Étude des phénomènes d'extraction en français parlé sur deux corpus de dialogue oral finalisé

Application à la communication orale homme-machine

Jean-Yves Antoine — Jérôme Goulian

VALORIA, Université de Bretagne Sud (EA 2593)
Rue Yves Mainguy, F-56 000 Vannes
{Jean-Yves.Antoine,Jerome.Goulian}@univ-ubs.fr

RESUME. *Cet article présente une étude détaillée des phénomènes d'extraction menée sur deux corpus de dialogue oral finalisé. Après avoir présenté le cadre de notre étude, nous détaillons l'ensemble des observations quantitatives effectuées (fréquences d'occurrence, répartition des extractions par procédé et par fonction syntaxique des éléments détachés, etc.). Nous donnons alors les conclusions que l'on peut tirer de cette étude du point de vue de l'ingénierie linguistique : traitement des structures non-projectives et des formes faibles de variabilité linéaire, influence du domaine applicatif considéré etc.*

ABSTRACT. *This paper presents a corpus analysis that concerns the study of extractions and other word-order variations on two task-oriented spoken dialogue corpora. We first present the context of this work, which concerns the problem of word-order variation and its processing. We then detail the main results of this corpus analysis. We finally present several conclusions that this study should provide to natural language processing (projectivity, weak word-order variation, task influence).*

MOTS-CLES : *analyse de corpus ; communication orale homme-machine ; français parlé ; ordre des mots ; extraction ; projectivité ; influence de la tâche.*

KEY WORDS: *corpus analysis ; spoken man-machine dialogue ; French spoken language ; word order ; extraction ; projectivity ; task influence*

1. Analyse de corpus et CHM orale

La communication orale homme-machine (CHM orale par la suite) a atteint une maturité que traduit l'apparition récente d'applications réelles telles que, par exemple, le système automatique de routage téléphonique grand public mis en place par AT&T (Lokbani & White, 1998), ou encore le système de réservation par téléphone des chemins de fer néerlandais faisant suite au projet européen ARISE (den Os *et al*, 1999). D'une manière générale, l'ensemble des traitements automatiques impliqués dans la CHM orale a connu au cours des dernières années des progrès significatifs.

En dépit de ces indéniables réussites, force est néanmoins de constater que la portée des résultats obtenus reste limitée. En particulier, on remarquera que la plupart des systèmes de dialogue développés à l'heure actuelle ne concernent qu'un seul domaine applicatif, celui du renseignement aérien (ATIS[1]) ou ferroviaire. Il s'agit là de tâches très finalisées mettant en jeu des vocabulaires de taille modeste (quelques milliers de mots). Cette spécialisation a permis la mise en œuvre d'approches très pragmatiques[2] ne reposant sur aucune analyse détaillée des énoncés oraux. Si ces méthodes se sont révélées robustes face au traitement de la parole spontanée, rien ne nous garantit cependant qu'elles se révèleraient toujours aussi efficaces sur des domaines d'application plus riches. D'une manière générale, la question de la généralisation des techniques utilisées en CHM orale à d'autres domaines n'est ainsi pas tranchée (Hirschman, 1998) et laisse éventuellement la porte ouverte à des changements de paradigmes sensibles.

La question de la généralité des méthodes utilisées en CHM orale constitue ainsi une interrogation centrale du domaine (Hirschman, 1998). Aussi regrettera-t-on de ne disposer que de peu d'outils méthodologiques pour la résoudre. Par exemple, il est indéniable que les grands programmes d'évaluation — tels que ceux de la (D)ARPA américaine (Pallett *et al*., 1994) par exemple — qui ont été mis en œuvre au cours de la dernière décennie ont eu une influence très positive sur les recherches en CHM orale. Cependant, en se limitant à un champ applicatif précis (ATIS) 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 du système étudié[3]. Ils ne présentent ainsi aucun caractère prédictif et sont par conséquent dans l'incapacité de nous guider sur les évolutions futures du domaine (Antoine et Caelen, 1999).

Parallèlement, les corpus oraux recueillis en CHM orale sont utilisés essentiellement pour l'apprentissage des systèmes de dialogue et de leurs composants langagiers (modèle de langage, compréhension de la parole, gestion du dialogue). Cet apprentissage suit le plus souvent une procédure (automatique ou

---

[1] ATIS : Air Transport Information Systems

[2] Citons par exemple, en compréhension automatique de la parole, les approches sélectives (Minker *et al*., 1999) consistant à ne détecter et ne considérer dans l'énoncé que certains segments - clés nécessaires à l'élaboration d'une requête d'interrogation d'une base de données (sens dit « utile » de l'énoncé).

[3] Pour prendre comme illustration le sujet qui nous intéressera dans cet article, ces programmes d'évaluation ne peuvent nous renseigner, quantitativement ou qualitativement, sur le comportement du système en présence de phénomènes d'extraction. L'utilité *immédiate* d'un tel diagnostic n'est certes pas évidente dans le cas des systèmes actuels, basés sur des approches sélectives et/ou stochastiques. Elle nous semble cependant nécessaire à la généralisation de la CHM orale à des domaines moins finalisés qui requièrent la mise en œuvre de techniques d'analyses plus détaillées.

non) itérative d'amorçage (*bootstrap*) qui consiste à effectuer une première modélisation grossière qui sera progressivement raffinée de manière empirique au vu des erreurs constatées du système. La technique du *bootstrap* repose ainsi sur une approche par essai-erreur ne laissant pas de place à une analyse détaillée des usages mis en jeu dans la situation considérée. C'est pourquoi notre connaissance des caractéristiques linguistiques du dialogue oral finalisé est encore relativement limitée. Ces connaissances sont encore plus parcellaires en ce qui concerne l'étude de la variabilité (influence de la tâche, du contexte d'interaction, etc.) de ces caractéristiques[4].

La situation actuelle de la CHM orale semble ainsi paradoxale : d'un côté, on dispose de systèmes qui sont de plus en plus efficaces sur un type de tâches bien défini. De l'autre, en adoptant une approche orientée « traitement de l'information » — qui a son intérêt — au détriment d'une approche plus linguistique, la CHM orale ignore en partie le matériau sur lequel elle travaille. Ignorant les phénomènes linguistiques qui caractérisent le langage oral en situation de dialogue homme-machine, elle est ainsi dans l'incapacité de s'interroger sur la pertinence réelle des paradigmes qu'elle a développés.

Dans cet article, nous souhaitons montrer comment la linguistique de corpus peut apporter une réponse à cette situation d'aveuglement paradoxale. Nous considérons en effet qu'une analyse précise, rigoureuse, qualitative mais aussi quantitative, de corpus oraux issus de diverses situations interactives (linguistique variationniste), est à même de fournir une caractérisation linguistique utile à la fois au prototypage des systèmes de dialogue et à la conduite des recherches futures dans le domaine.

À titre illustratif, nous allons nous intéresser dans cet article à un type de phénomène linguistique particulier. Il s'agit des extractions ou plus généralement l'ensemble des phénomènes concernant la variabilité de l'ordre des mots dans l'énoncé. Cette étude sera menée sur deux corpus de français parlé correspondant à un genre particulier, le dialogue oral finalisé, et correspondant à deux tâches différentes (réservation aérienne et renseignement touristique). Dans un premier temps, nous nous attacherons à l'analyse, d'un point de vue purement linguistique, de ces deux corpus. Puis, nous discuterons des implications de cette étude sur l'ingénierie des langues appliquée à la CHM orale. Au préalable, nous allons revenir sur les problèmes posés par le traitement automatique des phénomènes d'extraction.

2. Le problème de la variabilité de l'ordre des mots

2.1. *Variabilité et ingénierie des langues*

La question de l'ordre variable des mots et son corollaire, celui des dépendances discontinues, a toujours constitué 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 dépendances face aux grammaires de constituants (Covington, 1990 ; voir aussi (Hudson, 2000) pour un article introductif récent). De même, le traitement des structures discontinues constitue un des fondements d'un formalisme tel que les

[4] On citera tout de même à ce sujet (Morel *et al*., 1989).

grammaires syntagmatiques liées par la tête (Pollard et Sag, 94). Il a été également étudié (Rambow et Joshi, 94) dans le cadre des grammaires d'arbres adjoints (TAG).

D'une manière générale, on peut distinguer deux niveaux de variabilité de l'ordre des mots dans l'énoncé (Holan *et al*. 2000) :

— d'une part, une variabilité faible autorisant une position variable des constituants (continus)[5] de l'énoncé. Ces mouvements n'induisent aucune discontinuité dans la structure de dépendance de l'énoncé. C'est par exemple le cas du mouvement du groupe prépositionnel *pour Rio Sao Paulo* dans l'exemple suivant :

bon sinon ***pour Rio Sao Paulo*** *je pense qu'il y a pas mal de vols* (AF.II.8.C68)[6]

— d'autre part une variabilité forte qui se traduit par un relâchement des contraintes de continuités autorisant la production d'énoncés qualifiés de **non-projectifs**. C'est par exemple le cas de l'extraction des mots-questions (*wh-questions*) dont voici un exemple en anglais :

who *do you think that Mary claims that Sarah likes* (Hudson, 2000)

Cette variabilité forte peut être également 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 subordonnée relative *qui est nouveau* de son antécédent *un tarif encore plus intéressant sur Londres* :

(…) vous savez on a ***un tarif encore plus intéressant sur Londres*** *maintenant* ***qui est nouveau*** (AF.II.33.O17)

Ces extractions discontinues se retrouvent également à l'oral dans le cas du déplacement d'un verbe voire d'une proposition exprimant une modalité. Ainsi, dans l'exemple ci-dessous, l'extraction à droite[7] de la proposition *je crois* casse la continuité de la complétive *se presser pour réserver* :

Mais il faut quand même assez se presser ***je crois*** *pour réserver* (AF.I.67.C20)

Cette distinction entre variabilité forte et faible semble a priori primordiale. Les phénomènes d'extraction interrogent en effet essentiellement l'ingénierie des langues par leurs conséquences en matière de non-projectivité. Ainsi, certains formalismes tels que les grammaires de liens reposent sur le postulat de la

[5] Voir (Holan *et al*., 2000) pour une définition non " syntagmatique " de cette variabilité faible (*freedom of constituent order within a continuous head domain*)

[6] Exemple issu du corpus Air France présenté au § 3.1. (2° partie du corpus, dialogue numéro 8, tour de parole C68, i.e. 68ème prise de parole du client).

[7] Du point de vue de l'intention du locuteur, ce détachement peut être interprété comme une incise. On ne relève cependant aucune pause ou autre marque prosodique supportant une telle interprétation pour cet exemple. À un niveau purement syntaxique, un parseur traitera de toute manière cet exemple comme un énoncé avec extraction non projective.

projectivité du langage étudié (Sleator et Temperley 1991). On sait par ailleurs que la variabilité faible peut être modélisée dans un simple cadre hors-contexte (Holan 2000). Peut-on pour autant éluder les problèmes que pose la variabilité faible à l'analyse automatique ? Ce serait ignorer l'augmentation d'ambiguïté structurale à laquelle elle conduit. Or, les systèmes de traitement du langage parlé sont très sensibles au degré d'ambiguïté de l'analyse : ils ne travaillent en effet pas sur un seul énoncé bien formé, mais au mieux sur une dizaine d'hypothèses issues de la reconnaissance de la parole (*N-best sequences)*. Ces séquences présentent par ailleurs des inattendus structurels dus, soit à l'énoncé lui-même, soit aux erreurs de la reconnaissance. Toute augmentation artificielle de l'ambiguïté[8] ne peut ainsi que favoriser l'augmentation du taux d'erreur du système. C'est pourquoi notre étude portera à la fois sur l'observation de la variabilité faible et la variabilité forte.

Nos motivations relevant de l'ingénierie linguistique (Cunnigham, 1999), cette étude sera essentiellement de nature quantitative. Notre objectif n'est en effet pas de nous interroger sur des exemples théoriques intéressants mais artificiels, mais bien sur les phénomènes attestés en corpus présentant la plus grande fréquence d'occurrence. De ce point de vue, les études linguistiques sur la variabilité de l'ordre des mots ne répondent que partiellement à nos interrogations.

2.2. *Variabilité du langage parlé et études linguistiques*

De nombreux linguistes se sont intéressés à la question de la variabilité de l'ordre des mots. On sait ainsi que la variabilité forte s'observe en premier lieu sur les langages dits à ordre variable (russe, finnois, tchèque par exemple) tandis qu'à l'opposé, les langues à ordre fixe sont essentiellement concernées par la variabilité faible. (Holan *et al.*, 2000) montre ainsi que la complexité des structures non-projectives d'un langage très fixe tel que l'anglais est limitée, alors qu'elle est en théorie infinie en tchèque. Le français est également considéré comme une langue à ordre fixe (Covington, 1990). À notre connaissance, cette conclusion ne concerne cependant que le français écrit. On sait qu'il n'existe pas de frontière claire entre oral et écrit, et que cette transition doit plutôt être appréhendée sous la forme d'un continuum de genres (Biber, 1988 ; Biber *et al.*, 1999 ; Bilger et Blanche-Benveniste, 1999). Nos travaux concernent cependant un genre de dialogue oral excessivement spontané[9] qui ne peut être identifié à un genre écrit. Il semble donc nécessaire de procéder à une analyse détaillée de la variabilité dans ce genre d'oral spontané avant de lui appliquer des conclusions portant sur le français écrit.

Plusieurs études linguistiques (Gadet, 1989 ; Blanche-Benveniste et al. 1990) nous ont donné une connaissance approfondie sur la variabilité de l'ordre des mots en français parlé. Ainsi, les différents procédés qui président aux phénomènes d'extraction (inversion, double-marquage, présentatif…) sont bien identifiés. À de rares exceptions, ces études linguistiques ne présentent cependant qu'un caractère descriptif et explicatif, sans rendre compte de l'importance relative de ces différents phénomènes dans les productions orales.

De même, la linguistique de corpus ne s'est pas intéressée à notre connaissance à l'étude de la fréquence d'apparition des extractions dans les productions orales, ceci

[8] ou de la *perplexité* pour ce qui concerne les méthodes de traitements stochastiques.

[9] Plus précisément, les corpus étudiés relèvent de deux genres oraux définis par (Biber 88) : la conversation en face à face et la conversation par téléphone.

sans doute parce que cette quantification est difficilement automatisable. Par ailleurs, la plupart des études sur le français parlé[10] portent sur des récits ou des interview oraux — discours planifiés ou non préparés, émissions de radio ou de télévision (Biber, 1988) — et non pas sur des genres relevant de la conversation orale spontanée (Kerbrat-Orecchioni, 1999). C'est donc un champ d'investigation relativement inexploré que nous abordons ici. Cette étude se fonde néanmoins sur les études linguistiques mentionnées précédemment, comme le montre le paragraphe suivant.

3. Corpus étudiés et méthodologie d'analyse

3.1. *Corpus Air France et Murol*

Notre étude a porté sur deux corpus distincts, correspondant à deux situations interactives différentes (tableau 1) :

Tableau 1. *Description des corpus d'étude.*

Corpus	Nb de dialogues	Nb 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

— 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

[10] Citons néanmoins l'important travail de (Kerbrat-Orecchioni, 1990).

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.

3.2. *Recensement des phénomènes*

3.2.1. *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.

3.2.2. *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),

(…) **ces deux rues piétonnes** hein e alors p/ e pour vous **y** rendre (MU.4.O39)

et de **postposition** (détachement à droite) :

(…) *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 :

(...) ***sur Héraklion*** *on n'a qu'un seul tarif spécial* (AF.II.17.O14)
(...) ***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* :

(...) ***le visa*** *on* ***l****'a au consulat* (AF.I.48.C6)

Cette reprise peut également être réalisée à l'aide de l'expresion *c'est* ou plus généralement *çà* + <verbe> :

(…) *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 :

(...) ***j'ai quelqu'un qu****'est allé prendre des billets charters pour moi* (AF.I.43.C9)
(...) ***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 :

(...) ***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**. :

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**.

3.3. *Limites méthodologiques de l'étude : discussion*

La pertinence de toute analyse de corpus dépend fortement de la représentatitivté 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.

3.3.1. *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[11] (plus de 200 locuteurs différents), garantit de notre point de vue la pertinence des observations et conclusions réalisées.

[11] 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 représentatitivité de ce corpus.

3.3.2. *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 § 3.1.). 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) :

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 signiticative 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 influer 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 (§ 5.).

3.3.3. *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.

4. 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.

4.1. *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 spontanés. Le tableau 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.

Tableau 2. *Fréquence d'apparition des extractions sur chaque corpus (nombre moyen de tours de parole présentant au moins une extraction)*

Corpus	fréquence moyenne	écart-type	fréquence min sur un dialogue	Fréquence max sur un dialogue
Air France	13,6 %	10,5 %	0, 0 %	30,8 %
Murol	26, 6 %	10,2 %	8,3 %	40,7 %

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 3). On ne constate en effet aucune différence significative entre les deux populations (test de Student : $T = 0{,}101$; $T_{inv}(0{,}101) = 0{,}920$).

Tableau 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)*

Corpus Air France	**Clients**	**Accueil**
moyenne (*écart-type*)	13,2 % (*10,5 %*)	13,8 % (9,*9 %*)
Corpus Air France	**Particulier**	**Agence de voyage**
moyenne (*écart-type*)	14,9 % (*6,9 %*)	10,1 % (*8,2 %*)

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 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'habituation 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.

4.2. *Extractions orales et ordre canonique sujet-verbe-objet*

Intéressons-nous ainsi aux caractéristiques des extractions observées. Le tableau 4 présente la répartition de ces observations en fonction du sens du détachement.

Tableau 4. *Répartition des extractions en fonction du sens du détachement (pourcentage moyen de phénomènes d'un sens donné)*

Corpus	antéposition	postposition	écart-type[12]
Air France	82,5 %	17,5 %	20,4 %
Murol	85,5 %	14,5 %	8,7 %

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 ; $T_{inv}(0,407) = 0,685$).

Tableau 5. *Répartition des extractions suivant la fonction de l'élément détaché (pourcentage moyen de phénomènes d'un type donné)*

Corpus		sujet	argument	modifieur	associé
Air France	moyenne	30,7 %	12,0 %	27,4 %	30,0 %
	(*écart-type*)	(*29, 6 %*)	(*15,5 %*)[13]	(*26,5 %*)	(*24,5 %*)
Murol	moyenne	25,4 %	5,3 %	23,5 %	45,8 %
	(*écart-type*)	(*7,7 %*)	(4,1 %)	(*16,1 %*)	(*13,3 %*)

[12] L'écart-type des deux variables *antéposition* et *postposition* est bien entendu le même puisque P(antéposition) = = 1 - P(postposition)

[13] Cette valeur d'écart-type supérieure à celle de la moyenne de la variable observée ne doit pas étonner : on observe une distribution non symétrique des données, qui ne suit donc pas une loi normale.

Plus intéressante est l'observation de la répartition des extractions suivant la fonction syntaxique de l'élément déplacé (tableau 5). On remarque ici une prédominance nette des fonctions sujet, modifieur et associé, par opposition aux arguments qui semblent moins prêter à extraction. L'étude statistique de ces répartitions indique que cette moindre extraction des arguments est clairement significative[14], 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[15]. Nous reviendrons au paragraphe 5 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, modifieurs 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.

Tableau 6. *Répartition des extractions d'éléments sujets en fonction du sens du détachement (pourcentage moyen de phénomènes d'un sens donné)*

Corpus	antéposition	postposition	écart-type[16]
Air France	80,6 %	19,4 %	20,4 %
Murol	90,6 %	9,4 %	8,7 %

D'autre part, on remarque que la majeure partie des sujets déplacés correspond à une situation d'antéposition (tableau 6). Cette antéposition n'altère en rien l'ordre SVO, d'autant plus que les extractions du sujet se manifestent presque toujours[17] par une reprise sous la forme d'un double marquage ou d'un présentatif (tableau 7), comme dans les exemples ci-dessous :

Double-marquage ***moi-même je*** *suis Florence* (AF.II.11.C09)
Double-marquage *donc euh le* ***le problème c'est*** *de transformer un Bruxelles Marseille en Bruxelles Nice* (AF.II.6.C15)
Présentatif ***ce n'est pas ces 9 francs*** *de taxe* ***qui*** *ont* ***qui*** *vous ont chiffonné* (AF.I.44.O30)

[14] 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.
[15] 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 : Tsuj/mod = 0,911 et $T_{inv}(0,911) = 0,363$; Tsuj/ass = 1,059 et $T_{inv}(1,059) = 0,291$. Même test sur le corpus Murol : Tsuj/mod = 0,565 et $T_{inv}(0,565) = 0,580$; Tsuj/ass = 1,797 et $T_{inv}(1,797) = 0,091$ (cas proche du seuil de criticité à 10 %). Un test de Wilcoxon-Mann-Withney confirme la significativité de la prédominance des associés : Zsuj/ass = 3,135 et Z(0,01) = 2,576.
[16] L'écart-type des deux variables *antéposition* et *postposition* est bien entendu le même puisque P(antéposition) = = 1 - P(postposition)
[17] D'une manière générale, nous ne discuterons pas dans ce texte de la validité statistique — évidente — d'observations aussi tranchées.

Tableau 7. *Répartition des extractions d'éléments sujets en fonction du procédé utilisé (pourcentage moyen de phénomènes d'un sens donné)*

Corpus	double-marquage + présentatif	autres procédés	écart-type[18]
Air France	95,4 %	4,6 %	0,8 %
Murol	100,0 %	0,0 %	0,0 %

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 8).

Tableau 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)*

Corpus	% d'extractions avec ordre SVO respecté	% d'énoncés avec ordre SVO respecté
Air France	90,3 %	98,7 %
Murol	92,3 %	98,0 %

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.

4.3. *Extractions orales et projectivité*

Comme le montre le tableau 9, le nombre de discontinuités dues aux extractions est très limité.

Tableau 9. *Part relative des extractions non projectives (pourcentage par rapport à l'ensemble des phénomènes observés et par rapport à l'ensemble des énoncés)*

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 %*)

[18] L'écart-type des deux variables *double-marquage + présentatif* et *autres procédés* est également le même.

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 incises 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 (voir le § 6.1. pour une discussion sur ce sujet).

4.4. *Fonctions et procédés*

Pour terminer cette analyse globale, nous avons étudié la répartition des différents procédés utilisés pour marquer le détachement. Le tableau 10 donne une synthèse de ces observations. On observe tout d'abord que ces résultats sont relativement stables d'un corpus à l'autre.

Tableau 10. *Répartition des extractions suivant le procédé utilisé (pourcentage moyen de procédés d'un type donné)*

Corpus		**Inversion**	**double-marquage**	**présentatif**	**éléments binaires**
Air France	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 %*)

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 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 χ^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 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).

Tableau 11. *Répartition des extractions d'éléments arguments en fonction du procédé utilisé (pourcentage moyen de phénomènes d'un sens donné)*

Corpus	double-marquage + présentatif	inversion et énoncés binaires	Ecart-type
Air France	67,3 %	32,7 %	30,2 %
Murol	77,3 %	22,7 %	11,9 %

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 modifieurs 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 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 13a). Les différences de fréquences moyennes observées ne sont en effet pas statistiquement significatives (test du χ^2 sur une répartition équiprobable des 3 procédés : CHI_{murol} = 0,738 ; $CHI_{AirFrance}$ = 0,957).

Tableau 12. *Répartition des extractions d'éléments modifieurs et associés en fonction du procédé utilisé (pourcentage moyen de phénomènes d'un sens donné)*

Corpus	**modifieurs**		**associés**	
	inversion	**autres procédés**	**inversion**	**autres procédés**
Air France	96,8 %	3,2 %	100 %	0,0 %
Murol	92,9 %	7,1 %	100 %	0,0 %

Il en va de même avec les doubles-marquages (tableau 13b) qui semblent faire indifféremment appel à une reprise par pronom ou à une reprise en *ça* (test du χ^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 (modifieurs 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).

Tableau 13. *Répartition des doubles-marquages et présentatifs*

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 %

L'étude de l'influence du domaine de la tâche sur les extractions en situation de dialogue oral va précisément renforcer ce constat.

5. 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 § 4) 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.

5.1. *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 1 : 26,6 % contre 13,6 % pour le corpus Air France). Un test de Wilcoxon-Mann-Withney 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 (§ 3.3), 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 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. § 3.1), 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[19]. 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 14 donne les fréquences d'apparition des phénomènes d'extraction observés sur les deux sous-corpus ainsi obtenus. Un test de Wilcoxon-Mann-Withney 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.

Tableau 14. *Fréquence d'apparition des extractions sur chaque sous-corpus Murol (nombre moyen de tours de parole présentant au moins une extraction)*

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 %

[19] 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).

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 14).

5.2. *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 détachés. Le tableau 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[20]. On relève toutefois que l'extraction plus fréquente — en termes de répartition relative des procédés — d'associés dans le corpus Murol est statistiquement significative[21] (cf. § 4.2.).

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.

5.3. *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 10). Tout au plus observe-t-on une variation d'usage légèrement significative des procédés d'inversion (test de Wilcoxon-Mann-Withney : $Z = 1{,}897$; $Z(0{,}1) = 1{,}645$; $Z(0{,}05) = 2{,}326$). 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[22] qui sont utilisés pour mettre en œuvre l'extraction d'éléments d'un type donné.

---

[20] Outre les arguments étudiés ci-après, seuls les éléments sujets s'approchent d'un seuil de criticité statistiquement acceptable : test de Wilcoxon-Mann-Withney : $Z = 1{,}526$; $Z(0.1) = 1{,}645$.

[21] Par ailleurs, test de Wilcoxon-Mann-Withney sur une distribution identique des associés entre les corpus Murol et Air France : $Z = 3{,}075$; $Z(0{,}01) = 2{,}576$.

[22] De même, nous n'avons trouvé aucune différence significative de réalisation des présentatifs (structures avec *avoir* / clivages) et des doubles-marquages (reprise par pronom / *ça*).

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.

6. Extractions orales et ingénierie linguistique

Cette étude linguistique de corpus tend à décrire l'extraction comme un phénomène quantitativement important dans le dialogue oral, mais cependant relativement bien encadré quant à ses modes de réalisation. Il apparaît ainsi que la variabilité de l'ordre des mots est un problème qu'on ne saurait ignorer dans la perspective d'une CHM en français parlé robuste mais que l'on peut aborder a priori dans une perspective assez générale (indépendante de la tâche). Nous allons revenir en détail sur les différentes conclusions apportées par cette étude à l'ingénierie linguistique.

6.1. *Traitement des énoncés non projectifs (variabilité forte)*

La question du traitement automatique des énoncés discontinus a été largement étudiée en français comme dans d'autres langues. Au vu des observations réalisées sur ce corpus, et dans une perspective ingénierique[23], 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-projectives ne concernent en effet qu'un nombre très marginal d'énoncés.

Qu'ils reposent sur des modèles stochastiques de langage (Allen 1998 ; De Mori, 1995) ou sur des approches sélectives à base d'ilôts-clefs (Minker *et al*., 1999), 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. La recherche d'un dialogue oral plus riche et plus coopératif nécessitera certainement le développement d'analyseurs linguistiques beaucoup plus fins. Au vu de nos observations, il semble néanmoins que le traitement de la variabilité forte ne saurait être un critère de choix des modèles appelés à être utilisés à l'avenir. Nous travaillons ainsi sur un système de compréhension automatique de la parole (Goulian, 2001) utilisant — après une première étape de segmentation en chunks (Abney, 1991) — une grammaire de liens basée sur une contrainte de planarité incompatible avec le traitement des énoncés non-projectifs (Sleator et Temperley, 1991).

Comme nous l'avons déjà noté (cf. § 4.3), les extractions ne sont pas les seules sources de discontinuité 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.

---

[23] C'est-à-dire guidée par la recherche du taux d'erreur le plus faible et non pas de la meilleure couverture linguistique théorique d'un modèle.

Le traitement de ces procédés oraux doit cependant s'envisager plutôt à l'aide de méthodes spécifiques que dans le cadre général du traitement des discontinuités. Par exemple, des techniques robustes de détection de patterns (Heeman et Allen, 1999 ; Kurdi, 2000) permettent de corriger en pré-analyse une grande majorité de reprises et répétitions orales. Le traitement des incises — qui ne partagent bien souvent aucune dépendance syntaxique avec le reste de l'énoncé — soulève quant à lui des questions qui vont au delà de la « simple » problématique des structures non projectives.

6.2. *Traitement de la variabilité faible*

Paradoxalement, nos observations suggèrent que le traitement de la variabilité faible constituera une problématique de recherche de plus en plus prégnante en CHM orale. Deux éléments semblent appuyer cette prévision.

En premier lieu, les systèmes de dialogue oral ont été jusqu'ici utilisés dans des cadres applicatifs très finalisés, c'est-à-dire qu'ils travaillent généralement sur des énoncés oraux présentant une ambiguïté structurelle très limitée[24]. L'analyse de ces énoncés est donc le plus souvent dirigée par des considérations pragmatiques peu sensibles à l'ordre des mots dans l'énoncé. La généralisation du dialogue oral à des cadres applicatifs plus riches se traduit par une augmentation sensible de cette ambiguïté[25]. Afin d'y faire face, le recours à une analyse plus fine de l'énoncé, prenant en considération sa structure et l'ordre des mots ou des syntagmes dans la phrase, s'impose. Se pose alors la question de la capacité des modèles développés à modéliser et résoudre les problèmes posés par la variabilité faible. On notera à ce sujet que cette variabilité pose déjà parfois problème à des systèmes sélectifs utilisés dans un cadre très finalisé (Minker *et al*., 1999).

En second lieu, les systèmes de dialogue oral actuels se contentent d'extraire peu ou prou 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. Cette information utile est le plus souvent portée par le verbe, le sujet, les arguments voire les modifieurs de l'énoncé. Au contraire, les associés jouent un rôle de complément de phrase qui est rarement utilisé, sauf éventuellement pour l'étude des modalités. Une prise en compte de l'information pragmatique portée par certains de ces associés serait très utile à une meilleure conduite du dialogue. Il se trouve que ces éléments sont, comme nous l'avons vu, les plus sujets à détachement.

6.3. *Influence de la tâche*

Nos observations suggèrent une influence très limitée de la tâche sur les phénomènes d'extraction. Comme nous l'avons noté précédemment, le degré de finalisation de la tâche peut avoir cependant un impact indirect en nécessitant une prise en compte fine de l'ordonnancement linéaire dans le traitement de l'énoncé.

La variabilité d'occurrence des extractions observée entre nos corpus, mais également entre chaque dialogue, donne une idée de l'importance de ce traitement dans la perspective d'une analyse robuste. À l'opposé, la relative stabilité des

[24] Nous ne considérons pas ici l'ambiguïté due aux résultats issus de la reconnaissance de parole.

[25] Voir par exemple l'augmentation très sensible de la perplexité des modèles de langage entre le domaine ATIS (renseignement aérien) et l'application *Broadcast News* (informations générales)

procédés utilisés laisse espérer un certain degré de généralité pour une modélisation des phénomènes d'extraction.

7. Conclusion

Nous avons présenté dans cet article une analyse linguistique de corpus portant sur l'étude de la variabilité de l'ordre des mots en situation de dialogue oral finalisé. Par delà cette étude des phénomènes d'extraction — peu abordés à la fois dans leur dimension interactive et quantitative — nous avons cherché à montrer l'utilité de la linguistique de corpus pour la conduite des recherches en CHM orale. Plusieurs enseignements ont ainsi pu être tirés de cette étude quant au traitement des variabilités fortes et faibles de l'ordre de mots en français parlé spontané.

Ce type 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. Cet investissement se justifie cependant parfaitement par les enseignements qu'il peut apporter. De notre point de vue, la linguistique de corpus représente ainsi un moyen fort efficace de répondre au — relatif — aveuglement dans lequel la CHM orale risque de se fourvoyer à terme.

À l'heure actuelle, nous poursuivons nos études sur l'extraction sur plusieurs nouveaux corpus correspondant à des tâches de renseignement touristique (dialogue homme-homme réel et non plus simulé comme dans le cas du corpus Murol), de renseignement administratif et de réservation hôtelière. Nous entamons par ailleurs des études parallèles portant sur les phénomènes de reprise/répétitions et d'incises.

8. Remerciements

Les auteurs tiennent à remercier Agnès Hamon et Valérie Monbet, du laboratoire de Statistique Appliquée de l'université de BREtagne Sud (SABRES, Vannes) 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. Nous remercions également les relecteurs anonymes de cet article, pour leurs commentaires très éclairants.

9. Bibliographie

[ABN 91] Abney S., " Parsing by chunks ", in Berwick, Abney et Tenny (ed.), *Principle-based parsing*, Kluwer Ac., Amsterdam, 1991.

[ALL 96] Allen J., Miller B., Ringger E.K., Sikorski T. " Robust understanding in a dialogue system ", *34th meeting of the Association for Computational Linguistics*, 1996.

[ALL 98] Allen J., Natural Language Understanding, Benjamins Cummings, 2° édition, 1998, chapitre VIII : Statistical methods.

[ANT 99] Antoine J.-Y., Caelen J., " Pour une évaluation objective, prédictive et générique de la compréhension en CHM orale ", *Langues*, vol. 2, n° 2, p. 130-139, juin 1999.

[ANT 01] Antoine J.-Y., Goulian J., " Word order variations and spoken man-machine dialogue in French : a corpus analysis on the ATIS domain ", *Corpus Linguistics'2001*, UCREL, Lancaster, 2001. PAGES A COMPLETER

[BES 95] Bessac M. et Caelen J., " Analyses pragmatiques, prosodiques et lexicales d'un corpus de dialogue oral homme-machine ", *JADT'95*, Rome, Italie, 1995, p. 363:370.

[BIB 88] Biber D., Variation across speech and writing, Cambridge University Press, Cambridge, 1988.

[BIB 99] Biber D., Johansson S., Leech G., Conrad S., Finegan E., Longman grammar of spoken and written English, Londres, Longman, 1999.

[BIL 99] Bilger M., Blanche-Benveniste C., "Français parlé - oral spontané : quelques réflexions", *Revue Française de Linguistique Appliquée*, vol. 4, n° 2, p. 21-30, 1999.

[BLA 90] Blanche-Benveniste C., Bilger M., Rouget C. et van den Eynde K., Le français parlé : études grammaticales, CNRS Editions, Paris, 1990.

[BLA 97] Blanche-Benveniste C., Approches de la langue parlée en français, Coll. *L'essentiel Français*, Ophrys, Paris, 1997.

[CAE 97] Caelen J. *et al*., " Les corpus pour l'évaluation du dialogue homme-machine ", *JST-FRANCIL'97*, Avignon, 1997, p. 215-222.

[CHE 98] Cheyer A., Julia L., Martin J.C., " A unified framework for constructing multimodal experiments and applications ", *CMC'98*, in Bunt, H., Beun, R.J. & Borghuis, T. (Eds.). *Lecture notes in Artificial Intelligence*. http://www.limsi.fr/Individu/martin/publications/download/cmc98-1.ps

[COV 90] Covington M., " Parsing discontinuous constituents in dependency grammar ", *Computational Linguistics*, vol. 16, n° 4,1990, p. 234-236.

[CUN 99] Cunningham H. (1999), A definition and short history of Language Engineering, *Natural Language Engineering*, 5(1), pp. 1-16.

[DEM 95] De Mori R., "Modèles stochastiques de langage ", *Ecole d'été « Fondements et perspectives en traitement automatique de la parole »*, Marseille, 1995, p. 109-118.

[DUD 88] Dudewicz E. J., Mishra S. N., Modern mathematical statistics, *Wiley series in probability and mathematical statistics*, John Wiley & Sons, New-York, 1988.

[DYB 00] Dybkjaer L., Bernsen N.O.. "Usability issues in spoken dialogue systems", *Natural Language Engineering*, 2000, vol 6, n° 3-4, p. 243-272.

[FRA 97] Fraser N., "Assessment of interactive systems", in Gibbon D., Moore R., Winski R. (eds.), "EAGLES Handbook of standards and ressources for spoken language systems", chap. III. 1997.

[GAD 89] Gadet F., Le français ordinaire, Colin, Paris, 1989.

[GAD 92] Gadet F., Le français populaire, PUF, Paris, 1992.

[GOOS 93] Goosse A. (ed.), Le Bon usage, 13° édition, De Boeck Université / Duculot, Paris, 1993.

[GOU 01] Goulian J., "Compréhension automatique de la parole combinant syntaxe locale et sémantique globale pour une CHM portant sur des tâches relativement complexes", *TALN'2001*. Tours, 2001 (à paraître).

[GUS 00] Gustafson J., Bell L. "Speech technology on trial : experience from the August system", *Natural Language Engineering*, 2000, vol 6, n° 3-4, p. 273-286.

[HEE 99] Heeman P.A. et Allen J.F., "Speech repairs, intonational phrases and discourse markers : modeling speakers's utterances in spoken dialogues", *Computational Linguistics*, 1999, vol 25, n° 4, p. 527-572.

[HIR 98] Hirschman L., "Language understanding evaluations : lessons learned from MUC and ATIS", *LREC'98*. Grenade, Espagne, 1998, p. 117-122.

[HOL 00] Holan T., Kubon, Oliva K., Plátek M., "On complexity of word order", *TAL*., vol. 41, n° 1, 2000, p. 273-300, Hermès, Paris.

[HUD 00] Hudson R., "Discontinuity", *TAL*. vol. 41, n° 1, 2000, p. 15-56, Hermès, Paris.

[KAH 00] Kahane S., "Extractions dans une grammaire de dépendance lexicalisée à bulles", *TAL*. vol. 41, n° 1, 2000, p. 211-244, Hermès, Paris.

[KER 90] Kerbrat-Orecchioni C., Les interactions verbales, vol. 1, Colin, Paris, 1990.

[KER 99] Kerbrat-Orecchioni C.., "L'oral dans l'interaction : une liberté surveillée", *Revue Française de Linguistique Appliquée*, vol. 4, n° 2, p. 41-55, 1999.

[KUR 00] Kurdi M.Z.. A semantic based approach for spontaneous spoken dialogue understanding, *NLP'2000*, Patras, Grèce, 2000.

[LOK 98] Lokbani M. N. & White S., " La reconnaissance de la parole ", *La Recherche*, n° 319, 1998, p. 82.

[MIN 99] Minker W., Waibel A., Mariani J., *Stochastically based semantic analysis*, Kluwer ac., Amsterdam, 1999.

[MOR 89] Morel M.-A. *et al*., Analyse linguistique de corpus, Publications de la Sorbonne Nouvelle, Paris, 1989.

[OS 99] den Os E., Boves L., Lamel L., Baggia P., " Overview of the ARISE project " , *Eurospeech'99*, Budapest, Hongrie, Septembre 1999, p. 1527-1530.

[PAL 94] Pallett D.S, *et al*., " Benchmark tests for the ARPA Spoken Language Program " *ARPA Workshop on Spoken Language Technology*, 1994, p. 5-36.

[POL 94] Pollard C., Sag I., Head-driven Phrase Structure Grammar, University of Chicago Press, Chicago, 1994.

[RAM 94] Rambow O., Joshi A. , " A formal look at dependency grammars and phrase-structure grammars with special considerations of word-order phenomena ", in Wanner L. (ed.), *Current issues in Meaning-Text Theory*, Pinter, Londres,1994.

[SAB 94] Sabah G., " Projet DALI ", *rapport d'activité GDR-PRC Communication Homme-Machine*, p. 71-88, 1994.

[SLE 91] Sleator D. D. K., Temperley D., " Parsing English with a link grammar " , *rapport de recherche CMU-CS-91-196*, School of Computer Science, Carnegie Mellon University, Pittsburgh, 1991.

[TES 59] Tesnière L., Eléments de syntaxe structurale, Klincksiek, Paris, 1959.