie, Bekerley.
- Shriberg, E., Bates, R., et Stolcke, A. [1997]. A prosody-only decision-tree model for disfluency detection. In *Proceedings of Eurospeech'97* (Vol. 5, p. 2383–2386). Rhodes, Grèce.
- Siroux, J., Guyomard, M., Jolly, Y., Multon, F., et Remondeau, C. [1995]. Speech and tactile-based GEORAL system. In *Proceedings of Eurospeech'95*. Madrid, Espagne.
- Sleator, D., et Temperley, D. [1991]. *Parsing English with a Link Grammar* (Rap. Tech.). CMU, États-Unis : CMU-CS-91-196.
- Smaïli, K., Zitouni, I., Charpillat, F., et Haton, J.-P. [1997]. An hybrid language model for a continuous dictation prototype. In *Proceedings of Eurospeech'97* (p. 2723–2726). Rhodes, Grèce.
- Srinivas, B. [1997]. Performance Evaluation of Supertagging for partial Parsing. In *Proceedings of the 5<sup>th</sup> International Workshop on Parsing Technologies*. Philadelphia, États-Unis.
- Stallard, D., et Bobrow, R. [1992]. Fragment Processing in the DELPHI System. In *Proceedings DARPA Workshop on Speech and Natural Language* (p. 305–310).
- Stolcke, A., et Segal, J. [1994]. Precise n-gram probabilities from stochastic context-free grammars. In *Proceedings of the 32<sup>nd</sup> Meeting of the Association for Computational Linguistics, ACL'94* (p. 74–79). Las Cruces.
- Stolcke, A., et Shriberg, E. [1996]. Statistical language modeling for speech disfluencies. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'96* (Vol. 1, p. 405–408). Atlanta, GA.
- Tapanainen, P., et Järvinen, T. [1997]. A non-projective dependency parser. In *Proceedings of the 5<sup>th</sup> Conference on Applied Natural Language Processing (ANLP'97)* (p. 64–71). Washington, D.C.
- Tapias Merino, D. [2001]. Speaker compensation in automatic speech recognition. In J.-C. Junqua et G. van Noord (Eds.), *Robustness in language and speech technology* (p. 47–100). Kluwer Academic.
- Temem, J., Lamel, L., et Gauvain, J. [1999]. The MASK demonstrator : an emerging technology for user-friendly passengers kiosk. In *Proceedings World Congress on Railway Research*. Tokyo, Japon.
- Tesnière, L. [1959]. *Éléments de syntaxe structurale*. Paris : Klincksiek.
- Traum, D., et Heeman, P. [1997]. Utterance units in spoken dialogue. In E. Maier, M. Mast, et S. LuperFoy (Eds.), *Dialogue Processing in Spoken Dialogue System* (p. 125–140). Heidelberg : Springer-Verlag. (Lectures Notes in Artificial Intelligence)

- Valli, A., et Véronis, J. [1999]. étiquetage grammatical des corpus de parole : problèmes et perspectives. *Revue Française de Linguistique Appliquée*, IV(2), 113–133. (dossier : l'oral spontané)
- Vergne, J. [2000]. *Etude et modélisation de la syntaxe des langues à l'aide de l'ordinateur—Analyse syntaxique automatique non combinatoire*. Thèse d'habilitation. Université de Caen.
- Veth, J. de, Cranen, B., et Boves, L. [2001]. Acoustic features and distance measure to reduce vulnerability of ASR performance due to the presence of a communication channel and/or background noise. In J.-C. Junqua et G. van Noord (Eds.), *Robustness in Language and Speech Technology* (p. 9–45). Dordrecht, Pays-Bas : Kluwer Academic Publishers.
- Villaneau, J., Antoine, J.-Y., et Ridoux, O. [2002]. LOGUS : un système formel de compréhension du français parlé spontané – présentation et évaluation. In *Actes de TALN'2002* (p. 165–174). Nancy, France.
- Viterbi, A. [1967]. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, IT-13, 260–269.
- Véronis, J. [2000]. Annotation automatique de corpus. In J.-M. Pierrel (Ed.), *Ingénierie des langues* (p. 111–130). Hermès.
- Waibel, A., Finke, M., Gates, D., Galvadà, M., Kemp, T., Lavie, A., Maier, M., Mayfield, M. A., L. and, Rogina, I., Shima, K., Sloboda, T., Woszczyna, M., Zeppenfeld, T., et Zahn, P. [1996]. JANUS-II-Translation of Spontaneous Conversational Speech. In *Proceedings of ICASSP'96* (p. 409–412).
- Ward, W. [1990]. The CMU Air Travel Information Service : Understanding Spontaneous Speech. In *Proceedings ARPA Workshop on Speech and Natural Language* (p. 127–129).
- Ward, W. [1991]. Understanding spontaneous speech : the PHOENIX system. In *Proceedings of ICASSP'91* (p. 365–367).
- Webber, B. [1999]. Computational Aspects of Discourse and Dialogue. In *The handbook of discourse analysis*. Blackwell Publishers Ltd.
- Yamada, M., Itoh, F., Sakai, K., Komori, Y., Ohora, Y., et Fujita, M. [1995]. A spoken dialogue system with active/non active word control for cd-rom information retrieval. *Speech Communication*, 15(355–365).
- Zanten, G. van, Bouma, G., Sima'an, K., Noord, G. van, et Bonnema, R. [1999]. Evaluation of the NLP component of the OVIS2 spoken dialogue system. In F. van Eynde, I. Schuurman, et N. Schelkens (Eds.), *Computational Linguistics in the Netherlands 1998* (p. 213–229). Amsterdam.
- Zechner, K. [1998]. Automatic Construction of Frame Representations for Spontaneous Speech in Unrestricted Domains. In *Proceedings of COLING/ACL 98* (p. 1448–1452). Montreal, Canada.
- Zechner, K. [2001]. Automatic Generation of Concise Summaries of Spoken Dialogues in Unrestricted Domains. In *Proceedings of the 24<sup>th</sup> ACM-SIGIR International*

*Conference on Research and Development in Information Retrieval*. New Orleans, LA.

- Zechner, K., et Waibel, A. [1998]. Using Chunk Based Partial Parsing of Spontaneous Speech in Unrestricted Domains for Reducing Word Error Rate in Speech Recognition. In *Proceedings of COLING/ACL 98* (p. 1453–1459). Montreal, Canada.
- Zhang, R., Black, E., et Finch, A. [1999]. using detailed linguistic structure in language modeling. In *Proceedings of Eurospeech'99* (Vol. 4, p. 1815–1818). Budapest, Hongrie.
- Zue, V. [1997]. Conversational interfaces : Advances and challenges. In *Proceedings of Eurospeech'97* (Vol. 1, p. KN9–KN18). Rhodes, Grèce.
- Zue, V., Seneff, S., Glass, J., Hetherington, L., Hurley, E., Meng, H., Pao, C., Poliforni, J., Schloming, R., et Schmid, P. [1997]. From interface to content : Translingual access and delivery of on-line information. In *Proceedings of Eurospeech'97* (p. 2227–2230). Rhodes, Grèce.
- Zue, V., Seneff, S., Poliforni, J., Phillips, M., Pao, C., Goddeau, D., Glass, J., et Brill, E. [1994]. PEGASUS : A spoken language interface for on-line air travel planning. *Speech Communication*, 15, 331–340.

# Annexes



## Annexe A

# Étude des phénomènes d'extraction en français parlé sur deux corpus de dialogue oral finalisé

Nous reproduisons dans cette annexe une version courte de notre article publié dans la revue *Traitement Automatique des Langues* [Antoine et Goulian, 2001a]. Nous y présentons le cadre de notre étude et détaillons l'ensemble des observations quantitatives effectuées (fréquences d'occurrence, répartition des extractions par procédé et par fonction syntaxique des éléments détachés, etc.).

## A.1 Corpus étudiés et méthodologie d'analyse

### A.1.1 Corpus Air France et Murol

Notre étude a porté sur deux corpus distincts, correspondant à deux situations interactives différentes (tableau A.1) :

Corpus	Nb de dialogues	Nb de tours de parole	Nb de mots	Degré de finalisation	Degré d'interactivité
Air France	103	5 149	49 703	élevé	assez élevée
Murol	9	1 078	13 500	modéré	très élevée

TAB. A.1 – Description des corpus d'étude

- Le corpus **Air France** (AF), recueilli par l'équipe de Marie-Annick Morel à l'Université de la Sorbonne Nouvelle, puis retravaillé par Pierre Nerzic dans le cadre du projet DALI (Sabah, 1994), réunit un ensemble de conversations téléphoniques entre un centre de réservation aérienne et différents clients, qui peuvent être des particuliers ou des personnels d'agence de voyage. Tout en s'inscrivant dans le cadre applicatif du renseignement aérien, la tâche concernée est relativement plus complexe que celle étudiée dans ATIS. Le degré de finalisation de cette application reste cependant élevé. Le degré d'interactivité est celui de la conversation téléphonique. Le dialogue reste cependant contenu, l'hôtesse étant tenue à une certaine réserve. Un des objectifs du projet DALI était d'étudier les reformulations sur des conversations réelles. Le corpus sur lequel nous avons travaillé ne reprend donc que les dialogues initialement recueillis comportant une reformulation. Il est composé de 103 dialogues représentant 5149 tours de parole.
- Le corpus **Murol** du laboratoire CLIPS-IMAG (Bessac et Caelen, 1995) réunit un ensemble de conversations téléphoniques simulées entre deux compères jouant respectivement le rôle d'un touriste et d'un employé d'un syndicat d'initiative (renseignement touristique). La conversation peut porter sur des problèmes de localisation dans la ville concernée, de recherche d'activité sportive ou culturelle et aussi de restauration. Le domaine de la tâche est donc sensiblement moins finalisé que dans le corpus précédent. Pour chaque dialogue simulé, un scénario a été conçu afin de favoriser l'apparition de situations de négociation. Il s'agit donc de dialogues relativement longs présentant une interactivité très marquée (chevauchements très fréquents, par exemple). La partie analysée du corpus comporte quatre dialogues représentant 1078 tours de parole. Ces quatre dialogues longs ont été partagés en 9 sous-dialogues correspondant à des sous-tâches bien spécifiques (actualisation interactive d'un plan de la ville puis renseignement touristique et établissement d'un programme d'activités pour une demi-journée).



Nous avons procédé à un recensement très précis des phénomènes d'extraction sur ces deux corpus. Nous allons tout d'abord présenter la méthodologie mise en œuvre pour cette étude, puis d'en discuter les limites éventuelles.

## A.1.2 Recensement des phénomènes

### A.1.2.1 Fréquences d'occurrence

L'étude que nous avons menée a consisté à mesurer la fréquence d'apparition des différents types d'extractions dans les corpus étudiés. À la différence de l'écrit, il est difficile de segmenter les productions orales en phrases (Blanche-Benveniste *et al.*, 1990). Le choix de l'unité de segmentation sur laquelle seront basées nos observations se pose donc plus difficilement qu'à l'écrit. Dans l'optique d'un traitement automatique, nous avons choisi le tour de parole comme unité de comptabilisation. Par la suite, nous parlerons indifféremment d'énoncé ou de tour de parole.

Suivant les cas de figures, les phénomènes observés ont été détectés semi-automatiquement ou entièrement manuellement. Deux experts ont procédé par validation croisée à la détection et la comptabilisation de ces phénomènes. En règle générale, la présence de marqueurs linguistiques divers permet à l'expert de détecter avec sûreté les phénomènes d'extraction. Dans les cas moins clairs (adverbes, fausses conjonctions), nous nous en sommes remis au Grevisse (Goosse, 1993) qui présente l'avantage de définir une norme issue de l'observation et non d'une approche puriste de la langue. Les différentes formes d'interrogations (par *est-ce que* ou intonatives sans inversion du sujet) n'ont pas été comptabilisées. Il s'agit en effet de procédés réguliers que l'on doit considérer comme faisant partie intégrante de la grammaire du français parlé (Blanche-Benveniste *et al.*, 1990). Notons enfin qu'un énoncé peut présenter des phénomènes d'extraction multiples. Les fréquences d'apparition sont calculées en terme de pourcentage d'énoncés porteurs du phénomène considéré. La dispersion des données est analysée en terme d'écart-type de classes, chaque classe correspondant à un dialogue. Nous avons par ailleurs procédé à une analyse systématique de la validité statistique des résultats obtenus. Lorsque nos observations semblaient clairement significatives, nous avons procédé à une vérification statistique par test de la variance de Fisher-Snedecor et test de la moyenne de Student (Dudewicz et Mishra, 1988). Dans les cas où la validité des résultats était moins tranchée, nous avons procédé à une seconde vérification à l'aide d'un test non paramétrique (test de Wilcoxon-Mann-Whitney), la normalité de la distribution de nos observations n'étant pas garantie.

### A.1.2.2 Phénomènes étudiés

Chaque phénomène observé a été caractérisé suivant plusieurs dimensions. Tout d'abord, nous avons considéré le **sens de l'extraction**. Nous distinguons ainsi les détachements à gauche des détachements à droite (Gadet, 1992 : 74-75). Par commodité d'expression, nous qualifierons par la suite ces deux types d'extraction d'**antéposition** (détachement à gauche),

(E.1.1) (...) **ces deux rues piétonnes** hein e alors p/ e pour vous **y** rendre  
(MU.4.O39)

et de **postposition** (détachement à droite) :

(E.1.2) (...) *parce que c'est en sens unique le boulevard Voltaire* hein  
(MU.4.O55)

Ensuite, nous avons étudié le type de procédé mis en jeu par l'extraction. Quatre types principaux de procédés ont été retenus à la suite de (Gadet, 1992) :

- les **inversions**, qui correspondent à une modification de la position d'un groupe ne se manifestant par aucune autre marque linguistique. Par exemple :

(E.1.3) (...) *sur Héraklion on n'a qu'un seul tarif spécial* (AF.II.17.O14)

(E.1.4) (...) *jusqu'à 21 h 30 euh vous pouvez manger au restaurant*  
(MU.1.O70)

- les **dislocations**, que nous nommerons **doubles-marquages** à la suite de (Blanche-Benveniste, 1997). Dans ce cas, le déplacement est marqué par un clitique de reprise rappelant le groupe ainsi que sa fonction. Dans l'exemple qui suit, l'argument *le visa* est ainsi repris par le clitique *le* :

(E.1.5) (...) *le visa on l'a au consulat* (AF.I.48.C6)

Cette reprise peut également être réalisée à l'aide de l'expression *c'est* ou plus généralement *çà* + <verbe> :

(E.1.6) (...) *donc l'office du tourisme c'est c'est à côté du stade*  
(MU.2.C110)

- les **présentatifs**, dans lesquels un élément initial (*c'est* ou les introductifs construits avec le verbe *avoir* : *il y a* / *j'ai* / *on a* etc. ) introduit explicitement la partie de l'énoncé détachée et est suivi d'une (fausse) subordonnée introduite par *qui* ou *que*. Par exemple :

(E.1.7) (...) *j'ai quelqu'un qu'est allé prendre des billets charters pour moi* (AF.I.43.C9)

(E.1.8) (...) *c'est bien le renseignement que vous vouliez avoir*  
(AF.II.10.O27)

On remarquera que ce type d'extraction met fréquemment en jeu une structure d'énoncé clivée.

- les énoncés **binaires**, où l'élément détaché perd toute dépendance claire avec le reste de l'énoncé, souvent du fait d'une ellipse plus ou moins évidente à détecter. Par exemple :

(E.1.9) (...) *le tarif vacances euh toute modification entraîne des frais*  
(AF.II.26.010)

Pour chaque extraction, nous caractérisons également la fonction de l'élément détaché. Nous avons distingué quatre types de fonctions principales : le **sujet**, qui joue un rôle

particulier en français, les autres **arguments** gouvernés par le verbe, les **modifieurs** (encore appelés adjoints) qui sont dominés par le verbe mais ne peuvent être considérés comme appartenant à la valence de ce dernier, et enfin les **associés** (Blanche-Benveniste, 1997) que l'on peut définir comme des compléments de phrase.

Les exemples d'extractions donnés précédemment concernent les fonctions sujet, argument et modifieur. Nous ne détaillerons ici que le cas des associés, catégorie moins étudiée que les précédentes. Notons que certains associés ainsi détachés ne jouent qu'un rôle très limité dans la phrase. C'est le cas dans l'exemple ci-dessous où, à côté de l'extraction de la locution *en fait*, le pronom *là* réfère vaguement à la situation et est considéré comme **associé explétif** :

(E.1.10) *en fait là c'est pour retrouver un prix d'un voyage déjà effectué*  
(AF.I.4.C3)

Enfin, nous notons pour chaque extraction si celle-ci conduit à un énoncé **projectif**.

### A.1.3 Limites méthodologiques de l'étude : discussion

La pertinence de toute analyse de corpus dépend fortement de la représentativité des données étudiées. Les corpus de français parlé consacrés à des situations d'interaction spontanée étant relativement rares, il ne nous a pas été possible de disposer de ressources linguistiques répondant parfaitement aux objectifs de cette étude. Il est donc essentiel de discuter des limites méthodologiques éventuelles de ce travail avant d'en donner les résultats.

#### A.1.3.1 Taille des corpus

Une première remarque concerne la taille relativement modeste des corpus utilisés (50000 et 13500 mots). Outre la difficulté d'obtenir des corpus de dialogue oral plus conséquents, cette limitation s'explique avant tout par le caractère non automatisable de cette étude. Afin de prévenir tout biais d'analyse consécutif à un éventuel manque de données, nous avons soumis l'ensemble de nos résultats à une validation statistique poussée. Cet effort méthodologique, associé à la diversité des locuteurs du corpus Air France<sup>1</sup> (plus de 200 locuteurs différents), garantit de notre point de vue la pertinence des observations et conclusions réalisées.

#### A.1.3.2 Nature des dialogues

Une autre limite provient de la nature des dialogues étudiés. Dans le cas du corpus Air France, on peut s'interroger sur l'incidence d'une analyse restreinte à une sélection de dialogues comportant une reformulation (cf § A.1.1). Plus précisément, les dialogues retenus dans le projet DALI comportent une reformulation par le client de la requête initiale, suite à une invite du personnel d'accueil. Par exemple (dialogue AF.I.63) :

---

<sup>1</sup>On regrettera de ce point de vue la diversité plus limitée des locuteurs du corpus Murol, qui sont au nombre de huit. Comme nous le verrons, la remarquable cohérence statistique des résultats obtenus sur les deux corpus ne laisse cependant pas de doute sur la représentativité de ce corpus.

- O1 *Air France bonjour*
- C1 *oui bonjour je voudrais un renseignement e je voudrais savoir le prix d'un billet de Paris à Nice*
- O2 ***oui excusez-moi madame la ligne n'est pas très bonne je n'ai pas entendu votre question***
- C2 ***oui je voudrais savoir le prix d'un billet Paris Nice***
- O3 *Paris-Nice oui ce serait pour voyager à quelle période*

L'analyse systématique des reformulations, ainsi que celle des tours de parole les suivant immédiatement, montre que cette tournure n'a pas d'influence significative sur les phénomènes d'extraction. On constate que le locuteur peut adopter deux stratégies énonciatives différentes lors de la reformulation. Soit il conserve le même ordonnancement linéaire entre la requête initiale et la reformulation, comme dans l'exemple précédent. Dans le cas d'une requête avec détachement, il reprendra alors le même type d'extraction (dialogue AF.I.65) :

- O1 *Air France bonjour*
- C1 *alors bonsoir madame **ce matin** j'ai appelé et j'ai réser j'ai fait une réservation pour Barcelone et je voudrais savoir si c'est e si c'est e Orly sud ou nord*
- O2 ***excusez-moi madame je n'ai pas très bien compris votre question***
- C2 ***ce matin** j'ai appelé pour l'Espagne*

Ces reprises d'extraction sont cependant marginales et n'ont pas d'influence signitica-tive sur nos observations quantitatives.

Soit le client adopte un style télégraphique sur plusieurs tours de parole, qu'on ne peut analyser en terme d'éléments binaires (dialogue AF.I.58) :

- O1 *Air France bonjour*
- C1 *bonjour madame s'il vous plait je voudrais savoir vos vols pour le Japon Paris Japon Paris Tokyo quel jour vous l'avez et le plus direct possible*
- O2 ***oui excusez-moi je n'ai pas bien entendu votre question pourriez vous me la reposer s'il vous plait***
- C2 ***les vols***
- O3 *oui*
- C4 ***Paris Tokyo***
- O3 *oui*
- C4 ***les plus courts***

Là encore, l'emploi d'une telle stratégie est trop marginale — de l'ordre d'un cas pour mille énoncés — pour pouvoir influencer sur l'importance quantitative des phénomènes d'extraction.

Dans le cas du corpus Murol, c'est le caractère simulé du dialogue qui peut susciter l'interrogation. Le protocole d'élaboration de ce corpus reposait cependant sur des scénarii très peu contraints qui laissent à penser que ces recommandations ne sont pas de nature à altérer la naturalité des dialogues recueillis. Le problème posé par la simulation ne concerne donc pas la validité des observations effectuées, mais plutôt la représentativité

du corpus vis-à-vis du domaine applicatif étudié (renseignement touristique). Nous reviendrons sur cette question à l'occasion de l'étude différentielle des résultats obtenus sur les deux corpus (§ A.3).

### **A.1.3.3 Dialogue homme-homme vs. dialogue homme-machine**

Enfin, ces corpus correspondent à un dialogue homme-homme et non pas à un dialogue homme-machine. On peut se demander s'il s'agit du modèle le plus pertinent pour l'analyse des usages à destination de la communication homme-machine. De nombreuses études contradictoires ont été menées sur ce sujet. (Allen *et al.*, 1996) montrent par exemple que la technique du Magicien d'Oz ne permet pas toujours une interprétation claire des données. De même, les concepteurs d'interface homme-machine ont montré que les approches par amorçage sont susceptibles d'induire certains biais dans l'analyse des besoins. L'utilisation de dialogues homme-machine pour le prototypage des systèmes — technique d'amorçage (Fraser, 1997) — ne saurait donc être considérée comme optimale. Enfin, l'utilisation de corpus pilotes. En l'absence de solution idéale, l'utilisation parallèle de plusieurs approches semble raisonnable (Cheyer *et al.* 1998, Gustafson et Bell, 2000).

Dans le cadre de cette étude, nous avons choisi de travailler sur des corpus pilotes (Caelen *et al.*, 1997) homme-homme qui permettent de cerner les pratiques des utilisateurs qui devraient être modélisées par le système. L'utilisation de tels corpus est essentielle pour garantir l'utilisabilité des systèmes de dialogue oral (Dybkjaer et Bernsen, 2000). Cette idéalisation du dialogue homme-machine par l'interaction humaine a certes ses limites. Elle nous paraît néanmoins aussi justifiée que l'approche consistant à s'en remettre aux capacités de l'être humain — étudiée par exemple dans (Morel *et al.*, 1989) — à adapter son comportement langagier face à un système informatique. Le succès limité des bornes de réservation SNCF ou de l'annuaire électronique est là pour nous le rappeler.

Ainsi, il nous semble qu'en dépit de leurs spécificités, les corpus Air France et Murol sont suffisamment représentatifs d'un dialogue oral finalisé pour garantir la validité des observations que nous allons maintenant présenter en détail.

## **A.2 Résultats : l'extraction, un phénomène incontournable à l'oral**

Ce paragraphe présente les principales observations que l'on peut tirer de cette étude de corpus. Dans un premier temps, nous allons étudier ces résultats dans toute leur généralité. Nous reviendrons ensuite sur les données rendant compte des variabilités dues au domaine d'application considéré.

### **A.2.1 Importance relative des phénomènes d'extraction**

Si le français est considéré comme une langue à ordre fixe, on aurait cependant tort de considérer les extractions comme marginales dans le cas de dialogues oraux spon-

tanés. Le tableau A.2, qui présente la fréquence d’occurrence moyenne des extractions sur nos deux corpus, montre au contraire que ce phénomène peut-être très répandu. Par exemple, environ un quart des énoncés du corpus Murol comprennent au moins un élément détaché. On constate par ailleurs une différence assez sensible dans les fréquences d’apparitions relevées pour le corpus Air France (AF) et le corpus Murol. Nous reviendrons sur ce point ultérieurement.

Corpus	FRÉQUENCE			
	moyenne	écart-type	minimum (sur un dialogue)	maximum (sur un dialogue)
Air France	13,6 %	10,5 %	0 %	30,8 %
Murol	26, 6 %	10,2 %	8,3 %	40,7%

TAB. A.2 – Fréquence d’apparition des extractions sur chaque corpus (nombre moyen de tours de parole présentant au moins une extraction)

On notera cependant que ces observations présentent une forte diversité d’un dialogue à l’autre. Dispersion des données dont témoignent les variances relevées sur les deux corpus. Cette forte variation d’usage peut provenir du contexte dialogique (insistance dans les situations de négociation ou d’incompréhension, par exemple), mais aussi peut-être également d’un comportement langagier propre à chaque locuteur.

Afin de rechercher une explication à cette variabilité, nous avons réalisé plusieurs études différentielles sur le statut du locuteur. Celles-ci ont été conduites sur le corpus Air France, pour lequel nous disposons d’un nombre de locuteurs important. La première analyse a consisté à distinguer les clients du personnel d’accueil. Ceux-ci ne partageant pas les mêmes buts dans l’interaction, il est en effet plausible qu’on puisse observer des différences dans l’utilisation des extractions. Cette hypothèse est rejetée par l’étude des fréquences respectives d’apparition des extractions (tableau A.3). On ne constate en effet aucune différence significative entre les deux populations (test de Student :  $T = 0,101$  ;  $T_{inv}(0,101) = 0,920$ ).

Corpus Air France moyenne (écart-type)	Clients 13,2 % (10,5 %)	Accueil 13,8 % (9,9 %)
Corpus Air France moyenne (écart-type)	Particuliers 14,9 % (6,9 %)	Agence de voyage 10,1 % (8,2 %)

TAB. A.3 – Fréquence d’apparition des extractions suivant le statut des locuteurs dans le corpus Air France (nombre moyen de tours de parole présentant au moins une extraction)

Nous avons également cherché des différences d’usage entre, d’une part les clients habitués à la tâche que sont les agences de voyage, et d’autre part les clients non experts que sont les particuliers (tableau A.3). Là encore, aucune différence statistiquement significative n’a pu être observée entre les deux populations (test de Student :  $T =$

0,628 ;  $T_{inv}(0,101) = 0,532$ ). Une étude plus approfondie des différents procédés utilisés, ou encore des fonctions syntaxiques mises en jeu, ne permet pas plus de détecter une influence de l’habitude à la tâche (Antoine et Goulian, 2001).

La cohérence de ces résultats ne permet donc pas de trouver une explication satisfaisante à cette variabilité entre dialogues. Aussi retiendrons-nous avant tout que l’extraction est un procédé largement répandu en français oral, et qu’il est à ce titre digne d’intérêt dans une perspective computationnelle.

Une étude plus fine des phénomènes observés montre cependant que cet usage respecte globalement une certaine rigidité dans l’ordre des mots du français.

## A.2.2 Extractions orales et ordre canonique sujet-verbe-objet

Intéressons-nous ainsi aux caractéristiques des extractions observées. Le tableau A.4 présente la répartition de ces observations en fonction du sens du détachement<sup>2</sup>.

Corpus	antéposition	postposition	écart-type
Air France	82,5 %	17,5 %	20,4 %
Murol	85,5 %	14,5 %	8,7 %

TAB. A.4 – Répartition des extractions en fonction du sens du détachement (pourcentage moyen de phénomènes d’un sens donné)

On observe que l’antéposition est largement plus fréquente que la postposition. Comme le note (Gadet, 1992), l’antéposition est moins contrainte et semble plus adaptée à la mise en relief du thème de l’énoncé, d’où son usage plus fréquent. On notera enfin que cette distribution reste remarquablement stable d’un corpus à l’autre (test de Student :  $T = 0,407$  ;  $T_{inv}(0,407) = 0,685$ ).

Corpus		sujet	argument	modifieur	associé
Air France	moyenne	30,7 %	12,0 %	27,4 %	30,0 %
	(écart-type)	(29,6 %)	(15,5 %)	(26,5 %)	(24,5 %)
Murol	moyenne	25,4 %	5,3 %	23,5 %	45,8 %
	(écart-type)	(7,7 %)	(4,1 %)	(16,1 %)	(13,3 %)

TAB. A.5 – Répartition des extractions suivant la fonction de l’élément détaché (pourcentage moyen de phénomènes d’un type donné)

Plus intéressante est l’observation de la répartition des extractions suivant la fonction syntaxique de l’élément déplacé (tableau A.5). On remarque ici une prédominance nette des fonctions sujet, modifieur et associé, par opposition aux arguments<sup>3</sup> qui semblent

<sup>2</sup>L’écart-type des deux variables *antéposition* et *postposition* est bien entendu le même puisque  $P(\text{antéposition}) = 1 - P(\text{postposition})$

<sup>3</sup>La valeur d’écart-type pour la variable arguments supérieure à celle de la moyenne ne doit pas étonner : on observe une distribution non symétrique des données, qui ne suit donc pas une loi normale.

moins prêter à extraction. L'étude statistique de ces répartitions indique que cette moindre extraction des arguments est clairement significative<sup>4</sup>, alors qu'on ne décèle aucune différence significative entre les autres fonctions syntaxiques, sauf en ce qui concerne la prépondérance des extractions d'associés dans le corpus Murol<sup>5</sup>. Nous reviendrons au paragraphe A.3 sur les variations de répartition observées entre les deux corpus. Celles-ci ne remettent de toute manière pas en cause cette présence relativement faible des arguments dans les détachements.

Ces observations semblent cohérentes d'un point de vue linguistique. D'une part, modificateurs et associés sont soumis à moins de contraintes d'ordonnancement, en ce sens que leur déplacement n'altère pas l'ordre canonique SVO que suit le français.

Corpus	antéposition	postposition	écart-type
Air France	80,6 %	19,4 %	20,4 %
Murol	90,6 %	9,4 %	8,7 %

TAB. A.6 – Répartition des extractions d'éléments sujets en fonction du sens du détachement (pourcentage moyen de phénomènes d'un sens donné)

D'autre part, on remarque que la majeure partie des sujets déplacés correspond à une situation d'antéposition (tableau A.6). Cette antéposition n'altère en rien l'ordre SVO, d'autant plus que les extractions du sujet se manifestent presque toujours<sup>6</sup> par une reprise sous la forme d'un double marquage ou d'un présentatif (tableau A.7), comme dans les exemples ci-dessous :

Double-marquage	<i>moi-même je suis Florence</i> (AF.II.11.C09)
Double-marquage	<i>donc euh le le problème c'est de transformer un Bruxelles Marseille en Bruxelles Nice</i> (AF.II.6.C15)
Présentatif	<i>ce n'est pas ces 9 francs de taxe qui ont qui vous ont chiffonné</i> (AF.I.44.O30)

<sup>4</sup>Par exemple sur le corpus Air France, test de Student sur une répartition identique des détachements de sujet et des détachements d'arguments :  $T = 3,652$  ;  $T(0,1) = 2,600$ .

<sup>5</sup>Cette prépondérance est d'ailleurs assez proche de la limite de significativité. D'une manière générale, sur le corpus Air France, test de Student sur une répartition identique des détachements par couples de fonctions syntaxiques :  $T_{\text{suj/mod}} = 0,911$  et  $T_{\text{inv}}(0,911) = 0,363$  ;  $T_{\text{suj/ass}} = 1,059$  et  $T_{\text{inv}}(1,059) = 0,291$ . Même test sur le corpus Murol :  $T_{\text{suj/mod}} = 0,565$  et  $T_{\text{inv}}(0,565) = 0,580$  ;  $T_{\text{suj/ass}} = 1,797$  et  $T_{\text{inv}}(1,797) = 0,091$  (cas proche du seuil de criticité à 10 %). Un test de Wilcoxon-Mann-Withney confirme la significativité de la prédominance des associés :  $Z_{\text{suj/ass}} = 3,135$  et  $Z(0,01) = 2,576$ .

<sup>6</sup>D'une manière générale, nous ne discuterons pas dans ce texte de la validité statistique — évidente — d'observations aussi tranchées.



Corpus	double-marquage		
	+ présentatif	autres procédés	écart-type
Air France	95,4 %	4,6 %	0,8 %
Murol	100%	0 %	-

TAB. A.7 – Répartition des extractions d’éléments sujets en fonction du procédé utilisé (pourcentage moyen de phénomènes d’un sens donné)

À l’opposé, toute extraction d’un argument est susceptible de modifier l’ordre SVO de l’énoncé. Au total, on remarque que la majeure partie des extractions observées dans nos corpus oraux respecte cet ordre canonique (tableau A.8).

Corpus	% d’extractions	% d’énoncés
	avec ordre SVO respecté	avec ordre SVO respecté
Air France	90,3 %	98,7 %
Murol	92,3 %	98,0 %

TAB. A.8 – Part relative des extractions conservant l’ordre canonique SVO (pourcentage par rapport à l’ensemble des phénomènes observés et rapport à l’ensemble des énoncés)

Globalement, rares sont ainsi les énoncés oraux qui se caractérisent par un ordre SVO modifié. La norme de l’écrit est donc sauve : la grande variabilité d’ordonnancement observée en français oral ne saurait s’affranchir aisément de certaines contraintes fondamentales. Un dernier élément renforce ce constat : la comptabilisation des structures non projectives dans nos corpus.

### A.2.3 Extractions orales et projectivité

Comme le montre le tableau A.9, le nombre de discontinuités dues aux extractions est très limité.

Corpus	% d’extractions non-projectives	% d’énoncés discontinus
	moyenne (écart-type)	moyenne (écart-type)
Air France	2,3 % (7,4 %)	0,4 % (0,9 %)
Murol	0,5 % (0,6 %)	0,2 % (0,2 %)

TAB. A.9 – Part relative des extractions non projectives (pourcentage par rapport à l’ensemble des phénomènes observés et par rapport à l’ensemble des énoncés)

Au total, les détachements conduisant à des énoncés non-projectifs représentent moins de 0,5 % des énoncés de nos corpus oraux. Ces résultats restent remarquablement stables d’un corpus à l’autre (test de Student :  $T = 0,261$  ;  $T_{inv}(0,261) = 0,795$ ). Ainsi, nous n’avons pu observer que de rares exemples de relativisation à dépendance non bornée (Kahane, 2000). De même, rares sont les exemples de discontinuité dus à

une extraction de mot question (*wh-question*). Cette situation se comprend aisément lorsqu'on sait que le français oral fait avant tout usage de l'interrogation à intonation, procédé qui conserve l'ordonnancement linéaire de l'énoncé affirmatif, alors que l'interrogation à inversion est très rarement utilisée (Gadet, 1989) :

Inversion     **quand** pensez-vous pouvoir venir nous rencontrer  
 Intonation    vous pensez pouvoir venir nous rencontrer **quand**  
 Est-ce que    **quand** est-ce que vous pensez pouvoir venir nous rencontrer

D'un point de vue quantitatif, les discontinuités dues à l'interrogation sont donc à attendre surtout avec des structures en *est-ce que*, ainsi qu'avec certaines interrogatives indirectes.

Au final, on peut tenir les extractions non-projectives pour marginales sur ce genre de français parlé. Il existe d'autres sources de discontinuités à l'oral, parmi lesquelles les incisives et certaines formes de reprises. Nous entamons à ce sujet une nouvelle étude de corpus afin de quantifier l'importance de ces phénomènes en dialogue oral. On peut cependant se demander si la modélisation de ces phénomènes relève de traitements spécifiques ou rentre encore dans la problématique des grammaires non-projectives.

#### A.2.4 Fonctions et procédés

Pour terminer cette analyse globale, nous avons étudié la répartition des différents procédés utilisés pour marquer le détachement. Le tableau A.10 donne une synthèse de ces observations. On observe tout d'abord que ces résultats sont relativement stables d'un corpus à l'autre.

Corpus	Inversion	double-marquage	présentatif	éléments binaires
A. F. moyenne	60,6 %	24,9 %	13,2 %	1,3 %
(écart-type)	(30,2 %)	(25,5 %)	(22,3 %)	(8,6 %)
Murol moyenne	67,8 %	16,8 %	14,4 %	1,0 %
(écart-type)	(11,8 %)	(9,2 %)	(6,9 %)	(0,7 %)

TAB. A.10 – Répartition des extractions suivant le procédé utilisé (pourcentage moyen de procédés d'un type donné)

Comme pour la répartition par fonction syntaxique, on relève par contre une forte dispersion des résultats au sein d'un même corpus. Il n'en reste pas moins que l'inversion constitue le procédé majoritairement utilisé. Cette prédominance est statistiquement significative (test du Student sur une distribution identique des inversions et des doubles marquages :  $T_{AF} = 4,473$  ;  $T_{Murol} = 4,118$  ;  $T(0,01) = 2,600$ ). À l'opposé, on ne distingue pas de différence statistiquement significative entre doubles marquages et présentatifs, procédés (test de Student :  $T_{AF} = 1,366$ ,  $T_{inv}(1,366) = 0,174$  ;  $T_{Murol} = 0,031$  ;  $T_{inv}(0,031) = 0,975$ ). On constate enfin que les éléments binaires restent très marginaux. De part leur structure éclatée faisant appel à la résolution d'ellipses, les énoncés binaires correspondent au procédé le plus difficile à appréhender d'un point de

vue cognitif : une fois encore, un certain nombre de contraintes semble être à même de limiter la variabilité d’ordonnancement du français parlé.

Si l’inversion semble prédominante dans nos corpus, il est intéressant de procéder à une analyse plus discriminante étudiant l’usage de chaque procédé pour le déplacement de chaque classe de fonction syntaxique. Nous avons déjà vu (tableau A.7) que la fonction *sujet* donne quasiment toujours lieu à une extraction marquée (double-marquage et présentatif). Ce constat se retrouve de façon plus modérée avec la fonction *argument*. Dans ce cas, double-marquage, présentatif et inversion se partagent à peu près équitablement les cas d’utilisation (test du  $\chi^2$  sur une répartition équiprobable des 3 procédés :  $CHI_{murol} = 0,782$  ;  $CHI_{AirFrance} = 1,000$ ). Il n’en reste pas moins que les procédés marqués sont globalement majoritaires (tableau A.11), comme le montre un test de Student sur une distribution identique des procédés marqués et non marqués :  $T_{murol} = 3,139$  et  $T(0,01) = 2,921$  ;  $T_{AirFrance} = 2,552$ ,  $T(0,05) = 1,972$  et  $T(0,01) = 2,600$ ).

Corpus	double-marquage	inversion et	Ecart-type
	+ présentatif	énoncés binaires	
Air France	67,3 %	32,7 %	30,2 %
Murol	77,3 %	22,7 %	11,9 %

TAB. A.11 – Répartition des extractions d’éléments arguments en fonction du procédé utilisé (pourcentage moyen de phénomènes d’un sens donné)

Cet usage majoritaire de procédés d’extraction marqués pour ces deux fonctions peut s’expliquer là encore par la contrainte qu’impose l’ordre canonique SVO : la reprise par un clitique ou la mise en œuvre d’une structure à présentatif clairement identifiable a pour effet de corriger en partie ce changement non normatif d’ordonnancement.

À l’opposé, la position des modificateurs et des associés est beaucoup plus libre. Aussi n’utilise-t-on alors que très rarement, voire jamais, un dispositif marqué pour leur extraction (tableau A.12).

Enfin, on notera que la réalisation des présentatifs fait appel indifféremment à des introductifs du type *il y a* ou aux structures clivées *c’est... qui* (tableau A.13 a). Les différences de fréquences moyennes observées ne sont en effet pas statistiquement significatives (test du  $\chi^2$  sur une répartition équiprobable des 3 procédés :  $CHI_{murol} = 0,738$  ;  $CHI_{AirFrance} = 0,957$ ).

Corpus	modificateurs		associés	
	inversion	autres procédés	inversion	autres procédés
Air France	96,8 %	3,2 %	100 %	0 %
Murol	92,9 %	7,1 %	100 %	0 %

TAB. A.12 – Répartition des extractions d’éléments modificateurs et associés en fonction du procédé utilisé (pourcentage moyen de phénomènes d’un sens donné)

Il en va de même avec les doubles-marquages (tableau A.13 *b*) qui semblent faire indifféremment appel à une reprise par pronom ou à une reprise en *ça* (test du  $\chi^2$  sur une répartition équiprobable des 3 procédés :  $CHI_{murol} = 0,828$  ;  $CHI_{AirFrance} = 0,992$ ).

En guise de conclusion partielle, on peut conclure que le genre de français parlé étudié sur les corpus Murol et Air France se caractérise par une fréquence élevée de dispositifs à extraction, mais qu'en retour ceux-ci restent contraints par l'ordre canonique SVO standard. Ceci explique :

- d'une part que les extractions concernent avant tout des éléments libres de l'énoncé (modificateurs et associés) ou ne modifiant pas l'ordre SVO (sujets antéposés)
- d'autre part que les procédés mis en œuvre se caractérisent par une relative simplicité structurelle (extractions non-projectives rares).

Corpus	présentatif		Double-marquage		
	avoir	c'est	pronom	ça	autres
Air France	39,3 %	60,7 %	38,0 %	60,3 %	1,7 %
Murol	43,3 %	56,7 %	40,0 %	58,6 %	1,4 %

TAB. A.13 – Répartition des doubles-marquages et présentatifs

L'étude de l'influence du domaine de la tâche sur les extractions en situation de dialogue oral va précisément renforcer ce constat.

### A.3 Extractions et domaine d'application de la CHM orale

Un des objectifs assignés à cette analyse de corpus était l'étude de l'influence du contexte d'interaction sur les phénomènes d'extraction. Les résultats obtenus montrent qu'a priori, cette influence se manifeste plus en termes quantitatifs que qualitatifs. En effet, une remarquable stabilité (voir § A.2) peut être observée quant à la répartition des procédés relevant de l'extraction. Plusieurs différences significatives ont pu néanmoins être relevées entre les deux corpus.

#### A.3.1 Interactivité et extractions orales

Le corpus Murol se caractérise tout d'abord par une fréquence d'apparition moyenne des phénomènes d'extraction plus élevée (voir tableau A.1 : 26,6 % contre 13,6 % pour le corpus Air France). Un test de Wilcoxon-Mann-Whitney montre que cette différence est statistiquement significative ( $Z = 3,548$  ;  $Z(0,01) = 2,576$ ).

Cette variabilité très sensible d'un corpus à l'autre montre ainsi que les phénomènes d'extraction peuvent atteindre une importance prépondérante en français parlé. Comme nous l'avons déjà noté (§ A.1.3), il serait risqué d'identifier cette variabilité à une influence du domaine d'application, certaines réserves pouvant être émises quant à la représentativité du corpus Murol vis-à-vis du domaine du renseignement touristique.

Il semble de même difficile d'identifier cette variabilité au degré de finalisation de la tâche concernée. Comme nous l'avons noté (cf. § A.1.1), les deux corpus étudiés se distinguent en effet par le degré de finalisation de leur tâche, mais aussi par le niveau d'interactivité du dialogue observé<sup>7</sup>. Les sources de variabilités éventuelles sont donc multiples, et une analyse différentielle du corpus Murol semble plutôt privilégier une interprétation en faveur du degré d'interactivité du dialogue. Les dialogues simulés du corpus Murol suivaient en effet un scénario pré-établi qui définissait de manière relativement lâche deux tâches successives pour les interlocuteurs : tout d'abord l'actualisation d'un plan de la ville (tâche très finalisée) et ensuite le renseignement touristique proprement dit (finalisation modérée). Chaque dialogue a été séparé en sous-dialogues correspondant à ces deux tâches.

Le tableau A.14 donne les fréquences d'apparition des phénomènes d'extraction observés sur les deux sous-corpus ainsi obtenus. Un test de Wilcoxon-Mann-Whitney montre que la légère différence obtenue n'est pas statistiquement significative ( $Z = 1,225$ ;  $Z(0,1) = 1,645$ ). On ne peut ainsi trouver de différence d'usages significative dans un cas où le degré de finalisation varie mais pas le niveau d'interactivité (ni même d'ailleurs les locuteurs!). Le degré d'interactivité semble donc être un facteur plus important de variabilité, ce qui constitue d'ailleurs une conclusion relativement intuitive. Des études complémentaires portant sur deux nouveaux corpus devraient nous permettre de mieux expliquer cette variabilité à l'aide d'une analyse factorielle de données.

Sous-corpus Murol	fréquence moyenne	écart-type	maximum sur un dialogue
Actualisation de plan	20,5 %	11,1 %	35,0 %
Renseignement touristique	28,5 %	10,1 %	40,7 %

TAB. A.14 – Fréquence d'apparition des extractions sur chaque sous-corpus Murol (nombre moyen de tours de parole présentant au moins une extraction)

Rappelons enfin que ces observations se traduisent également par une forte dispersion des données. Si des tendances lourdes peuvent être observées sur de grands corpus de tailles significatives, une forte variabilité inter-individuelle ou inter-dialogique reste donc présente. Ce constat milite également en faveur d'une prise en considération accrue des phénomènes d'extraction : dans un dialogue au moins, près de 4 énoncés sur 10 présentent un phénomène d'extraction (tableau A.14).

### A.3.2 Variabilité d'occurrence et fonctions concernées

Une différence globale d'usage des extractions ayant ainsi été observée, il est intéressant d'étudier comment cette variation se répartit suivant la fonction des éléments

---

<sup>7</sup>Ce degré d'interactivité ne semble pas lié à la tâche, mais plutôt à la situation d'enregistrement : relation client relativement formelle dans le cadre du corpus Air France, à comparer à l'interaction plus libre du corpus Murol (simulation de dialogue entre compères, suivant une méthodologie proche de celle du magicien d'Oz).

détachés. Le tableau A.5 étudié précédemment présente la répartition des extractions par fonction respectivement sur les deux corpus Murol et Air France. Si on peut observer de légères variations entre ces deux distributions, celles-ci ne sont pas statistiquement significatives dans leur globalité<sup>8</sup>. On relève toutefois que l'extraction plus fréquente — en termes de répartition relative des procédés — d'associés dans le corpus Murol est statistiquement significative<sup>9</sup> (cf. § A.2.2).

Ainsi, l'augmentation de la fréquence d'occurrence des extractions ne modifie pas en profondeur la nature des éléments déplacés : chaque type d'élément est concerné par cet accroissement et les arguments restent toujours peu fréquents comparativement aux autres procédés. On constate tout de même que les associés participent fortement à cet usage accru des procédés d'extraction : il s'agit là encore d'un résultat assez intuitif. D'une part, les associés sont a priori les éléments les plus libres de l'énoncé. D'autre part, ils remplissent souvent un rôle de marqueurs pragmatiques particulièrement sensibles à la mise en relief. De ce point de vue, l'étude de leurs détachements est intéressant dans l'optique d'une modélisation améliorée du dialogue.

### A.3.3 Variabilité d'occurrence et procédés

Nous avons également étudié les conséquences des variabilités observées entre les deux corpus du point de vue de la répartition des procédés utilisés. Il apparaît là encore qu'une différence de fréquence d'usage n'a pas d'influence sensible sur la nature des procédés utilisés (tableau A.10). Tout au plus observe-t-on une variation d'usage légèrement significative des procédés d'inversion (test de Wilcoxon-Mann-Whitney :  $Z = 1,897$  ;  $Z(0,1) = 1,645$  ;  $Z(0,05) = 2,326$ ). Cette variation peut être directement imputée au détachement plus important d'associés, ces éléments ne donnant lieu qu'à extraction par inversion.

D'une manière générale, ce sont donc toujours les mêmes procédés<sup>10</sup> qui sont utilisés pour mettre en œuvre l'extraction d'éléments d'un type donné.

En résumé, cette étude différentielle montre que si les phénomènes d'extraction peuvent présenter une variation sensible d'utilisation d'une situation d'interaction orale à une autre, leur réalisation repose sur des moyens remarquablement homogènes. Une fois encore, les contraintes normatives qui jouent sur la mise en œuvre des extractions semblent garantir une certaine stabilité de production. Cette observation a bien entendu toute son importance du point de vue de la généralisation des méthodes d'ingénierie linguistique utilisées en CHM orale.

**Remerciements** Nous tenons à remercier Agnès Hamon et Valérie Monbet, du laboratoire de Statistique Appliquée de l'université de BREtagne Sud (SABRES, Vannes)

---

<sup>8</sup>Outre les arguments étudiés ci-après, seuls les éléments sujets s'approchent d'un seuil de criticité statistiquement acceptable : test de Wilcoxon-Mann-Whitney :  $Z = 1,526$  ;  $Z(0,1) = 1,645$ .

<sup>9</sup>Par ailleurs, test de Wilcoxon-Mann-Whitney sur une distribution identique des associés entre les corpus Murol et Air France :  $Z = 3,075$  ;  $Z(0,01) = 2,576$ .

<sup>10</sup>De même, nous n'avons trouvé aucune différence significative de réalisation des présentatifs (structures avec *avoir* / *clivages*) et des doubles-marquages (reprise par pronom /  $\zeta a$ ).

pour la vérification des tests statistiques présentés dans cet article. Nous remercions également le laboratoire CLIPS-IMAG (Grenoble) et son directeur Jean Caelen pour la mise à disposition du corpus MUROL.





## Annexe B

Segmentation des énoncés oraux –  
liste complète des étiquettes et des  
règles utilisées

Nous présentons dans cette annexe un exemple du lexique et les quelques règles de désambiguïsation appliquées lors du processus d'étiquetage grammatical des énoncés oraux. Nous donnons également la liste complète des groupes élémentaires que nous cherchons à caractériser. Pour chacun des groupes élémentaires retenus, nous indiquons les expressions régulières permettant de les identifier.

## B.1 Exemples du lexique

Notre lexique est organisé de façon à regrouper ensemble les formes (mots) qui correspondent à chaque étiquette grammaticale retenue. Au sein d'un même ensemble, les mots sont listés selon leur forme lemmatisée et leurs formes fléchies. Chacun des ensembles est compilé en un transducteur dont l'alphabet est composé d'un seul caractère (lettres, chiffres, espace) permettant d'associer à chaque forme présente l'étiquette correspondante à l'ensemble considéré (la fonction identité est appliquée sur les mots de l'énoncé). Une forme ambiguë est présente dans plusieurs ensembles distincts. L'accès au lexique est réalisé par l'application du transducteur déterministe encodant l'ensemble du lexique (union des différents ensembles). Les formes ambiguës se voient ainsi attribuer l'ensemble des étiquettes qui lui sont associées. Nous donnons ci-dessous un court exemple du lexique concernant les verbes conjugués avant compilation.

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%          définition du lexique -- VERBES CONJUGUES
%%
%%  noms des macros utilisées :
%%    - racine
%%    - terminaisons (des verbes conjugués)
%%      forme d'un verbe conjugué : [racine-n°x,terminaisons-n°x]
%%    - terminaisons-lemmes
%%      forme de l'infinitif : [racine-n°x,terminaisons-lemmes-n°x]
%%
%% words('mot') : macro dénotant l'expression régulière [m,o,t]
%%
%% La nomenclature utilisée (numérotation) suit celle du Petit Robert
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%% -----
%% VERBES du 1er GROUPE
%% -----

macro(terminaisons-lemmes1,
[e,r]).

...

```

```

macro(terminaisons1a,
{
[a,s^],
[a,i,{[e,n,t],s,t}^],
[a,s,s,e,n,t],
[e,{[n,t],s}^],
[e,r,{[a,s^],[a,i],[a,i,e,n,t],[a,i,{s,t}],[i^,e,z],[i^,o,n,s],[o,n,t]}],
[i^,{[e,z],[o,n,s]}],
[â,{t,[m,e,s]}],
[è,r,e,n,t]
}
).

```

%%-----

```

macro(terminaisons3a,
{
[e,{[n,t],s}^],
[e,r,{[a,s^],[a,i],[a,i,e,n,t],[a,i,{s,t}],[i^,e,z],[i^,o,n,s],[o,n,t]}],
[i^,e,z],
[i,o,n,s],
[è,r,e,n,t]
}
).

```

```

macro(racinev3a,
{
words(['acquiesc']),
words(['agac']),
words(['agenc']),
words(['amorc']),
words(['annonc']),
words(['avanc']),
words(['balanc']),
...
}).

```

%%-----

```

macro(terminaisons3b,
{
[e,a,s^],
[e,a,i,{[e,n,t],s,t}^],
[e,a,s,s,e,n,t],

```

```
[e,r,{[a,s^],[a,i],[a,i,e,n,t],[a,i,{s,t}],[i^,e,z],[i^,o,n,s],[o,n,t]}],  
[e,â,{t,[m,e,s]}],  
[è,r,e,n,t]  
}  
).
```

```
macro(racinev3b,  
{  
words(['abrég']),  
words(['afflig']),  
words(['allong']),  
words(['aménag']),  
words(['arrang']),  
words(['asperg']),  
words(['boug']),  
...  
}).
```

```
%% -----
```

```
macro(racinev1a,  
{  
words(['abandonn']),  
words(['abhorr']),  
words(['abîm']),  
...  
words(['gagn']),  
words(['gard']),  
words(['gar']),  
...  
words(['visit']),  
words(['zon']),  
words(['zyeut'])  
}  
).
```

```
macro(liste_v1ergrp,  
determinize(  
{  
[racinev1a,terminaisons1a],  
[racinev3a,terminaisons3a],  
[racinev3abis,terminaisons3abis],  
[racinev3b,terminaisons3b],  
[racinev4a,terminaisons4a],
```

```

[racinev4abis,terminaisons4abis],
[racinev5,terminaisons5],
[racinev5bis,terminaisons5bis],
[racinev6,terminaisons6],
[racinev3b6bis,terminaisons3b6bis],
[racinev8,terminaisons8],
[racinev8bis,terminaisons8bis],
[racinev8b,terminaisons8],
vb-aller
}
)).

```

```

%% -----
%% VERBES du 2ème et du 3ème GROUPE
%% -----

```

```

macro(terminaisons-lemmes1,
[i,r]).

```

...

```

macro(terminaisons2,
{
[i,{s,t}],
[i,s,s,{[e,n,t],[i^,o,n,s],[i^,e,z],e}],
[i,s,s,a,i,{[e,n,t],s,t}^],
[î,{[m,e,s],[t,e,s]}],
[i,r,e,n,t],
[i,r,{[a,s^],[a,i],[a,i,e,n,t],[a,i,{s,t}],[i^,e,z],[i^,o,n,s],[o,n,t]}],
[i^,{[e,z],[o,n,s]}]
}
).

```

```

macro(racinev2,
{
words(['about']),
words(['abrut']),
words(['accompl']),
words(['affaibl']),
...
words(['redéfin']),
words(['réfléch']),
words(['répart']),
words(['rétabl']),

```

```
words(['réuss']),
words(['sais'])
}).
```

```
% -----
```

```
...
```

```
% -----
```

```
macro(liste_v2et3emegrp,
determinize(
{
[racinev2,terminaisons2],
[racinev11,terminaisons11],
[racinev12,terminaisons12],
[racinev14,terminaisons14],
[racinev16,terminaisons16],
[racinev16bis,terminaisons16bis],
[racinev17,terminaisons17],
[racinev18,terminaisons18],
[racinev21,terminaisons21],
...
[racinev56,terminaisons56],
[racinev57,terminaisons57],
[racinev58,terminaisons58],
[racinev60,terminaisons60]
}
)).
```

```
%% -----
```

```
macro(liste_v_conjugue,
determinize(
{
liste_auxilliaire,
liste_v1ergrp,
liste_v2et3emegrp
}
)).
```

## B.2 Règles de désambiguïsation utilisées

Afin de privilégier un taux de précision élevé de l'étiquetage des énoncés oraux<sup>1</sup>, seules quelques règles de désambiguïsation sont appliquées. Ces règles permettent de lever l'ambiguïté sur quelques unes des formes fréquentes à l'oral (la forme *quoi* par exemple) et les ambiguïtés permettant de simplifier la segmentation (ambiguïté entre verbe et modal par exemple). Les ambiguïtés résiduelles sont levées par les niveaux de traitement supérieurs de l'analyse (segmentation, rattachement sémantico-pragmatique).

1. ambiguïté de la forme *puis* entre modal conjugué ou forme considérée comme conjonction de coordination :
  - > décision de l'étiquette **modal** si *puis* est précédé d'un pronom interrogatif (*quand, combien, comment, etc.* ou relatif (*où, lequel, ...*)).
  - > aucune décision dans le cas contraire.
2. ambiguïté de la forme *quoi* entre pronom interrogatif et son utilisation comme élément phatique :
  - > décision de l'étiquette **pronom** si *quoi* est précédé de l'auxiliaire être conjugué (*c'est quoi les tarifs*).
  - > aucune décision dans le cas contraire.
3. ambiguïté des modaux (ou assimilés) entre leurs rôles de modal et de verbe principal. (*je veux une chambre* vs. *je veux réserver une chambre*)
  - a) -> décision de l'étiquette **modal** si la forme ambiguë est suivie de la succession d'étiquettes suivante exprimée au moyen de l'expression régulière<sup>2</sup> (présence éventuelle de clitiques entre le modal et le verbe à l'infinitif) :  
[ **pronom personnel**<sup>^</sup>, **pronom réfléchi**<sup>^</sup>, **verbe infinitif** ].
  - b) -> affectation de l'étiquette **verbe conjugué** dans le cas contraire
4. ambiguïté entre adjectif et son utilisation comme élément phatique (exemple : la forme *bon*)
  - > décision de l'étiquette **adjectif** si la forme ambiguë est suivie d'un nom commun.
  - > aucune décision dans le cas contraire.
5. ambiguïté entre adjectif interrogatif et pronom interrogatif (exemple : la forme *quel, quel moyen de locomotion* vs *quel est le trajet*)

---

<sup>1</sup>Le contexte des règles appliquées peut être perturbé par la présence de répétitions ou de reprises.

<sup>2</sup>Le symbole  $A^{\wedge}$  dénote la présence facultative d'une occurrence de l'étiquette  $A$ .

-> décision de l'étiquette **adjectif interrogatif** si la forme ambiguë est suivie d'un nom commun.

-> aucune décision dans le cas contraire.

6. ambiguïté des formes *le, la, les* entre pronoms personnels et articles définis (singulier ou pluriel)
  - a) -> décision de l'étiquette **pronom** si les formes ambiguës sont précédées d'un pronom personnel de la liste : *je, j, tu, il*, etc. ou de la forme de négation *ne*
  - b) -> décision de l'étiquette **pronom** si les formes ambiguës sont suivies d'un verbe (verbes à l'infinitif, verbes conjugués, auxiliaires, modaux, participes présents, etc.)
  - c) -> affectation de l'étiquette **article défini** (singulier ou pluriel) dans le cas contraire
7. ambiguïté entre pronoms interrogatifs et pronoms relatifs (exemple : la forme *où*)
  - a) -> décision de l'étiquette **pronom relatif** si la forme ambiguë est précédée d'un nom.
  - b) -> décision de l'étiquette **pronom interrogatif** si la forme ambiguë termine l'énoncé. (*il se trouve où*)

-> aucune décision dans le cas contraire.
8. ambiguïté de la forme *si* entre adverbe et conjonction de subordination
  - a) -> décision de l'étiquette **adverbe** si la forme ambiguë est suivie d'un adverbe suivi ou d'un adjectif.
  - b) -> affectation de l'étiquette **conjonction de subordination** dans le cas contraire.
9. ambiguïté de la forme *s'* entre pronom réfléchi ou conjonction de subordination
  - a) -> décision de l'étiquette **pronom réfléchi** si la forme ambiguë est suivie du pronom *en* ou d'un verbe.
  - b) -> affectation de l'étiquette **conjonction de subordination** dans le cas contraire.
10. ambiguïté de la forme *en* entre pronom personnel et préposition
  - a) -> décision de l'étiquette **pronom personnel** si la forme ambiguë est suivie d'un verbe (verbes à l'infinitif, verbes conjugués, auxiliaires, modaux, etc.) ou si elle est précédée par un pronom réfléchi.
  - b) -> affectation de l'étiquette **préposition** dans le cas contraire.
11. ambiguïté entre adverbe et préposition (exemple : structures elliptiques *je réserve aussi le repas avec*)

-> décision de l'étiquette **adverbe** si la forme ambiguë termine l'énoncé.

-> aucune décision dans le cas contraire.
12. ambiguïté des formes *nous* et *vous* entre pronoms réfléchis et pronoms personnels
  - a) -> si l'ambiguïté est présente sur deux formes successives suivies d'un verbe conjugué, affectation de l'étiquette **pronom personnel** pour la première, de l'étiquette **pronom réfléchi** pour la seconde.
  - b) -> décision de l'étiquette **pronom réfléchi** si la forme ambiguë est précédée d'un pronom personnel et suivie d'un verbe conjugué.



c) -> décision de l'étiquette **pronom personnel** si la forme ambiguë est suivie d'un pronom réfléchi.

-> aucune décision dans les cas contraires

**Ordre d'application des règles.** Les contextes droit et gauche utilisés sont constitués de formes non ambiguës. L'ordre d'application des différentes règles est ainsi important. La désambiguïsation résulte de l'application d'une cascade ordonnée de transducteurs. Chaque transducteur encode une ou plusieurs des règles ci-dessus<sup>3</sup>. L'ordre des règles appliquées est indiqué ci-après :

- **premier transducteur** : règles 1, 2, 4, 5, 6a, 8a et 9a
- **second transducteur** : règles 6b, 7a, 8b, 9b, 10a et 12a
- **troisième transducteur** : règles 3a, 6c, 10b et 12b
- **quatrième transducteur** : règles 3b, 7b, 10c, 11 et 12c

### B.3 Liste complète et définitions des groupes élémentaires retenus

Les tableaux suivants répertorient les 32 groupes élémentaires retenus ainsi que leurs définitions données sous la forme d'expressions régulières<sup>4</sup>. Les expressions régulières sont fondées sur les étiquettes grammaticales répertoriées dans le tableau 5.2 (cf. chapitre 5 § 5.3.2.2 page 90)<sup>5</sup>. Les transducteurs réalisant la segmentation ont été compilés à partir de ces expressions. L'ordre d'application des transducteurs est donné au chapitre 5 figure 5.3 § 5.3.3.

Le tableau B.3 concerne les marqueurs des répétitions et des corrections, le tableau B.2 les groupes élémentaires grammaticaux retenus (cf. le tableau 5.1 chapitre 5 § 5.3.1 pour des exemples correspondant à ces segments) et le tableau B.3 les groupes correspondant aux expressions langagières indépendantes du domaine de l'application que nous avons retenues<sup>6</sup>.

---

<sup>3</sup>Pour une règle donnée, les décisions distinctes sont prises en cascades (une décision par transducteur appliqué).

<sup>4</sup>La syntaxe des expressions utilisées est la suivante :  $[A,B]$  dénote la concaténation des langages  $A$  et  $B$  et  $\{A,B\}$  l'union de  $A$  et  $B$ .  $A^\wedge$  indique que l'élément  $A$  est optionnel (un élément ou 0),  $A^+$  indique 1 ou plusieurs occurrences de  $A$  et  $A^*$  indique 0 ou plusieurs éléments  $A$ .  $@A$  regroupe l'ensemble de nos étiquettes jouant le rôle de  $A$ ;  $@DET$  représente par exemple l'ensemble des étiquettes jouant le rôle de déterminant : articles définis et indéfinis (singulier ou pluriel), adjectifs possessifs, adjectifs démonstratifs.  $mot/A$  indique un mot particulier étiqueté  $A$

<sup>5</sup>Dans un souci de lisibilité, nous transcrivons complètement les étiquettes grammaticales, codées par un ou deux caractères dans chaque transducteur.

<sup>6</sup>Pour ces derniers groupes élémentaires, nous ne donnons qu'une partie des expressions régulières utilisées.

Catégorie du groupe	Définitions – Expressions régulières
HES	[@INTERJ-HES*] succession de plusieurs interjections, appuis du discours ou formes utilisées comme éléments phatiques : <i>ben, euh, quoi</i> , etc.
COR	{ [ @INTERJ-HES* non/neg <sup>+</sup> @INTERJ-HES* ], [ @INTERJ-HES* non/neg* @ADV-COR @INTERJ-HES* ] } avec @ADV-COR : <i>pardon, enfin, en-fait</i> , etc.

TAB. B.1 – Groupes élémentaires identifiant les marqueurs des répétitions et des corrections

Catégorie du groupe	Définitions – Expressions régulières
Groupes adverbiaux GADV <sub>q</sub>  GADV <sub>qs</sub> GADV	{ @ADV-Q*, @ADV-QS } avec @ADV-Q : <i>très, trop, assez, aussi, trop-de, etc.</i> et @ADV-QS : <i>plus, moins</i> @ADV-QS { [ le/art <sub>d</sub> S Gadv <sub>qs</sub> adv ], [ pas/neg <sup>^</sup> , { Gadv <sub>q</sub> , Gadv <sub>qs</sub> }, adv ] }
Groupes adjectivaux GADJ <sub>num</sub> GADJ	[ adj <sub>num</sub> ] { [ pas/neg <sup>^</sup> , { Gadv <sub>q</sub> , Gadv <sub>qs</sub> } <sup>^</sup> , adj ], [ art <sub>d</sub> P, art <sub>d</sub> S, art <sub>ind</sub> P, adj <sub>num</sub> <sup>^</sup> , Gadv <sub>qs</sub> , adj ] }
Groupes nominaux  GN   GN <sub>NPROPRE</sub>	{ [ @DET <sup>^</sup> , Gadj <sup>^</sup> <sub>num</sub> , Gadj*, { nom, @PRONOM } ] [ @GADJ-tout <sup>^</sup> , @DET <sup>^</sup> , Gadj*, nom ] @DEICT, pr <sub>pers</sub> S, pr <sub>pers</sub> P, pr <sub>dem</sub> S, pr <sub>dem</sub> P } avec @PRONOM formes pronominales des numéraux et de certains adjectifs ( <i>premier, dernier, etc.</i> ) @GADJ-tout : groupes adjectivaux composés à partir des formes de l'adjectif <i>tout</i> @DEICT ensemble des déictiques (adverbes <i>ici, là, etc.</i> ) [ nomp ]
Groupes verbaux GV <sub>INF</sub> GV <sub>PPR</sub> GV <sub>PP</sub> GV <sub>PPMOD</sub> GV <sub>PPRMOD</sub> GV <sub>INT</sub> GV <sub>MOD</sub> GV	[ pr <sup>^</sup> <sub>ref</sub> , v <sub>inf</sub> ] { [ pr <sup>^</sup> <sub>ref</sub> , v <sub>ppr</sub> ], [ aux <sub>ppr</sub> , aux <sup>^</sup> <sub>pp</sub> , v <sub>pp</sub> ] } [ pr <sup>^</sup> <sub>ref</sub> , aux, aux <sup>^</sup> <sub>pp</sub> , v <sub>pp</sub> ] [ aux, aux <sup>^</sup> <sub>pp</sub> , mod <sub>pp</sub> , v <sub>inf</sub> ] [ { [ aux <sub>ppr</sub> , aux <sup>^</sup> <sub>pp</sub> ], pr <sup>^</sup> <sub>ref</sub> , }, mod <sub>ppr</sub> , v <sub>inf</sub> ] [ pour/prep <sup>^</sup> , pr <sub>int</sub> , pr <sup>^</sup> <sub>ref</sub> , v ] [ mod, mod <sub>inf</sub> , v <sub>inf</sub> ] { [ pr <sup>^</sup> <sub>ref</sub> , v ], aux }
Groupes prépositionnels  GP	{ [ prep, { GN, GV <sub>inf</sub> , Gadj } ], [ {pour, sans}/prep, y/adv <sup>^</sup> , GV <sub>inf</sub> ] , [ {par, pour}/prep, @DEICT ] }
pronom interrogatif adjectif interrogatif pronom réfléchi pronom relatif conjonction de subordination négation conjonction de coordination	mots formant un segment à eux seuls pr <sub>int</sub> adj <sub>int</sub> pr <sub>ref</sub> pr <sub>rel</sub>  conj <sub>sub</sub> neg  conj <sub>coo</sub>

TAB. B.2 – Définitions des groupes grammaticaux élémentaires retenus

Catégorie du groupe	Définitions – Expressions régulières
DATE	<pre>{ @JOUR<sub>adv</sub>,   [ @NomJour, prochain/adj ],   [ (le/art<sub>defS</sub>)<sup>^</sup>, { @NumMois, [ (sur)<sup>^</sup>lendemain/nom ], }   [ (le/art<sub>defS</sub>)<sup>^</sup>, @NomJour, @Num<sup>^</sup>Mois ],   [ [ (le/art<sub>defS</sub>)<sup>^</sup>, @NomJour<sup>^</sup>, @NumMois ]<sup>^</sup>, @NomMois ]   [ {mi, fin, début}/nom , @NomMois ] }</pre>
HEURE	<pre>[ @NumHeure, heure(s<sup>^</sup>)/nom,   { [ et/conjcoo, {quart, demi(e<sup>^</sup>) }/nom ],     [ moins/adv<sub>qs</sub>, { [ le/art<sub>dS</sub>, quart/nom ],                       @NumMinute- } ] ],   @NumMinute }<sup>^</sup>,   { [ du/prep, { matin, soir }/nom ],     [ de/prep, l/art<sub>dS</sub>, après-midi/nom ] }<sup>^</sup>] avec @NumHeure : 0, 1, .. , 24, midi, minuit @NumMinute- : 5, 10, .. , 25 @NumMinute : 0, 1, .. , 59</pre>
ADRESSE	<pre>{ [ adj<sub>num</sub><sup>^</sup>, @NomRue, { de, des, du }/prep<sup>^</sup>, @DET<sup>^</sup>,   { nomp, nom, [ nom, nomp ] } ]   [ art<sub>dS</sub><sup>^</sup>, @AdjOrd, arrondissement/nom ] } avec @NomRue : avenue, rue, boulevard, square, etc. @AdjOrd : premier, deuxième, .., vingtième</pre>
DISTANCE	<pre>[ adj<sub>num</sub>, (kilo)<sup>^</sup>mètre(s)<sup>^</sup>/nom ]</pre>
PRIX	<pre>[ adj<sub>num</sub>, euro(s)<sup>^</sup>/nom ]</pre>
GPDATE	<pre>[ @prepDate, Date ] avec @prepDate : vers, de, du, pour, etc.</pre>
GPHEURE	<pre>[ @prepHeure, Date ] avec @prepHeure : vers, à, pour, etc.</pre>
GPA DRESSE	<pre>[ @prepAdresse, Adresse ] avec @prepAdresse : de, du, à, au, près-de, dans, etc.</pre>
GPDISTANCE	<pre>[ @prepDistance, Distance ] avec @prepDistance : de, à, etc. etc.</pre>
GPPRIX	<pre>[ @prepPrix, Prix ] avec @prepPrix : pour, avec, de, etc.</pre>

TAB. B.3 – Quelques exemples des définitions pour les groupes élémentaires retenus concernant les expressions langagières indépendantes de l'application

## Annexe C

# Construction semi-automatique de la grammaire de liens

La présente annexe retrace la démarche incrémentale ayant permis la construction semi-automatique de la grammaire de liens utilisée. Les différentes étapes progressives pour la construction de la grammaire sont les suivantes :

1. Expertise du domaine d'application. Le chapitre C.1 est un extrait des documents ayant été rédigés à la suite de l'expertise du domaine d'application considéré : le renseignement touristique. Cette expertise non exhaustive a été réalisée manuellement à partir des ressources disponibles au laboratoire (1000 énoncés du corpus PARISCORP [Bonneau-Maynard et Devillers, 1998] recueillis dans le cadre de l'Action de Recherche Concertée B2 de l'AUF<sup>1</sup>). Cette expertise a eu pour objectifs (cf. chapitre 4 §4.1) :
  - la structuration du domaine d'application afin de rendre plus aisée la construction du dictionnaire sémantico-pragmatique,
  - l'identification des schémas et des concepts impliqués.
2. Déduction, à partir de l'expertise réalisée, des étiquettes sémantico-pragmatiques utilisées par la grammaire de liens. Nous en donnons la liste dans le chapitre C.2.
3. Écriture du dictionnaire sémantico-pragmatique. Nous donnons à titre d'exemple un court sous-ensemble du dictionnaire sémantico-pragmatique ayant été déduit (écriture manuelle) à partir de l'expertise réalisée dans le chapitre C.3. C'est à partir de ce dictionnaire (une réécriture semi-formelle des relations identifiées, cf. 6 § 6.2.4) qu'a été construite semi-automatiquement la grammaire de liens. Le dictionnaire sémantico-pragmatique est l'interface privilégiée permettant d'enrichir la couverture sémantico-pragmatique du système.
4. Génération de la grammaire de lien. La grammaire de lien est compilée automatiquement de manière "aveugle" (cf. chapitre 6 §6.2.4) à partir du dictionnaire précédent. Nous donnons au chapitre C.4 un sous-ensemble exemple de la grammaire de liens actuelle.

## C.1 Expertise du domaine d'application

Ce chapitre est un extrait des documents ayant été rédigés pour l'identification des concepts et la structuration du domaine. Nous avons tout d'abord procédé par la caractérisation des *objets* du monde de l'application évoqués dans le corpus d'étude. Nous regroupons dans un premier temps les propriétés (spécifiques et générales) de chacun des objets repertoriés. Nous structurons cette présentation en proposant une première classification des objets en fonction d'une requête de demande d'information particulière. La seconde partie présente les différentes requêtes de demande d'information pouvant être formulées en précisant les sous-ensembles des objets que chacune d'entre elles peut prendre en argument.

---

<sup>1</sup>Agence des Universités Francophones

### C.1.1 Les objets du monde de l'application

L'étude du corpus PARISCORP a permis de répertorier 43 concepts pour décrire les différents objets de l'application. Nous décrivons dans ce chapitre les propriétés spécifiques de chacun des objets identifiés (que nous appellerons concepts de propriétés). Ces objets possèdent toutefois des propriétés communes que nous commençons par évoquer.

### C.1.2 Caractéristiques générales

La principale caractéristique commune à l'ensemble des objets répertoriés provient de la manière dont ils peuvent être *référéncés* dans l'énoncé. Nous retiendrons dans la suite de ce document quatre façons de référencer un objet dans un énoncé :

1. explicitement par un nom (exemple : *l'hôtel Caumartin*),
2. indirectement par une propriété liée à l'objet (exemple : *l'hôtel le plus proche*),
3. par l'intermédiaire d'un autre objet (exemples : *le plan de la gare Saint Lazarre* ou *la chambre de l'hôtel Caumartin*). Nous donnons la liste de ces objets au paragraphe C.1.3.2,
4. par une combinaison de 3. et (1. ou 2.).

Les formulations 1. et 3. nécessitent des concepts de propriété généraux (et donc des étiquettes sémantico-pragmatiques) pour lier l'objet à sa référence (cf. § C.2.). Nous qualifierons dans la suite de ce document d'objets *définis* les objets porteurs dans l'énoncé d'une telle propriété<sup>2</sup>. Cette distinction est importante pour déterminer l'ensemble des objets arguments des différents schémas que nous retiendrons.

Le paragraphe suivant décrit l'ensemble des propriétés spécifiques liées aux objets du monde de l'application.

### C.1.3 Liste des objets répertoriés

Nous avons choisi pour la présentation des différents objets répertoriés une classification selon un critère particulier : celui de leur localisation. Nous distinguons ici les objets *localisables* (ceux qui disposent d'une adresse physique dans la ville de Paris et pouvant donc être argument d'une requête de demande de *localisation*<sup>3</sup> ou être qualifiés par un critère exprimant une distance en ville) de ceux qui ne sont pas localisables par une adresse. D'autres regroupements en classe peuvent être effectués entre ces objets ; ces regroupements seront présentés dans la section suivante en fonction des requêtes de demandes d'informations retenues.

---

<sup>2</sup>À noter que le groupe nominal *l'hôtel* par exemple fait partie de l'ensemble des objets *définis*. Cependant dans ce cas précis, aucune étiquette sémantico-pragmatique ne sera nécessaire pour représenter cette information : le caractère indéfini ou défini porté par le déterminant est identifié au moment de l'analyse en *chunk* (cf. chapitre 5.).

<sup>3</sup>Exemple d'énoncé exprimant cette requête : *donnez moi l'adresse de l'hôtel le plus proche.*

Nous présentons les propriétés de chacun des objets de la manière suivante :

< **Nom du concept** >  $\left[ \begin{array}{l} \text{Nom du concept} : [\text{exemples d'éléments du lexique}] \\ \text{propriété associée} : [\text{pouvant exprimer la propriété}] \end{array} \right]$

Dans l'ensemble de ce chapitre, les étiquettes avec astérisque (\*) désignent des propriétés pouvant être instanciées zéro ou plusieurs fois dans le concept ; celles avec le signe (+) des propriétés pouvant être instanciées zéro et **au plus** une fois ; les autres devant être obligatoirement instanciées. Le signe | représente l'opérateur logique OU.

### C.1.3.1 Classification des objets selon le critère de localisation

#### 1. Objets localisables

##### – **Vue d'ensemble**

La figure C.1 représente sous forme d'arborescence les objets détaillés par la suite. Seules les feuilles de l'arbre (en gras) sont des concepts particuliers de l'application. Chaque noeud de l'arbre est un nom générique qui dénote l'ensemble des concepts qu'il englobe.

##### – **Les lieux de transport** : < Lieux Transport >

Quatre concepts sont regroupés dans cet ensemble : < **Gare** > et < **Aéroport** > (ne possédant pas de propriétés spécifiques) ; < **Station** > et < **Ligne** > ayant respectivement les structures suivantes :

< **Station** >  
 [ Type transport urbain : [ **taxis** | RER | **bus** | **métro** ] ]  
 < **Ligne** >  
 [ Type transport urbain : [ RER | **bus** | **métro** ] ]

##### – **Les établissements** : < Etablissements >

Les établissements sont répartis en trois catégories (les établissements financiers, de loisirs et ceux à caractère commercial –répartis eux mêmes en trois catégories : culturels, de tourisme (logement et restauration) et simple commerces–) regroupant au total 14 concepts.

- < Etablissements Financiers >  
 < **E.Financiers** > (aucune propriété spécifique)
- < Etablissements Loisirs >  
 < **E.Loisirs** > (aucune propriété spécifique)



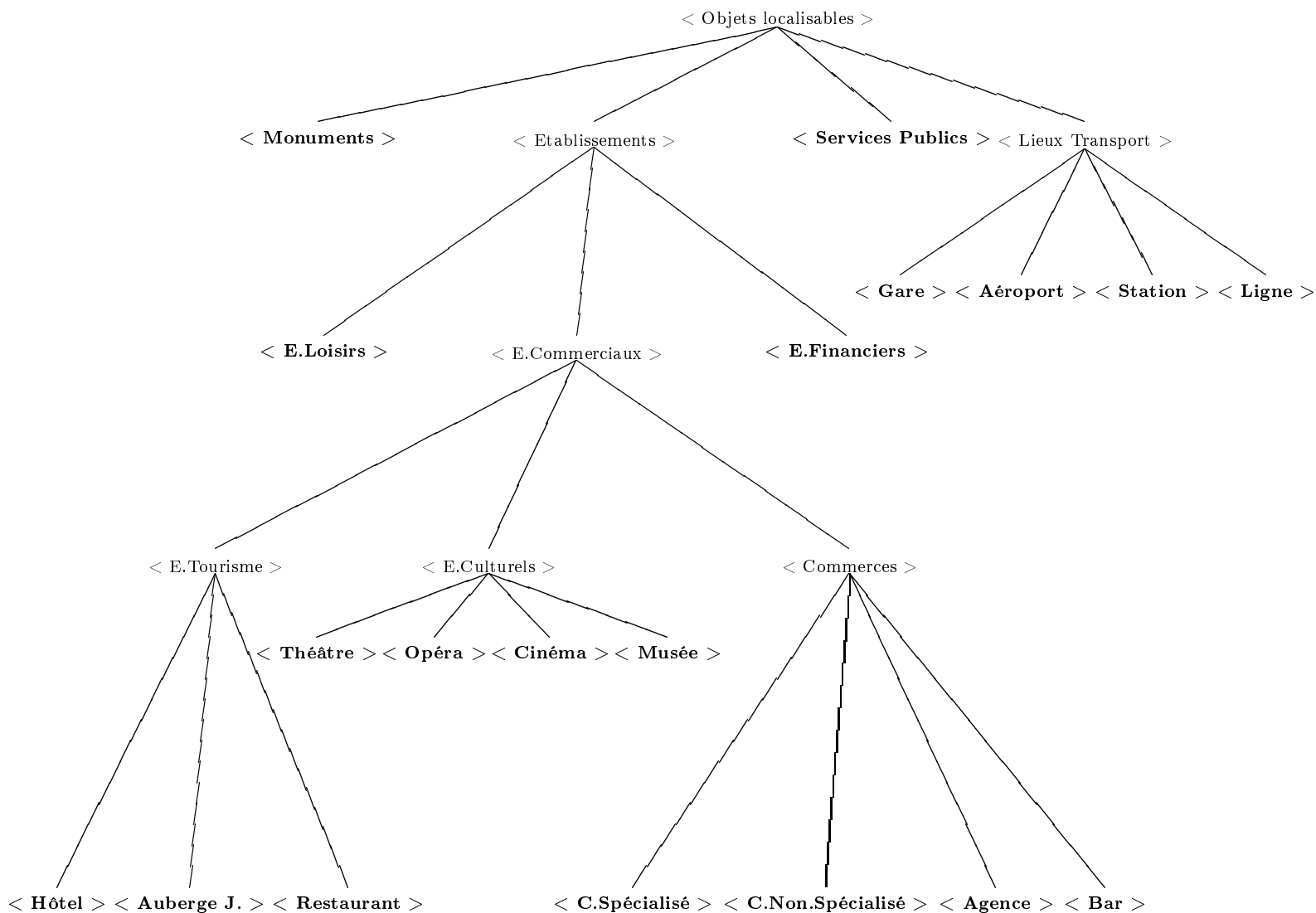


FIG. C.1 – Objets Localisables. Les feuilles de l’arbre (en gras) représentent les concepts ; les noeuds dénotent l’ensemble des concepts qu’ils englobent.

- < Etablissements Commerciaux >

- < E.Tourisme >

< **Restaurant** >

[ rapidité<sup>+</sup> : [ normale | rapide ]  
type repas proposé\* : [ menu | plat du jour | carte ]  
spécialité\* : [ indien | chinois | viande | ... ] ]

< **Hôtel** >

[ catégorie<sup>+</sup> : [ 1 | 2 | 3 | 4 | 5 ] (étoiles)  
services\* : [ restaurant | petit déjeuner |  
demi pension | sauna ] ]

< **Auberge Jeunesse** > (aucune propriété spécifique).

- < Etablissements Culturels >

< **Théâtre** >

[ Objet programme\* : < **Spectacle défini** > ]

< **Opéra** >

[ Objet programme\* : < **Oeuvre Opéra définie** > ]

< **Musée** >

[ Objet programme\* : < **Exposition définie** >  
thème\* : [ biologie | histoire | ... ] ]

< **Cinéma** >

[ compagnie<sup>+</sup> : [ Gaumont | Paramount | ... ]  
caractéristique salle\* : [ accès handicapés | ... ]  
Objet programme\* : < **Film défini** > ]

Les concepts généraux associés : < **Spectacle** >, < **Exposition** > et < **Film** > sont présentés dans le paragraphe suivant (objets non localisables directement par une adresse en ville). Le concept < **Oeuvre Opéra** > est par ailleurs un sous concept de < **Spectacle** >.

- < Commerces ><sup>4</sup>

< **C.Spécialisé** > (aucune propriété spécifique)

< **C.Non.Spécialisé** >

[ article vendu\* : [ vêtements | chaussures | ... ] ]

---

<sup>4</sup>Les deux premiers concepts de cet ensemble se distinguent uniquement par la façon dont ils peuvent être évoqués dans l'énoncé : par un nom "type" de commerce dans le premier cas (une pharmacie, une boulangerie, etc.), par la ou les marchandise(s) vendue(s) pour le second.

< **Agence** >

[ Type agence : [ voyage | immobilière | ... ]

< **Bar** >

[ spécialité\* : [ bière | vin | ... ] ]

– Les “services” (urgences, offices de tourisme, etc.) **offerts par une ville** :

< **Services Publics** > (aucune propriété spécifique)

– Les **principaux monuments de la ville** (historiques ou curiosités) :

< **Monuments** > (aucune propriété spécifique)

## 2. Objets non localisables

– **Vue d’ensemble**

La figure C.2 représente sous forme d’arborescence les objets détaillés par la suite. Seules les feuilles de l’arbre (en gras) sont des concepts particuliers de l’application. Chaque noeud de l’arbre est un nom générique qui dénote l’ensemble des concepts qu’il englobe.

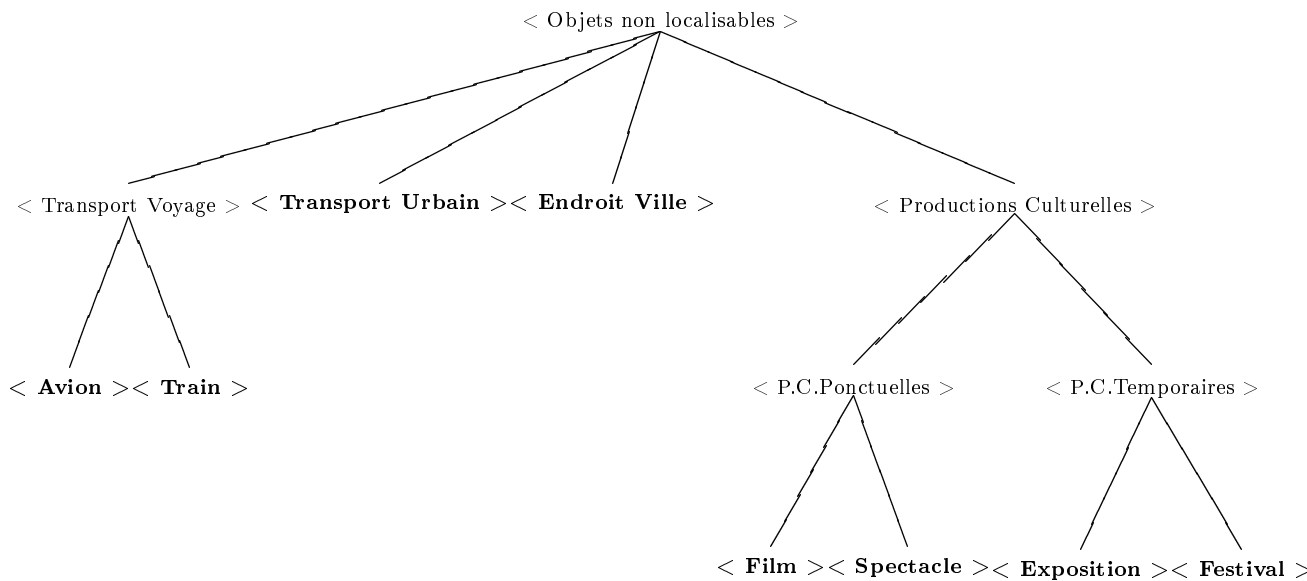


FIG. C.2 – Objets non localisables (par une adresse en ville). Les feuilles de l’arbre (en gras) représentent les concepts ; les noeuds dénotent l’ensemble des concepts qu’ils englobent.

– **Les moyens de transport pour les voyages** : < Transport Voyage >

< **Avion** >

[ Type avion<sup>+</sup> : [ airbus | concorde ] ]

< **Train** >

[ Type train<sup>+</sup> : [ grandes lignes | banlieues | TGV ] ]

– **Les moyens de transport urbain** : < **Transport Urbain** >

< **Transport Urbain** >

[ Type transport urbain : [ bus | métro | RER | taxi ] ]

– **Les productions culturelles** : < Productions Culturelles >

Les productions culturelles sont réparties en deux catégories : les productions appelées *temporaires* (i.e. qui durent un certain temps sur plusieurs jours : les expositions et les festivals) et les productions dites *ponctuelles* : les spectacles et films de cinéma).

– < P. C. ponctuelles >

< **Film** >

[ version<sup>+</sup> : [ V.O | V.F ]  
genre\* : [ western | polar | ... ] ]

< **Spectacle** >

[ Type de spectacle : [ < **Concert** > | < **Oeuvre Opéra** > |  
< **Pièce Théâtre** > ] ]

Avec les concepts < **Concert** >, < **Oeuvre Opéra** > et < **Pièce Théâtre** > définis de la manière suivante :

< **Concert** >

[ Type musique<sup>+</sup> : [ classique | rock | ... ]  
Auteur<sup>+</sup> : (nom)  
Interprète<sup>+</sup> : (nom) ]

< **Oeuvre Opéra** >

[ Type oeuvre<sup>+</sup> : [ classique | opérette ]  
Auteur<sup>+</sup> : (nom)  
Interprète voix\* : (nom)  
Chef d'orchestre<sup>+</sup> : (nom) ]

< **Pièce Théâtre** >

Type pièce <sup>+</sup> :	[ comédie   drame   ... ]
Auteur <sup>+</sup> :	(nom)
Acteurs* :	(nom)
Metteur en scène <sup>+</sup> :	(nom)

- < P. C. temporaires >

< **Festival** >

Thème <sup>+</sup> :	(...)
Organisation* :	(nom)
Programme* :	< <b>Spectacle défini</b> >

< **Exposition** >

Thème* :	[ peinture   sculpture   ... ]
Auteur <sup>+</sup> :	(nom)

- < **Endroit Ville** > (lieux dans une ville : boulevard, square, rue, etc.) (aucune propriété spécifique)

### C.1.3.2 Les objets devant être identifiés par référence à un autre objet

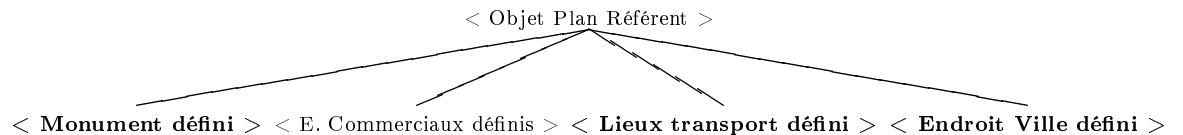
Nous répertorions dans ce paragraphe les 10 objets ne pouvant être évoqués dans un énoncé qu'en faisant référence (implicitement ou explicitement) à un autre objet.

1. Un plan (d'un lieu, d'un établissement) : < **Plan** >

< **Plan** >

[ Objet Référent : < Objet Plan Référent > ]

L'arborescence suivante représente les concepts appartenant à l'ensemble noté < Objet Plan Référent > ; ces concepts ont été présentés dans les sections précédentes.



2. Une chambre (d'un hôtel) : < **Chambre** >

< **Chambre** >

Pte type chambre <sup>+</sup> :	[ simple   double ]
nombre lits <sup>+</sup> :	[ 1   2   ... ]
bains <sup>+</sup> :	[ commun   salle de bain   douche ]
catégorie <sup>+</sup> :	[ basse   moyenne   haute ]
Objet Référent :	< Objet Chambre Référent >

< Objet Chambre Référent >

On a < Objet Chambre Référent > :

< **Hôtel défini** > < **Auberge Jeunesse définie** >

3. Une table (d'un restaurant) : < **Table** >

< **Table** >

nb personnes* :	[ 1   2   ... ]
Objet Référent :	< <b>Restaurant défini</b> >

4. Un repas (dans un restaurant) : < **Repas** >

< **Repas** >

type menu <sup>+</sup> :	[ menu   plat jour ]
Objet Référent :	< <b>Restaurant défini</b> >

5. Une visite (d'un lieu culturel) : < **Visite** >

< **Visite** >

caractéristique <sup>+</sup> :	guidée
Objet Référent :	< Objet Visite Référent >

On a : < Objet Visite Référent > :

< Objet Visite Référent >

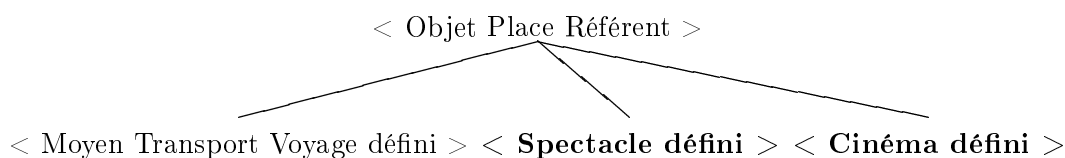
< **Exposition définie** > < **Musée défini** > < **Monuments défini** >

6. Une place (de cinéma, de train, etc.) : < **Place** >

< **Place** >

catégorie <sup>+</sup> :	[ première   deuxième ]
Objet Référent :	< Objet Place Référent >

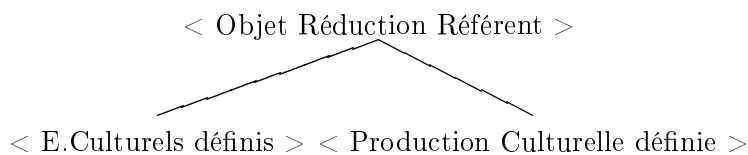
Avec < Objet Place Référent > :



7. Une réduction (pour une manifestation culturelle) : < **Réduction** >

< **Réduction** >  
[ Type Réduction<sup>+</sup> : [ vermeil | étudiant | ... ]  
Objet Référent : < Objet Réduction Référent > ]

Avec < Objet Réduction Référent > :



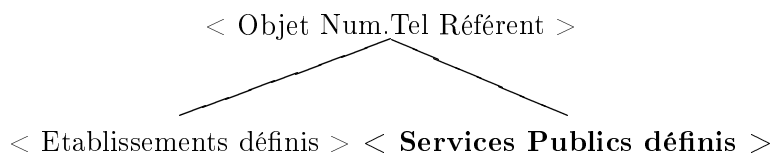
8. Une salle (de spectacles) : < **Salle** >

< **Salle** >  
[ Objet Référent : < E.Culturels définis > ]

9. Un numéro de téléphone : < **Num. Tel** >

< **Num.Tel** >  
[ Objet Référent : < Objet Num.Tel Référent > ]

Avec < Objet Num.Tel Référent > :



10. Un endroit dans une gare : < **Endroits Gare** >

< **End. Gare** >  
[ Type<sup>+</sup> : [ consignes | guichet | ... ]  
Objet Référent : < **Gare définie** > ]

### C.1.3.3 Objets divers

Compte tenu de la classification choisie pour la présentation des différents objets du monde de l'application (regroupement Objets localisables / Objets non localisables),

certaines objets (rares dans le corpus) n'ont pas pu être listés. Il s'agit par exemple d'objets comme les *cabines téléphoniques*, jamais localisées par une adresse dans une requête d'un locuteur, mais disposant de propriétés susceptibles d'être évoquées.

< **Cabines Téléphoniques** >

[ type Cabine<sup>+</sup> : [ pièces | carte ] ]

**C.1.4 Les schémas retenus**

L'ensemble des énoncés du corpus d'étude ont pu être couverts par les 8 actes dialogiques présentés au chapitre 4. Nous n'avons ainsi retenu dans un premier temps que ces actes dialogiques. Nous présentons dans ce paragraphe en détail les différents *schémas* (actes de demande d'information et de confirmation) retenus afin d'identifier les différents concepts *arguments* associés. Chacun des actes présentés portent sur un sous-ensemble des objets définis dans le paragraphe précédent. Nous présentons sous forme arborescente ces sous-ensembles en utilisant les notations adoptées précédemment. Chaque acte de demande d'information ainsi que leurs concepts arguments associés est représenté dans ce paragraphe de la manière suivante :

Nom de l'acte

[ Nom du concept argument : < ensembles des concepts pouvant être instanciés > ]

**1. les actes de demandes d'information**

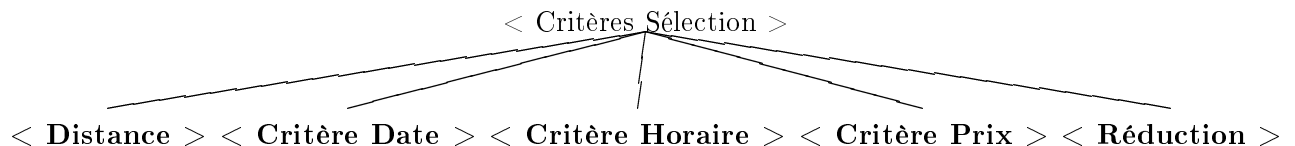
**C.1.4.1 Sélection**

acte-sélection :

[ objet à sélectionner : < **Objet** >  
critères-sélection\* : < Critères Sélection > ]

Tous les objets présentés précédemment peuvent potentiellement être sélectionnés selon certains critères. L'ensemble < **Objet** > dénote ainsi l'ensemble des objets généraux non définis qui ont été présentés.

< Critères Sélection > est l'ensemble des concepts suivants :



Cet ensemble de concepts exprime également les différentes manières de référencer indirectement un objet particulier.



(a) critère de proximité / éloignement par rapport à un autre objet :

< **Distance** >

[ Objet Référent : < Objet Distance Défini > ]

< Objet Distance Défini > regroupe tous les objets *localisables* définis ainsi que chaque lieu défini de la ville (rue, square, etc.) :

< Objet Distance défini >

< Objets localisables définis > < **Endroit Ville défini** >

Un critère de sélection sur une distance peut être formulé dans un énoncé de deux manières différentes :

- par un identificateur de proximité (“*proche de*” par exemple),
- par une limite inférieure et/ou supérieure de distance à respecter. Dans ce cas deux étiquettes sémantico-pragmatiques seront éventuellement nécessaires pour identifier la limite inférieure de la limite supérieure :

< **Limite Distance** >

[ limite inf distance<sup>+</sup> : (*distance inf*) ]  
[ limite sup distance<sup>+</sup> : (*distance sup*) ]

(b) critère sur les dates (d’ouverture ou de fermeture d’un établissement, de début ou de fin d’un évènement) :

< **Critère Date** >

[ Type critère date : [ **fermeture** | **ouverture** | ... ]

De la même manière que pour le concept précédent, les différentes manières d’exprimer un critère sur une date nécessiteront éventuellement des étiquettes sémantico-pragmatiques particulières (par exemple si l’utilisateur évoque une période comprise entre une date  $x$  et une date  $y$ , cf. § C.2).

(c) critère sur les horaires :

< **Critère Horaire** >

[ type horaire : [ **fermeture** | **ouverture** | ... ]  
[ Moment Journée : [ **matin** | **après-midi** | ... ]

Nous donnons dans le chapitre C.2 les étiquettes nécessaires en fonction des expressions des horaires.

- (d) critère de prix : < **Critère Prix** > (aucune propriété spécifique)
- (e) critère d'une présence de réduction particulière, exprimé par le concept < **Réduction** > présenté auparavant.

#### C.1.4.2 Demande de localisation

acte-localisation :

[ objet à localiser : < **Objet Localisable défini** > ]

L'ensemble < **Objet localisable défini** > correspond aux objets définis associés aux objets de la figure C.1 auxquels s'ajoutent les concepts < **Plan** > et < **Endroits Gare** > pouvant faire l'objet d'une demande de localisation.

#### C.1.4.3 Demande d'information sur une distance

demande-info-distance :

[ objet 1 défini : < **Objet Distance défini** > ]  
 [ objet 2 défini : < **Objet Distance défini** > ]

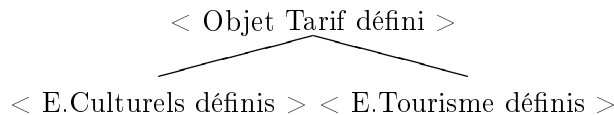
L'ensemble de concepts < **Objet Distance défini** > (regroupant les objets *localisables* définis et chaque lieu défini – rue, square, ... – de la ville) est l'ensemble présenté § C.1.4.1.

#### C.1.4.4 Demande d'information portant sur un tarif

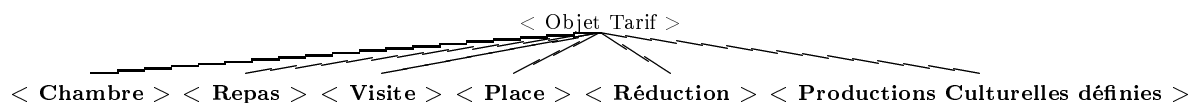
demande-info-tarif :

[ objet tarif : [ < **Objet Tarif défini** > | ]  
 < **Objet Tarif** > ]

Ces deux ensembles de concepts se distinctent selon la manière dont les objets sont identifiés. L'ensemble < **Objet Tarif défini** > correspond aux établissements dont on veut connaître les prix généraux (hôtels, théâtres, ... ) :



L'ensemble < **Objet Tarif** > est celui des objets définis (par référence à un autre objet) dont on cherche à connaître le prix (une chambre d'un hôtel particulier, une visite d'un musée, etc.) ou encore le montant pour une réduction dans un établissement donné, un film particulier dans un cinéma donné, etc.



#### C.1.4.5 Demande d'information portant sur les horaires

demande-info-horaire :

$$\left[ \begin{array}{l} \text{type horaire} : \quad [ \text{fermeture} \mid \text{ouverture} ] \\ \text{objet horaire} : \quad [ \langle \text{Objet Horaire défini} \rangle \mid \\ \quad \text{date}^+ : \quad \quad \quad (date) \end{array} \right]$$

On a ici  $\langle \text{Objet Horaire défini} \rangle$  :

$$\begin{array}{c} \langle \text{Objet Horaire défini} \rangle \\ \swarrow \quad \searrow \\ \langle \text{Etablissements définis} \rangle \quad \langle \text{Productions Culturelles définies} \rangle \end{array}$$

#### C.1.4.6 Demande d'information portant sur les dates

demande-info-date :

$$\left[ \begin{array}{l} \text{Type date} : \quad [ \text{fermeture} \mid \text{ouverture} \mid \dots ] \\ \text{objet date} : \quad [ \langle \text{Objet Date défini} \rangle \mid \end{array} \right]$$

Avec  $\langle \text{Objet Date défini} \rangle : \langle \text{Objet Horaire défini} \rangle$ , ensemble précédemment décrit.

#### C.1.4.7 Demande d'information portant sur un trajet

demande-info-trajet :

$$\left[ \begin{array}{l} \text{Moyen Locomotion}^* : \quad [ \langle \text{Transport Urbain} \rangle \mid \\ \quad \quad \quad \langle \text{Transport Urbain défini} \rangle \mid \\ \quad \quad \quad \text{à pied} \mid \text{en voiture}, \dots ] \\ \text{Objet Trajet} : \quad \quad \quad \langle \text{Objet Trajet défini} \rangle \end{array} \right]$$

Le concept  $\langle \text{Transport Urbain} \rangle$  a été présenté § C.1.3.1. On a par ailleurs :

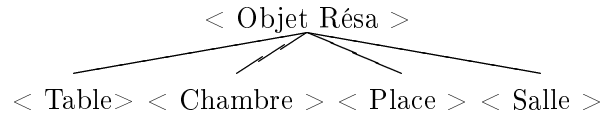
$\langle \text{Objet Trajet défini} \rangle$

$$\left[ \begin{array}{l} \text{départ} : \quad \quad \langle \text{Lieux définis} \rangle \\ \text{correspondance}^* : \quad \langle \text{Lieux définis} \rangle \\ \text{arrivée} : \quad \quad \langle \text{Lieux définis} \rangle \end{array} \right]$$

L'ensemble  $\langle \text{Lieux définis} \rangle$  correspond à :

$$\begin{array}{c} \langle \text{Lieux définis} \rangle \\ \swarrow \quad \downarrow \quad \searrow \\ \langle \text{Etablissements définis} \rangle \quad \langle \text{Monuments définis} \rangle \quad \langle \text{Lieux Transport définis} \rangle \quad \langle \text{Endroits Ville définis} \rangle \end{array}$$





### C.1.4.13 Demande de réservation d'un voyage

demande-réservation-voyage :

[	trajet :	< <b>Objet Voyage défini</b> >	]
	date :	< <b>Dates Voyage</b> >	
	heure :	< <b>Heures Voyage</b> >	
	transport :	[ < avion >   < train > ]	
	Moyen payment <sup>+</sup> :	[ carte bleue   chèque   ... ]	]

Le concept < **Objet Voyage défini**> décrit le voyage souhaité entre différentes villes :

< <b>Objet Voyage défini</b> >	[	ville départ :	(nom ville)	]
		ville correspondance* :	(nom ville)	
		ville arrivée :	(nom ville)	]

Les concepts représentant les dates et les heures du voyage incluent de la même manière une date (resp. une heure) de départ et d'arrivée.

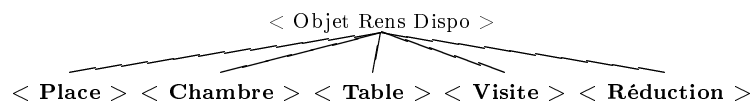
## 2. Les actes de demande de confirmation

À chacune des 13 demandes d'informations présentées jusqu'à présent correspondent des demandes de confirmation (demande de confirmation qu'un film donné dure bien 2 heures par exemple). Ces requêtes nécessiteront des étiquettes sémantico-pragmatiques spécifiques afin d'identifier :

- l'acte de demande d'information qui demande à être confirmé,
- et la réponse proposée (cf. § C.2).

Les autres actes de demande de confirmation correspondent aux requêtes de demande de renseignements sur la disponibilité d'objets (disponibilité de places pour un concert par exemple) et la présence ou l'absence de propriétés spécifiques sur un objet.

L'ensemble des concepts objets dont on peut chercher à tester la disponibilité est donné par l'ensemble suivant :



Chacun des concepts de propriété que nous avons présenté au paragraphe C.1.3 peut potentiellement faire l'objet d'un acte de demande de vérification de présence ou d'absence de la propriété en question.

## C.2 Liste des étiquettes pragmatiques retenues pour l'analyse des dépendances entre segments

Nous donnons dans ce chapitre la liste complète des 144 étiquettes sémantico-pragmatique auxquelles nous avons abouti à partir de l'expertise précédente. Celles-ci correspondent principalement aux concepts de propriétés et aux concepts arguments identifiés.

### C.2.1 Étiquettes exprimant les actes dialogiques retenus

- Ouverture-dialogue
- Fermeture-dialogue
- Reponse (2 étiquettes : Accord, Désaccord)
- Assertif
- Demande-explicitation (+ 1 étiquette *Objet-explicitation* reliant le groupe verbal identifiant l'acte et l'élément sur lequel porte la requête<sup>5</sup>.)
- Correction
- Demande-information (cf. paragraphe suivant)
- Demande-confirmation (cf. paragraphe suivant)

### C.2.2 Schémas et principaux concepts arguments associés

Le tableau C.1 répertorie les étiquettes impliquées dans les actes de demande d'information. La dernière colonne du tableau donne des exemples très simples de chunks pouvant être reliés par ces étiquettes. Ces exemples sont donnés à titre illustratif : plusieurs autres liens peuvent être instanciés entre les chunks qui nous servent d'exemple. L'instanciation des liens entre chunks dépend des autres chunks présents conjointement dans l'énoncé et de leur position relative. Pour plus de lisibilité, nous représentons simplement les chunks en encadrant entre crochets les mots de l'énoncé le composant sans noter les étiquettes grammaticales caractérisant ce chunk (cf. chapitre 5).

Le tableau C.2 répertorie les étiquettes impliquées dans les actes de demande de confirmation.

---

<sup>5</sup>Cette étiquette relie par exemple les chunks suivants : [répéter] et [les horaires].

Schémas	Concepts arguments associés	exemples
Sélection	OBJET-SELECT CRITERESELECTLoc CRITERESELECTDate CRITERESELECTHoraire CRITERESELECTPrix CRITERESELECTRéduction	[donnez] [la liste] [la liste] [des hôtels] [la liste] [proche de ...] [mardi] [quel est] [à 16 heures] [quel est] [quel est] [pour pas cher] [la liste] [avec réduction]
Info-Localisation	OBJET-LOC	[quelle est] [l'adresse] [l'adresse] [de l'hôtel]
Info-Distance	OBJET-DISTa OBJET-DISTb	X [est] il [loin] de Y [X] est il [loin] de Y X est il [loin] [de Y]
Info-Tarif	OBJET-TARIF	[quel est] [le prix] [la prix] [de la chambre]
Info-Horaire	OBJET-HORAIRE	[quels sont] [les horaires] [horaires] [du cinéma]
Info-Date	OBJET-DATE	[quel jour] [ferme] [ferme] [le musée]
Info-Trajet	LIEU(-DEPART, -DEST, -ETAPE) MOYEN-LOC	je [voudrais] [aller] [pour aller] [à, via, de ..] [pour aller] à X [en métro]
Info-Transport		[quel métro] [pour aller]
Info-Durée	OBJET-DUREE	combien [de temps] [dure] [dure] [la visite]
Info-Prestations ou Info-Programme	ETABLISSEMENT REFERENT	[quel est] [le programme] [programme] [du ciné]
Réservation-Etablissement	OBJET-RESA DATE-RESA HEURE-RESA MOYEN-PAIEMENT	[voudrais] [réserver] [réserver] [une place] [réserver] [pour demain] [réserver] [à 16 heures] [réserver] [par carte]
Réservation-Voyage	VILLE(-DEPART, -DEST, -ETAPE) DATE-RESA HEURE-RESA MOYEN-LOC MOYEN-PAIEMENT	[réserver] [un billet] [réserver] [pour, de, via ...]  [un billet] [de train]

TAB. C.1 – Étiquettes des liens correspondant aux principaux schémas et concepts arguments des actes de demande d'information

Schémas	Concepts arguments associés	exemples
Confirmation Conf-Loc  Conf-Distance  Conf-Tarif  (...) (cf. tableau C.1)	ReponseLoc-CONF  ReponseDistance-CONF  ReponseTarif-CONF	c'[est] [bien] ...   c'est [bien] 10 euros [le prix] [est ce que] c'est [10 euros]
Rens-Disponibilité	OBJET-RENS-DISPO	[y a t il] des places [de libre] [y a t il] [des places]
Rens-Propriété	Propriété-RENS	[est ce que] vous [faites] des réductions [y a t il] [la télé]

TAB. C.2 – Étiquettes des liens correspondant aux principaux schémas et concepts arguments des actes de demande de confirmation

### C.2.3 Autres concepts arguments

Les tableaux précédents ont répertorié les principaux concepts arguments impliqués dans les schémas retenus. Le tableau C.3 regroupe les étiquettes retenues pour l'expression des autres concepts (contraintes liées aux requêtes : expression de périodes horaire, dates de voyage etc.)



Concepts arguments	exemples
LimiteTarif	[entre] 10 [et] 20 euros
Limite-Inf-Tarif	entre [10] [et] 20 euros
Limite-Sup-Tarif	entre 10 [et] [20 euros]
LimiteDistance (...)	
Periode	
Date-Début-Periode	[réserver] [du 10 mai]
Date-Fin-Pistance	[réserver] du 10 [au 20 mai]
PeriodeHoraire (...)	
Date(-Départ, -Arrivée)	
Horaire(-Départ, -Arrivée)	
Ident-Date	c'[est] [un mardi]
Ident-Num-jour	c'[est] [le 10]
Type-Horaire	[horaire] [de fermeture]
Moment-Journée	[3 heures] [du matin]
Ident-Catégorie-Tarif	c'[est] [cher]
Ident-Loc	je [suis] [à la gare]

TAB. C.3 – Étiquettes des liens correspondant aux concepts arguments

### C.2.4 Concepts de propriété

Les tableaux C.4 et C.5 regroupent l'ensemble des étiquettes retenues pour l'expression des concepts de propriété. Le tableau C.4 répertorie des étiquettes exprimant des propriétés générales sur les objets. Le tableau C.5 liste l'ensemble des étiquettes correspondant aux propriétés spécifiques à chacun des objets identifiés.

Concepts de propriétés	exemples
Référence-Objet-Ident	[l'hôtel] [Caumartin]
Objet-Référent	[la chambre] [de l'hôtel]
PtePrix	[un restaurant] [pas cher]
PteDate	[un musée] ouvert [le 8 mai]
PteHoraire	[un musée] ouvert [à 8 heures]
PteDistance	[un musée] [près de la gare]
Quantité	j'en [veux] [trois]

TAB. C.4 – Étiquettes des liens correspondant aux concepts de propriétés généraux

Concepts de propriétés	objets concernés
Type-Agence	agence (de voyage, immobilière)
Spécialité-Bar	bar (à vin)
Type(-Train, -Avion)	avion, train (de banlieue)
Ident(-Train, -Avion, -TransportUrbain)	avion, train, le bus (10)
Verion, Genre	film (en VO, policier)
TypeMusique, Interprete	concert
TypOeuvre, Auteur, ChefOrchestre, InterpreteVoix	œuvre opéras
TypePièce, Auteur, MiseEnScène, Acteur	pièce de théâtre
ThèmeFestival, OrganisationFestival	festival
ThèmeExpo	exposition
TypeTransportUrbain	staion, ligne (de bus)
Rapidité, TypeRepasProposé, Spécialité	restaurant
Catégorie, Services	hôtel
TypeChambre, NbLits, Bain, CatégorieChambre	chambre
Thème	opéra
Compagnie, PropriétéSalle, PropriétéSon	cinéma
ArticlesVendus	commerce
NbPersonnes	table d'un restaurant
TypeMenu	repas
CaractéristiquesVisite	visite d'un musée
CatégoriePlace	place (en première)
TypeRéduction	réduction (étudiante)
TypeCarte	carte (bancaire, routière)
TypeCabineTel	cabine téléphonique (à cartes)
PositionVoyage	places (fenêtre, hublots)
TypeVoitureVoyage	places (fumeur)

TAB. C.5 – Étiquettes des liens correspondant aux concepts de propriété

### C.2.5 Autres étiquettes

La tableau C.6 regroupe enfin l'ensemble des étiquettes retenues pour le traitement de divers phénomènes syntaxiques (coordinations, ellipses, relatives, négation, etc.) et les quelques étiquettes non spécifiques au domaine d'application n'ayant pas encore été évoquées (rôle des adverbes dans le relâchement de contraintes par exemple).

Étiquettes	exemples
Politesse	[///] [merci]
Modifieur	[près de la gare] [si possible]
Négation	[pas] [près de la gare]
Agent	[je] [cherche]
Patient	[donnez] [moi]
COO	[un hôtel] [et] [un restaurant]
COO(Tarif, Durée, etc.-	(cf. chapitre 6)
COR	[un hôtel] [non] [un restaurant]
COR(Tarif, Durée, etc.-	(cf. chapitre 6)
Antécédent	[un hôtel] [qui] est ...
Schéma	je [voudrais] [ceux] du musée aussi
Objet	je voudrais [ceux] [du musée] aussi

TAB. C.6 – Autres Étiquettes

### C.3 Exemple du dictionnaire sémantico-pragmatique

Nous donnons dans ce chapitre un exemple simplifié du dictionnaire sémantico-pragmatique à partir duquel est compilée la grammaire de liens utilisée. L'écriture de ce dictionnaire est fondée sur la structuration des concepts décrite au chapitre C.1. Les entrées du dictionnaire correspondent aux chunks caractérisés par la segmentation du corpus d'étude, par exemple :

- GVint(être)
- GN(horaire.defini)
- GP(de,cinema.defini)
- GN(étoile)
- Coo(et)
- ...

Chaque entrée du dictionnaire a la structure suivante :

```
%%           Exemple de la structure de chaque entrée du dictionnaire

GN(X)      %% identifiant de l'entrée
           SC: SC1
                %% liste des étiquettes correspondant aux schémas
                %% pouvant être instanciés à partir du chunk GN(X)

           CA: CA1 CA2
                %% liste des étiquettes correspondant
                %% aux concepts arguments
```

```

CP: CP1 CP2 CP3
    %% liste des étiquettes correspondant
    aux concepts de propriétés

Mod: Mod1
    %% liste des étiquettes de type Modifieurs

S: S1
    %% liste des liens syntaxiques
    (Antécédent par exemple)

```

Prenons deux exemples simples correspondant aux chunks GN(horaire.defini) et GP(de,cinéma.defini) :

```

GN(horaire.defini)
  SC: Info-Horaire
  CA: OBJET-Horaire
  CP: Type-Horaire

```

```

GP(de,cinéma.defini)
  CA: OBJET-Horaire OBJET-Loc OBJET-Date ... ETABLISSEMENT-REFERENT
  CP: Référence-Objet-Ident Objet-Référent Pteprix ... Compagnie ...

```

L'entrée du dictionnaire correspondant à GN(horaire.defini) indique que ce groupe nominal peut faire l'objet de trois types de liens :

- Info-Horaire : étiquette correspondant à un schéma et pouvant être instanciée entre GN(horaire.defini) et GVint(être) par exemple : [quels sont] [les horaires]
- OBJET-Horaire : étiquette correspondant à un concept argument et pouvant être instanciée entre GN(horaire.defini) et GP(de,cinéma.defini) par exemple : [les horaires] [du cinéma]
- Type-Horaire : étiquette correspondant à un concept de propriété et pouvant être instanciée entre GN(horaire.defini) et GP(de,fermeture.defini) par exemple : [les horaires] [de fermeture]

L'objet *cinéma* fait partie de l'ensemble de concepts que nous avons dénoté au chapitre C.1 < Etablissements-Culturels >. À ce titre la structuration du domaine nous indique que le groupe prépositionnel GP(de,cinéma.defini)<sup>6</sup> peut faire l'objet d'actes de demande d'information portant sur les horaires, les tarifs, une localisation, etc. et donc peut être relié avec d'autres segments de l'énoncé par les étiquettes OBJET-Horaire, OBJET-Tarif, OBJET-Loc, etc. De la même manière, la liste CP (concept de propriété) correspondant au chunk GP(de,cinéma.defini) est complétée par les différentes étiquettes correspondant aux concepts de propriétés correspondant à l'objet cinéma.

---

<sup>6</sup>Comme d'ailleurs le groupe nominal GN(cinéma.defini).

Une première version du dictionnaire a ainsi été écrite manuellement à partir de l'expertise du domaine réalisée. Ce dictionnaire a été ensuite progressivement complété en examinant les structures de liens produites sur les énoncés du corpus Pariscorp (le système utilise successivement les différentes grammaires de lien compilées à partir des versions successives du dictionnaire (cf. chapitre 6 § 6.2.4)).

## C.4 Exemple de la grammaire de liens

Nous donnons enfin dans ce chapitre un sous-ensemble de la grammaire de lien utilisée. Afin d'illustrer le processus de compilation, nous donnons ci-dessous l'extrait du dictionnaire sémantico-pragmatique ayant conduit à la grammaire de lien présentée.

```
%% Extrait du dictionnaire sémantico-pragmatique ayant permis
%% la génération automatique de la grammaire de lien présentée.
```

```
GN(horaire.défini)
```

```
    SC: Info-Horaire
```

```
    CA: OBJET-HORAIRE
```

```
    CP: Type-Horaire
```

```
GP(de,fermeture.défini)
```

```
    CP: Type-Horaire
```

```
GP(de,cinéma.défini)
```

```
    CA: OBJET-HORAIRE OBJET-LOC ETABLISSEMENT-REFERENT
```

```
    CP: Référence-Objet-Ident Objet-Référent
```

```
GAdv(éventuellement)
```

```
    Mod: Modifieur
```

```
GN(restaurant.défini)
```

```
    CA: OBJET-LOC OBJET-HORAIRE
```

```
    CP: Spécialité Rapidité PtePrix
```

```
GVint(etre)
```

```
    SC: Info-Horaire Info-Tarif Sélection
```

```
%% Coo
```

```
Coo(et)
```

```
    Att: Horaire
```

```
    GP(de,cinéma.défini)
```

GP(de,musée.défini)

Coo(et)

Att: Tarif

GP(de,cinéma.défini)

GP(de,musée.défini)

On notera dans ce dictionnaire deux entrées correspondant à la coordination logique *et*. Ces entrées indiquent à la fois l'attribut impliqué (i.e. la caractéristique que devra respecter les deux segments qui seront reliés par la coordination : ils doivent ici pouvoir faire l'objet d'une requête Horaire ou Tarif) et l'ensemble des segments répondant à chaque caractéristique. Le processus de compilation crée à la fois les règles pour la coordination et enrichit les règles des segments impliqués.

Les règles principales de compilation sont les suivantes :

- Les liens obligatoires sont ceux qui correspondent au plus haut niveau dans la hiérarchie (SC > CA > CP > Mod).
- Il ne peut y avoir qu'un seul lien de plus haut niveau de hiérarchie par règle. Par exemple, pour l'entrée GP(de,cinéma.défini), il n'y aura qu'une seule étiquette correspondant au concept argument par règle (le concept argument est dans ce cas le concept de plus haut niveau car il n'y a pas de schémas pour cette entrée). Toutes les combinaisons des concepts de propriété sont en revanche possibles simultanément (les étiquettes sont facultatives).

Nous donnons ci-après La grammaire de lien générée<sup>7</sup>. Les notations adoptées dans les règles se lisent de la manière suivante :

- $X^+$  : lien à droite
- $X^-$  : lien à gauche
- $\{X^-\}$  : lien facultatif
- $@X^-$  : plusieurs liens de même type autorisés

GN(horaire.défini):

(Info-Horaire- & {@OBJET-HORAIRE-}) or  
(Info-Horaire- & {@OBJET-HORAIRE+}) or  
(Info-Horaire+ & {@OBJET-HORAIRE+}) or  
({@OBJET-HORAIRE-} & Info-Horaire-) or  
({@OBJET-HORAIRE-} & Info-Horaire+) or  
({@OBJET-HORAIRE+} & Info-Horaire+);

GP(de,fermeture.défini):

(Type-Horaire-) or

---

<sup>7</sup>Nous ne donnons pas dans ce document toutes les règles correspondant à l'entrée GP(de,cinéma.défini).

(Type-Horaire+);

GP(de, cinéma.défini):

(OBJET-HORAIRE- & {Référence-Objet-Ident-} & {Objet-Référent-}) or  
(OBJET-HORAIRE- & {Référence-Objet-Ident-} & {Objet-Référent+}) or  
(OBJET-HORAIRE- & {Référence-Objet-Ident+} & {Objet-Référent+}) or  
(OBJET-HORAIRE+ & {Référence-Objet-Ident+} & {Objet-Référent+}) or  
(OBJET-HORAIRE- & {Objet-Référent-} & {Référence-Objet-Ident-}) or  
(OBJET-HORAIRE- & {Objet-Référent-} & {Référence-Objet-Ident+}) or  
(OBJET-HORAIRE- & {Objet-Référent+} & {Référence-Objet-Ident+}) or  
(OBJET-HORAIRE+ & {Objet-Référent+} & {Référence-Objet-Ident+}) or  
({Référence-Objet-Ident-} & OBJET-HORAIRE- & {Objet-Référent-}) or  
({Référence-Objet-Ident-} & OBJET-HORAIRE- & {Objet-Référent+}) or  
({Référence-Objet-Ident-} & OBJET-HORAIRE+ & {Objet-Référent+}) or  
({Référence-Objet-Ident+} & OBJET-HORAIRE+ & {Objet-Référent+}) or  
({Référence-Objet-Ident-} & {Objet-Référent-} & OBJET-HORAIRE-) or

...

({Objet-Référent-} & OBJET-HORAIRE+ & {Référence-Objet-Ident+}) or  
({Objet-Référent+} & OBJET-HORAIRE+ & {Référence-Objet-Ident+}) or  
({Objet-Référent-} & {Référence-Objet-Ident-} & OBJET-HORAIRE-) or  
({Objet-Référent-} & {Référence-Objet-Ident-} & OBJET-HORAIRE+) or  
({Objet-Référent-} & {Référence-Objet-Ident+} & OBJET-HORAIRE+) or  
({Objet-Référent+} & {Référence-Objet-Ident+} & OBJET-HORAIRE+) or  
(OBJET-LOC- & {Référence-Objet-Ident-} & {Objet-Référent-}) or  
(OBJET-LOC- & {Référence-Objet-Ident-} & {Objet-Référent+}) or  
(OBJET-LOC- & {Référence-Objet-Ident+} & {Objet-Référent+}) or  
(OBJET-LOC+ & {Référence-Objet-Ident+} & {Objet-Référent+}) or  
(OBJET-LOC- & {Objet-Référent-} & {Référence-Objet-Ident-}) or  
(OBJET-LOC- & {Objet-Référent-} & {Référence-Objet-Ident+}) or  
(OBJET-LOC- & {Objet-Référent+} & {Référence-Objet-Ident+}) or  
(OBJET-LOC+ & {Objet-Référent+} & {Référence-Objet-Ident+}) or  
({Référence-Objet-Ident-} & OBJET-LOC- & {Objet-Référent-}) or

...

({Référence-Objet-Ident+} & {Objet-Référent+} & OBJET-LOC+) or  
({Objet-Référent-} & OBJET-LOC- & {Référence-Objet-Ident-}) or  
({Objet-Référent-} & OBJET-LOC- & {Référence-Objet-Ident+}) or

...

({Objet-Référent-} & {Référence-Objet-Ident+} & OBJET-LOC+) or  
({Objet-Référent+} & {Référence-Objet-Ident+} & OBJET-LOC+) or  
(ETABLISSEMENT-REFERENT- & {Référence-Objet-Ident-} & {Objet-Référent-}) or  
(ETABLISSEMENT-REFERENT- & {Référence-Objet-Ident-} & {Objet-Référent+}) or  
(ETABLISSEMENT-REFERENT- & {Référence-Objet-Ident+} & {Objet-Référent+}) or  
(ETABLISSEMENT-REFERENT+ & {Référence-Objet-Ident+} & {Objet-Référent+}) or

...

({Référence-Objet-Ident+} & ETABLISSEMENT-REFERENT+ & {Objet-Référent+}) or  
 ({Référence-Objet-Ident-} & {Objet-Référent-} & ETABLISSEMENT-REFERENT-) or  
 ...  
 ({Objet-Référent-} & {Référence-Objet-Ident+} & ETABLISSEMENT-REFERENT+) or  
 ({Objet-Référent+} & {Référence-Objet-Ident+} & ETABLISSEMENT-REFERENT+) or  
 (COOHoraire- & {Référence-Objet-Ident+}) or  
 ({Référence-Objet-Ident+} & COOHoraire+) or  
 (COOHoraire- & {Objet-Référent+}) or  
 ({Objet-Référent+} & COOHoraire+) or  
 (COOTarif- & {Référence-Objet-Ident+}) or  
 ({Référence-Objet-Ident+} & COOTarif+) or  
 (COOTarif- & {Objet-Référent+}) or  
 ({Objet-Référent+} & COOTarif+);

GAdv(éventuellement):

(Modifieur-) or  
 (Modifieur+);

GN(restaurant défini):

(OBJET-LOC- & {Spécialité-} & {Rapidité-} & {PtePrix-}) or  
 (OBJET-LOC- & {Spécialité-} & {Rapidité-} & {PtePrix+}) or  
 (OBJET-LOC- & {Spécialité-} & {Rapidité+} & {PtePrix+}) or  
 (OBJET-LOC- & {Spécialité+} & {Rapidité+} & {PtePrix+}) or  
 ...  
 (OBJET-LOC- & {Rapidité-} & {Spécialité-} & {PtePrix+}) or  
 (OBJET-LOC- & {Rapidité-} & {Spécialité+} & {PtePrix+}) or  
 (OBJET-LOC- & {Rapidité+} & {Spécialité+} & {PtePrix+}) or  
 (OBJET-LOC+ & {Rapidité+} & {Spécialité+} & {PtePrix+}) or  
 ...  
 (OBJET-LOC+ & {PtePrix+} & {Spécialité+} & {Rapidité+}) or  
 (OBJET-LOC- & {PtePrix-} & {Rapidité-} & {Spécialité-}) or  
 (OBJET-LOC- & {PtePrix-} & {Rapidité-} & {Spécialité+}) or  
 (OBJET-LOC- & {PtePrix-} & {Rapidité+} & {Spécialité+}) or  
 (OBJET-LOC- & {PtePrix+} & {Rapidité+} & {Spécialité+}) or  
 (OBJET-LOC+ & {PtePrix+} & {Rapidité+} & {Spécialité+}) or  
 ({Spécialité-} & OBJET-LOC- & {Rapidité-} & {PtePrix-}) or  
 ({Spécialité-} & OBJET-LOC- & {Rapidité-} & {PtePrix+}) or  
 ({Spécialité-} & OBJET-LOC- & {Rapidité+} & {PtePrix+}) or  
 ({Spécialité-} & OBJET-LOC+ & {Rapidité+} & {PtePrix+}) or  
 ({Spécialité+} & OBJET-LOC+ & {Rapidité+} & {PtePrix+}) or  
 ({Spécialité-} & OBJET-LOC- & {PtePrix-} & {Rapidité-}) or  
 ...  
 ({Spécialité-} & {Rapidité-} & {PtePrix+} & OBJET-LOC+) or  
 ({Spécialité-} & {Rapidité+} & {PtePrix+} & OBJET-LOC+) or







({Rapidité-} & OBJET-HORAIRE+ & {Spécialité+} & {PtePrix+}) or  
 ({Rapidité+} & OBJET-HORAIRE+ & {Spécialité+} & {PtePrix+}) or  
 ({Rapidité-} & OBJET-HORAIRE- & {PtePrix-} & {Spécialité-}) or  
 ({Rapidité-} & OBJET-HORAIRE- & {PtePrix-} & {Spécialité+}) or  
 ({Rapidité-} & OBJET-HORAIRE- & {PtePrix+} & {Spécialité+}) or  
 ({Rapidité-} & OBJET-HORAIRE+ & {PtePrix+} & {Spécialité+}) or  
 ({Rapidité+} & OBJET-HORAIRE+ & {PtePrix+} & {Spécialité+}) or  
 ({Rapidité-} & {Spécialité-} & OBJET-HORAIRE- & {PtePrix-}) or  
 ({Rapidité-} & {Spécialité-} & OBJET-HORAIRE- & {PtePrix+}) or  
 ({Rapidité-} & {Spécialité-} & OBJET-HORAIRE+ & {PtePrix+}) or  
 ({Rapidité-} & {Spécialité+} & OBJET-HORAIRE+ & {PtePrix+}) or  
 ({Rapidité+} & {Spécialité+} & OBJET-HORAIRE+ & {PtePrix+}) or  
 ...  
 ({Rapidité-} & {PtePrix+} & OBJET-HORAIRE+ & {Spécialité+}) or  
 ({Rapidité+} & {PtePrix+} & OBJET-HORAIRE+ & {Spécialité+}) or  
 ({Rapidité-} & {PtePrix-} & {Spécialité-} & OBJET-HORAIRE-) or  
 ({Rapidité-} & {PtePrix-} & {Spécialité-} & OBJET-HORAIRE+) or  
 ({Rapidité-} & {PtePrix-} & {Spécialité+} & OBJET-HORAIRE+) or  
 ({Rapidité-} & {PtePrix+} & {Spécialité+} & OBJET-HORAIRE+) or  
 ({Rapidité+} & {PtePrix+} & {Spécialité+} & OBJET-HORAIRE+) or  
 ({PtePrix-} & OBJET-HORAIRE- & {Spécialité-} & {Rapidité-}) or  
 ({PtePrix-} & OBJET-HORAIRE- & {Spécialité-} & {Rapidité+}) or  
 ({PtePrix-} & OBJET-HORAIRE- & {Spécialité+} & {Rapidité+}) or  
 ({PtePrix-} & OBJET-HORAIRE+ & {Spécialité+} & {Rapidité+}) or  
 ({PtePrix+} & OBJET-HORAIRE+ & {Spécialité+} & {Rapidité+}) or  
 ({PtePrix-} & OBJET-HORAIRE- & {Rapidité-} & {Spécialité-}) or  
 ...  
 ({PtePrix+} & {Spécialité+} & OBJET-HORAIRE+ & {Rapidité+}) or  
 ({PtePrix-} & {Spécialité-} & {Rapidité-} & OBJET-HORAIRE-) or  
 ({PtePrix-} & {Spécialité-} & {Rapidité-} & OBJET-HORAIRE+) or  
 ({PtePrix-} & {Spécialité-} & {Rapidité+} & OBJET-HORAIRE+) or  
 ({PtePrix-} & {Spécialité+} & {Rapidité+} & OBJET-HORAIRE+) or  
 ({PtePrix+} & {Spécialité+} & {Rapidité+} & OBJET-HORAIRE+) or  
 ...  
 ({PtePrix-} & {Rapidité-} & {Spécialité+} & OBJET-HORAIRE+) or  
 ({PtePrix-} & {Rapidité+} & {Spécialité+} & OBJET-HORAIRE+) or  
 ({PtePrix+} & {Rapidité+} & {Spécialité+} & OBJET-HORAIRE+);

GVint(etre):

(Info-Horaire-) or  
 (Info-Horaire+) or  
 (Info-Tarif-) or  
 (Info-Tarif+) or  
 (Sélection-) or

(Sélection+);

%% Coo

Coo(et):

(COOHoraires- & COOHoraires+ & Horaires+) or

(COOHoraires- & Horaires- & COOHoraires+) or

(COOTarif- & COOTarif+ & Tarif+) or

(COOTarif- & Tarif- & COOTarif+);

## Annexe D

# Évaluation par défi – Énoncés dérivés

Nous reproduisons dans cette annexe 3 séries d'énoncés dérivés constituant un sous-ensemble du jeu de test à partir duquel notre système a été évalué dans le cadre de DEFI. Chaque série a été produite par un participant différent à partir de l'énoncé initial suivant :

*donnez moi les horaires de fermeture du musée Grévin et du musée du Louvre*

Ces séries reflètent ainsi la diversité des phénomènes introduits en fonction des centres d'intérêts de chacun des concepteurs.

## D.1 Série 1

1. donnez moi les horaires de fermeture du Grévin et du Louvre
2. à quelle heure ferment le musée Grévin et le musée du Louvre
3. le musée Grévin et le musée du Louvre ils ferment à quelle heure
4. je voudrais les tarifs et les horaires de fermeture du musée Grévin et du musée du Louvre
5. donnez moi les horaires de fermeture des musées Grévin et Louvre
6. la fermeture du musée Grévin et du musée du Louvre c est à quelle heure
7. est ce que vous auriez pu me fournir les horaires de fermeture du musée Grévin et du musée du Louvre
8. le musée Grévin ses horaires de fermeture c est quoi et le musée du Louvre
9. pour la fermeture du musée Grévin et du musée du Louvre les horaires c est quoi
10. vous avez les horaires de fermeture du musée Grévin et du Louvre ils ferment à quelle heure le Grévin et le Louvre
11. donnez moi les tarifs non pardon les horaires de fermeture du musée Grévin et du musée du Louvre
12. donnez moi les horaires de fermeture le dimanche du musée Grévin et du musée du Louvre
13. les horaires de fermeture vous les avez ceux du musée Grévin et du musée du Louvre
14. je voudrais savoir les horaires de fermeture du musée Grévin et du musée du Louvre est ce que vous les avez

## D.2 Série 2

1. ben donnez moi les horaires de fermeture du musée Grévin et euh du musée du Louvre
2. donnez moi les horaires de fer euh fermeture du musée Grévin et du musée du Louvre

3. donnez moi les horaires de fermeture du musée Grévin et du musée du musée du Louvre
4. donnez moi les horaires d ouverture non de fermeture du musée Grévin et du musée du Louvre
5. donnez moi les horaires de fermeture du musée d Orsay euh Grévin et du musée du Louvre
6. donnez moi les horaires de fermeture du musée euh musée Grévin et du musée Grévin euh non du Louvre
7. donnez moi les horaires fermeture du musée Grévin et du musée du Louvre
8. donnez moi les horaires de fermeture tu musée Grévin et du musée du Louvre
9. les moi horaires de fermeture du musée Grévin et du musée du Louvre
10. donnez moi pour le musée Grévin et le musée du Louvre les horaires de fermeture
11. pour le musée Grévin et le musée du Louvre c est quoi les horaires de fermeture
12. horaires de fermeture s il vous plaît du musée Grévin et du Louvre
13. donnez moi les horaires de fermeture du musée Grévin et puis si c est possible aussi du musée du Louvre
14. donnez moi mais horaires de fer euh fermeture du musée Grévin et du musée du Louvre
15. donnez moi pour le musée Grévin et le musée du Louvre euh et bien les horaires de fermeture

### D.3 Série 3

1. à quelle heure ferment le musée Grévin et le musée du Louvre
2. quels sont les horaires des musées euh celui du Louvre
3. et pour le musée Grévin il ferme à quelle heure
4. quels sont les horaires possibles pour euh les visites
5. et on peut le visiter jusqu à quelle heure ce musée
6. quel est le musée qui ouvre le plus tôt entre le musée Grévin et le musée du Louvre
7. non pas le musée Grévin celui du Louvre je voudrais les horaires s il vous plaît
8. bonjour euh je voudrais les les ho à quelle heure enfin les horaires d ouverture du Louvre s il vous plaît
9. pour visiter c est jusqu à quelle heure
10. donnez moi les euh de fermeture du loué Grévin et du musée du Louvre
11. donnez moi les horaires de Louvre du musée musée et du et du Louvre
12. je voudrais si possible les informations sur les horaires pour euh les visites dans les musées

13. horaires de fermeture

14. non non à quelle heure il ouvre pas la fermeture

15. est ce que le musée est ouvert demain matin







## Résumé

Nous présentons une stratégie robuste d'analyse des énoncés oraux pour la compréhension hors-contexte de la parole en dialogue homme-machine finalisé. Nous faisons l'hypothèse qu'une analyse détaillée des énoncés oraux (associant syntaxe et sémantique) est essentielle au traitement correct des énoncés et est la condition nécessaire au développement d'applications non plus limitées à des cadres très finalisés mais faisant preuve d'une certaine genericité. Dans le système proposé (ROMUS), appliqué au renseignement touristique, une analyse syntaxique partielle de surface (cascades de transducteurs) permet tout d'abord la segmentation de l'énoncé en groupes minimaux élémentaires. Cette étape, générique, est motivée par la prise en compte explicite et intrinsèque des régularités observées dans les productions orales. Une analyse globale des dépendances sémantico-pragmatiques entre les segments (grammaires de liens) permet ensuite de déduire la représentation sémantique de l'énoncé.

**Mot clés : Traitement Automatique des Langues, Dialogue homme-machine, Compréhension Automatique de la Parole, TAL robuste, Théorie des Langues et Automates, Grammaires de Liens, Linguistique de Corpus**

## Abstract

This PhD focusses on speech understanding in man-machine communication. We discuss the issue of how a speech understanding system can be made robust against spontaneous speech phenomena as well as achieving a detailed analysis of spoken French. We argue that a detailed linguistic analysis (with both syntax and semantics) is essential for correctly process spoken utterances and is also a necessary condition to develop applications that are not entirely dedicated to a very specific task but present sufficient genericity. The system presented (ROMUS) implements speech understanding in a two-stage process. The first one achieves a finite-state shallow parsing that consists in segmenting the utterance into basic units (spoken adapted chunks). This stage is generic and is motivated by the regularities observed in spoken French. The second one, a Link Grammar parser, looks for inter-chunks dependencies in order to build a rich representation of the semantic structure of the utterance.

**Keywords : Natural Language Processing, man-machine Dialog, Speech Understanding, Robust Parsing, Transducers, Link Grammars, Corpus Linguistics**