

**UNIVERSITÉ JOSEPH FOURIER - GRENOBLE I  
SCIENCES & GÉOGRAPHIE**

N° attribué par la bibliothèque

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

**THÈSE**

pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ JOSEPH FOURIER - GRENOBLE I**

**Discipline : Informatique Système et Communication**

présentée et soutenue publiquement

par

***Yannick FOUQUET***

le 26 octobre 2004

# **Modélisation des attentes en dialogue oral**

---

**JURY**

M. Christian Boitet	<i>Président</i>
Mme. Anne Nicolle	<i>Rapporteur</i>
M. Andreas Herzig	<i>Rapporteur</i>
M. Denis Vernant	<i>Examineur</i>
M. Jean Caelen	<i>Directeur</i>



## Remerciements

*Je tiens à remercier ici tous ceux qui, de près ou de loin, par leurs conseils, leurs encouragements ou simplement leur présence, ont contribué à l'élaboration, l'enrichissement et l'amélioration de ce manuscrit :*

*M. Jean Caelen pour son accueil, ses conseils, ses critiques, ses encouragements, sa confiance en mes travaux et l'indépendance qu'il a su me laisser dans leur réalisation.*

*M. Christian Boitet et Mme. Anne Nicolle pour leurs remarques qui m'ont permis d'améliorer ce manuscrit, et pour avoir accepté de participer à mon jury à titre respectivement de président et rapporteur.*

*M. Andréas Herzig et M. Denis Vernant pour avoir accepté respectivement de rapporter cette thèse et d'en être examinateur.*

*Le laboratoire CLIPS dont les équipes GEOD et MULTICOM avec qui j'ai particulièrement apprécié de travailler ou simplement converser dans une ambiance chaleureuse, en particulier Anne-Claire, Solange, Jean-François, Laurent, Audrey, Daniel, Anne, Richard, Doms, Eric, Brigitte B., Brigitte M., Caroline, Francis, Camille, et tous les autres membres de ces deux équipes pour leur aide précieuse et/ou leurs discussions (transcriptions, annotations, analyses, relectures, magiciens d'Oz, enregistrements vidéos, soutien moral, discussions diverses, gâteaux, etc.) mais aussi Michel, Bernard, Valérie, Nicole, Annie, et tous ceux avec qui j'ai pu partager des moments enrichissants et conviviaux.*

*Je tiens à remercier également tous ceux qui ont dû me supporter ou m'encourager et m'enrichir durant toute cette thèse et même avant : mes amis, ma famille, et particulièrement Révol qui n'aura malheureusement pas pu assister à la soutenance de son filleul, et mes parents pour m'avoir incité à « faire des études ».*

*Quant à Marjolaine...je ne le lui dirai jamais assez.*



## Table des matières

Introduction	13
Problématique	19
1. Pourquoi l'oral ?	20
2. Pourquoi des dialogues finalisés ?	22
3. Attentes du point de vue du locuteur	24
4. Méthodologie de recherche : démarche incrémentale	25
<b>A. Contexte pratique et théorique</b>	<b>29</b>
Chapitre I : contexte pratique d'étude	31
1. Communication humaine : ERIM et l'aide au traducteur	31
2. Communication humaine médiatisée : NESPOLE! et l'annotation en actes	32
3. Communication Homme-Machine : PVE et le modèle générique de dialogue	35
Chapitre II : Cadre théorique en modélisation du dialogue oral	39
1. Philosophie analytique et acte du discours	42
1.1. Théorie descriptive d'Austin	43
1.2. Théorie explicative de Searle	45
1.3. Théorie illocutoire de Vanderveken (formalisation des précédentes)	47
1.4. Une taxonomie des actes de langage	50
1.5. Une direction d'ajustement	52
1.6. Limites	54
2. Ethnométhodologie	55
3. Psychologie cognitive - Cognisciences	57
3.1. La coopérativité selon Grice	59
3.2. La pertinence selon Sperber & Wilson	62
3.3. Une formalisation : DRT & SDRT	65
4. Linguistique	68
Conclusion : ce qui est utile aux attentes	71

## **B. Etude théorique** \_\_\_\_\_ **73**

### Chapitre I : Modélisation et traitement des attentes en dialogue – aspect théorique \_ 75

1. les attentes en dialogue : notions et modélisation _____	76
1.1. La notion d'attentes en dialogue _____	76
1.2. Modélisation des attentes en dialogue _____	79
1.3. Les attentes du locuteur _____	81
2. Spécification des attentes _____	83
2.1. Une notation en acte de dialogue issue des actes de langage _____	83
2.2. Spécification des attentes en dialogue _____	89
2.3. Attentes et psycholinguistique _____	89
3. Application _____	94
3.1. Combinaison des 3 approches _____	94
3.2. Mise en œuvre _____	99
Conclusions théoriques _____	104

### Chapitre II : Approche à base de corpus \_\_\_\_\_ 105

1. Corpus-pilote homme-homme pour Homme-Homme : Nespole! _____	107
1.1. Protocole Nespole! _____	107
1.2. Description du corpus NESPOLE! _____	109
2. Corpus-pilote homme-homme pour Homme-Machine : PVE _____	110
2.1. Protocole PVE _____	110
2.2. Description du corpus-pilote PVE _____	112
3. Corpus-simulé homme-machine version 0 : Pré-tests PVE _____	115
3.1. Protocole corpus-simulé PVE version 0, pré-tests _____	115
3.2. Description du corpus-simulé PVE version 0, prétests _____	116
4. Corpus simulé homme-machine version 1 : TestsPVE _____	119
4.1. Protocole TestsPVE _____	119
4.2. Corpus simulé PVE version 1, tests _____	120
Conclusion sur les corpus _____	124

<b>C. Expérimentation et tests</b>	<b>125</b>
Chapitre I : Plate-forme d'expérimentation (Magicien d'oz)	127
1. Magicien d'Oz	128
1.1. Magicien d'Oz pour du dialogue	128
1.2. Magicien d'Oz pour du dialogue oral	129
2. Architecture générale du système	130
3. Reconnaissance de parole, un modèle de langage spécifique.	132
3.1. Corpus & vocabulaire pour l'apprentissage d'un modèle de langage.	133
3.2. Apprentissage d'un modèle de langage.	134
3.3. Evaluation des modèles de langage.	136
3.4. Au niveau de l'ergonomie de l'interface	137
4. Synthèse de parole française	138
5. Magicien d'Oz et collecte (aide à l'annotation)	139
6. Magicien d'Oz et contrôleur de dialogue	141
6.1. Plate-forme mettant en jeu un magicien d'Oz	141
6.2. Composants de la plate-forme et choix ergonomiques associés	143
Conclusion sur le système	146
Chapitre II : Modélisation stochastique	147
1. L'importance des rôles en dialogue oral finalisé	148
2. Méthodologie pour la prédiction des actes vs attentes	149
2.1. Prédiction des actes	149
2.2. Génération et gestion des attentes	151
3. Description statistique des corpus	154
3.1. Corpus de dialogues homme-homme NESPOLE!	154
3.2. Corpus de dialogues homme-machine PVE	155
4. Tests en dialogue homme-homme : Sur le corpus NESPOLE!	156
4.1. Prédiction d'actes	156
4.2. Génération et gestion des attentes	158
4.3. Test : Comparaison des deux approches	159
Conclusion en dialogue homme-homme	161
5. Tests en dialogue homme-machine : Dans le corpus PVE	162

5.1. Statistiques de succession	162
5.2. Attentes	163
5.3. Comparaison, taux de bonne prédiction	165
Conclusion en dialogue home-machine	167
Conclusion d'une modélisation stochastique	168
<b>Conclusions et perspectives</b>	<b>171</b>
Limites	173
Perspectives (mise en œuvre)	174
Perspectives (modèle)	176
Perspectives (communication non verbale)	177
Compréhension	178
Attendus	178
Attentes	178
Bilan général	180
<b>Références bibliographiques</b>	<b>181</b>
Références non citées	190
<b>Annexes</b>	<b>193</b>
Statistique de cooccurrence des bi-grammes {Acte, Acte-suivant}	195
Statistiques sur les attentes	197
Analyse des résultats d'attentes acte par acte pour les différents corpus	198
1. Actes posant des attentes	199
2. Actes répondant à des attentes	203
Instructions à suivre pour le système PVE	207
Système de Reconnaissance + Synthèse	209
Interface « Client »	210
Dialogue complet vu de l'annotateur	210
Enoncés proposés au contrôleur	211
Liste des mots phonétisés et classes	214
Quatrième de couverture (résumé, summary)	220



## Liste des figures

<b>Figure 1</b> : Architecture générale en dialogue homme-machine [Caelen 92] _____	23
<b>Figure 2</b> : Méthodologie incrémentale suivie en dialogue homme-homme _____	25
<b>Figure 3</b> : Méthodologie incrémentale suivie en dialogue homme-machine _____	26
<b>Figure 4</b> : Architecture de traduction de parole par langage pivot ([Besacier 01]) _____	34
<b>Figure 5</b> : Architecture du système de dialogue dans le projet PVE (en grisé, les données spécifiques à la tâche) _____	36
<b>Figure 6</b> : Modèle du code (vue simplifiée) _____	40
<b>Figure 7</b> : Modèle du code _____	40
<b>Figure 8</b> : Modèle projectif du dialogue informatif [Vernant 92] _____	54
<b>Figure 9</b> : Modèle de communication humaine, [Kerbrat-Orrechioni 80] _____	58
<b>Figure 10</b> : Exemple d'analyse conversationnelle, modèle genevois [Lehuen 97]. _____	69
<b>Figure 11</b> : Exemple de dialogue géré par le logiciel MINIDIAL [Lehuen 97] _____	70
<b>Figure 12</b> : Attentes et linguistique : écart incident à l'axe régissant _____	72
<b>Figure 13</b> : Réussite (selon conditions) et satisfaction (selon réponse) d'un acte. _____	76
<b>Figure 14</b> : Possibilités de réponses à un interacté, inspiré de [Vernant 97b]. _____	80
<b>Figure 15</b> : Actes orientés vers soi ou vers autrui _____	87
<b>Figure 16</b> : Attentes vues comme une liste de réponses possibles _____	89
<b>Figure 17</b> : Ebauche du modèle des attentes _____	94
<b>Figure 18</b> : Les probabilités comme écart d'incidence _____	103
<b>Figure 19</b> : Architecture du système de dialogue. _____	130
<b>Figure 20</b> : Architecture détaillée, système de dialogue à base de magiciens d'Oz _____	131
<b>Figure 21</b> : Architecture du système de reconnaissance vocale _____	132
<b>Figure 22</b> : Du signal de parole aux actes (et attentes) _____	139
<b>Figure 23</b> : Interface du Magicien d'Oz pour l'annotation en acte de dialogue _____	140
<b>Figure 24</b> : Interface du Magicien d'Oz pour le contrôleur de dialogue _____	142
<b>Figure 25</b> : Comparaison de divers <i>corpus</i> pour les attentes après une demande d'information de l'agent _____	168
<b>Figure 26</b> : Comparaison de divers <i>corpus</i> pour les attentes après une demande d'information du client _____	169
<b>Figure 27</b> : Architecture du système - serveurs opportunistes _____	174



## Liste des tableaux

<b>Tableau 1</b> : Deux suites possibles à un même énoncé _____	32
<b>Tableau 2</b> : Modalités de Greimas _____	41
<b>Tableau 3</b> : Distinguo sens littéral / sens en contexte _____	46
<b>Tableau 4</b> : Conditions de succès / satisfaction _____	49
<b>Tableau 5</b> : Modalités de Greimas pour nos attentes _____	71
<b>Tableau 6</b> : Classification des actes de [Vilnat 97]. _____	84
<b>Tableau 7</b> : Classification des actes de dialogue de [Ozkan 94] _____	85
<b>Tableau 8</b> : Classification des actes de Caelen en regard de Searle _____	85
<b>Tableau 9</b> : Notre taxonomie des actes de dialogue avec des exemples _____	86
<b>Tableau 10</b> : Synthèse de notre taxonomie des actes de dialogue _____	87
<b>Tableau 11</b> : Notre taxonomie des actes de langage comparée à l'état de l'art ____	88
<b>Tableau 12</b> : Réactions de l'allocutaire face à un énoncé _____	92
<b>Tableau 13</b> : Exemple de dialogue avec gestion des attentes _____	101
<b>Tableau 14</b> : Répartition des 4633 actes du corpus Nespole! en fonction des rôles. _____	109
<b>Tableau 15</b> : Exemple d'annotation, corpus pilote pour du dialogue humain (Nespole!) _____	109
<b>Tableau 16</b> : Variantes et paramètres pour le magicien d'Oz (PVE) _____	111
<b>Tableau 17</b> : Répartition des 1138 actes du corpus-pilote PVE en fonction des tâches et du rôle. _____	114
<b>Tableau 18</b> : Répartition des 361 actes du corpus de prétests PVE en fonction des tâches et du rôle. _____	118
<b>Tableau 19</b> : Les six tâches retenues dans l'expérimentation en magicien d'Oz (PVE) _____	119
<b>Tableau 20</b> : Répartition des 1626 actes du corpus de tests PVE en fonction des tâches et rôles. _____	123
<b>Tableau 21</b> : Description du corpus PVE en fonction des tâches. _____	124
<b>Tableau 22</b> : Evolution de la plate-forme magicien d'Oz _____	141

<b>Tableau 23</b> : Liste des actes utilisés pour les tests _____	149
<b>Tableau 24</b> : Taux de prédiction d'acte par la méthode statistique dans PVE (un acte par tour) _____	150
<b>Tableau 25</b> : Répartition par rôle des 4454 actes du corpus d'apprentissage de Nespole!. _____	154
<b>Tableau 26</b> : Répartition par rôle des 74 actes du corpus de pré-tests de Nespole!. _____	154
<b>Tableau 27</b> : Répartition par rôle des 189 actes du corpus de test de Nespole!. _____	154
<b>Tableau 28</b> : Répartition en tâches et rôles des 1544 actes du corpus d'apprentissage stochastique PVE. _____	155
<b>Tableau 29</b> : Répartition en tâches et rôles des 443 actes du corpus de test stochastique PVE. _____	155
<b>Tableau 30</b> : Fréquence d'apparition des actes de l'agent dans les 30 dialogues _____	156
<b>Tableau 31</b> : Fréquence d'apparition des actes du client dans les 30 dialogues _____	156
<b>Tableau 32</b> : Pourcentages d'actes suivant par rapport à chaque acte (Nespole!) _____	157
<b>Tableau 33</b> : Pourcentages d'attentes par rapport à chaque acte posant un but (Nespole!) _____	158
<b>Tableau 34</b> : Matrice de cooccurrence des bigrammes {Acte-courant, Acte-suivant} dans le 31ième dialogue de notre corpus (Nespole!) _____	159
<b>Tableau 35</b> : Taux (%) de prédictions d'acte et de prédiction d'attentes et nombre à prédire (Nespole !). _____	160
<b>Tableau 36</b> : Pourcentages d'actes suivants par rapport à chaque acte (PVE) _____	163
<b>Tableau 37</b> : Pourcentages d'attente par rapport à chaque acte (PVE) _____	164
<b>Tableau 38</b> : Statistiques de bigrammes (%) et attentes en fonction de chaque acte (PVE) _____	165
<b>Tableau 39</b> : Description du corpus _____	166
<b>Tableau 40</b> : Résultats de prédiction d'acte versus attentes (PVE) _____	166
<b>Tableau 41</b> : Taux (%) de prédictions d'acte et de prédiction d'attentes et nombre à prédire. _____	167

## **Introduction**

La machine ne pourra pas remplacer l'homme dans toutes les situations. Comme le disent Nicolle et Luzzati : « *il s'agit de prendre les machines pour ce qu'elles sont, sans faire d'anthropomorphisme, car chacun préfère dialoguer avec des humains pour ce qui est de la conversation courante. Mais dans beaucoup de domaines, les machines jouent des rôles que les humains ne peuvent pas jouer : elles ne s'ennuient jamais quand elles répètent la même chose, elles ont une mémoire sûre et sans limite pratique, elles calculent plus sûrement, elles font moins d'erreurs. Or les limites à leurs usages, les appréhensions et difficultés des usagers, tiennent souvent aux difficultés de l'interaction.* » [Nicolle & Luzzati 99]

Le dialogue a un certain nombre de spécificités. Sur l'aspect linguistique, on peut constater que limiter le domaine d'application ne réduit pas les phénomènes linguistiques à traiter, que ce soit pour les caractéristiques de la langue ou les traitements de l'implicite inhérent à tout dialogue. S'il a souvent été dit qu'il était possible de restreindre fortement l'ensemble des phénomènes langagiers dans des dialogues orientés par la tâche, cette affirmation est loin d'être confirmée, comme le note [Sabah 97]. Diverses caractéristiques linguistiques sont incontournables pour que la communication soit réellement naturelle. Plutôt qu'avoir une grande couverture de la langue (au niveau du vocabulaire ou de la syntaxe), le système de dialogue homme-machine devrait traiter les aspects inévitables du langage comme les anaphores, les ellipses, les mécanismes de référenciation et comprendre deux énoncés identiques apparaissant dans des contextes distincts. Pour ce dernier cas, l'interprétation contextuelle doit pouvoir faire des inférences (pour tenter de soulever les implicites) et utiliser le contexte (pour tenter de trouver le sens réel qui peut être différent du sens littéral).

Contrairement à ce qui est couramment admis pour la compréhension de textes, le traitement de dialogue nécessite une souplesse et une tolérance importante aux inattendus, comportements hors des normes langagières classiques. En effet, outre les fautes de frappe et d'orthographe pour l'écrit, ou de diction pour l'oral, le dialogue fait apparaître de nombreux usages de formes syntaxiques erronées ou imprévues. On y retrouve les reprises, hésitations ou autocorrections, mais aussi les structures elliptiques qui ne peuvent s'interpréter qu'en fonction du contexte des interventions précédentes. La syntaxe des énoncés oraux diffère de celle de l'écrit. De nombreuses structures considérées comme asyntaxiques à l'écrit sont possibles à l'oral. En particulier, un énoncé oral peut être incomplet. Autoriser une part d'implicite dans ce que dit l'interlocuteur impose de savoir le traiter en dégageant le sujet général de la discussion et déduisant de ces informations le but et le plan éventuel de l'interlocuteur. En outre, il doit pouvoir réagir en fonction du comportement langagier de l'interlocuteur. La compréhension et la gestion de dialogue doivent donc être dynamiques afin de toujours garder le contrôle et répondre correctement.

Le dialogue oral impose une bonne gestion du canal de communication. En particulier, les aspects métadialogiques entraînent des sous-dialogues de répétition, de demande de confirmation, de traitement de contestation de l'interlocuteur sur les résultats de reconnaissance vocale ou de la compréhension ou même de la synthèse, de mise en attente, de maintien de dialogue et de relance dès que le traitement d'une

intervention devient trop long ou lorsque l'utilisateur ne répond pas à une sollicitation du dialogue.

Traiter le dialogue oral, humain ou homme-machine nécessite de bien entendre et reconnaître ce qui a été prononcé, mais une simple reconnaissance n'est qu'une brique que l'on peut éventuellement remplacer par l'écrit par exemple. Outre les aspects propres à l'oral, le dialogue nécessite une bonne compréhension des énoncés et une bonne interprétation dans le contexte d'élocution. Cette interprétation peut permettre ensuite une bonne gestion de l'interaction.

L'**interprétation contextuelle** intervient après le niveau de compréhension textuelle. Elle tente de représenter d'un point de vue pragmatique le schéma sémantique de l'énoncé en fonction de l'état de la tâche ou du dialogue. Il est alors nécessaire non seulement de déterminer précisément les références aux actions et aux objets spécifiés dans l'énoncé, mais aussi de dégager les intentions du locuteur et de distinguer les actes de langage qui portent sur la tâche (question ou réponse à propos de la tâche) ou sur le dialogue (demande de reformulation, accusé de réception). Un bon typage des actes est important mais non suffisant. « *Un acte de langage indirect est l'acte de langage produit (indirectement) par le biais de l'accomplissement d'un autre acte de langage (« direct » par opposition au premier).* » [Herzig & Longin 02] Un système incapable de reconnaître un acte indirect, celui porteur du sens véritablement voulu ne pourra pas interpréter correctement un énoncé. Par exemple, un énoncé tel que « Peux-tu ouvrir ce fichier ? » doit s'interpréter comme une demande d'ouverture d'un fichier x et non comme une question sur la possibilité d'une telle ouverture. L'interprétation devra donc découvrir l'acte indirect de l'énoncé et déterminer les référents afin que la gestion de l'interaction puisse se faire au mieux. Cela fera appel à la sémiotique et en particulier à la pragmatique.

La sémiotique est communément (Morris, 1938) divisée en trois parties :

- La **syntaxe** traite des relations qui existent entre les signes en vertu de leur forme apparente,
- La **sémantique** traite des relations entre les signes et leur signification
- La **pragmatique** traite des relations entre les signes, leur signification et l'usage qu'on en fait en parlant.

Un système d'interprétation doit donc posséder des connaissances nombreuses : des connaissances sémantiques, syntaxiques et lexicales mais aussi des connaissances statiques sur la tâche et l'application, des connaissances sur la structure du dialogue et des connaissances sur les interlocuteurs.

Suite à l'interprétation contextuelle de l'énoncé, il faut déterminer les actions à effectuer pour la tâche et celles qui concernent l'avancement du dialogue. Le **contrôle du dialogue** tente de répondre au mieux aux énoncés de l'utilisateur. Il est donc important pour le système de déterminer le but que cherche à atteindre l'utilisateur, c'est-à-dire la tâche qu'il souhaite réaliser. Pour une bonne gestion de l'interaction, il peut être important de bien reconnaître les plans du locuteur mais aussi de pouvoir faire des raisonnements valides. Mais le choix de la réponse la plus appropriée ne peut se contenter de cela. Il doit tenir compte d'une stratégie à employer. En particulier, pour améliorer l'efficacité du dialogue, la machine a intérêt à adopter une stratégie de coopération le plus souvent possible. La coopération consiste à répondre à l'acte indirect du locuteur. Par exemple, avec un énoncé tel que « Peux-tu ouvrir ce fichier ? », répondre « oui » répond à l'acte direct qui est ici une question sur la possibilité d'ouverture d'un fichier. Ce « oui » est donc non coopératif. L'acte indirect ici est une demande d'action d'ouverture du fichier. Effectuer cette commande d'ouverture est donc coopératif. En revanche, effectuer cette commande nécessite de déterminer le référent effectif de « ce fichier » et de l'action d' « ouverture de fichier ».

La tendance générale est de fonder la gestion de l'ensemble du dialogue sur un processus de planification unique, s'appuyant soit sur un automate précompilé de dialogue (modèle statique comme Halpin [Rouillard 00]) soit sur un ensemble de règles conduisant à un modèle dynamique. Il existe plusieurs niveaux de planification. Le premier niveau concerne le métadialogue incluant les phases d'ouverture et de clôture du dialogue. Le deuxième niveau peut s'appuyer sur une gestion plutôt opportuniste. Une telle gestion ne rejette pas les connaissances que doit posséder un système de contrôle performant. Ces connaissances sont de deux types. Les connaissances statiques, telles que le modèle de la tâche ou de l'application et le modèle de dialogue, sont nécessaires pour déterminer les buts possibles et les moyens possibles de les atteindre. Les connaissances dynamiques, telles que l'état de la tâche ou l'historique du dialogue, permettent de traiter les ellipses, les références, etc.



La présente thèse propose une étude des attentes du locuteur sous diverses faces. Le document est organisé en six chapitres.

A la suite de la définition de notre problématique, le premier chapitre expose le contexte de notre étude qui se répartit en trois projets du laboratoire. Ces projets concernent le dialogue humain, le dialogue humain médiatisé par la machine et le dialogue homme machine.

Le deuxième chapitre présente un état de l'art dans le domaine du dialogue. Sa particularité est de puiser dans quatre disciplines les fondements de notre modèle. Nous aborderons donc la philosophie du langage, l'ethnométhodologie, la psychologie cognitive et la linguistique.

Le chapitre III propose une modélisation des attentes issue des différentes disciplines abordées dans le chapitre II. Nous commencerons par leur notation ainsi que leur justification sur le plan psycho-linguistique. Nous les estimerons et aborderons l'intégration de probabilités.

La quatrième chapitre présente l'approche à base corpus utilisée dans le cadre de notre étude. Nous y décrirons donc les différents protocoles employés ainsi que les corpus obtenus, transcrits et analysés.

Le chapitre suivant présente l'expérimentation qui met en jeu une plate-forme de simulation en magicien d'Oz. L'architecture de cette plate-forme sera détaillée, mettant en relief les aspects oraux du système et les aspects de simulation ainsi que les choix techniques et ergonomiques qui ont été faits.

La chapitre six présente alors l'évaluation suite à l'expérimentation. Il propose une modélisation stochastique qu'il valide dans le cadre du dialogue humain et dans celui du dialogue homme-machine. Les résultats seront alors discutés afin de conclure sur l'intérêt de notre approche.



## **Problématique**

Une problématique intéressante en Dialogue Homme-Machine est d'obtenir un degré de généralité élevé afin d'adapter le système aisément au profil de l'utilisateur et à de nouvelles tâches. Ainsi nous voyons la généralité dans la modélisation du dialogue comme une relative indépendance à l'utilisateur d'une part et surtout à la tâche d'autre part. Il s'agit donc dans un premier temps de dégager les composantes indépendantes et celles liées à la tâche. Nous verrons ici pourquoi nous restreindre au dialogue oral finalisé dans lequel un participant cherche à atteindre un but : réaliser une tâche. Nous aborderons alors les attentes vues du côté du locuteur puis notre méthodologie.

## **1. Pourquoi l'oral ?**

La parole est aujourd'hui étudiée pour différents domaines tels que la reconnaissance vocale, la communication homme-machine ou la communication humaine médiatisée par la machine à travers les projets que nous décrirons plus loin. Mais nombre de systèmes, surtout de reconnaissance vocale, sont créés à partir de textes. C'est le cas notamment des systèmes commerciaux de dictée vocale actuels, qui utilisent des centaines d'heures de parole pour apprendre à reconnaître des phonèmes, mais ne se fondent que sur l'étude de textes écrits issus des journaux *Le Monde* pour apprendre à reconnaître les phrases prononcées. Or, nous savons que l'individu ne parle pas comme il écrit. L'étude de textes semble donc inadaptée pour l'élaboration de systèmes de dialogue oral. Une originalité du CLIPS est de se pencher sur l'étude de dialogues écrits à travers les news où le langage se rapproche du dialogue oral. Mais la meilleure approche reste celle de dialogue oral en situation, afin de bien simuler la situation future. Dans le cadre de la traduction, nous pouvons préciser que la traduction de parole ne représente pas les mêmes problèmes que la traduction de textes. En effet, la parole dépend beaucoup plus de la situation. Elle a un aspect social en plus du contenu purement linguistique. Elle offre des phénomènes langagiers spécifiques et dynamiques et respecte moins les caractéristiques formelles que l'écrit. Le texte offre, lui, une formation plus régulière avec notamment les ponctuations que l'on ne retrouve pas expressément à l'oral.

Par ailleurs, les contraintes dialogales sont différentes de la simple dictée vocale. L'émission de parole est mal formée avec les hésitations (« um », « hmm », etc.), les répétitions (« Alors je, je voudrais, ce que je voudrais »), les faux départs (« je vous propose de nous rencontrer mar, heu, mercredi »). De plus, les énoncés, mal construits, contiennent plusieurs idées (« non, mardi ce n'est pas possible pour moi, pourquoi pas, mercredi matin, mercredi le 12 »).

Enfin, contrairement aux systèmes de dictée vocale qui préconisent une diction lente dans un environnement calme, un système de dialogue oral doit pouvoir gérer une reconnaissance vocale imparfaite, parfois perturbée par une élocution rapide, souvent perturbée par un environnement bruyant (la toux, le rire, la sonnerie du téléphone, la fermeture d'une porte, etc.). L'étude d'un dialogue oral nous permettra de modéliser les particularités du dialogue d'une part et les particularités de l'oral d'autre part. Notre originalité tiendra ici dans le fait que nous étudierons aussi le dialogue dans une langue non maternelle. En effet, les recherches actuelles portent essentiellement sur l'étude de dialogues en langue maternelle. Si celle-ci montre déjà les particularités liées au dialogue (structure, phénomènes susdits), elle

ne montre, en revanche, pas les phénomènes particuliers liés à une mauvaise maîtrise de la langue. Les phénomènes liés au dialogue se retrouveront amplifiés. Et se rajouteront des maladroites qui pourront plus ou moins entraver la compréhension du dialogue. C'est ainsi que, dans notre corpus, nous avons pu remarquer une digression du but original afin de résoudre une incompréhension sur un terme (exemple ci-dessous). Mais d'autres phénomènes occasionneront moins de gêne et seront ignorés de l'interlocuteur qui aura compris l'idée ou jugera ce manque d'information comme peu pertinent. Ce sont les fautes typiques d'accord (« combien *des* étoiles »), de conjugaison (« il faut que vous *appelez* »), de syntaxe (« pouvez-vous lire *bien* ? »), d'élision (« entre deux et trois étoiles »), de mot en trop (« il y a *des* différents niveaux de pistes ») ou de mauvais terme (« vous venez *par* voiture ? ») ou de termes linguistiquement possibles mais qui ne se disent pas (« le numéro *téléphonique* »).

Pour ces fautes-ci, l'interlocuteur a tendance à corriger la faute dans sa représentation de l'énoncé.

Mais si le problème est jugé plus important, un sous-dialogue s'instaurera afin de résoudre ce problème. C'est le cas du sous-dialogue suivant :

« C: j'aurais voulu savoir le/ les arrhes qu'il fallait verser avant  
 A: vous désirez savoir s'il vous plaît ? le ? j'ai pas compris, excusez-moi/  
 C: le/ les arrhes, le e l/ l'avance qu'il faut faire  
 A: ah l'agence ? e(m)  
 C: l'avance pardon  
 A: ah l'avance ok, »

Dans cet échange, l'agent semble ne pas avoir compris le terme arrhes. Le client sait, au vu des différents échanges précédents, que l'agent ne maîtrise pas le français. Il commence malgré tout à se répéter comme dans le cas d'une incompréhension liée à l'audition. L'agent précisant qu'il n'a pas compris (et non pas mal entendu), le client suppose une incompréhension sur le terme « arrhes » et cherche à le décrire brièvement. S'ensuit une deuxième incompréhension mais celle-ci uniquement auditive au vu de la réponse de l'agent qui répète ce qu'il a compris. La construction du sens se termine par l'approbation de l'agent sur la représentation du sens du terme.

Cet exemple illustre également l'importance du contexte dans le dialogue.

## 2. Pourquoi des dialogues finalisés ?

L'objectif principal est la mise en œuvre de systèmes qui analysent et interprètent les énoncés humains (en langue naturelle), afin d'aider l'utilisateur du système à atteindre son but (c'est-à-dire la réalisation d'une tâche ou la résolution d'un problème en utilisant une langue naturelle). Le but n'est pas de réaliser des systèmes capables de participer à une conversation quelconque. Nous parlerons donc de systèmes de dialogue finalisé (task oriented). Une première définition de dialogue finalisé a été donnée par [Borillo 93]. C'est pour lui un « *échange communicationnel, effectué à distance et visant un objectif – demande de renseignement, résolution de problème – objectif présent dans la conscience des deux protagonistes, même si ceux-ci ne prennent pas une part égale dans l'échange et n'y apportent pas une égale compétence ou un même niveau de connaissance* ». La machine a un rôle collaboratif (en instaurant une relation « opérateur-tâche »).

Le langage humain est très complexe. Malgré son utilisation constante il est ardu de le « maîtriser » réellement. Le dialogue est à la fois « *une des capacités qui nous est les plus familières* » et « *quelque chose que l'on ne maîtrise pas* » vraiment (au niveau lexical, syntaxique, sémantique et pragmatique) [Bilange 92], [Carré et al. 91] C'est pour cette raison qu'une minorité de systèmes utilisent la parole comme moyen de communication, car il subsiste encore des problèmes pour l'implémentation des données linguistiques (entre autres).

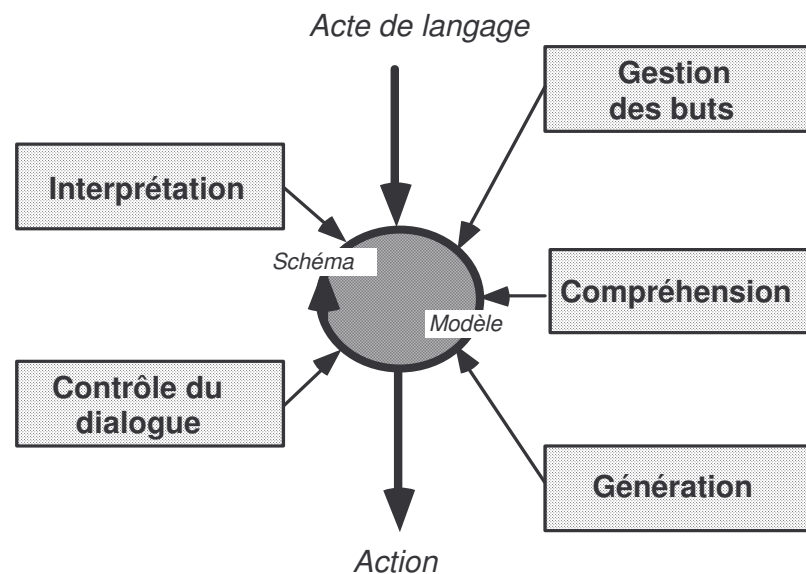
Cela est dû, principalement, au locuteur qui utilise dans son expression, des stratégies variables qu'il adapte au cours du dialogue (en fonction du but à atteindre) et au fait qu'il est difficile de planifier des données comprenant de nombreuses ambiguïtés, irrégularités, etc. Cela montre aussi la difficulté qui existe pour faire collaborer deux disciplines telles que l'informatique et la linguistique, car « *la linguistique vise à décrire la richesse de la langue, alors que les réalisations informatiques tendent inéluctablement à une simplification réductrice* » [Luzzati 95].

C'est pour cette raison que la plupart des études sur le dialogue se font à travers l'observation de **dialogues finalisés** entre participants humains. Car c'est « *un genre de discours assez bien spécifié fonctionnant sur un certain nombre de paramètres qui réduisent sa portée mais également réduisent sa complexité et sa diversité* » [Borillo 93].

« *De toute façon, on ne sait vraiment réaliser de tels modules de gestion que dans le cadre de dialogues finalisés [Pierrel 87], où l'influence de la tâche a une importance essentielle. [Sabah 89]* »

C'est ainsi dans ce cadre là que nous aborderons notre notion d'attente car nous pensons que la situation de dialogue finalisé est plus propice à l'émergence d'attentes claires que des dialogues informels. Nous aborderons alors la notion du but et la finalité des dialogues étudiés. Ainsi, « Avez-vous l'heure ? » peut être considéré comme l'amorce d'un petit dialogue finalisé dont le but est pour le locuteur d'avoir l'heure. A l'inverse, une conversation de salon n'est pas considérée comme finalisée car elle n'a pas de but précis.

Dans ce cadre de dialogues finalisés, Caelen [Caelen 92] propose une architecture générale pour le dialogue homme-machine comme nous pouvons le voir sur la Figure 1.



**Figure 1:** Architecture générale en dialogue homme-machine [Caelen 92]

Dans cette architecture, l'action, réponse de la machine est déterminée à partir de données issues de différents modules que nous détaillerons plus loin. Eventuellement multimodale, cette action modifie l'état de la situation et des connaissances. Le noyau fonctionnel, structure de donnée commune, reçoit l'acte de langage du locuteur et enrichit son schéma au fur et à mesure des connaissances apportées par les autres modules. Il fait notamment intervenir le modèle de la tâche défini comme « *script sélectionné ou appris en fonction des connaissances sur l'utilisateur et des connaissances pragmatiques (mondes d'arrière plan, situation, etc.)* » [Caelen 92]. Nous verrons plus loin une architecture plus évoluée adoptée dans le cadre du projet PVE.

Nous pensons que la prise en compte des attentes du locuteur peut faciliter l'interprétation de l'énoncé et le contrôle du dialogue. Nous verrons dans cette thèse l'indexation des attentes sur la gestion des buts.

### **3. Attentes du point de vue du locuteur**

Depuis quelques années, certains auteurs comme [Lehuen 97], [Vilnat 97] ou [Lemeunier 00] ont introduit la notion d'attente dans leurs modèles de dialogue. En revanche, ces modèles restent très liés à la tâche puisqu'ils proposent des attentes vues du côté de la machine. Ainsi, à tout moment, le système doit connaître ses propres attentes (donc en fonction du moment où il se situe dans le déroulement de la tâche). L'originalité de notre approche est de regarder ces attentes du côté du locuteur (et non plus de la machine). Repérer les attentes d'un être humain est loin d'être chose aisée mais cela rend possible l'indépendance par rapport à la tâche. C'est, pour nous, une solution qui peut nous permettre d'envisager un dialogue générique adaptable à une tâche particulière. Ainsi, en plus d'offrir une meilleure gestion du dialogue dans le cas du dialogue homme machine et un meilleur suivi du dialogue en dialogue homme homme médiatisé par la machine, cette solution apporte une généralité qui nous semble très importante dans un contexte économique où les systèmes de dialogue devraient se répandre de plus en plus. Dès lors, on pourrait enfin sortir des conditions de laboratoire afin de s'adapter aux conditions réelles et ainsi éviter de devoir créer un modèle par tâche au profit d'un modèle générique s'adaptant à la tâche. Nos objectifs de départ sont donc de modéliser les attentes du locuteur en fonction du contexte et de la tâche en cours. Cela apporte trois originalités. La première est celle décrite plus haut, considérer les attentes du côté du locuteur. La deuxième est de pouvoir ainsi modéliser un contexte social. La troisième est de s'affranchir de la tâche contrairement aux modèles actuels qui définissent les attentes comme ce qui est attendu par la machine (ce qui implique une très forte imbrication avec la tâche) et laissent de côté le contexte social trop difficile à modéliser.

Dans cette thèse, nous aborderons le cœur de notre approche dont une originalité est de prendre en compte les attentes des interlocuteurs. Nous présenterons ensuite notre modèle (avec une mise en œuvre à travers les choix de notation) ainsi que l'analyse de corpus qui lui est associée. Puis, nous montrerons l'apport des attentes et leur gestion en comparaison avec une approche plus classique de prédiction d'acte [Fouquet 02]. Nous présenterons alors la mise en œuvre de notre modèle à travers une expérimentation avec magiciens d'Oz pour le test et l'enrichissement du modèle.



#### 4. Méthodologie de recherche : démarche incrémentale

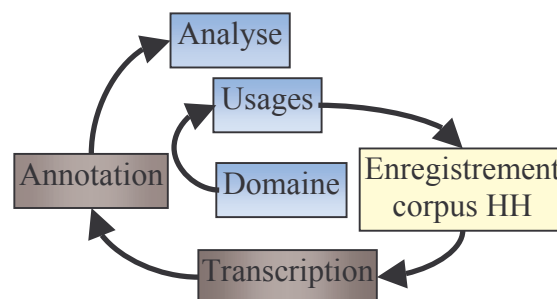
Pour notre travail, nous avons suivi une méthodologie incrémentale, classique en interaction homme-machine.

Cependant, notre théorie des attentes, qui sous-tend cette méthodologie, est valable en dialogue humain comme en dialogue homme machine. Nous faisons l'hypothèse suivante :

Un individu en situation de dialogue (oral ou non, humain ou homme-machine) finalisé aura certaines attentes par rapport à son allocataire<sup>1</sup> suite à son énoncé.

Nous allons chercher, tout au long de cette thèse, à valider cette hypothèse en dialogue humain comme en dialogue homme-machine. Nous comparerons également les attentes dans ces deux types de situation.

Pour valider notre théorie des attentes en dialogue humain comme en dialogue homme-machine et comparer les deux situations, nous avons d'abord appliqué notre méthodologie incrémentale au dialogue humain selon le schéma de la Figure 2 :

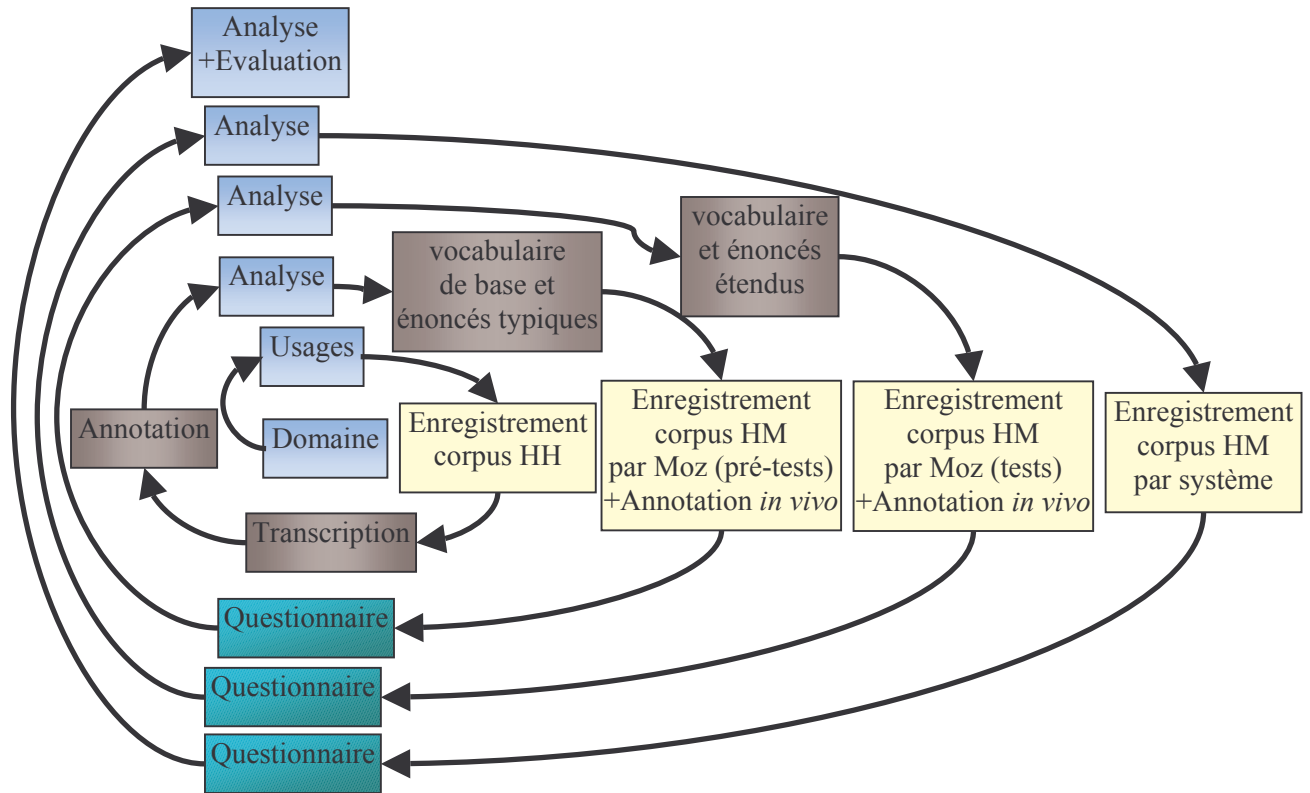


**Figure 2** : Méthodologie incrémentale suivie en dialogue homme-homme

Suivant le schéma de la Figure 2, nous avons défini le domaine de l'application cible. Nous avons opté pour le renseignement touristique dans un projet (NESPOLE!) que nous décrirons plus loin. Après une brève analyse d'usages, nous avons collecté un corpus de dialogues humains via un outil d'audioconférence (Netmeeting<sup>TM</sup>). Ce corpus a ensuite été transcrit puis annoté afin d'en faire une analyse pour nos *attentes*.

<sup>1</sup> Le récipiendaire de l'énoncé oral du locuteur

Dans le cadre du dialogue homme-machine, la méthodologie suit le même chemin mais se poursuit jusqu'à l'application informatique du système de dialogue final. Le cycle vu ci-dessus (Figure 2) se répète donc afin d'obtenir la méthodologie incrémentale que nous pouvons voir sur la Figure 3, ci-dessous :



**Figure 3** : Méthodologie incrémentale suivie en dialogue homme-machine

Nous avons donc commencé comme pour le dialogue humain par définir le domaine de l'application cible, un assistant virtuel dans le cadre d'un projet (PVE) que nous décrirons plus loin. La deuxième étape a consisté en une observation des usages des systèmes existants, avec et sans ressources informatiques. S'ensuit la collecte par le canal téléphonique d'un corpus de dialogues humains réels mettant en jeu des secrétaires et leurs interlocuteurs. La transcription et l'annotation de ce corpus ont permis une analyse de celui-ci sur différents points (notamment les *attentes*, mais aussi un ensemble d'énoncés types, une architecture type des dialogues, etc.).

À partir des transcriptions, le vocabulaire de base dans le domaine de l'application cible a été élaboré, puis une base d'énoncés typiques a été constituée en suivant les travaux de [Morel 88] et [Bilange 92]. La mise en place d'une plate-forme de type magicien d'Oz alimentée par ces données a permis un premier enregistrement de corpus de dialogues homme-machine (pré-tests) et une annotation *in vivo* des énoncés par des actes de dialogue. Après chaque enregistrement, on a fait passer un questionnaire concernant notamment les critiques et points à améliorer au niveau de l'ergonomie de chaque point d'entrée de la plate-forme (client ou agent-simulé).

Une analyse du corpus obtenu a permis d'affiner les énoncés au dialogue homme-machine et de rajouter les énoncés non prévus par l'analyse homme-homme mais apparus dans l'analyse homme-machine et considérés comme importants. Une fois la plate-forme améliorée, on a pu effectuer nos tests et enregistrer un corpus de dialogue homme-machine (tests) dans le contexte de l'application visée. L'annotation des énoncés en actes de dialogue a de nouveau été effectuée *in vivo*, puis vérifiée *a posteriori*. A la suite du test, les sujets ont passé un entretien et rempli un nouveau questionnaire concernant la qualité perçue du système, les points à améliorer, etc.

L'analyse du corpus en termes d'attentes et l'analyse des entretiens a permis l'élaboration d'un système de dialogue qu'il conviendra ensuite d'évaluer. Une nouvelle expérimentation sera donc mise en place. Elle permettra un nouvel enregistrement de corpus de dialogues homme-machine mais cette fois-ci passant par le système et non plus par un compère simulant le système.



## **A. Contexte pratique et théorique**



## **Chapitre I : contexte pratique d'étude**

Le présent chapitre décrit le contexte pratique de notre étude à travers trois projets qui concernent du dialogue homme-homme ou du dialogue homme-machine.

### **1. Communication humaine : ERIM et l'aide au traducteur**

Nous proposons dans cette thèse une notion d'attente qui puisse sa source dans la communication humaine. Dans cette dernière, nous observerons le cas des incompréhensions soulevées par une maîtrise imparfaite de la langue à travers une analyse sous l'angle psycholinguistique.

Le projet soutenu par la région Rhône-Alpes, ERIM (Environnement Réseau pour l'Interprétariat Multimodal) [Fafiotte 00] a pour objectif l'aide à l'interprétariat par la réalisation d'outils d'aide à la traduction en ligne. L'idée est de créer un environnement permettant d'aider l'interprète en vue de lui donner la capacité de s'adapter plus rapidement à une conversation qu'il prend en cours (notamment en cas de changement d'interprète). Nous pouvons aussi imaginer une situation dans laquelle les utilisateurs ne font appel à un interprète que ponctuellement. Dans le cadre d'ERIM, cependant, nous considérerons que les locuteurs conversent dans une langue commune (par exemple en anglais pour un japonais et un français) et qu'ils ne feront appel à l'interprète que lorsque la situation l'exigera, c'est à dire en cas d'incompréhension majeure. Ce que nous appellerons dans ce cas machine de traduction, sera alors le poste intermédiaire qui pourra fournir une aide aux interlocuteurs et à l'interprète.

## 2. Communication humaine médiatisée : NESPOLE! et l'annotation en actes

En traduction de parole, comme en communication homme-machine, modéliser le dialogue est très utile. Pour la reconnaissance vocale, cela permet de restreindre l'ensemble des solutions. Pour la traduction, cette modélisation permet de réduire les ambiguïtés. Nous pensons qu'une bonne modélisation du dialogue n'est possible qu'à travers la modélisation des *attentes* dans le dialogue. Les attentes se définissent comme l'ensemble des réponses possibles à une intervention dans un contexte donné ; par exemple les attentes à une question de confirmation sont oui/non/ne sais pas. Cela peut être également une question de clarification en cas d'incompréhension comme dans le **Tableau 1** ci-dessous :

A : Habitez-vous à Grenoble ? B : Oui	A : Habitez-vous à Grenoble ? B : Pardon, vous avez-dit Grenoble ? A : Oui B : Non, j'habite dans la banlieue
--	---

**Tableau 1** : Deux suites possibles à un même énoncé



Le problème devient plus complexe dans le cas des questions indirectes qu'il ne faut pas considérer comme des questions de confirmation, par exemple,

A : Avez-vous l'heure ?

B : Il est 13 heures

et aussi dans le cas d'un échange de propositions portant sur le monde, par exemple,

A : Dans quelle rue habitez-vous ?

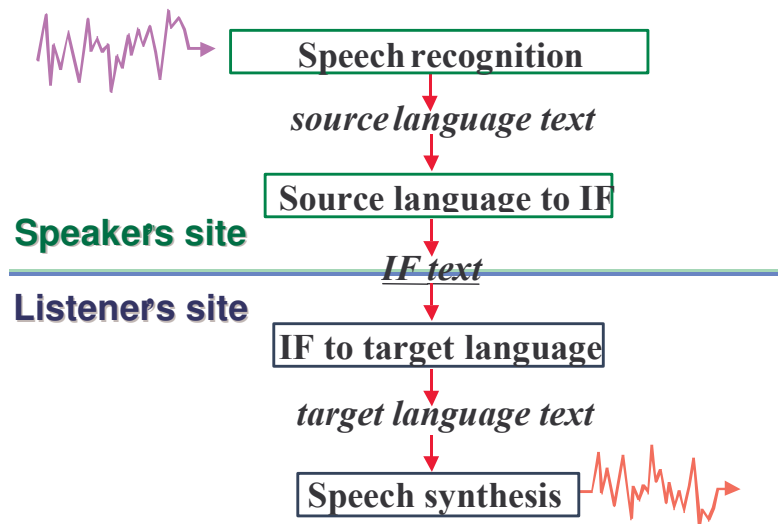
B : Paradis

qui sous-tend que Paradis est le nom d'une rue dans ce contexte, car dans ce cas l'attente est un nom propre (qu'il ne faudra donc pas traduire dans la conversation par exemple).

Le projet européen NESPOLE ! (NEgociating through SPOken Language in E-commerce) [Besacier & al. 01] est un consortium explorant les futures applications de la traduction de parole dans le cadre du commerce et des services électroniques. C'est un projet de traduction automatique de dialogues multilingues dans le domaine du renseignement touristique. Les partenaires étaient *ITC/IRST* de Trento (Italie), *ISL Labs.* de *UKA* (Karlsruhe, Allemagne), *CMU* (Pittsburgh, USA), *Aethra* (une compagnie italienne spécialisée dans les logiciels de vidéoconférence), *APT* : une agence de tourisme dans la région du Trentin (Italie) et le laboratoire *CLIPS* (Grenoble, France).

Le scénario pour la première démonstration de NESPOLE! comportait un agent parlant italien, situé dans une agence de tourisme en Italie et un client situé ailleurs (parlant anglais, allemand ou français) utilisant un simple terminal (PC avec cartes son et vidéo, logiciel de vidéoconférence comme NetMeeting™). Ce choix dépendait de la technologie actuelle. Dans un futur proche, le téléphone portable de troisième génération peut être également utilisé comme terminal.

Le client souhaite organiser un voyage dans la région du Trentino et consulte le site Internet de l'agence de tourisme (APT) afin d'obtenir des informations. Si le client désire connaître davantage d'informations à propos d'un sujet particulier ou préfère un contact plus direct, un service de traduction de parole lui permet d'interagir dans sa propre langue avec l'agent italien d'APT. Une session de vidéoconférence peut alors être ouverte entre le client et l'agent et le dialogue débute entre eux. Le scénario commence avec l'hypothèse que le touriste a déjà visité le site *APT www.trentino.to*. Comme le confirment nos intuitions mais aussi une courte analyse des courriels reçus par APT à propos de demandes d'informations générales, le touriste n'a pas parcouru le site Internet ni lu les pages en détail et souhaite demander des informations sur certains détails d'un sujet spécifique. L'architecture employée pour la traduction de parole est proposée sur la **Figure 4** ci-après :



**Figure 4** : Architecture de traduction de parole par langage pivot ([Besacier 01])

Dans cette architecture, le locuteur A parle devant un système de reconnaissance. L'hypothèse de reconnaissance alimente ensuite le système d'analyse qui traduit l'énoncé en une forme intermédiaire commune à tous les participants du projet. Cette forme intermédiaire (IF) est une représentation sous forme d'acte de parole du contenu de l'énoncé. Elle est alors transférée vers le générateur de langue à partir d'IF qui opère une transformation de la forme intermédiaire pour aboutir à un énoncé dans la langue cible. Cet énoncé est ensuite synthétisé dans la langue cible. Le locuteur A parle donc dans une langue source qui est d'abord traduite en une langue intermédiaire (pivot) avant d'être traduite en langue cible, langue dans laquelle l'allocutaire B entendra l'énoncé.

Pour améliorer les systèmes de traduction de parole encore expérimentaux, les verrous scientifiques et technologiques à lever sont la robustesse, le facteur d'échelle, la portabilité multi-plate-forme et l'interaction multimodale avec un contenu multimédia.

Dans ce projet, nous avons enregistré un corpus [Burger & al. 01] de dialogues entre un agent italien parlant français pour l'occasion et des clients français. Une première analyse, sous l'angle psycholinguistique, de ces dialogues transcrits nous a permis de relever les difficultés de la langue et en particulier à l'oral dans le cadre du dialogue. Notre conclusion porte alors sur l'importance d'analyser (et gérer ensuite) les dialogues à un niveau pragmatique. L'annotation de ces dialogues nous a permis de valider notre approche intégrant les attentes.

### **3. Communication Homme-Machine : PVE et le modèle générique de dialogue**

Le dialogue homme-machine, forme d'interaction entre l'homme et la machine, est très différent du dialogue humain. L'homme et la machine ont une sémiotique non partagée. Leurs connaissances sont différentes. Le champ sémantique est plus restreint, le lexique plus contrôlé. Leurs rôles et leurs compétences sont dissymétriques. La machine n'est pas cognitive, ni sociale. Le langage de l'utilisateur est plus direct qu'avec un interlocuteur humain. Ses phrases sont plus courtes et d'une complexité syntaxique relativement pauvre. Ce type de dialogue contient beaucoup moins d'implicatures conversationnelles (ellipses, non-dits, implicites, etc. ; cf. chapitre II.3.1 La coopérativité selon Grice p54) qu'un dialogue humain.

En revanche, un système de dialogue homme-machine est une interface entre l'utilisateur qui peut converser en langue naturelle et le système. Il offre donc l'accès au système avec un minimum de contraintes et doit être capable alors de 'comprendre' ce que dit son interlocuteur, c'est à dire de créer une représentation interne en fonction de connaissances issues de nombreux domaines étudiés par différentes disciplines comme nous le verrons dans l'état de l'art (linguistique, pragmatique, informatique, sciences cognitives, psycholinguistique, etc.)

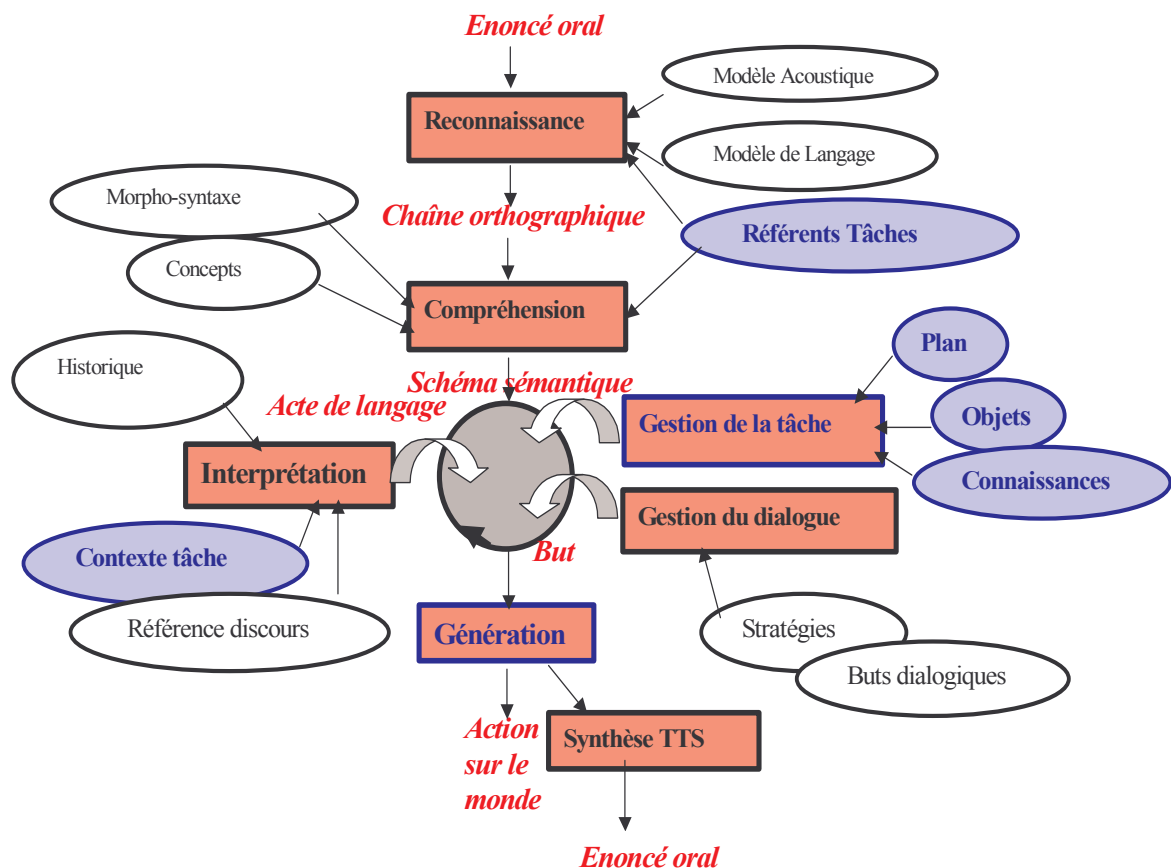
Le projet RNRT PVE (Portail Vocal d'Entreprise) a pour objectif de concevoir et de réaliser un système de dialogue homme-machine dont le modèle est générique (c'est à dire indépendant de la tâche à réaliser) donc adaptable à diverses tâches. L'objectif est de traiter des dialogues courts (2 à 3 minutes) et complexes (contenant des phénomènes courants comme l'anaphore, l'ellipse, etc.) en langage naturel. Le support choisi est la communication intra-entreprise via un portail accessible par la voix (le téléphone). Ce portail pourrait être comparé à une sorte d'assistant virtuel permettant de résoudre diverses tâches habituellement dévolues à des secrétaires. Suivant alors une relation utilisateur-système de type « maître-serviteur », l'assistant doit pouvoir résoudre des tâches aussi diverses que la mise en relation avec un membre de l'entreprise, la prise de rendez-vous, la planification de réunion avec vérification des emplois du temps, la réservation de salle de réunion ou plus rarement la consultation d'un agenda commun, l'envoi de documents, etc. Les interactions avec le système se font par un dialogue vocal en langue naturelle.

Dans un tel système, le modèle de dialogue est donc de première importance et nous pensons qu'une bonne gestion des attentes de l'appelant est une condition nécessaire au succès du dialogue. Dans ce projet, nous avons pu mettre en œuvre

l'idée de généricité (indépendance par rapport à la tâche) et valider une plate-forme générique pour élaborer des dialogues vocaux entre un système et un utilisateur.

Le but final est de modéliser un générateur de système de dialogue suffisamment générique pour convenir à un large éventail d'entreprises et surtout de tâches. Un autre objectif est la démonstration de la possibilité d'interaction vocale par le biais du dialogue en langue naturelle.

Les participants à ce projet étaient : l'entreprise Qualipse, le laboratoire CLIPS, le bureau d'étude Amoweba et la compagnie IBM France. L'entreprise Qualipse (ex Neurosoft) a développé le prototype de dialogue. Le laboratoire CLIPS a effectué la collecte de dialogues réels, l'analyse de ces dialogues, la modélisation de ces dialogues et l'évaluation ergonomique du prototype. Le bureau d'étude Amoweba (ex Novadis Service) spécialisé dans l'ingénierie cognitive a réalisé les études d'usages dans diverses entreprises. La compagnie IBM France a fourni une plate-forme de développement pour VoiceXML permettant de s'affranchir des parties reconnaissance de parole et synthèse vocale.



**Figure 5 :** Architecture du système de dialogue dans le projet PVE (en gris, les données spécifiques à la tâche)

Ce projet propose de lever différents verrous propres au dialogue oral. Ainsi, pour la partie reconnaissance de parole, il a fallu aborder les aspects multilocuteur et multilingues, canal téléphonique, robustesse et couverture linguistique (en particulier, les noms propres). Ensuite, au niveau de l'analyse sémantique certaines difficultés ont dû être levées comme la richesse des concepts, la complexité des énoncés, le sens commun ou la couverture sémantique. Ces différentes couvertures (linguistique, sémantique) sont cruciales dans l'indépendance d'un système à la tâche. Pour l'analyse pragmatique, nous nous sommes attachés aux problèmes de représentation du discours à travers la résolution des référents à la tâche et au discours ainsi qu'aux connaissances mutuelles. Au niveau de la gestion de la tâche, nous avons étudié les approches à base de planification à mettre en comparaison avec celles prônant l'action située. Dans le cadre de la gestion du dialogue, nous nous sommes intéressés à la pertinence des stratégies à employer ainsi qu'à la généralité du modèle de contrôle de dialogue. La génération offre des intérêts de recherche au niveau du degré de force illocutoire et des effets perlocutoires. Enfin, au niveau de la synthèse vocale nous avons approfondis les problèmes de prosodie dans le dialogue.

L'indépendance totale du modèle de dialogue par rapport au contexte de la tâche est impossible pour plusieurs raisons. Les référents de la tâche interviennent dans les phases de reconnaissance vocale et de compréhension sémantique. Pour cette dernière, il n'existe pas de module à large couverture. Enfin le contexte de la tâche intervient dans l'interprétation pragmatique.

Notre but fut donc d'élaborer un modèle le plus indépendant possible par rapport au contexte de la tâche notamment par l'utilisation de la théorie des actes de langage et de celle des attentes, toutes deux indépendantes de la tâche. Le contrôleur du dialogue ne doit pas être conçu spécifiquement pour l'organisation d'un ensemble de tâches, mais comme un composant indépendant qui contrôle le dialogue en choisissant la stratégie la mieux adaptée à la réalisation des buts dialogiques. Ce contrôleur reçoit en entrée les données issues du gestionnaire de tâche. Les dictionnaires et ontologies sont spécifiques à l'univers du domaine.

L'intérêt d'un modèle générique réside bien sûr dans sa ré-utilisabilité. Si l'on dispose d'un tel système, réaliser une application particulière se réduit à développer un gestionnaire de la tâche et à adapter les ressources linguistiques correspondantes.

Ma contribution dans ce projet fut (selon la méthodologie décrite en Figure 3) :

- ✓ Analyse du type de corpus pertinent pour l'élaboration du modèle de dialogue
- ✓ Collecte d'un corpus de dialogues réels dans le cadre de l'entreprise.
- ✓ Sélection des dialogues les plus pertinents et classement par tâche.
- ✓ Elaboration d'une norme de transcription.
- ✓ Transcription de ces dialogues.
- ✓ Elaboration d'une grille d'annotation
- ✓ Annotation de ces dialogues en terme de buts illocutoires, stratégies, actes et échanges
- ✓ Elaboration d'une structure d'échanges générique pour chaque dialogue
- ✓ Elaboration d'une structure d'échanges générique à tous les dialogues. Utilisation d'un modèle d'analyse de dialogue (structural avec l'école de Genève Roulet & Moeschler avec composante dynamique Luzzati & Bilange)
- ✓ Elaboration du vocabulaire utilisé dans le corpus homme-homme
- ✓ Elaboration des énoncés typiques dans le corpus homme-homme
- ✓ Elaboration d'une expérimentation de type Magicien d'Oz (version 0) pour recueillir du corpus homme-machine adapté à la tâche à réaliser et valider certaines hypothèses avant la modélisation effective.
- ✓ Elaboration du vocabulaire utilisé dans le corpus homme-machine
- ✓ Elaboration des énoncés typiques dans le corpus homme-machine
- ✓ Validation des premières hypothèses
- ✓ Elaboration d'une expérimentation de type Magicien d'Oz (version 1) pour recueillir du corpus homme-machine adapté à la tâche à réaliser et valider certaines hypothèses avant la modélisation
- ✓ Elaboration d'un questionnaire
- ✓ Dépouillement du questionnaire et des résultats
- ✓ Calcul des Attentes
- ✓ Modélisation du dialogue avec toutes ces données : opportuniste pour être dynamique et flexible

## **Chapitre II : Cadre théorique en modélisation du dialogue oral**

Le dialogue oral en parole spontanée s'éloigne sur bien des points du langage écrit. En particulier la complexité structurelle de l'oral est plus pauvre, les phrases plus courtes usant de moins de relatives ou de subordonnées mais plus de juxtapositions. De plus, la communication directe permet de faire des références implicites à travers de nombreux déictiques, anaphores ou ellipses. Enfin, l'oral présente de nombreuses agrammaticalités telles que les répétitions, hésitations, corrections, inachèvement, etc.

De nombreux travaux se fondent sur du dialogue écrit supprimant ainsi ces phénomènes ainsi que la phase de reconnaissance vocale et ses erreurs. Ils permettent cependant d'étudier le dialogue et même certains phénomènes de l'oral (moins présents à l'écrit).

Le dialogue homme-machine fait intervenir différentes disciplines telles que la philosophie du langage, les cognisciences, l'ethnoscience (et en particulier l'ethnométhodologie) et les technosciences (l'informatique).

Les premiers modèles sont inspirés du modèle du code de Shannon & Weaver qui pose la communication en termes de codage du côté de l'émetteur, transfert d'information par un canal de communication et décodage du côté du récepteur (Figure 6) :



**Figure 6** : Modèle du code (vue simplifiée)

Cette théorie suppose alors un modèle sans perte d'information dans la transmission. La pensée (représentation conceptuelle) de l'émetteur est transmise sans perte de sorte que la pensée reçue est identique à celle émise (Figure 7).



**Figure 7** : Modèle du code

Or le processus de parole est non séquentiel. La parole ne traduit pas exactement la pensée et le décodage ne reformule pas la pensée comme elle a été émise. La communication verbale n'est pas un simple transfert de pensées par des mots. Elle implique une interprétation du nouvel environnement acoustique. Il se produit donc quelque chose de plus qu'un simple décodage du sens linguistique. Nous montrerons dans le présent chapitre différentes approches du dialogue que nous tenterons de combiner par la suite. Nous commencerons par la philosophie du langage avec la notion d'acte de langage. Les approches issues des ethnoscience et des cognisciences seront ensuite présentées pour finir par l'approche linguistique.

Selon la sémiotique [Greimas 80,90], le dialogue est un procès (processus) qui fait sens. Tout procès est vu comme un parcours narratif qui vise un gain qui correspond à *avoir plus* ou *être mieux*. Les interlocuteurs cherchent donc à remplir une attente. Chaque étape du dialogue est donc motivée par un *vouloir avoir plus* ou



un *vouloir être mieux*. Greimas définit les interactions dialogales à travers quatre carrés sémiotiques définissant quatre modalités (**Tableau 2**).

ontique	actionnelle & volitive	épistémique	déontique
avoir	faire	savoir	devoir
être	vouloir	croire	pouvoir

**Tableau 2** : Modalités de Greimas

Les modalités **actionnelles et volitives** définissent le *vouloir*, le *faire* et leur contraire (v, ¬v, f, ¬f) permettant de dénoter l'intention et l'action.

Les modalités **ontiques** définissent l'*avoir*, le *être* et leur contraire (a, ¬a, e, ¬e). Supposant un but à atteindre, nous laisserons ces modalités de côté dans cette thèse.

Les modalités **épistémiques** définissent le *croire*, le *savoir* et leur contraire (c, ¬c, s, ¬s) permettant de représenter les croyances et les connaissances.

Les modalités **déontiques** définissent le *pouvoir* et le *devoir* et leur contraire (p, ¬p, d, ¬d).

Ces modalités peuvent alors être combinées pour représenter les énoncés des interlocuteurs. On pourra relever parmi ces croisements certains qui seront à la base de nos travaux. Ce seront par exemple, *vouloir faire*, *ne pas vouloir faire*, *vouloir faire faire*, *faire faire* (directif de Searle), *faire croire* (par exemple mentir), *faire devoir* (obliger), *faire pouvoir* (donner un choix), *faire savoir* (informer), etc.

Ainsi, un procès émergera par un *vouloir*. Par exemple un *vouloir faire savoir* se manifestant sous la forme *faire savoir* et la réaction de l'allocutaire permettra la continuation du procès. Un exemple simple de procès montre l'intérêt du point de vue sémiotique :

A veut avoir X donc A veut faire Y

si A peut faire Y alors A fait Y et A a X

sinon A fait faire Y à B, A fait devoir Y à B donc B doit faire Y et A a X

La conversation s'amorce donc par l'espérance d'un gain. La communication relève alors d'une « tentative d'ajustement où l'on doit ajouter au transport de l'information le jeu des rôles et des actes par quoi les interlocuteurs se reconnaissent comme tels, agissent comme tels et fondent ainsi des communautés linguistiques dans un monde humain » [Wittgenstein 58]

## 1. Philosophie analytique et acte du discours

La philosophie du langage dont la philosophie analytique est issue s'intéresse à l'individu placé en situation de communication sur un plan intentionnel et actionnel.

La pragmatique est l'étude du rôle que jouent les facteurs contextuels dans la communication verbale. Elle étudie donc le sens des énoncés en contexte en décrivant les inférences nécessaires pour accéder au sens communiqué par l'énoncé (différent du sens littéral). Elle peut donc résoudre les implicatures. Elle se rajoute à la couche de sémantique qui se préoccupe du sens de l'énoncé indépendamment du contexte. En effet, elle s'intéresse aux conséquences des énoncés en sus du contenu, faisant intervenir les notions d'intention et d'effet. Elle peut alors traiter des cas complexes qui ne sont solubles que par la prise en compte du contexte. C'est le cas notamment des anaphores, qui reprennent un élément du contexte précédent sans le répéter (ex: je *le* veux), des ellipses, qui suppriment des mots utiles mais sous-tendus, des déictiques qui servent à désigner un objet déterminé dans la situation (ex: je veux *celui-ci*) ou même de phrases incomplètes.

Introduite par les philosophes du langage ordinaire [Austin 69] et [Searle 70], la pragmatique use d'unités atomiques nommées actes de langage selon l'idée que « *parler c'est agir sur le monde et sur les autres, c'est une forme de vie* » [Wittgenstein 58]. Ces actes sont caractérisés par leurs aspects actionnel, intentionnel, conventionnel et contextuel. Elle combine ces actes pour définir l'intervention comme unité monologique puis l'échange comme unité dialogique.

Concernant les actes, [Bilange 92] soutient que si les reconnaître et les générer suffit pour dialoguer correctement alors la procédure de dialogue peut se concevoir simplement comme une sorte d'automate réglant les enchaînements des actes de langages entre eux.

La théorie des actes se situe dans la théorie des intentions. Si A énonce une proposition, c'est qu'il a l'intention de la produire. Si B veut comprendre cette proposition, il doit interpréter l'intention qui l'a sous-tendue, dans le contexte où elle a été produite. C'est alors qu'intervient la notion de but (état mental ou situationnel que désire atteindre le conversant) qui motive le dialogue et l'oriente. « *En tant qu'activité orientée vers une finalité, le dialogue amène au fur et à mesure les partenaires de l'interaction à progresser vers un but qui signe l'accord (ou le désaccord) entre les locuteurs* » [Vernant 94]. Cette activité orientée vers un but se manifeste à travers les actes de dialogue accomplis par les locuteurs.

Frege introduit la notion de force assertive selon l'idée que l'objet de la logique ne doit pas être la proposition mais le jugement de cette proposition. Ainsi, il étudie la proposition et l'assertion de cette proposition, cette dernière constituant et assurant la vérité de la proposition. Soit une proposition  $p =$  « le petit chat est mort ». On peut la considérer simplement sans valeur de jugement. On peut aussi l'asserter, l'affirmer, donnant ainsi une valeur de jugement à la proposition.

### **1.1. Théorie descriptive d'Austin**

Cette notion de force assertive est le point de départ de la réflexion d'Austin [Austin 62, 70] qui élabore une théorie descriptive des actes de discours. En affirmant « Quand dire c'est faire », Austin introduit l'idée que le locuteur qui formule un énoncé ne le fait pas uniquement pour énoncer des propositions vraies, fausses ou à vérifier. Il s'exprime bien plus dans l'intention de réaliser une action et produire certains effets sur l'allocataire. L'énoncé résulte davantage d'usages constatifs (l'usage scientifique qui décrit le phénomène et s'attache à sa vérité, ex : le ciel est bleu) et performatif (la magie sociale qui transforme le discours en action et s'attache davantage au succès de la proposition). Il met donc l'accent sur la pragmatique en partant du principe que tout énoncé est un acte de langage. Pour formaliser ces interventions, l'auteur a alors distingué deux types d'énoncés :

Les **constatifs** permettent de décrire un état ou une situation à travers des énoncés assertifs.

Les **performatifs** servent à créer une situation, obtenir une action.

Ces derniers contiennent l'action qu'ils dénotent. Par exemple, dire « Je vous déclare mari et femme » c'est faire l'action du mariage à condition que l'on ait la légitimité et les conditions pour la faire (prêtre, maire, adjoint au maire par délégation de pouvoir, etc.). Le déclarant doit donc avoir le pouvoir social de dire quelque chose qui sera un acte social. Ainsi produire cet énoncé permet de modifier le monde (les deux personnes sont désormais mariées). Un performatif est donc une action sociale soumise à des contraintes (on ne marie pas des pingouins, les mariés doivent être majeurs, non-mariés, de sexe différent [actuellement en tous cas], etc.). Or les deux types d'énoncés sont très imbriqués. Un constatif peut recevoir des caractéristiques d'un performatif. « Le ciel est bleu » peut être transformé en « j'affirme que le ciel est bleu ». De même, un performatif présuppose un constatif. « Je vous avertis que le taureau va foncer » présuppose un taureau. Pour tout acte de discours il faut alors distinguer trois dimensions (dans chacune desquelles on peut retrouver des fonctions de constatif et de performatif).

Tout énoncé est donc un acte de parole à trois composantes :

(a) le **locutoire** (le *dire*) : l'énoncé proprement dit. On se place au niveau sémantique de la signification à travers les notions de référence et de prédication. C'est l'acte de dire.

(b) l'**illocutoire** (le *faire*) : l'action réalisée par l'énoncé. C'est l'acte effectué en disant quelque chose qui introduit la notion de force illocutoire. On se place au niveau pragmatique car on s'attache à l'énoncé produit dans une situation donnée avec des interlocuteurs donnés.

(c) le **perlocutoire** (l'*effet produit* sur son interlocuteur) : l'action produite sur l'interlocuteur. C'est l'acte effectué par le fait de dire. On s'intéresse alors à l'auditeur avec des notions d'objectifs et d'effets. On fournit les connaissances pour permettre d'agir.

Sur un même locutoire (ex : une porte et la fermeture de cette porte) peuvent donc se rattacher plusieurs actes illocutoires (ex : constatif « la porte est fermée », ordre « ferme la porte », performatif « j'affirme que la porte est fermée », etc.) et l'acte perlocutoire permet de définir les intentions (e.g. : « ferme la porte » indique que l'on veut que la porte soit fermée, mais aussi que ce soit l'allocutoire qui ferme la porte).

Un même énoncé peut donc poursuivre plusieurs buts :

- un but illocutoire : l'effet espéré de l'action (que la porte soit fermée).
- un but perlocutoire : l'effet espéré sur l'interlocuteur (qu'il obéisse et aille fermer la porte).

Les deux premières dimensions répondent à des conventions strictes. En effet, la sémantique répond à un lexique et une grammaire conventionnels. De même l'ordre est conventionnalisé : A donne un ordre à B donc il en a le droit et B doit obéir.

En revanche la dimension perlocutoire ne répond pas à des conventions. Elle n'est pas calculable a priori car elle dépend de nombreux paramètres, comme la relation entre les interlocuteurs. L'effet n'est pas assuré. Lorsque A donne un ordre à B, B peut ne pas obéir aveuglément et cette limite à l'obéissance reste très personnelle. A peut donc être sûr que son ordre sera exécuté alors que B peut oublier, négliger ou refuser cet ordre.

## 1.2. Théorie explicative de Searle

Searle pense que les unités premières de signification dans l'usage et la compréhension du langage ne sont pas des propositions isolées ni des conditions de vérité, mais plutôt des actes de discours du genre illocutoire [Austin 62] tels que des assertions, questions, promesses, ordres et déclarations. En parlant, nous entendons accomplir des actes illocutoires. Cela fait partie de ce que nous voulons dire et communiquer aux interlocuteurs.

Searle introduit la théorie des intentions comme généralisation de la théorie des actes de langage. Ainsi, pour A, énoncer une proposition résulte de l'intention de la produire. Pour B, comprendre cette proposition revient à interpréter l'intention qui l'a sous-tendue, dans le contexte où elle a été produite. Cette théorie situe donc la communication dans une perspective pragmatique.

La taxonomie d'Austin a donc été reprise et affinée par Searle [Searle 72] considérant, selon un principe d'exprimabilité, qu'un locuteur en parlant effectue simultanément quatre actions :

(a) un acte d'**énonciation** (le *dire*) : l'action même de parler, de dire quelque chose, énoncer les mots. C'est l'acte d'énonciation par le fait d'articuler, d'activer ses mâchoires et de produire du son, résultat d'une activité phonétique, grammaticale et sémantique.

(b) un acte **locutoire** (*dire en disant*) : l'acte propositionnel qui construit du sens ; référer, prédiquer.

(c) un acte **illocutoire** (le *faire*) : l'action qui est réalisée par le fait de dire quelque chose. Cette action consiste à informer, affirmer, suggérer, questionner, promettre, ordonner, demander, conseiller, etc. Elle est affectée d'un certain degré de force illocutoire (ou fonction communicative) qui caractérise la fonction de l'acte et d'un contenu propositionnel qui indique ce qui est affirmé, suggéré, demandé, etc.

(d) un acte **perlocutoire** (*faire croire en disant*) : l'action qui décrit la production intentionnelle d'effets sur l'interlocuteur. Par exemple, les actes perlocutoires de convaincre, effrayer, insulter, etc. C'est l'idée d'agir sur l'interlocuteur. Ainsi l'énoncé : « passe-moi le sel » ou « ferme la porte » montrent l'intention de voir si l'interlocuteur obéit, « haut les mains » se fait dans l'intention de neutraliser et faire peur, etc. On s'intéresse alors à l'effet produit sur l'interlocuteur

Ce principe d'exprimabilité permet une description des actes de langage mais aussi leur régulation. Searle inclut cette théorie des intentions dans la pragmatique, incluse elle-même dans une théorie du langage et une théorie de l'action.

Par extension, le terme acte de langage regroupe ces quatre actions. Toutefois, dans le contexte de ce travail, ce terme se réfère le plus souvent à l'acte illocutoire qui contient l'action proprement dite et les intentions qui l'ont motivée (buts). En fait, jusqu'au niveau illocutoire (c), un acte de langage peut se représenter sous la forme  $Fp$  où  $p$  représente le contenu propositionnel et  $F$  la force illocutoire. « je t'ordonne de me passer le sel » se décompose en une force directive  $F$  marquée par « je t'ordonne » et un contenu propositionnel  $p$  marqué par « me passer le sel ».

Jusqu'au niveau illocutoire (c), le rôle des interlocuteurs n'intervient pas. La notion de locuteur n'est prise en compte qu'en termes de locuteur abstrait et non de locuteur parlant. Searle a donc introduit la notion de sens littéral par opposition au sens en contexte. Ce sens littéral est adapté à l'interprétation d'un acte direct comme l'exemple précédent. Mais ce sens littéral ne permet pas l'interprétation des actes indirects dont l'énonciation n'est pas littérale comme dans l'exemple du Tableau 3.

Énoncé	Contexte	Acte résultant
Il fait froid ici	Fenêtre ouverte	Ordre (fermer la fenêtre)
	Fenêtre fermée	Assertion (froid)

**Tableau 3** : Distinguo sens littéral / sens en contexte

L'énoncé « Il fait froid ici », dans une pièce sans moyen de chauffage, n'a pas la même interprétation selon que la fenêtre est ouverte ou fermée. Dans le premier cas, cet énoncé peut s'analyser comme une demande (indirecte) du locuteur que son allocutaire ferme la fenêtre (acte d'ordre avec force faible). Dans le deuxième cas, c'est plutôt une simple assertion (directe) exprimée littéralement. L'énoncé ne peut donc pas être interprété sans l'action sous-tendue (ici, l'action de fermer la fenêtre) ni le contexte d'énonciation (fenêtre ouverte ou fermée).

Il est nécessaire, selon une stratégie inférentielle, d'associer à l'acte ses conditions de réalisation (situation, monde, arrière-plan, etc.), le pourquoi, les intentions de l'énonciateur, le but poursuivi, etc. Cet acte littéral associé au contexte permet par inférence de répondre à l'acte indirect. Ainsi un énoncé tel que « Avez-vous l'heure ? » qui peut avoir pour réponse directe « oui » ou « non » sera analysé par inférence en acte indirect « Quelle heure est-il ? » qui aura davantage pour réponse « il est 18h00 ».

Searle répertorie cinq actes que nous détaillerons plus loin : assertif, directif, commissif, expressif, déclaratif. Il s'intéresse alors au but de l'acte, à son mode d'accomplissement, au degré de sincérité, au degré de force. Il introduit également quatre types de conditions de félicité (réussite ou succès) d'un acte :

Les conditions de contenu propositionnel précisent qu'il « *doit exister une procédure, reconnue par convention, dotée par convention d'un certain effet, et comprenant l'énoncé de certains mots par de certaines personnes dans de certaines circonstances* ».

Les conditions essentielles ou conditions sur le monde définissent qu'il « *faut que, dans chaque cas, les personnes et les circonstances particulières soient celles qui conviennent pour qu'on puisse invoquer la procédure en question* ».

Les conditions préliminaires ou conditions sur la situation, précisent que « *la procédure doit être exécutée par tous les participants, à la fois correctement et intégralement* ».

Les conditions de sincérité définissent l'état psychologique dans lequel le locuteur doit être s'il est sincère. « *Lorsque la procédure –comme il arrive souvent– suppose chez ceux qui recourent à elle certaines pensées ou certains sentiments, lorsqu'elle doit provoquer par la suite un certain comportement de la part de l'un ou de l'autre des participants, il faut que la personne qui prend part à la procédure (et par là l'invoque) ait, en fait, ces pensées ou ces sentiments, et que les participants aient l'intention d'adopter le comportement impliqué.* » De plus, « *ils doivent se comporter ainsi, en fait par la suite* ».

### **1.3. Théorie illocutoire de Vanderveken (formalisation des précédentes)**

Vanderveken [Vanderveken 88] a formalisé une logique illocutoire du discours, essentiellement monologique, poursuivant ainsi les travaux de Searle. Selon lui, les actes illocutoires sont « *les unités premières de signification dans l'usage et la compréhension des langues naturelles* ». Ainsi, tout énoncé contient un marqueur de force illocutoire. Il établit donc une taxonomie des actes de discours dans laquelle les « *actes illocutoires élémentaires de la forme  $F(P)$  sont exprimés dans les langues naturelles par des énoncés élémentaires de la forme  $f(p)$  où  $f$  est un marqueur de forme illocutoire et  $p$  une clause* » [Vanderveken 88, p21]. Il distingue six types d'énoncé.

L'énoncé **déclaratif** permet de dire comment sont les choses (« La porte est ouverte »).

L'énoncé **conditionnel** permet de signifier ce qui se passerait si certains faits existaient (« J'irais bien en montagne, si le temps était plus propice »). Ce type d'énoncé a été abandonné par la suite au profit de l'énoncé **performatif** permettant de faire des déclarations, c'est à dire agir par les mots (« Je ferme la porte »).

L'énoncé **impératif** donne des directives à l'allocutaire (« Ferme la porte »).

L'énoncé **interrogatif** permet de poser des questions (« Etes-vous certain ? »).

L'énoncé **exclamatif** sert à exprimer les états mentaux des locuteurs (« Je suis triste »).

Enfin, l'énoncé **optatif** exprime les souhaits du locuteur (« J'aimerais qu'il pleuve »).

Avec ces six types d'énoncés, il propose cinq usages.

L'usage **assertif** consiste à exprimer une proposition P avec le but illocutoire de représenter l'état des choses précisé par P.

L'usage **engageant** (ou commissif) consiste à exprimer une proposition P avec le but illocutoire de s'engager à accomplir l'action future contenue dans P.

L'usage **directif** consiste à exprimer une proposition P avec le but illocutoire impératif (c'est à dire d'essayer de faire en sorte que l'allocutaire accomplisse l'action future contenue dans P).

L'usage **déclaratif** consiste à exprimer une proposition P avec le but illocutoire d'accomplir l'action contenue dans P par l'énonciation elle-même.

Enfin, l'usage **expressif** consiste à exprimer une proposition P avec le but illocutoire exclamatif (c'est à dire manifester l'état mental contenu dans P).

Vanderveken s'intéresse alors à l'aspect vériconditionnel de la signification de l'énoncé et introduit une distinction entre conditions de satisfaction et conditions de succès munissant les actes illocutoires d'intentionnalité.



Une **condition de satisfaction** porte sur le contenu de l'énoncé et correspond aux conditions de vérité de leur contenu propositionnel. L'énoncé « Il pleut » est vrai ou faux. Une assertion sera satisfaite si elle est vraie. Une promesse sera satisfaite si elle est tenue. Un conseil sera satisfait s'il est suivi. Cette condition vérifie donc que les effets de l'acte illocutoire sont vrais dans le monde.

Les **conditions de succès** portent plutôt sur les actes et correspondent aux croyances et aux intentions du locuteur. Un acte sera un succès si l'engagement est tenu ou si la description est exacte. Ce succès dépendra des six composantes de chaque force : le but illocutoire (définissant les relations entre les mots et les choses) qui dénote l'intention de réaliser l'acte illocutoire, le mode d'accomplissement (moyens et manières d'accomplir un acte, par ex. il faut avoir autorité pour commander ou donner une option de refus lors d'une demande), le contenu propositionnel (qui doit être tenu pour vrai), le degré de sincérité (qui reflète les attitudes psychologiques ; le locuteur croit le contenu propositionnel), le degré de force illocutoire (un degré de puissance de conditions de sincérités) et certaines conditions préparatoires (des vérités sur le contexte et sur l'arrière-plan ; le locuteur a des raisons de croire en la vérité du contenu propositionnel).

Succès \ Satisfaction	Oui	Non
Oui	vraie et j'y crois => vraie	J'y crois mais faux => erreur
Non	vraie mais je n'y crois pas => erreur du mensonge	faux et je n'y crois pas => mensonge

**Tableau 4** : Conditions de succès / satisfaction

Ainsi, une assertion telle que « le soleil brille » est satisfaite si elle est vraie (le soleil brille effectivement) ou non satisfaite si elle est fautive (il pleut par exemple). Les conditions de succès vont permettre de définir alors la volonté de coopération du locuteur. La même assertion sera un succès si le locuteur la croit vraie mais ne sera pas un succès s'il ment (donc s'il la croit fautive). En croisant ces deux conditions on obtient donc des situations peu analysées auparavant et pourtant très possibles. La vérité, l'erreur ou le mensonge apparaissent évidemment mais l'erreur du mensonge devient possible. Ainsi lorsque le locuteur ment (il ne dit pas ce qu'il croit) et commet une erreur (ce qu'il croit est faux), cette erreur du mensonge aboutit à une assertion vraie.

Vanderveken introduit alors l'implication cognitive et d'engagement selon l'idée que le locuteur sait quel acte il tente d'accomplir et quels actes il s'engage à accomplir par ce fait : « *Quand une force illocutoire  $F$  contient une autre  $F'$ , nous le savons et chaque acte de discours de la forme  $F(P)$  nous engage à accomplir l'acte correspondant  $F'(P)$ . De même, quand une proposition  $P$  en implique cognitivement une autre  $Q$  (en ce sens qu'on ne peut l'exprimer sans savoir a priori qu'elle implique l'autre  $Q$ ), il en résulte des engagements illocutoires forts; les actes illocutoires de forme  $F(P)$  dont la force est primitive engagent le locuteur à accomplir l'acte correspondant de forme  $F(Q)$  quand  $Q$  satisfait les conditions sur le contenu propositionnel. Ainsi nous ne pouvons affirmer ou déclarer la conjonction  $P \& Q$  sans affirmer ou déclarer  $P$ . Nous sommes capables de faire pas mal d'inférences pratiques et théoriques valides en parlant. L'engagement illocutoire fort d'un locuteur est décidable et bien fondé alors que l'implication propositionnelle classique ne l'est pas. »*

#### 1.4. Une taxonomie des actes de langage

Une taxonomie des actes de langage a été introduite par Searle [Searle 69], puis affinée par Vanderveken. Nous présenterons ici cette taxonomie en la mettant en correspondance avec une notation introduite par Caelen dérivée des modaux de Greimas. A partir de celle-ci, nous avons nous-même établi une taxonomie adaptée à notre notion d'attentes. Nous la présenterons en détails dans la partie II.

L'**acte assertif** est utilisé pour décrire un état de fait existant. Le locuteur exprime comment sont les choses. Le but est de rendre le contenu propositionnel (la proposition) conforme au monde. L'acte assertif permet donc d'indiquer une connaissance ou du moins de **révéler les croyances** du locuteur. Selon la notation de Caelen, cet acte sera de la forme  $F^s$  pour *Faire savoir*. Le locuteur fait savoir une information. « Il fait beau aujourd'hui ». Nous y retrouverons alors tous les énoncés d'affirmation, d'information, d'assertion, de confirmation, de constatation, de présentation, de description, de commentaires, d'explication, d'illustration, d'exposition, d'épellation, de répétition, de rectification, de concession, de début d'action ou de changement d'activité mais aussi les énoncés de négation, de contestation, de critique, de restriction et d'abandon. La condition de sincérité est donc que le locuteur croit en la proposition qu'il énonce.

L'**acte directif** a pour but d'obliger l'interlocuteur à réaliser une action future. Le locuteur tente de faire faire les choses par autrui. Le but est alors de rendre le monde conforme au contenu propositionnel dans lequel est comprise l'action future à réaliser par l'interlocuteur. L'acte directif permet d'**exprimer les souhaits** ainsi que la volonté du locuteur. On y retrouve donc les énoncés contenant des verbes modaux (vouloir, pouvoir, devoir, falloir) ou le verbe aller au futur. Plus généralement, seront directifs tous les énoncés contenant un ordre, une autorisation, une invitation, un conseil, une suggestion, un avertissement, un défi, une question, une interrogation, une demande d'information, une demande de précision, une demande de confirmation, une requête, une répétition, une relance, une insistance ou une supplication. Cet acte aura différentes notations. Il pourra être  $F^f$  (*Faire faire*) pour faire faire une action : « Attendez un instant, s'il vous plaît ». Il pourra être  $F^d$  (*Faire devoir*) quand l'obligation est forte : « passe-moi le sel ». Il sera enfin  $F^{fs}$  (*Faire faire savoir*) pour faire faire savoir une information : « Peux-tu me donner l'heure, s'il te plaît ? ». La condition de sincérité est donc que le locuteur souhaite que son interlocuteur fasse l'action contenue dans la proposition énoncée.

L'**acte commissif** (ou promissif ou engageant) est un engagement personnel à réaliser une action future. Le locuteur promet de faire quelque chose. C'est donc un directif dans lequel l'interlocuteur est le locuteur lui-même. Le but est donc le même mais envers lui-même, c'est-à-dire de rendre le monde conforme au contenu propositionnel dans lequel figure son action à réaliser. L'acte promissif permet alors de **révéler l'intention** du locuteur. Il sera, pour nous, de la forme  $F^p$ , pour *Faire pouvoir* et regroupera des énoncés qui permettent d'ouvrir le dialogue mais aussi ceux de promesse, d'offre ou de justification personnelle. La condition de sincérité est que le locuteur ait l'intention de faire l'action contenue dans la proposition énoncée.

L'**acte expressif** a pour but illocutoire d'exprimer l'état psychologique associé. Il n'a pas de but de conformation car il y a déjà congruence entre les mots et la réalité. Il regroupe les énoncés d'expression de souhait, de remerciement, de plaisir, de salutation, d'excuse, d'hypothèse, de spéculation, de présupposition, d'anticipation, de satisfaction, de félicitations, d'hésitation, de résignation, d'étonnement, de regret, de déception, de plainte, de menace, de juron et d'insulte. Le contenu propositionnel est considéré comme présupposé et le locuteur se réjouit ou déplore sa véracité. L'acte expressif permet alors de **révéler l'état psychologique** du locuteur. Il sera noté  $F^{\emptyset}$  pour un *Faire savoir* à direction d'ajustement vide.

D'après Caelen, cet acte est très peu présent en dialogue homme-machine. En dialogue humain, en revanche, il est très présent à l'oral à travers les phatiques. La condition de sincérité est que le locuteur exprime une propriété attribuée à l'un des interlocuteurs.

L'**acte déclaratif** rend effectif son contenu propositionnel. Le locuteur modifie le monde par cette déclaration. Le but est donc autant de rendre conforme le monde aux mots que les mots au monde. Il nécessite alors des conditions extralinguistiques qui confèrent au locuteur le pouvoir de modifier le monde par le seul accomplissement de cet acte.

Par exemple, « je vous déclare mari et femme » n'a de valeur de modification du monde que selon certaines conditions. Le locuteur doit avoir le pouvoir de faire cette déclaration (prêtre, maire, adjoint au maire par délégation de pouvoir, capitaine du navire ou consul lors d'une cérémonie). Les interlocuteurs doivent être au nombre de deux, non mariés auparavant. Par ce seul acte dans ces conditions, les interlocuteurs seront mariés. Ce seront donc tous les énoncés de début, de fin ou d'abandon d'action. Ils seront notés  $F^a$  par Caelen. Nous le notons simplement F. Ils n'auront pas de condition de sincérité.

### **1.5. Une direction d'ajustement**

L'approche issue de la philosophie du discours est essentiellement monologique car elle ne prend pas en compte la dimension dialogique à la différence des jeux de langages de Wittgenstein. En effet, il n'y a pas de notion d'interlocution entre un locuteur et un allocutaire. L'approche propose davantage une relation dirigée, par les effets perlocutoires, entre le locuteur et un auditeur qui peut éventuellement se transformer ensuite en locuteur. Le rôle de l'auditeur est secondaire car il doit juste comprendre les intentions du locuteur. Cependant, Searle introduit une distinction implicite pour l'ajustement du monde aux mots. En effet, il sépare les directifs pour lesquels autrui doit agir et les commissifs pour lesquels c'est le locuteur qui doit agir. C'est pourquoi Vernant [Vernant 94] propose une autre taxonomie explicitant cette distinction pour toutes les catégories. Il propose alors cinq directions d'ajustement.

L'ajustement des **mots au monde** est un ajustement où les mots servent à décrire le monde et les croyances par rapport à ce monde. Ils s'ajustent à la réalité. On y retrouve alors les actes **assertifs**. Ces actes seront des constatifs (ex : vous êtes attentifs) lorsque l'agent est différent du sujet. Ces constatifs seront alors statifs (ex :

la porte est ouverte) ou factifs (ex : le bateau coule). Les actes assertifs seront descriptifs (ex : je vous parle de dialogue) lorsque l'agent est le sujet. Ces descriptifs seront alors expressifs (ex : je suis triste) ou comportatifs (ex : j'ai ouvert la porte). Notons que les expressifs n'ont, pour Searle, pas de direction d'ajustement car il y a déjà congruence entre mot et monde et que l'on ne peut savoir si ce que l'on dit est vrai ou faux. Selon Vernant, on peut le savoir non pas par comparaison avec un état mental (qui reste invisible) mais par comparaison avec le comportement. Vanderveken omet les directions mot-mot et monde-monde mais insiste sur une autre direction, la direction vide d'ajustement. Cette direction contient les actes dont le but est expressif. En les accomplissant, le locuteur ne cherche pas à établir une quelconque correspondance entre les mots et le monde. Il exprime simplement son état mental (état, gratitude, orgueil, etc.).

A l'inverse, l'ajustement du **monde aux mots** contient les énoncés qui modifient le monde par les simples mots. Ces énoncés ont pour but de transformer le monde par l'action future du locuteur (dans le cas d'un engagement) ou de l'interlocuteur (dans le cas d'une directive) afin que ce monde corresponde au contenu propositionnel. On y retrouve ainsi les actes engageants qui seront soit **directifs** (ex : soyez attentifs) si l'agent est différent du sujet, soit commissifs (ex : je vous parlerai de dialogue) si l'agent est le sujet.

L'ajustement des **mots aux mots** est un ajustement métadiscursif. On y retrouve alors les citatifs (ex : vous m'avez demandé si...) lorsque l'agent est différent du sujet et les expositifs (ex : je conclurai en disant) lorsque l'agent est le sujet.

Enfin, le **double ajustement des mots par rapport au monde** contient les actes **déclaratifs** pour lesquels l'agent est différent du sujet. Le but de ces actes est de faire en sorte que le monde corresponde au contenu propositionnel en affirmant que ce contenu est vrai. Nous aurons alors les déclarations comme « le séminaire est terminé » qui ajustent le monde aux mots puisque par ce simple énoncé le séminaire se termine, mais aussi les mots au monde puisque le séminaire est effectivement terminé. Le fait même de l'utilisation de l'énoncé transforme le monde pour le faire correspondre aux mots.

L'ajustement du **monde au monde** n'étant pas du langage, il n'aura aucun sens ici.

Avec cette notion d'ajustement, Vernant propose un modèle projectif pour le dialogue. Dans celui-ci, les interlocuteurs communiquent en direction d'un but (**Figure 8**). Le dialogue est donc une série d'ajustements qui permettent de progresser vers un accord (la résolution d'un but). La convergence n'étant pas toujours possible, il offre la possibilité, pour chacun des participants, de diverger au point d'abandonner le but si le dialogue s'éloigne trop de la ligne directrice.

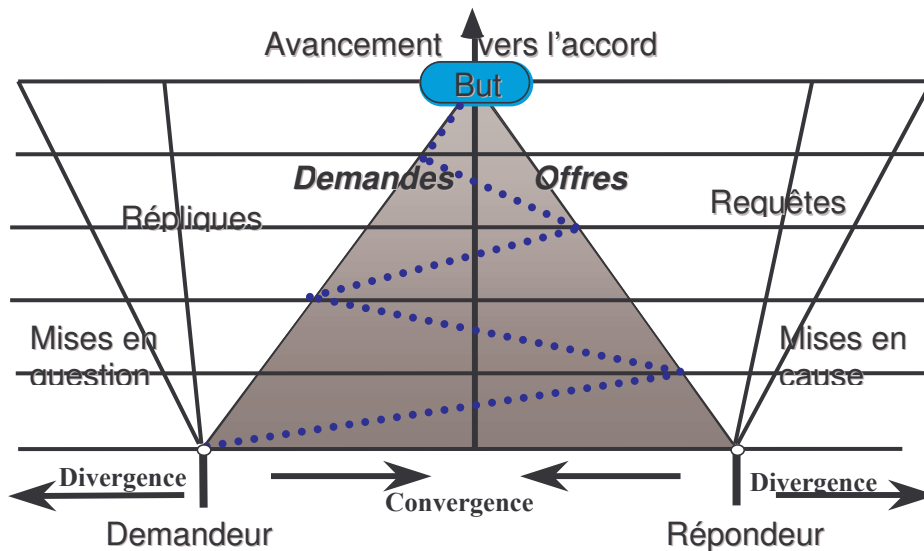


Figure 8 : Modèle projectif du dialogue informatif [Vernant 92]

### 1.6. Limites

Les théories d'origine philosophique ne sont pas linguistiques. Elles ne permettent donc pas de passer directement de l'analyse de surface des énoncés à la détermination de leur force illocutoire. Par exemple, les modes grammaticaux (indicatif impératif, interrogatif, optatif, etc.) ne sont pas des forces illocutoires. L'indicatif peut correspondre à un assertif (« il pleut »). Mais il peut correspondre également à un déclaratif (« la séance est ouverte »), un directif indirect (« il fait froid »), un expressif (« je suis triste »). Ainsi, la modalité est un indice mais ne suffit pas toujours à déterminer l'acte. Un directif n'est pas toujours exprimé par un mode interrogatif. Ainsi le travail d'interprétation consiste-t-il à analyser des indices linguistiques et d'autres indices afin de déterminer l'acte produit.

De plus, la logique illocutoire reste essentiellement monologique. Austin et Searle ne prennent pas en compte l'interaction. Or, un acte de discours ne prend sens que dans le contexte du dialogue, où la même forme peut recevoir plusieurs interprétations. Une extension est nécessaire pour aborder les problèmes du dialogue (notamment sa dynamique) et de l'intersubjectivité.

## 2. Ethnométhodologie

L'ethnométhodologie est une branche des ethnoscience qui étudie la communication dans une perspective sociale. L'idée sous-jacente est que les individus interagissent dans un cadre normalisé et suivent des règles et des conventions socioculturellement bien définies.

[Garfinkel, Heritage, Schegloff 50] ont montré que nos raisonnements sont socialement normalisés selon un principe d'identité. L'ethnométhodologie étudie donc l'interaction dans un monde social, c'est-à-dire avec des principes de nature sociale. Elle considère un raisonnement normalisé par le milieu socioculturel. Elle suppose ensuite des rôles qui vont induire des comportements et une hiérarchie sociale abordant le concept de dominance à travers des droits et des devoirs, les uns par rapport aux autres, que l'on peut outrepasser. Elle pose enfin le principe d'interaction par inter-compréhension à travers une connivence de nature sociale.

Ainsi, les interlocuteurs doivent partager des connaissances, se situer l'un par rapport à l'autre et agir pour faire une chose ensemble (convaincre, séduire, se valoriser, etc.). Cela implique donc une certaine motivation des deux parties, l'une pour dire quelque chose, l'autre pour montrer son écoute. Cette réciprocité des motivations est fondée sur l'intercompréhension. L'intercompréhension se constitue lorsque le locuteur A définit son projet d'action (son intention) en fonction de la réaction qu'il **attend** de son allocataire B afin de réaliser son but. Ainsi, si A propose une conversation à B, A anticipe que son projet de conversation, une fois compris, sera accepté par B comme la raison et la motivation de son action future (sa réponse). Il en va de même pour B au tour de parole suivant.

A court terme, cette réciprocité des motivations permet un réglage des tours de parole par paires adjacentes. Selon un principe d'action-réaction, si A pose une action de question, B proposera une réaction de réponse ou de réfutation de la question. Si un silence, un élément anormal ou non attendu se présente, d'autres types de raisonnement seront utilisés afin de vérifier des hypothèses (B est peut-être sourd, endormi ou grincheux).

A long terme, c'est une réciprocité des perspectives qui règle les niveaux supérieurs d'organisation de l'interaction liés à une conception hiérarchique de l'action. Ce principe fonde la complémentarité ou la symétrie des rôles des partenaires. De lui résulte la stratégie dans l'interaction, issue d'un accord entre les partenaires selon la théorie des faces proposée par Goffman. Selon [Goffman 73, 74], tout être social possède deux « faces ».

La **face négative** correspond à ce qu'il décrit comme "les territoires du moi" – « *territoire corporel, spatial ou temporel ; biens et réserves, matérielles ou cognitives.* ». Elle correspond à la face cachée, c'est à dire aux véritables caractéristiques de l'être.

La **face positive** correspond *grosso modo* « *au narcissisme, et à l'ensemble des images valorisantes que les interlocuteurs construisent et tentent d'imposer d'eux-mêmes dans l'interaction.* ». Elle correspond donc à la face que l'on montre de soi-même, c'est à dire l'image que l'on affiche à autrui.

Tout être social va donc, dans l'interaction, essayer de cacher sa face négative et de montrer sa face positive, et éventuellement de percer la face négative de l'interlocuteur.

Pour Goffman, « *l'idée selon laquelle on procède à une présentation de son moi aux autres n'est guère originale ; ce qu'il faut souligner en revanche, c'est que l'on peut analyser la structure même du moi en fonction des dispositions pour donner ces représentations* ». En jouant sur un principe de réversibilité, on peut s'avancer à dire que la représentation est une symbolisation des infrastructures des éléments participant à la mise en scène, et du système des interactions qui les lient. C'est ainsi que l'auteur distingue le personnage de l'acteur. Le personnage correspond au moi profond de l'individu. Dans l'acteur, il extériorise ce moi en l'adaptant à une mise en scène particulière en cherchant dans l'acte à protéger, préserver ou renforcer son moi profond. L'acteur est un échafaudage de la construction publique du personnage. Il place formellement le concept d'interaction au centre de son œuvre et le rend quasi-synonyme de relation. L'interaction est un processus de synchronisation de la relation entre plusieurs êtres. Chacun réévalue en permanence son comportement dans la situation, au regard de ce qui s'y passe. « *L'interactant socialisé traite l'interaction verbale comme n'importe quel autre type d'interaction, comme une chose qui mérite des précautions rituelles. C'est parce qu'il se réfère automatiquement à la face qu'il sait comment se conduire vis-à-vis d'une conversation. C'est en se demandant sans cesse et à tout coup : 'est-ce que, en faisant ou en ne faisant pas cela, je risque de perdre la face ou de la faire perdre aux autres ?' qu'il décide à chaque moment, consciemment ou non, de sa conduite.* »

Selon le degré de maîtrise du code rituel, il est possible de comprendre la signification des propos tenus par les acteurs des entretiens. En effet, un acteur ne maîtrisant pas le code rituel aura tendance à être plus spontané, mais dans le même temps, cette absence de maîtrise liée à l'obsession de la reconnaissance et de la face orientera son discours pour tenter de cacher sa méconnaissance.



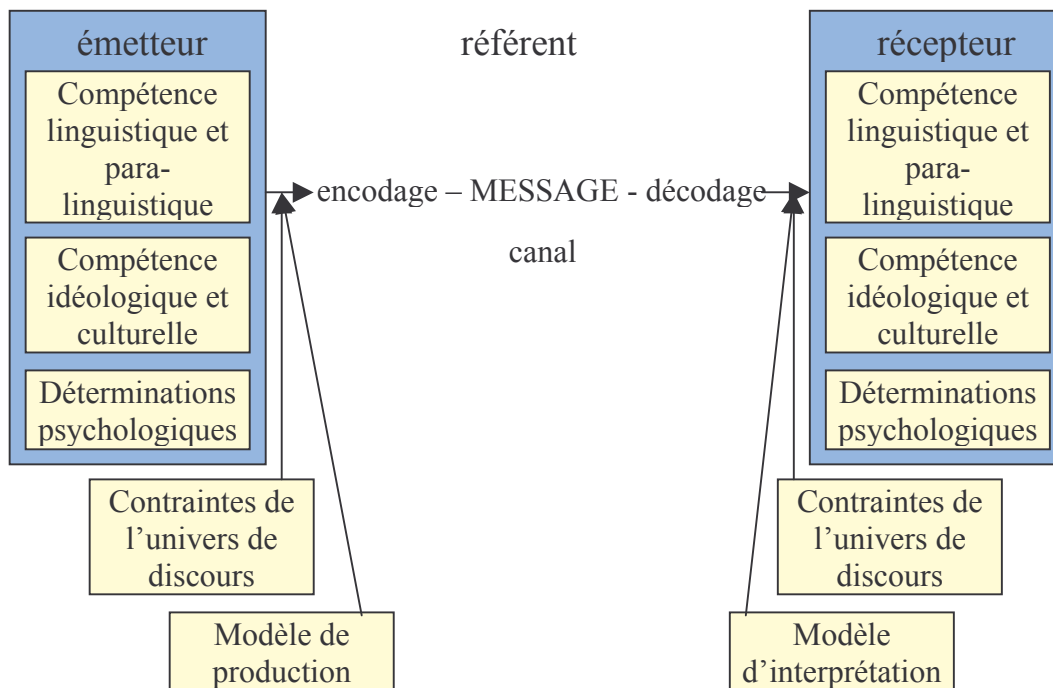
L'objectif de Goffman est alors d'isoler certains cadres de base de la compréhension [Goffman 91], en faisant l'hypothèse qu'il existe un nombre limité de schémas d'interprétation primaires et de transformations possibles de ces cadres de base, communs à ceux qui vivent dans une même culture. Il distingue actes et intentions, cadres de base naturels, physiques, non intentionnels et non animés, et cadres sociaux, intentionnels, incluant volonté, finalité, intelligence et humanité. Ces cadres se transforment continuellement, modifiant l'interprétation en contexte. La distinction pourra donc être faite entre le réel et la simulation, par exemple entre la vraie mise en garde et la blague. Elle permet donc d'appréhender la dissimulation, le faire-semblant, l'ironie ou la mauvaise foi, c'est-à-dire la coopération et le refus de coopération. En revanche, une divergence d'interprétation, différence de cadrage, peut aboutir aux incidents, quiproquos, etc. Ces divergences sont d'autant plus fréquentes que la part du non-dit est grande car ce non-dit englobe le sous-entendu, l'implicite, l'allusion, les habitudes partagées, les savoirs tacites, etc.

La notion de face renvoie à celle de rôle, c'est à dire à la position d'où on parle et que l'on doit conserver. Suchman [Suchman 87] va plus loin encore en introduisant la notion d'action située. Dans un tel type d'action, l'accent est mis sur la prise de décision en situation, faisant par cela intervenir plus fortement le contexte. Elle montre ainsi que par la grande dynamique du dialogue, les interactants s'adaptent et ajustent en permanence, ce qui rend très difficile la prédiction d'un acte de parole par un tiers, car ce tiers n'a pas accès à la double interprétation du locuteur et de l'allocutaire. Elle fonde alors sa description de la conversation sur des stratégies opportunistes, qui ne sont *instanciables* que localement et en tous cas non planifiables à long terme. Par cela, elle remet en cause un grand nombre de théories fondées de près ou de loin sur la logique des intentions. Malheureusement, son apport, très riche par ailleurs, reste plus critique que constructif, car elle ne propose pas de modèle formalisable pour le dialogue.

### **3. Psychologie cognitive – Cognisciences**

Les cognisciences s'intéressent au fonctionnement de l'individu, c'est-à-dire, dans la communication, aux aspects liés à la perception, à la planification et au raisonnement cognitifs. Les premiers modèles dérivent de la théorie du code de Shannon et Weaver [Shannon 59]. Cette théorie suppose une phase de codage puis de décodage des informations linguistiques et extra-linguistiques. Ce codage serait issu de conventions, de connaissances partagées et de suppositions mutuelles. Le modèle de

Kerbrat-Orrechioni, dérivé des travaux de Jakobson [Jakobson 60], reprend ce principe en complétant les phases de codage et décodage (**Figure 9**).



**Figure 9** : Modèle de communication humaine, [Kerbrat-Orrechioni 80]

L'information est encodée sous forme de message et part de l'émetteur pour aller vers le récepteur via un canal de transmission. Cette information est décodée par le récepteur après transmission. Ainsi, certaines connaissances entrant en jeu en émission et réception du message sont partagées. En particulier, les connaissances linguistiques et paralinguistiques (gestes, attitudes, etc.) mais aussi culturelles, idéologiques. L'univers du discours est le même pour les deux parties. Ainsi, un émetteur dans un état d'embarras adaptera son message à cet état et ce message ne pourra être décodé et interprété correctement qu'à la condition de connaître l'état de l'émetteur. L'effort cognitif semble être aussi important pour coder que pour décoder. L'émetteur formule au mieux son idée tandis que le récepteur choisit la signification adaptée au contexte.

C'est sur cette formulation que nous nous pencherons maintenant à travers deux approches successives : Grice puis Sperber & Wilson. Ces deux approches ont un principe commun qui est que le locuteur choisit ses mots en fonction des destinataires à qui il a affaire. Il choisit la manière la plus pertinente de dialoguer pour une raison pratique, optimiser l'efficacité du dialogue. C'est une conséquence de la rationalité.

### 3.1. La coopérativité selon Grice

[Grice 68, 75] introduit la conversation (le dialogue i.e. toute forme d'interaction communicationnelle) comme objet propre. Selon lui, les individus sont des agents rationnels, donc toute forme de conversation est une activité rationnelle. Un échange dialogique suit donc une certaine logique fondée sur un principe général de coopérativité. « *Un locuteur L, par l'énonciation de x, a l'intention de produire un effet r sur un récepteur R grâce à la reconnaissance de cette intention* » [Grice 68]. Ainsi interviennent les notions d'intention et de buts reconnus et acceptés par les interlocuteurs. Grice définit quatre maximes de coopérativité à l'intention des interlocuteurs. Ainsi, s'ils veulent coopérer, les interlocuteurs doivent être sincères et pertinents. Voici ces quatre maximes.

La **maxime de qualité** définit la sincérité du locuteur à travers trois conseils : « *Que votre contribution soit véridique* » ou plus simplement dites le Vrai. « *N'affirmez pas ce que vous croyez être faux* » et « *N'affirmez pas ce pour quoi vous manquez de preuves* ». Cette maxime correspond au principe de sincérité de Searle.

La **maxime de quantité** définit la pertinence du propos, comme le fait de dire ce qu'il faut et pas plus : « *Que votre contribution contienne autant d'information qu'il est requis (pour les visées conjoncturelles de l'échange)* » et « *Que votre contribution ne contienne pas plus d'information qu'il n'est requis* ». Ainsi, le locuteur coopératif se devra d'être aussi informatif qu'il le faut mais pas plus qu'il ne faut. Il donnera la quantité d'information nécessaire et suffisante (suffisante pour éviter les dérives susceptibles de dévier l'échange vers des points de détail ou les effets indirects de penser qu'il y a une raison particulière à un tel excès d'information).

La **maxime de relation** définit l'adéquation du propos avec une seule règle : « *Parlez à propos* ». En effet, on ne peut changer de sujet que sous certaines conditions (on ne passe pas du coq à l'âne). Cette règle, pourtant concise, cache de nombreux verrous liés à la définition des différents genres et centres de pertinence possibles ou leur modification au cours d'un échange.

La **maxime de modalité** définit la clarté du propos (comment on doit dire ce que l'on dit) avec un simple conseil : « *Soyez clair* », « *Evitez de vous exprimer avec obscurité* », « *Evitez d'être ambigu* ». Ainsi, le locuteur coopératif se devra d'être le plus clair possible.

Suivre ces règles permet de se montrer coopératif tandis que les transgresser montre un refus de coopérer. Ainsi, l'ironie ou le mensonge ne sont pas coopératifs, et affirmer une chose dont on n'est pas sûr ne l'est pas non plus. L'application et la non-application de ces règles permettent un large éventail de jeux possibles avec celles-ci. Le locuteur peut alors refuser de jouer le jeu conversationnel (« - Pouvez-vous me donner l'heure ? - Je n'ai pas le temps »). Il peut être amené à violer involontairement une règle lorsqu'il n'arrive pas à toutes les concilier (on peut donner trop peu d'information pour ne pas dire plus que ce que l'on sait). Il peut également violer délibérément une règle introduisant une stratégie indirecte de communication (on peut faire comprendre ce que l'on veut dire sans devoir le dire). C'est ce que Grice nomme *implicature* conversationnelle, et qui permet un écart entre ce qui est communiqué (pragmatique) et ce qui est dit (sémantique). Le sens pragmatique n'est pas seulement un sens décodé (sémantique), mais il contient également un sens inférentiel (les actes indirects, les implicatures). Le problème de l'implicite peut être abordé avec l'exemple suivant :

A : ma voiture est en panne

B : il y a un garage à l'entrée du village

B propose une réplique appropriée car il interprète la déclaration de A comme une requête indirecte. Mais cette réplique ne sera appropriée que si A l'interprète comme une réponse appropriée à sa question déguisée. Mais B enfreindrait la règle de pertinence s'il pensait le garage fermé. Il dit donc implicitement que le garage est ouvert. En effet, il n'y a pas lieu de supposer qu'il n'observe pas les règles et supposer que le garage est ouvert est nécessaire pour donner sa réplique sans enfreindre la règle de pertinence. De plus, il pense que son interlocuteur peut déduire que le garage est ouvert, donc il ne l'explique pas.

Ces maximes, proposant des principes non pas moraux mais rationnels, définissent la communication coopérative comme étant la conséquence de la rationalité humaine. Elles sont fondées sur une raison pratique pour augmenter l'efficacité du dialogue. Grice introduit alors le terme d'**implicature**. Il le définit comme l'ensemble des conséquences de l'application ou du viol de ces maximes, incluant l'usage du langage mais aussi les ellipses, les non-dits, les implicites, etc.

De l'énoncé « le projet a quatre partenaires », l'application de la maxime de quantité (donner autant d'information qu'il est requis) permet d'impliciter que « le projet a exactement quatre partenaires ». De la suite d'énoncés « Je voudrais un billet pour Grenoble pour demain, départ vers 23h00 de Lyon », « Attention, le dernier tramway est à minuit », l'application de la maxime de relation permet au locuteur

d'inférer que son allocataire a donné une information pertinente à savoir que le train n'arrivera pas avant minuit et qu'il n'y aura plus de tramway pour se déplacer une fois arrivé à la gare de Grenoble. Une réponse vague, viol de la maxime de quantité, peut s'expliquer par le manque d'information et l'évitement du viol de la maxime de qualité. Une telle réponse permet donc d'impliciter le manque d'information. A l'inverse, les figures de rhétorique telles que l'ironie, la litote ou la métaphore sont des viols de la maxime de qualité puisque les énoncés seront faux, mais on peut aussi violer la maxime de quantité dans le but d'impliciter autre chose que l'énoncé, comme dans l'exemple de Grice [Grice 75] :

*« A rédige une recommandation pour un de ses élèves qui postule un poste en philosophie, et voici le contenu de la lettre : "Monsieur, M. X a une maîtrise remarquable de l'anglais, et il a été assidu à mes cours. Je vous prie d'agréer, etc." (Glose : A ne refuse pas de jouer le jeu, puisque, s'il ne voulait pas coopérer, il pourrait fort bien ne pas écrire du tout. Il a parfaitement la possibilité d'en dire plus, puisque celui qu'il recommande est un de ses étudiants; par ailleurs, il n'ignore pas qu'on lui demande plus d'information que cela. Il faut donc qu'il soit supposé vouloir transmettre une information qu'il répugne à exprimer noir sur blanc. Cette supposition ne tient que si l'on présume qu'il pense que M. X ne vaut rien en philosophie. C'est donc cela qu'il implicite.) »*

Ainsi, ces implicatures dépendent fortement du contexte et des locuteurs par le biais de leurs connaissances partagées, explicitement liées au contexte social, à des attentes réciproques, à des conventions d'usage, et non aux seuls mécanismes mentaux, largement décontextualisés, auxquels se réfère le programme cognitiviste.

Cette analyse va donc plus loin que les simples implications logiques car les conclusions que l'on peut tirer de l'énoncé dépendent fortement du contexte. Une proposition p implique une proposition q si et seulement si chaque situation qui vérifie p vérifie q. En revanche, une proposition p présuppose q si la simple assertion de p ou non p implique la vérité de q. Par exemple, « ça manque de sel » suppose que ce fait est vrai mais aussi que le locuteur souhaite le sel. **L'implicature** est donc fortement liée à l'idée de **présupposition**.

Frege [Frege 71] postule qu'une proposition contenant un nom propre ou une référence définie véhicule la présupposition que ce nom dénote quelque chose, sans que cette présupposition fasse partie des conditions de vérité de la proposition. Pour qu'une assertion soit vraie ou fausse, ses présuppositions doivent être vraies ou satisfaites.

Pour [Moeschler & Reboul 94], il est important de distinguer présuppositions, implications et assertions.

**La sémantique** oppose la présupposition à l'implication. Une proposition *P* **implique sémantiquement** une proposition *Q* si et seulement si, chaque situation qui rend *P* vraie rend *Q* vraie. Une proposition *P* **présuppose sémantiquement** *Q* si la simple assertion de *P* aussi bien que de non *P* implique la vérité de *Q*.

[Stalnaker 77] formule une **définition pragmatique de la présupposition** :

*« Une proposition P est une présupposition pragmatique d'un locuteur dans un contexte donné si le locuteur assume ou croit que P, assume ou croit que son interlocuteur assume ou croit que P, et assume ou croit que son interlocuteur reconnaît qu'il fait ces hypothèses, ou a ces croyances. »*

Ainsi, à un énoncé du type « Veux-tu du café ? », une réponse directe « Non, je n'en veux pas » est explicite. Mais une réponse « Je ne veux pas d'excitant » ou « je veux pouvoir dormir » n'est pas explicite. La réponse indirecte associée (« je ne veux pas de café ») pourra être implicite à partir d'un raisonnement issu de l'énoncé (« je ne veux pas d'excitant » ou « je veux dormir » mais aussi des connaissances partagées (« le café est un excitant », « le café empêche de dormir »).

### **3.2. La pertinence selon Sperber & Wilson**

Les maximes de Grice ont été regroupées par Sperber & Wilson à travers un seul principe de pertinence, adoptant une notion d'inférence dérivée de l'implicature.

D'une manière générale, les processus cognitifs humains sont organisés de façon à produire les effets cognitifs les plus grands possibles au prix d'un effort mental le plus réduit possible. Pour cela, l'individu doit porter son attention sur les informations les plus pertinentes parmi toutes celles dont il dispose, afin de minimiser le coût du traitement cognitif de son destinataire. Sperber et Wilson s'opposent alors au modèle du code présenté précédemment, le jugeant trop simpliste. Pour eux [Sperber 00], le modèle du code se résume en trois thèses :

1. Un code associe à tout sens une expression, et à toute expression un sens.
2. Les langues humaines sont des codes.
3. Le locuteur encode au moyen d'une expression le sens qu'il veut communiquer. L'auditeur décode l'expression et identifie ainsi le sens voulu selon des conventions, des connaissances partagées et des suppositions mutuelles.

Or le sens linguistique sous-détermine le vouloir-dire du locuteur. L'énoncé littéral ne suffit pas à faire passer le sens que veut faire passer le locuteur. Le processus d'inférence proposé par Grice est donc nécessaire. L'auditeur ne se contente pas de décoder le sens linguistique (voire les informations extralinguistiques) de l'énoncé, il infère le sens voulu par le locuteur. Il l'infère à partir de deux types d'éléments, qui sont d'une part ce sens linguistique, et d'autre part le contexte. Une inférence est un processus qui part de prémisses et aboutit à une conclusion. En ce sens, elle fonctionne comme un raisonnement, mais elle peut opérer de façon automatique et inconsciente. Dans ce processus d'inférence, l'auditeur est guidé par des considérations de pertinence.

L'hypothèse du modèle inférentiel de la communication humaine se résume en trois thèses :

1. Le communicateur produit un **indice** du sens voulu.
2. Le destinataire, dans tous les cas, **infère** ce sens voulu à partir de l'indice fourni et du contexte.
3. Un énoncé linguistique est un indice complexe du sens voulu par locuteur. Ce n'est pas un encodage de ce sens voulu.

Selon Sperber & Wilson, le contexte vient de trois sources :

1. L'interprétation des énoncés précédents.
2. L'environnement physique.
3. La mémoire à long terme vue comme une suite structurée de concepts (avec une adresse). Chaque adresse contient des « *relations logiques qu'un concept peut entretenir avec d'autres concepts (implications, contradictions, etc.)* », des informations encyclopédiques regroupant « *toutes les connaissances qu'a le sujet et qui tombent ou qui sont censées tomber sous le concept (elles permettent de préciser l'extension du concept)* » et des informations lexicales correspondant aux « *contreparties en langue naturelle du concept* ».

[Sperber & Wilson 90] appellent alors « principe de pertinence » l'idée fondamentale selon laquelle l'indice communiqué est assorti d'une garantie de pertinence. Ce principe de pertinence cognitive semble jouer un rôle essentiel dans la communication humaine en permettant d'orienter des processus inférentiels. Il résulte du rapport entre l'effet cognitif produit chez l'allocutaire et l'effort qu'il a fallu pour le produire. Il nécessite, comme pour Grice, les notions d'intention et de but reconnus et acceptés par les interlocuteurs. La communication a donc pour objet un ensemble

d'actions en vue de réaliser un but explicite (partagé) ou implicite (non partagé). Si ce but est implicite, chacun suit son propre but jusqu'à explicitation ou négociation pour atteindre, après d'éventuels réajustements, un but commun.

Les auteurs reprennent les maximes de quantité de Grice pour définir un modèle inférentiel fondé sur une certaine réciprocité des motivations logiques. Lorsque A dit P à B, B reconnaît l'intention de A contenue dans P, donc B répond R à A, et R contient l'intention de B. Par exemple, lorsque Anne dit « ferme la porte » à Bertrand, Bertrand reconnaît l'intention de Anne contenue dans cette phrase, il répond « d'accord » à Anne et cette réponse contient l'intention de Bertrand, celle de répondre favorablement à la demande d'Anne. Ainsi, Bertrand fait un choix pertinent.

Sperber & Wilson définissent alors la notion de choix pertinent. Un choix pertinent est un choix prenant en compte le monde d'arrière-plan mais aussi les environnements cognitifs des conversants. Ce choix, qui porte alors aussi bien sur l'individu que sur la situation ou le contexte, se fait par sélection d'éléments saillants afin de focaliser l'attention de l'allocutaire. Ces éléments, appelés faits, peuvent être perçus dans la situation comme dans le monde d'arrière-plan ou inférés par hypothèses.

Les auteurs définissent ainsi un processus de communication comme échange d'indices pertinents (afin de pouvoir servir à retrouver l'intention du locuteur) par présentation ostensive de faits liés à des intentions informative et communicative, l'acte d'ostension consistant à présenter un fait saillant ou rendre saillant un fait au moyen d'indices. Pour eux, la communication ostensive-inférentielle met donc en jeu deux intentions.

**L'intention informative** est l'intention d'informer l'allocutaire de quelque chose, notamment qu'il attend une réaction de celui-ci et que cette réaction doit être appropriée, c'est-à-dire qu'elle doit permettre de réaliser son but.

**L'intention communicative** est celle d'informer l'allocutaire de cette intention informative.

À la suite de Bange, [Lemeunier 01] élude cette distinction en considérant que « " *l'intention informative* " se réalise par " *l'intention communicative* " [...] c'est-à-dire que, dans une perspective actionnelle, elle a une place subordonnée par rapport au but visé. » [Bange 92] (p. 200). Dans tous les cas, les interlocuteurs doivent se montrer coopérants pour que la communication réussisse. La coopération serait donc une résultante cognitive plus que sociale.



### 3.3. Une formalisation : DRT & SDRT

Il est nécessaire d'aller au-delà d'une simple compréhension d'énoncés pour parvenir à une interprétation contextuelle en fonction, entre autres, de l'historique du dialogue et de la tâche, des intentions et présuppositions de l'utilisateur. La référence est un phénomène vaste et complexe, susceptible de couvrir le champ de nombreuses disciplines. Il s'agit de comprendre par quels mécanismes une expression linguistique réfère, c'est à dire permet d'accéder à des objets ou des événements construits ou extérieurs au discours dont cette expression est issue. Par exemple :

«A : Je peux vous proposer la salle Lafayette ou celle de 30 places

C : Réserve donc la seconde »

On peut observer une variété de phénomènes référentiels comme une référence directe à une salle à l'aide de son nom, ou la reprise d'un référent qui vient juste d'être mentionné (la seconde).

Les études de [Carbonell et al. 83] ont montré que nous employons volontiers des références pronominales en langue naturelle. Ces procédés anaphoriques<sup>2</sup>, tout comme les ellipses<sup>3</sup> ou les tropes<sup>4</sup> assurent la concision et la souplesse d'une langue naturelle. Il est donc intéressant, pour vérifier la validité de notre système de vérifier que ce corpus de dialogue homme/machine en langue naturelle recèle de tels éléments. A ce propos, Jacquet écrit « *On peut même penser que la perception d'autrui qui fonde l'une des régulations essentielles du dialogue est irréductible à la « perception » dont une machine est capable. (...) Cela dit, cette machine (faut-il le rappeler ?) est différente d'un autrui. (...) Peut-être que l'une des conditions indispensables au bon fonctionnement de l'illusion est le transfert des pronoms personnels ? [Jacquet 97]».*

Dans les applications de communication homme-machine, la pragmatique joue donc un rôle très important dans l'interaction entre l'homme et l'univers de l'application et pour interpréter complètement l'énoncé. Elle considère l'énoncé et son contexte. Elle permet notamment de résoudre les problèmes référentiels.

La référence repose sur une large palette de marqueurs possibles. On y retrouve les déterminants indéfinis (un, une, des), définis (le, la, les) ou démonstratifs (ce, cet, cette, ces), les adjectifs possessifs (son, sa, ses), les pronoms personnels (il,

---

<sup>2</sup> L'anaphore est un mot renvoyant à un membre de phrase déjà énoncé.

<sup>3</sup> L'ellipse est l'omission d'un ou de plusieurs mots d'un énoncé sans altération du sens de l'énoncé.

<sup>4</sup> Un trope est une figure rhétorique consistant à employer un mot ou une expression en les détournant de leur sens propre.

elle, ils, elles) et les pronoms possessifs (le mien, le tien, le sien) ou démonstratifs (celui-ci, celui de gauche). La façon d'interpréter ces formes correspond à des modes référentiels variables. On en distingue deux.

L'**anaphore** consiste à faire reposer l'interprétation d'une expression référentielle donnée sur un élément antérieur du dialogue. C'est le cas typique du pronom personnel à la troisième personne dont l'interprétation repose sur l'existence d'un antécédent.

Le **déixis** considère l'interprétation de cette expression en fonction du contexte immédiat d'énonciation, notamment le positionnement dans le temps (maintenant) ou dans l'espace (ici) du locuteur (je). C'est le cas de l'association du geste à la parole pour effectuer des références multimodales (pointage avec le doigt, croix ou cercle sur une carte, etc.).

La référence aux objets dans un dialogue homme-machine correspond à l'identification des éléments de la tâche sur lesquels doivent porter les actions ou interrogations exprimées par le locuteur. De manière générale, cette identification s'effectue dans le cadre d'un ensemble de connaissances partagées par le locuteur et le système de dialogue concernant l'existence effective des objets correspondant. L'interprétation traite donc les déictiques ou indexicaux<sup>5</sup>, les anaphores (co-référence), les ellipses, les phrases incomplètes, les présuppositions et des implicatures (avec l'aspect inférentiel : tout ce qui n'est pas codé dans la langue). Pour traiter tous ces points, l'interprétation de l'énoncé doit être reliée au contexte du dialogue par la construction dynamique de ce contexte au fil des énoncés.

La théorie de la représentation du discours (DRT) de [Kamp & Reyle 93] cherche à construire une représentation incrémentale (la structure de représentation du discours – DRS) distinguant l'ensemble des variables nécessaires à l'identification des référents et les contraintes portant sur ces variables. Ainsi, un nouvel énoncé dans le dialogue met à jour les référents du discours en fonction des expressions référentielles contenues dans l'énoncé. Il tente de résoudre les relations anaphoriques par le biais d'équations d'égalité. Il ajoute enfin de nouvelles contraintes associées aux prédicats nominaux ou verbaux contenus dans l'énoncé. Une DRS peut être traduite en une représentation logique du premier ordre. Elle ne permet pas de représenter les relations rhétoriques dans le discours. Par exemple :

---

<sup>5</sup> ensemble des termes qui tirent leur signification, au moins en partie, de la situation de la communication ; ce sont par exemple les pronoms de la première et deuxième personne, les références temporelles comme *demain*, *la semaine prochaine*, les références spatiales comme *ici*, *là*, etc.

- (1) Anne est venue hier
- (2) Elle a parlé longtemps
- (3) Elle a parlé de sa fille
- (4) Elle a parlé de Bertrand
- (5) Puis elle est partie

Une représentation en DRT peut conduire à une DRS du genre :

[x,y,z,u,v,w : x=Anne, y=venue, z= longtemps, u=sa fille, v=Bertrand, w=partie, etc.]

Dans cette représentation, le discours peut se poursuivre par un énoncé du type :

- (6) Elle a bien grandi

En effet, le référent pour « sa fille », u, est encore accessible après l'énoncé (5). Or, dans des conditions normales de discours, (6) ne serait pas autorisé en raison de l'organisation hiérarchique de ce discours, non prise en compte par la DRT. Il semble donc nécessaire de considérer les relations rhétoriques entre les énoncés afin de représenter l'enchaînement logique du discours. Dans cet exemple, (5) suit (2), (4) suit (3), mais (3) ne suit pas simplement (2). En fait, si X est la DRS représentant la succession (3)→(4), X peut être considérée comme subordonnée par rapport à (2) (qui se retrouve englobante). De même pour Y représentant la succession (2)→(5), englobée par (1).

Asher [Asher 96] et Lascarides [Asher & Lascarides 03] proposent d'encapsuler ces structures dans une structure plus globale du dialogue grâce à un processus dynamique de construction du sens. Ils introduisent alors la théorie formelle de représentation du discours structuré (SDRT), fondée sur la DRT, et représentant la structure du dialogue comme un composant essentiel à l'interprétation du discours et résultant de l'intégration des informations pragmatiques et sémantiques. La SDRT permet alors de décrire et d'expliquer de nombreux effets rhétoriques comme l'anaphore, l'ellipse, les ambiguïtés lexicales, et aussi les implicatures et les présuppositions.

[Asher & Lascarides 02] modélisent les buts associés à l'acte de langage. Pour eux, le but d'une question est de connaître la réponse. La finalité d'une requête est que celle-ci soit satisfaite. Mais la notion de but est différente de celle d'intention car le but est une propriété de l'énoncé là où l'intention est une propriété du locuteur.

[Xuereb 04] propose alors (travaux en cours) une application de la SDRT pour le dialogue homme-machine.

## 4. Linguistique

Les conversations de salon n'ont pas de but précis. Elles ont pourtant un but phatique ou communicationnel. Elles instaurent une relation entre les interlocuteurs et ont donc un rôle social non négligeable. Laissant de côté les aspects sociaux du dialogue, les linguistes ont davantage cherché à rendre compte de la dynamique du dialogue à travers des modèles structurels.

Les premiers modèles élaborés par l'école de Genève (Roulet [Roulet 85] et Moeschler [Moeschler 89]) proposent une structure hiérarchique et fonctionnelle des dialogues. Ceux-ci voient le discours comme cadre de structuration d'échanges linguistiques, considérant que ces structures sont largement déterminées par les contraintes intrinsèques au dialogue. Ils évitent ainsi les notions d'intention ou de présupposition qu'ils considèrent trop subjectives.

Ils élaborent donc *a posteriori* la structure de dialogues entre humains en utilisant des liens de composition, de succession, d'imbrication ou de subordination entre les différents constituants. Ils s'inspirent de la linguistique, mais aussi de la philosophie précédemment décrite (Austin, Searle, Grice) et de l'ethnométhodologie (Goffman) pour proposer une structure à quatre types de constituants.

Les **incursions** (ou transactions pour Bilange) sont les plus grandes unités de dialogue. Elles contiennent généralement un échange d'ouverture, une série d'échanges puis un échange de clôture et définissent ainsi les grandes phases du dialogue.

Les **échanges** constituent les plus petites unités de dialogue. Ils contiennent deux à trois interventions avec changement de locuteur. Pendant cette suite de tours de parole, un but discursif posé en début d'échange est maintenu jusqu'à sa satisfaction ou son abandon. Les auteurs distinguent les échanges confirmatifs pour l'ouverture et la clôture, et les échanges réparateurs (trois interventions). Ces derniers servent à neutraliser la menace potentielle (véhiculée par la première intervention [Goffman 73]) pour la transaction elle-même. Cette menace pour la face négative de l'interlocuteur pourra être neutralisée par ces échanges réparateurs. Cependant, les échanges confirmatifs et réparateurs peuvent ne contenir qu'une intervention lorsqu'il n'y a pas place à une réponse, lorsque la réponse est implicite ou lorsque la réponse est non-verbale [Coulthard 79].

Les **interventions** sont les constituants de l'échange. Sous-discours à fonction thématique, elles sont composées d'actes de langage produits par un seul

locuteur (tour de parole). Une intervention est initiative lorsqu'elle ouvre un échange ou rouvre un échange mis en attente. Mettant l'allocutaire en position de réaction, elle est généralement suivie d'une intervention réactive, elle-même suivie d'une évaluative (positive lorsque l'on s'avance vers le but, négative lorsque l'on s'en écarte) ou d'une clôture lorsque le but est atteint. La fonction évaluative a été rajoutée par Bilange [Bilange 92] et sert « à évaluer l'échange auquel elle appartient et donc implicitement à amorcer la clôture de l'échange »

Les **actes de langage** correspondent aux actes de parole de Searle [Searle 72] et constituent l'unité minimale d'énonciation (unité élémentaire de l'action). Ils sont donc le grain le plus fin permettant une analyse propositionnelle fine en termes de buts, intentions, etc. Selon les axes éponymes proposés par Bilange [Bilange 92], ils peuvent être directeurs lorsqu'ils explicitent le but du locuteur ou subordonnés lorsqu'ils explicitent son argumentation.

En résumé, Roulet & Moeschler proposent une grammaire de la forme :

Incursion -> Echange<sub>ouverture</sub> Echange<sub>continuation</sub>\* Echange<sub>clôture</sub>

Echange -> intervention<sub>initiative</sub> intervention<sub>réactive</sub> (intervention<sub>évaluative</sub>)

Intervention -> (acte<sub>subordonné</sub>) acte<sub>directeur</sub> (acte<sub>subordonné</sub>)

Cette grammaire définit les principales phases du dialogue. Elle omet les phases particulières d'abandon temporaire (incidence) ou définitif (rupture) qui sont considérés comme des échanges. L'incidence initialise un sous-but et met en attente le précédent (et son échange associé) tandis que la rupture abandonne le but précédent pour initier un sous-dialogue.

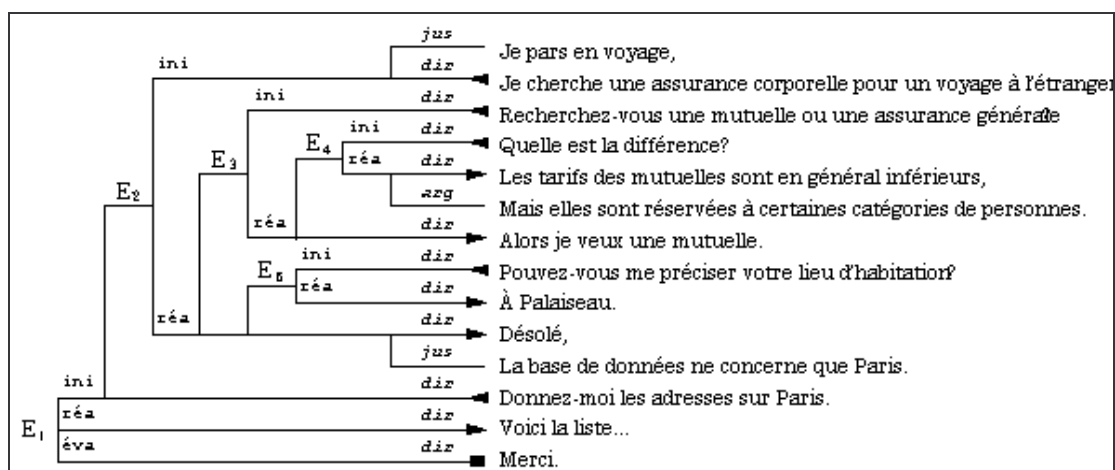


Figure 10 : Exemple d'analyse conversationnelle, modèle genevois [Lehuen 97].

L'inconvénient de cette approche descriptive réside dans son analyse *a posteriori* nécessitant le dialogue complet avant analyse. Comme la machine doit pouvoir répondre en temps réel, Luzzati propose un modèle dynamique dont l'idée principale réside dans la problématique des incidences qu'un dialogue homme-machine doit éviter. Considérant que la communication homme-machine ne produit que des dialogues à structures hiérarchiques, il propose donc une représentation calculatoire de la structure du dialogue tenant compte des incidences.

Il développe un modèle à deux axes. L'axe **régissant** (horizontal) correspond à la résolution de la tâche par les participants au dialogue. L'axe **incident** (vertical) correspond aux moments où, par incompréhension ou manque d'information, les interlocuteurs s'engagent dans des sous dialogues de clarification, précision, explication, reformulation, etc. Le but sera alors de retourner au plus vite sur l'axe régissant. Sur cet axe incident se retrouvent donc les incidences du dialogue mais aussi la rupture lorsque l'écart devient trop important pour réussir à terminer le dialogue avec la résolution de la tâche. Parfois les quiproquos entre l'homme et la machine sont tels que l'incompréhension est totale. Eviter le quiproquo nécessite donc d'instaurer des règles de conversation et des stratégies (ex : revenir au départ au bout d'une certaine incidence).

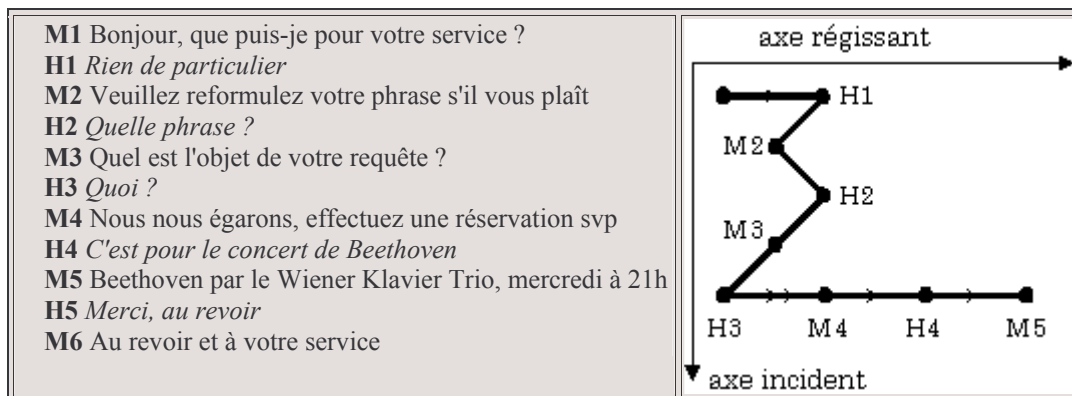


Figure 11 : Exemple de dialogue géré par le logiciel MINIDIAL [Lehuen 97]

Sur la **Figure 11**, une première incidence en M2 (demande de reformulation) est suivie d'une autre incidence H2 (demande de précision) reformulée en M3 et H3 (nouvelle incompréhension de M3) et enfin un retour express sur l'axe régissant.

Il distingue donc les questions et réponses principales et secondaires régissantes et les questions et réponses incidentes et cherche à répondre aux erreurs de compréhension à répétition qui finissent par bloquer le dialogue. Il contrôle alors des variables interactionnelles, qui mesurent les écarts à l'axe régissant, et force le retour sur cet axe en cas de trop fort éloignement.

## Conclusion : ce qui est utile aux attentes

Nous avons vu dans ce chapitre de nombreuses raisons de s'intéresser aux attentes des locuteurs dans le dialogue. Diverses théories peuvent être combinées pour construire une notion d'attente pour le dialogue. Nous nous appuyerons donc sur des travaux issus de différents domaines tels que la sémiotique, la philosophie analytique, la logique, l'ethnométhodologie, les cognisciences et la linguistique.

La sémiotique nous apportera la notation des actes avec les modaux de Greimas (faire, savoir, être, avoir, devoir, pouvoir, vouloir, etc.). La combinaison permet des actes de la forme faire, faire-savoir, faire-faire, faire-faire-savoir, faire-pouvoir, faire-devoir, etc. Comme nous nous plaçons dans l'action, les modalités 'vouloir', 'être' et 'avoir' ne permettront pas de représenter des actes, mais elles seront utilisées dans le dialogue. Le vouloir (et ses dérivés, vouloir-faire, vouloir-savoir, vouloir-avoir, vouloir-être) représentera par exemple l'intention du locuteur.

Ontique	Actionnelle & volitive	épistémique	déontique
<del>avoir</del>	faire	savoir	devoir
être	<del>vouloir</del>	croire	pouvoir

**Tableau 5** : Modalités de Greimas pour nos attentes

La philosophie du langage, par son approche pragmatique, nous permettra de représenter les attentes sous forme d'actes de dialogue dérivés des actes de langage. Nous avons vu dans ce chapitre l'intérêt d'une telle approche pragmatique par la nécessité de prendre le contexte en considération (notamment pour les actes indirects). Nous nous intéresserons en particulier à la force illocutoire, car c'est sur elles que porteront, selon nous, les attentes (le contenu propositionnel étant fixé). C'est elle qui nous permettra de considérer les attentes du côté locuteur et ainsi indépendamment de la tâche. Nous pensons par ailleurs que l'acte perlocutoire correspond à l'attente principale du locuteur.

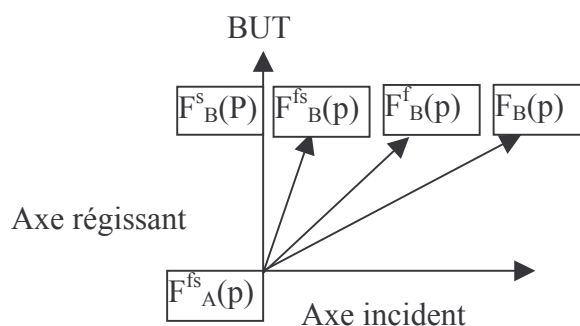
La notion de direction d'ajustement nous intéressera également, car selon nous, l'ajustement monde  $\rightarrow$  mots et le double ajustement posent des attentes. A l'opposé, les directions mots  $\rightarrow$  mots (méta-discursif) et mots  $\rightarrow$  monde (descriptif) ne posent pas d'attentes.

L'ethnométhodologie, à travers la théorie des faces, instaure des rôles. Nos attentes doivent s'ajuster en fonction de ces rôles. Elles s'ajusteront aussi à la stratégie employée. Ainsi, la notion de coopérativité a une importance capitale car elle détermine directement le type d'acte attendu. En effet, un locuteur coopératif a certaines attentes, un locuteur non-coopératif en a d'autres.

Sperber et Wilson critiquent Grice sur la notion d'implicature. Nous pensons que les attentes peuvent apporter une aide à la résolution de ces implicatures.

Les présupposés seront plus ardues à résoudre par les attentes. Les attentes sont assujetties à la notion de pertinence. Elles seront pour nous liées aux notions d'intention et de but. En effet, la naissance d'un but introduit des attentes. *A contrario* (et c'est pourquoi nous travaillerons sur des dialogues finalisés), l'absence de but aboutit, selon nous, à l'absence d'attentes. Cette notion de but est très importante dans notre approche. En effet, grâce au but, l'attente introduit l'allocutaire. Si  $\text{But}_A = s_A(p)$  alors  $\text{Attente}_A = F_B^s(p)$  c'est à dire si le but de A est de connaître p, alors sa première attente sera que B fasse connaître p.

La linguistique structurale introduit la notion d'incidences qui peuvent être plus ou moins importantes en fonction des attentes des interlocuteurs. Nous le verrons par ailleurs à travers une étude psycho-linguistique d'un corpus homme-homme. Dans un dialogue dirigé par les buts, nous pensons que l'acte le plus probable à la suite d'un acte est l'acte qui va tenter d'atteindre le but. Ainsi, nous pensons que les probabilités permettent de décrire l'écart d'incidence par rapport à l'axe régissant de Luzzati. Ainsi, l'acte le plus attendu reste-t-il sur l'axe régissant, tandis que le deuxième s'en écarte légèrement, et ainsi de suite jusqu'aux actes tellement peu attendus qu'ils peuvent remettre en cause le dialogue.



**Figure 12** : Attentes et linguistique : écart incident à l'axe régissant

Après avoir vu « Pourquoi modéliser les attentes », nous pouvons donc à présent aborder la question : « Comment modéliser les attentes ? »



## **B. Etude théorique**



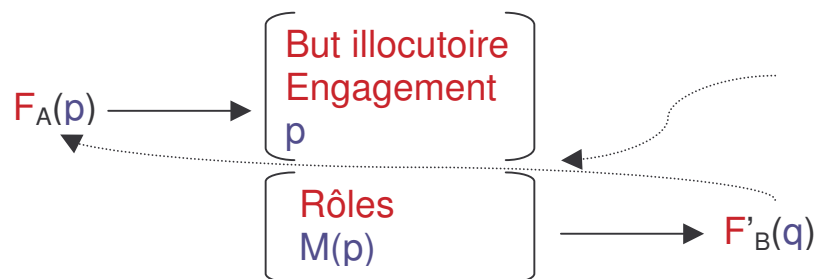
## **Chapitre I : Modélisation et traitement des attentes en dialogue – aspect théorique**

*« Afin d'établir un dialogue verbal homme-machine, avec une composante dialogale, il faut que la machine se fasse une image de son interlocuteur pour reconnaître ses intentions et ses stratégies. C'est ce qui fait la différence entre la communication homme-machine de type stimulus-réponse, ayant seulement un aspect dialogique, et une dialogue homme-machine plus complet, avec une composante dialogale. » [Vernant 94].*

## 1. les attentes en dialogue : notions et modélisation

### 1.1. La notion d'attentes en dialogue

La notion d'attentes a été introduite par [Bilange 92] qui la rattache aux mécanismes de prédiction des systèmes de compréhension orale. Elle est également sous-tendue par Trognon et Brassac [Trognon & Brassac 92] à travers leur interprétation dialogique de la logique illocutoire. En effet, dans cette conception, la force illocutoire d'un acte n'est effective qu'après la réponse de l'interlocuteur, car elle est fonction de sa réponse. Ainsi, la réaction à un énoncé donné est conditionnée par le type d'acte que l'énoncé accomplit. Ils étendent alors la logique dialogique de Vanderveken. Pour eux, la satisfaction d'un acte dépend de sa réponse, alors que sa réussite dépend des conditions dans lesquelles il a été accompli. Nous l'illustrons en **Figure 13**.



**Figure 13** : Réussite (selon conditions) et satisfaction (selon réponse) d'un acte.

Le locuteur A pose un acte  $F_A(p)$ . L'acte  $F_A(p)$  réussit lorsque les conditions sur le but illocutoire, sur l'engagement, sur le contenu propositionnel  $p$  sont remplies. Cette réussite dépend également des rôles des participants et de conditions du monde sur le contenu propositionnel. En revanche, cet acte de A n'est satisfait qu'avec l'acte de B,  $F'_B(q)$ , en réponse à celui-ci. Selon nous, il peut également être satisfait par une action extérieure. La force illocutoire de l'acte ne peut donc être déterminée qu'avec l'acte en réponse.

Prenons l'exemple d'un acte de A de contenu « il fait froid ! » avec pour contexte une porte ouverte. L'acte sera réussi si l'on a pu déterminer le but illocutoire  $b=\{\text{porte fermée}\}$ , si les conditions sur  $p$  sont réunies (qu'il fait effectivement froid), si les rôles des participants sont tels que A peut inciter (ou ordonner) B à fermer la porte, etc. En revanche l'acte ne sera satisfait que lorsque B, en réponse à A, fermera la porte ou qu'une action extérieure (un coup de vent, une tierce personne) fermera cette porte.

La satisfaction d'un acte se situe à un niveau pragmatique et entre dans le processus de compréhension en facilitant une interprétation des énoncés. Cette interprétation est généralement statique et s'appuie sur l'idée qu'un contexte préétabli suffit pour des dialogues orientés par la tâche. Elles sont donc établies à partir du modèle de tâche et ignorent le contexte réel de chaque dialogue.

*« Si l'on est capable de différencier les énoncés d'après les actes illocutoires qu'ils accomplissent et les effets perlocutoires qu'ils peuvent provoquer, on est non seulement capable de déterminer le type de réaction qu'un locuteur attend mais également en mesure de prédire le type de réaction qu'un énoncé peut provoquer. »* Pierrel & Romary dans [Pierrel 00].

Dans le cadre du dialogue, [Lehuen 97], [Lemeunier 00] ou [Vilnat 97] proposent d'introduire la notion d'attentes : pour Lehuen, ces attentes sont *« ce qui est attendu par le système au regard de l'état courant de la tâche »*. Il les définit donc en terme d' *« objets du domaine d'application et d'actes de dialogue orientés par la tâche »* ; pour Vilnat, ce sont *« ce que l'utilisateur est susceptible de dire (ce qui implique qu'il n'est pas obligé de le dire), en raison des buts et des plans en cours de développement »* ; pour Lemeunier, ce sont *« des entités pragmatiques manipulées par le système qui correspondent à ce que l'utilisateur est censé dire et ce à quoi le système est donc censé s'attendre »*. Ces attentes sont donc très liées à la tâche d'une part et considérées du point de vue du système d'autre part.

Selon Lemeunier [Lemeunier 99], les attentes ne sont pas liées directement à la planification. Les hiérarchies sont prévues et reconstruites au fur et à mesure de l'interaction comme le ferait une reconnaissance de plans. Les attentes ne sont pas une modélisation de la tâche, mais des représentations mentales des interactants du point de vue du système informatique. Son approche est donc plus cognitive, mais n'entre pas en contradiction avec la planification notamment avec la notion de plans partagés.

Pour Nicolle [Nicolle 99] en revanche, la notion d'attente dérive davantage des systèmes multi-agents. Elle abandonne donc la notion de planification au profit de modèles fondés sur des réactions plus élémentaires dirigées vers un but. Ainsi le comportement complexe vu de l'extérieur peut être simple du point de vue du sujet qui le réalise.

Nous pensons que trop contraindre la conversation est un handicap pour les *agents naturels*. C'est pourquoi nous nous intéressons aux travaux effectués en sciences humaines où les rituels et les indications ont dans la communication la même place que les règles strictes des informaticiens. Lemeunier [Lemeunier 00],

dans sa modélisation des attentes, utilise une mémoire interactionnelle correspondant aux aspects purement cognitifs de l'interaction courante, sans remémoration des dialogues antérieurs éventuels, et sans prendre en considération les aspects sociaux dont rend par exemple compte la théorie des faces de Goffman. D'autres travaux, surtout dans le domaine des systèmes multi-agents, s'intéressent pourtant à ces aspects [Chicoisne & Pesty 99]. Pour nous, le côté social de l'interaction communicative est primordial. C'est pourquoi nous l'aborderons à travers la théorie des faces de Goffman [Goffman 73, 74] présentée précédemment.

### **Attentes et planification**

La planification est vue, traditionnellement, comme une activité qui consiste à sélectionner une série d'actions, selon un ordre plus ou moins défini, afin de passer d'un état initial à un état final souhaité. Les agents essayent de reconnaître les plans de leurs interlocuteurs en vue de les utiliser lors de la construction des réponses. Bange décrit la planification comme suit :

*« Le terme de **schéma d'action** désigne le savoir pratique disponible sur le déroulement prévisible de l'action. C'est cela qu'on peut appeler aussi **plan** ou **projet d'action**, même s'il n'y a pas toujours une planification véritable. Dans la vie quotidienne, il nous suffit en effet souvent de dénommer l'activité que nous projetons et le savoir ainsi actualisé nous permet de prévoir et d'exécuter le détail des actions impliquées dans le plan. C'est seulement si la voie qui conduit au but n'est pas claire, si l'action projetée est trop complexe, qu'on a une véritable planification de l'action. L'acteur tente alors de s'en tirer avec des **schémas d'action connus** adaptés à la situation nouvelle par analogie, il pratique une sorte de bricolage de **morceaux stéréotypés**, ajustés les uns aux autres. » [Bange 92]*

L'exemple classique est le repas au restaurant. Nous commençons par demander une table, nous installer. Un serveur nous donne les menus, etc. Nous déroulons un plan classique. Si le plan est visible dans ce cas-ci, de nombreux plans le sont beaucoup moins. Par exemple :

Q1 = Vous payez par chèque ou en espèces ?

R1 = Je ne sais pas si j'ai pris un stylo ?

Q2 = La machine le remplit toute seule ...

Lorsque l'agent pose son énoncé Q2, il reconstruit le plan de son interlocuteur. Si celui-ci cherche un stylo, c'est probablement qu'il veut payer par chèque. Notons au passage que l'emploi de l'anaphore 'le' faisant référence au chèque fait appel aux connaissances partagées, c'est-à-dire qu'un chèque se remplit

(par défaut avec un stylo). L'interlocuteur devra donc résoudre cette anaphore pour comprendre cet énoncé Q2.

Q2' = vous avez une pièce d'identité ?

R2 = Non mais c'était pour signer ...

Lorsque cet agent continue son énoncé Q2', son interlocuteur reconstruit le plan qu'il imagine que l'agent s'est construit à son sujet. Dans sa réponse R2, l'interlocuteur montre qu'il sait Q2 (i.e. que la machine remplit les chèques). Notons par ailleurs que l'interprétation de R2 est aisée pour un être humain, mais très difficile pour un système qui pourra voir le 'non' comme une réponse à Q2'. C'est ainsi qu'Allen et Perrault posent l'importance des plans en considérant que « *ce modèle est en mesure de rendre compte des réponses qui fournissent davantage d'informations qu'il n'en est explicitement demandé, des réponses à des fragments de phrases, ainsi qu'à des actes de langage indirects.* » [Allen & Perrault 80]

La planification peut donc être dirigée par les buts. Elle nécessite alors une modélisation précise des plans et buts des interlocuteurs. Le plan sera alors vu comme une suite d'actions permettant de passer d'un état initial du monde vers l'état souhaité (le but). Ces plans dépendent donc fortement de la tâche. Les incidences en dehors de celle-ci ne peuvent pas être traitées, faute de plan accessible. La notion d'attente, telle qu'elle est vue jusqu'à présent, reste liée à la tâche. Elle peut selon nous être résolue par cette planification.

A l'inverse, la planification peut être dirigée par les données. Elle est alors linguistique et nécessite une modélisation pragmatique précise contenant une modélisation sémantique et le rattachement du contenu propositionnel aux actions possibles. On parle alors davantage de modèles opportunistes qui cherchent à gérer l'événement en fonction de la situation, c'est à dire par l'action située. Nos attentes se rapprochent davantage de ces modèles opportunistes.

En effet, les attentes du locuteur ne sont pas régies par des scénarii préalablement codés afin de pouvoir réagir à toutes les éventualités et notamment aux incidences (qui peuvent être hors tâche) dans le dialogue oral. Elles s'appuient en revanche sur la notion de but, et donc sur la planification dirigée par les buts mais elles restent indépendantes de la tâche par la gestion des buts illocutoires.

## **1.2. Modélisation des attentes en dialogue**

Actuellement, différents modèles de dialogue permettent d'améliorer la communication homme-machine. Ceux-ci ne tiennent pas compte des attentes du

locuteur. La modélisation des attentes dans le dialogue automatisé a des retombées diverses. Dans le dialogue homme-machine, la prédiction des interventions de l'utilisateur au cours du dialogue permet d'améliorer la compréhension, la pertinence des réponses de la machine, et de contraindre, dans le cadre d'un dialogue oral, la combinatoire des solutions de reconnaissance vocale. Dans le dialogue homme-homme multilingue médiatisé par la machine, le suivi de dialogue permet aussi de réduire la combinatoire des solutions de reconnaissance vocale mais également de réduire l'ambiguïté dans l'interprétation de l'énoncé puis dans sa traduction. Une bonne gestion des attentes reflète donc le niveau de « compréhension » de la machine. Elle permet alors d'en faire une évaluation et de bâtir une méthode scientifique pour la conception de dialogues homme-machine.

L'attente, en général, peut se définir au niveau d'un locuteur comme un état mental dans lequel sont codés les buts, les connaissances (sur la tâche), les règles de la conversation et les rôles des participants.

Par exemple, si un individu pose une question, il aura une attente de réponse informative c'est-à-dire une réponse dans le champ topique de la question. Mais la réponse peut-être aussi négative dans le cas d'un non-savoir, auquel cas la réponse est de type conversationnelle (excuse par exemple). Plus précisément, à une question telle que « *Vous reste-t-il des places ?* », on attend une réponse informative comme « *oui, voulez-vous un coin fumeur ?* » ou « *en première classe ?* » ou une réponse conversationnelle comme « *attendez, je vais me renseigner* » ou « *je ne sais pas, adressez-vous au guichet d'information* ». C'est ce que nous pouvons voir à travers un exemple en Figure 14, selon Denis Vernant.

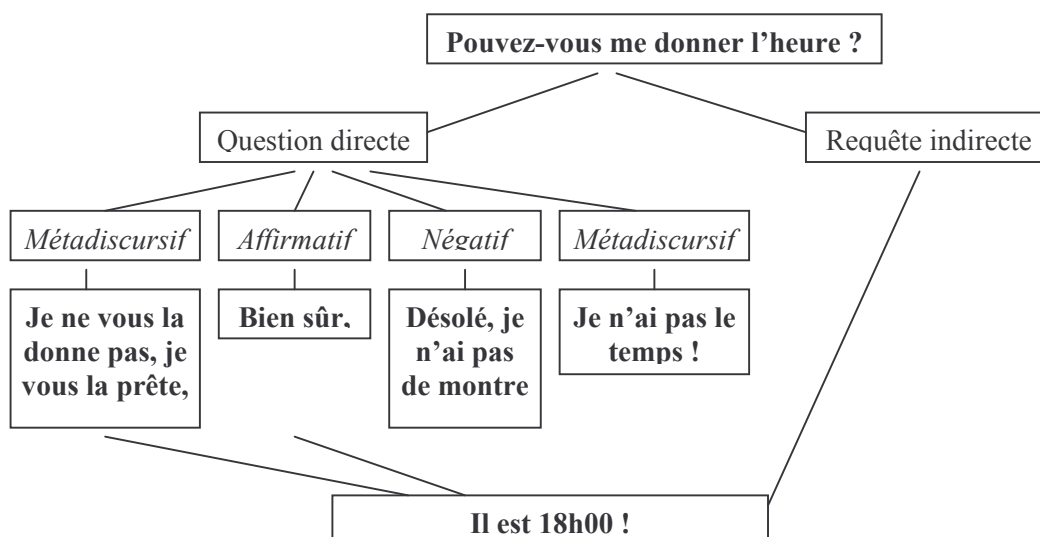


Figure 14 : Possibilités de réponses à un interacté, inspiré de [Vernant 97b].



Analysant plusieurs dialogues dans lesquels cette même question est posée, nous pouvons affirmer qu'à une question comme « *Pouvez-vous me donner l'heure, s'il vous plaît ?* », on attend diverses réponses comme « *oui bien sûr, il est 18h00* » ou « *non, désolé, je n'ai pas de montre* » voire certaines formes d'humour comme « *je ne vous la donne pas, je vous la prête* » ou « *je n'ai pas le temps* » (qui peut être une simple réponse négative selon le ton).

Dans ces exemples, différents types d'actes dialogiques sont possibles : affirmation avec attitude coopérative dans le topique de la question, incidence dans le topique de la question avec demande d'information complémentaire, phatique de mise en attente, clôture de conversation avec attitude coopérative dans le topique de la tâche, etc.

Selon nous, ces attentes doivent être modélisées sous divers angles. Le codage des buts nécessite une approche intentionnelle dérivée de la philosophie du langage à travers la théorie de l'illocutoire défendue par [Austin 70] et [Searle & Vanderveken 85]. Cette théorie est fondée sur l'idée que les énoncés ont pour fonction de servir à l'accomplissement d'un certain type d'acte, conventionnel, créateur de droits et de devoirs pour les interlocuteurs, appelé acte de langage (portant plusieurs informations sur l'énoncé : un contenu propositionnel et une force illocutoire). Le codage des connaissances sera dérivé des travaux de [Sperber & Wilson 89] sur la pertinence liée à une intention informative et une intention communicative. Enfin, les règles de conversations et les rôles des participants, non traités dans les systèmes actuels, seront représentés pour nous à travers la théorie des faces de [Goffman 73] et [Goffman 74]. Ainsi, nous aurons une approche pluridisciplinaire (utilisant les différents contextes, l'historique du dialogue, le champ de la tâche, les marqueurs de discours de [Colineau 97] et les règles du dialogue, etc.) triple permettant de rendre compte des différents aspects du dialogue. Nous prendrons en considération l'aspect intentionnel avec l'idée de dialogues dirigés par des buts que l'on s'engage à poursuivre. Mais nous tiendrons aussi compte des aspects cognitifs à travers la représentation des connaissances des participants et des aspects sociaux avec l'idée sous-jacente que l'homme est conditionné par la société (rituels, raisonnements normalisés, hiérarchie, etc.)

### **1.3. Les attentes du locuteur**

Nous considérons les attentes du point de vue du locuteur, comme les résultats espérés de ses actes.

En effet, un locuteur engagé dans un dialogue entend poursuivre ses buts et obtenir des effets positifs de ses actes. Il projette donc un certain plan sur son interlocuteur (qu'il peut considérer comme coopératif) et en attend en retour un certain comportement. Les attentes se construisent au fur et à mesure que le dialogue se déroule, en fonction des buts et des comportements des interlocuteurs, ce sont essentiellement des hypothèses réflexives construites par chacun des interlocuteurs. Par exemple, si le locuteur B ne se montre pas coopératif, la stratégie du locuteur A, guidée par ses attentes va certainement être davantage directive pour forcer B à adopter les buts de A. Ce qui se formalise par : si B est coopératif alors  $F_A(p) \rightarrow F_B(p)$ , si on constate  $F'_B(p)$  alors B n'est certainement pas coopératif.

Notre modélisation s'appuie donc sur l'idée de contexte. En effet, un certain acte (par exemple :  $F_A(p)$ ), devrait fournir, selon le contexte (ici B coopératif), tel autre acte en réponse ( $F_B(p)$ ) répondant aux attentes de A. Nous nous appuyerons, en premier lieu, sur l'idée de [Vernant 92] que le dialogue est dirigé par les buts de locuteurs. Ces buts doivent être à la fois atteints et satisfaits pour que ce dialogue aboutisse à un succès [Vanderveken 90]. Il nous semble alors nécessaire de représenter ces buts à travers un contexte intentionnel inspiré de [Colineau 97]. Dans ce même contexte, nous pourrions représenter les engagements des participants selon l'idée que les interactants poursuivent des buts et s'y engagent plus ou moins (selon par exemple leur coopération). Ensuite, un contexte cognitif nous semble pertinent afin de représenter les connaissances de chacun sur l'autre et sur la tâche. Nous nous inspirerons, notamment, ici, de la théorie de la pertinence de [Sperber & Wilson 89]. Enfin, l'aspect social du dialogue est souvent (en raison de la difficulté à le représenter) mis de côté lorsque l'on cherche à modéliser ceux-ci. Nous souhaitons, pour notre part, intégrer un contexte social afin de représenter les règles de conversation mais aussi les rôles des participants. En effet, pour nous, si dans un type de dialogue donné, les rôles peuvent être équivalents, nous montrerons plus loin que ce n'est pas toujours le cas. Dès lors, il nous semble important d'en tenir compte. Nous nous appuyerons, pour ceci, sur la théorie des faces de [Goffman 74] qui tente de décrire les règles de conversation à travers l'idée de faces à cacher ou à défendre.

Les attentes sont donc, pour nous, codées par l'**état mental** (d'un locuteur) qui contient les **buts** (et les engagements), les **connaissances** (sur le monde et sur la tâche), les **règles** de conversation, et les **rôles** des participants.

Nous nous appuyerons alors sur les recherches effectuées sur le niveau pragmatique des énoncés en utilisant une notation en actes de dialogues dérivée des actes de discours de [Vernant 97], des actes de langage de [Austin 62] et des modaux

de [Greimas 66]. En effet, nous pensons qu'un tel acte est suffisamment porteur de sens pour tenter de lui associer des attentes (tirées de cet acte et du contexte) afin de comprendre au mieux ce que souhaite le locuteur. Nous présenterons donc notre notation en actes de dialogue, puis celle, dérivée de la première, des attentes. Nous montrerons enfin l'intérêt de la modélisation des attentes du locuteur dans le cadre de dialogue réel.

## **2. Spécification des attentes**

### **2.1. Une notation en acte de dialogue issue des actes de langage**

La théorie de l'illocutoire ([Austin 70], [Searle & Vanderveken 85]) spécifie que tous les énoncés ont pour fonction de servir à l'accomplissement d'un certain type d'acte, conventionnel, créateur de droits et de devoirs pour les interlocuteurs, appelé acte de langage. Chaque énoncé porte plusieurs informations simultanément : non seulement un contenu propositionnel mais aussi une force illocutoire (qui se décline notamment pour les actes directifs en ordres, requêtes ou demandes). C'est l'association des deux entités, contenu propositionnel et intention du locuteur, qui détermine, en fonction de conventions, le type d'acte de dialogue réalisé.

Il existe de très nombreuses listes d'actes de dialogue, hiérarchisées ou non, pour représenter les propos des interlocuteurs dans un dialogue finalisé. La plupart d'entre elles utilisent comme critère de discrimination l'objet sur lequel porte l'acte : la tâche (requête, question, information), le dialogue (ouverture, clôture, satisfaction, contestation, acceptation, refus) et le contrôle du dialogue (demande de reformulation, récapitulation, relance).

L'acte de dialogue est défini comme une **action** intentionnelle qui a pour paramètre un **locuteur**, un **auditeur**, et un **contenu propositionnel**.

L'identification d'un acte de dialogue est toutefois une affaire de convention et de norme en regard de la tâche que le système doit réaliser. En effet, souvent, le locuteur veut dire quelque chose de plus que ce qu'il dit littéralement ; la phrase contient des marqueurs de force illocutoire propres à une classe d'actes, mais elle est énoncée dans le but d'accomplir un autre acte, appelé acte indirect [Searle 82]. Il est important de remarquer que l'acte indirect ajoute (mais ne remplace pas) une force illocutoire ou un contenu propositionnel à l'acte direct. Le locuteur reste ainsi généralement satisfait lorsque son interlocuteur interprète le sens second et réalise,

par exemple, une action en réponse à une requête indirecte comme « peux-tu me donner la liste des objets ? ». À l’instar de [Luzzati 89], nous nous fondons essentiellement sur les formes de surface des énoncés pour identifier les actes de dialogue. Concernant les taux d’identification d’acte de dialogue « *il est possible de continuer à améliorer les résultats sur un corpus donné. (...) on peut d’une part enrichir les marques linguistiques collectées, et d’autre part étudier de nouvelles stratégies de dialogue.* » [Colineau 97]. De plus, « *la tendance est favorable pour l’analyse de dialogues homme-machine. En effet, dès lors que le dialogue suit un déroulement relativement contraint, les locuteurs s’adaptent aux capacités langagières et discursives de la machine.* » [Colineau 97].

S’inspirant des théories de [Austin 62] et [Searle 72] sur les actes de langage, [Vanderveken 90] définit la force illocutoire  $F(p)$  comme un acte de force  $F$  s’appuyant sur le locutoire représenté par le contenu propositionnel  $p$ . Il découpe les actes illocutoires en cinq catégories (déclaratif, directif, assertif, promissif, expressif). Nous pensons que ce découpage peut être affiné à l’aide des modaux sémiotiques de [Greimas 66]. Ainsi, un acte sera, pour nous, un faire, faire-faire, faire-savoir, faire-croire, faire-faire-savoir, faire-devoir, faire-pouvoir, faire-faire-croire, etc.

Dans une notation plus simplifiée, nous ne garderons que les initiales, d’où une notation de la forme  $F^{fs}(p)$  pour un faire-faire-savoir de contenu  $p$ . Cette notation nous permet de mettre en valeur l’idée d’Austin, reprise par les philosophes du langage, que « dire c’est faire » (c’est-à-dire que lorsque l’on dit quelque chose, on exécute une action qui peut décrire ou modifier le monde). Ainsi, chaque acte débute par un faire, la précision de la force étant fournie en exposant. Nous aurons donc respectivement des actes de la forme  $F(p)$ ,  $F^f(p)$ ,  $F^s(p)$ ,  $F^c(p)$ ,  $F^{fs}(p)$ ,  $F^d(p)$ ,  $F^p(p)$ ,  $F^{fc}(p)$ , etc.

Nous reprendrons, pour notre notation, la classification de [Vilnat 97] dérivée de [Moeschler 85] (**Tableau 6**) :

Acte initiatif	Acte réactif
Assertion	évaluation positive ou négative
Demande d’information	réponse positive ou négative
Demande de confirmation	confirmation ou infirmation
Offre/requête	acceptation ou refus

**Tableau 6** : Classification des actes de [Vilnat 97].

Il est cependant nécessaire d'introduire l'action dans cette classification où elle n'avait pas place. [Ozkan 94] l'introduit en proposant une classification tirée des modaux de Greimas, mais orientée vers l'action (**Tableau 7**):

	Acte actionnel	Acte communicationnel
Acte assumé	Faire	Faire Savoir
Acte délégué	Faire Faire	Faire Faire Savoir

**Tableau 7** : Classification des actes de dialogue de [Ozkan 94]

Cette classification laisse de côté la notion d'offre. Souhaitant introduire cette notion d'offre, Caelen [Caelen 97] affine cette notation en proposant des actes de langage de la forme Fp, avec la classification suivante (**Tableau 8**) :

Acte	modaux	explication	Searle
F <sub>A</sub> p	faire action	les effets immédiats obtenus sont p	déclaratif
F <sub>F</sub> p	faire-faire	demande de faire p, les effets p sont différés	directif
F <sub>FS</sub> p	faire faire-savoir	demande sur p, la réponse est différée	
F <sub>S</sub> p	faire savoir p	les effets sont obtenus immédiatement	assertif
F <sub>S</sub> ∅	faire savoir ∅	Cas particulier du F <sub>S</sub> p où la direction d'ajustement est considérée vide (phatique)	expressif
F <sub>P</sub> p	faire pouvoir	offre un choix fermé ou ouvert, la réplique est attendue parmi p	promissif

**Tableau 8** : Classification des actes de Caelen en regard de Searle

Il évoque aussi la possibilité d'un F<sub>D</sub>p (faire devoir). Celui-ci, orienté vers l'action, est dérivé de Vernant et oblige une action dont l'effet sera p (directif). Il est important de ne pas le confondre avec le faire-devoir dérivé de Greimas, plus orienté vers les connaissances, et qui regroupe les répliques de confirmation, contestation, rectification et réparation.

Ces différentes notations ont été combinées avec celles vues dans l'état de l'art pour obtenir des actes de langage de la forme F(p) avec (Tableau 9) :

Acte	modaux	Explication	exemples
<b>F(p)</b>	<i>Faire</i>	<b>Action</b> dont les effets p obtenus sont immédiats	C'est noté ; C'est fait Je clique ; Au revoir
<b>F<sup>f</sup>(p)</b>	<i>Faire-faire</i>	<b>Demande d'action</b> , de faire, dont les effets p sont différés	Attendez un instant Cliquez sur ce lien
<b>F<sup>fs</sup>(p)</b>	<i>Faire-faire-savoir</i>	<b>Demande d'information</b> sur p dont la réponse est différée	Pouvez-vous me donner l'heure ? Quelle heure est-il ?
<b>F<sup>fd</sup>(p)</b>	<i>Faire-faire-devoir</i>	<b>Demande d'engagement</b> dont l'effet obtenu sera p	Vous le noterez, hein ? C'est sûr ?
<b>F<sup>fp</sup>(p)</b>	<i>Faire-faire-pouvoir</i>	<b>Demande de possibilité</b> dont les choix seront faits parmi p	Puis-je cliquer ici ?
<b>F<sup>s</sup>(p)</b>	<i>Faire-savoir</i>	<b>Information</b> dont les effets p obtenus sont immédiats	Les produits X sont formidables Il est 11h30
<b>F<sup>S</sup>∅(p)</b>		Cas particulier du phatique, la direction d'ajustement est vide	Mmm
<b>F<sup>d</sup>(p)</b>	<i>Faire-devoir</i>	<b>Engagement</b> de soi-même, obligation d'action de soi-même dont l'effet sera p	Je vais le noter C'est promis
<b>F<sup>p</sup>(p)</b>	<i>Faire-pouvoir</i>	<b>Possibilité</b> , invite, offre de choix fermé ou ouvert dont la réplique est attendue parmi p	Bonjour Merci En quoi puis-je vous aider ? Préférez-vous ceci ou cela ?

**Tableau 9:** Notre taxonomie des actes de dialogue avec des exemples

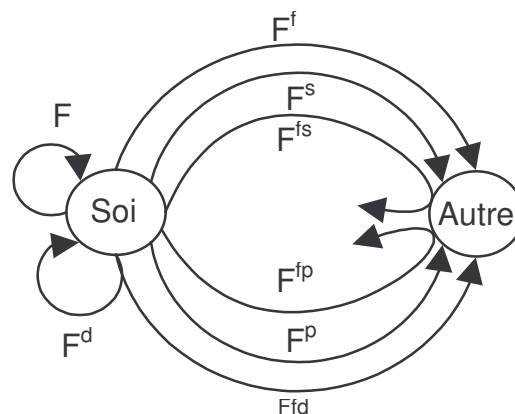
Ainsi seront repris les actes de Caelen issus de ceux d'Ozkan. Nous avons introduit l'acte F<sup>fd</sup>, demande d'engagement (incluant l'engagement de l'autre), qui reprend la demande de confirmation de [Moeschler 85]. Pour l'équilibre, la demande d'offre F<sup>fp</sup> a également été introduite mais nous verrons qu'elle est finalement très peu présente dans nos corpus.

	Acte assumé	Acte délégué
Acte actionnel	F : action	F <sup>f</sup> : demande d'action
Acte communicationnel	F <sup>s</sup> : information	F <sup>fs</sup> : demande d'information
Acte engageant actionnel	F <sup>d</sup> : engagement	F <sup>fd</sup> : demande d'engagement
Acte engageant statif	F <sup>p</sup> : possibilité, invite	F <sup>fp</sup> : demande de possibilité

**Tableau 10** : Synthèse de notre taxonomie des actes de dialogue

Dans le Tableau 10, nous avons réparti les actes en deux catégories. Les actes que le locuteur assume sont ceux qui apparaissent lorsque le locuteur effectue quelque chose. Les actes que le locuteur délègue sont ceux qui naissent lorsque le locuteur demande quelque chose à l'allocutaire. Ce quelque chose pourra être une action, une information, un engagement dans l'action ou un choix.

Les actes ont un lien avec autrui dans le sens où leur force illocutoire est relative à l'interlocuteur. Ils peuvent, en outre, se définir en fonction de soi et de l'autre (Figure 15) :



**Figure 15** : Actes orientés vers soi ou vers autrui

Nous pouvons alors constater que deux actes (l'action et l'engagement d'action) sont orientés vers soi-même. Ils n'engagent que soi (vis-à-vis d'autrui). Les autres actes sont orientés vers autrui. On fait savoir à l'autre. On demande à l'autre une action ou un engagement. On offre des choix à l'autre. Dans ces cas, il n'y a pas de retour direct sur soi. Lorsque l'on demande une information à l'autre ou lorsque l'on demande à l'autre un choix, on est orienté vers l'autre, mais avec un retour vers soi (une information F<sup>s</sup> ou un choix F<sup>p</sup> *a priori*).

Nous pouvons alors comparer ci-dessous notre taxonomie des actes et leur correspondance avec les modèles existants :

Searle	Vernant	Caelen	Fouquet	Exemple
Directif	Engageant directif	$F^f, F^d$	$F^f$	ordre, suggestion, conseil
		$F^{fs}$	$F^{fs}$	question
			$F^{fd}$	demande d'engagement
			$F^{fp}$	demande d'offre
Commissif	Engageant commissif d'état	$F^p$	$F^p$	promesse, offre, introduction
	Engageant commissif d'action		$F^d$	promesse d'action engageante
Expressif	Assertif descriptif expressif	$F^s\emptyset$	$F^s\emptyset$	description d'état
Déclaratif	Déclarations	$F^a$	F	action, conclusion
Assertif	Assertif descriptif comportatif			description d'action
		Assertif constatif	$F^s$	$F^s$
	Métadiscursif citatif	citation du discours d'autrui		
	Métadiscursif expositif	exposition de son propre discours		

**Tableau 11** : Notre taxonomie des actes de langage comparée à l'état de l'art

Nous pouvons remarquer notamment que le remerciement est pour nous non pas expressif (quoique réellement il le soit) mais davantage commissif dans le sens où il permet de clore l'échange et d'offrir l'ouverture d'un autre échange. En outre, la confirmation (faire devoir au sens de Greimas, l'interlocuteur ne pouvant pas ignorer cette confirmation) aura deux déclinaisons. Ce sera pour nous aussi un faire devoir à la condition qu'elle soit engageante. Dans le cas d'une confirmation non engageante (« C'est bien lundi ? », « oui »), ce sera simplement un faire savoir.

A partir de cette taxonomie, nos actes de dialogue tiendront compte du rôle du locuteur à travers sa représentation indicielle. Ils auront la forme suivante :

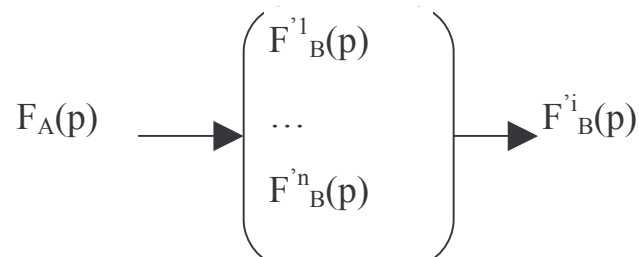
$F_A(p)$  lorsque le locuteur A pose un acte F (noté comme ci-dessus) de contenu propositionnel p.

Le locuteur A pourra alors poser un *Faire-faire-savoir* sur le contenu propositionnel p ( $F^fs_A(p)$ ) et le locuteur B pourra lui répondre par un *Faire-savoir* sur ce même contenu ( $F^s_B(p)$ ). Le contenu p pourra concerner l'heure par exemple.



## 2.2. Spécification des attentes en dialogue

Les attentes sont, pour nous, des actes qui respectent la notation de Frege reprise par [Vanderveken 90], affinés par les modaux de [Greimas 66]. Nous aurons donc des attentes de la forme :  $F_{A(p)}^{fs}$ ,  $F_{A(p)}^f$ ,  $F_{A(p)}^s$ , etc. Ainsi, lorsque le locuteur A propose un acte de faire-savoir ( $F_{A(p)}^{fs}$ ), les attentes seront, par exemple, une réponse de l'allocutaire ( $F_{B(p)}^s$ ) ou une demande de précision de celui-ci ( $F_{B(p)}^{fs}$ ). Les attentes correspondent alors dans ce cadre aux actes de B,  $\{F_{B(p)}^{i,n}\}$  possibles en réponse à l'acte du locuteur A,  $F_{A(p)}$ . Parmi ces actes possibles (les attentes) se trouvera peut-être l'acte que l'allocutaire choisira effectivement :  $F_{B(p)}^i(q)$ . Cette situation se résume au schéma suivant, **Figure 16** :



**Figure 16** : Attentes vues comme une liste de réponses possibles

Gérer les attentes des interlocuteurs permettra d'améliorer la prédiction des interventions non pas par la restriction des actes de dialogue possibles mais par une pondération de cette possibilité. Cette solution offrira alors la possibilité de tenir compte des interventions prévisibles mais sans écarter complètement les autres. L'effet de surprise pourra se mesurer si une intervention sort du cadre des attentes.

## 2.3. Attentes et psycholinguistique

De partance pour les Pyrénées, je me trouvais dans le train de Grenoble vers les Pyrénées. Celui-ci, en direction de Pau, était relié, comme souvent, à un autre qui devait le quitter à Narbonne pour Irun (Espagne). A côté de moi, deux jeunes filles espagnoles souhaitaient rejoindre l'Espagne. Dans la bonne voiture mais pas dans le bon train, il leur fallait comprendre qu'elles devaient monter dans le train de tête avant Narbonne. L'une d'entre elles parlait un peu le français, et cherchait à comprendre la situation. Constatant qu'en français (je ne parle pas l'espagnol) nous ne parvenions pas à nous comprendre, nous avons opté pour une autre langue intermédiaire : l'anglais. Ne maîtrisant pas parfaitement cette langue, j'ai pu constater

toute une pléthore de phénomènes linguistiques : mots anglicisés, descriptions afin d'expliquer des termes manquant à mon vocabulaire, gestes, etc. Ce genre de situation me paraît intéressant pour étudier ces phénomènes liés aux attentes intervenant dans un dialogue en langue intermédiaire.

Nous avons donc testé ces phénomènes comme des révélateurs des attentes sur le corpus collecté dans le cadre du projet NESPOLE!. L'agent italien chargé de dialoguer avec nous en français ne maîtrisait pas parfaitement la langue. Nous avons alors pu constater les mêmes phénomènes que sus-dits : les mots incompris (ex: les arrhes), les mots prononcés en langue natale, etc. Cette situation reflète bien les phénomènes que nous devons prendre en compte dans un système de dialogue oral.

La première preuve d'inhabileté dans la langue qui vient à l'esprit est, avec la prononciation, la tendance à franciser les termes dont on ne connaît pas la traduction. La réaction de l'interlocuteur va alors varier selon sa propre compétence langagière, la valeur inductive du contexte (du terrain commun élaboré [Lemeunier 00]), enfin et surtout l'importance de la compréhension de ce fragment dans la résolution de la situation. Ainsi, il peut choisir d'ignorer la confusion si celle-ci n'affecte pas la compréhension du message dans son ensemble. Mais il peut aussi amorcer une digression dans le but d'élucider l'intention du locuteur.

Nous pouvons ainsi répertorier différents types d'irrégularités d'expression. Les plus fréquentes sont celles qui poseront *a priori* le plus de problèmes aux systèmes de reconnaissance vocale. Ce sont toutes les erreurs qui ne peuvent être prises en compte par un tel système mais qui ne perturbent pas la suite du dialogue, l'interlocuteur ignorant ces lacunes qui ne l'empêchent pas de comprendre le message du locuteur. Les moins fréquentes sont les erreurs qui perturberont l'interlocuteur au point qu'il renonce provisoirement à son but initial pour résoudre cette incompréhension.

### **2.3.1. Irrégularités non perturbantes**

Ce sont, comme susdit, les irrégularités qui ne retardent pas le déroulement du dialogue initial. L'interlocuteur ne les relève pas. Il peut considérer avoir mal entendu. Il peut aussi penser que le locuteur n'a pas parlé correctement. Dès lors, deux situations peuvent se produire, amenant chacune à l'ignorance de cette faute. Dans le premier cas, l'interlocuteur corrige la faute. Dans le deuxième, il n'utilise pas le fragment incompris.

### **L'énoncé est interprété indépendamment du fragment**

Dans cette situation, l'interlocuteur, qui n'a pas compris un fragment de l'énoncé initial, exclut l'élément. C'est le cas notamment lorsque l'agent parle d'« excursions *mirées* »<sup>6</sup>. L'interlocuteur ne garde que l'idée d'excursions, jugeant la précision sur le terme *mirées* inutile. Lorsqu'elles interviennent sur des détails trop éloignés du but initial, ces incompréhensions ne donnent donc pas lieu à une négociation sur leur représentation. C'est le cas lorsque l'interlocuteur baisse son niveau d'attention sur une partie du dialogue qui l'intéresse moins. Remarquons au passage que ce genre de situation arrive tous les jours, lorsque, par exemple, nous répondons au téléphone, et que nous ne comprenons pas l'identité du locuteur qui se présente. Nous ne lui demanderons peut-être pas de préciser, parce que nous jugerons cette information sans intérêt.

### **L'énoncé est interprété comme s'il était correct**

Dans ce cas, beaucoup plus fréquent, l'interlocuteur ne relève pas la faute et la corrige dans sa représentation. Il interprète cet énoncé comme si celui-ci était correct. C'est ici que l'on retrouvera les fautes typiques d'accord (« combien *des* étoiles »), de conjugaison (« il faut que vous *appelez* »), de syntaxe (« pouvez-vous lire *bien* ? »), d'élision (« entre (*les*) deux et trois étoiles »), de mot en trop (« il y a *des* différents niveaux de pistes »), de mauvais terme (« vous venez *par* voiture ? ») ou de termes techniquement possibles mais qui ne se disent pas (« le numéro *téléphonique* »). Celles-ci, encore présentes lors de dialogues monolingues, ne posent pas de problème quant à l'intercompréhension des interactants, au sens de Lemeunier [Lemeunier 00], mais peuvent mettre en défaut nombre de systèmes de reconnaissance de parole.

#### **2.3.2. Irrégularités affectant la suite du dialogue**

Cette situation intervient lorsque l'interlocuteur ne comprend pas un terme. Cela peut arriver entre deux locuteurs français ou étrangers lorsqu'un mot est inconnu ou mal prononcé ou non approprié. L'incompréhension est totale lorsque la représentation de ce terme est primordiale pour la suite du dialogue. Celui-ci laisse place à une négociation sur l'explication du terme et reprendra son fil initial lorsque ce but de représentation commune sera atteint.

---

<sup>6</sup> Francisation de l'expression italienne signifiant *excursions ciblées*, *excursions à thème*.

C'est ce que nous allons voir à travers l'exemple ci-dessous :

« Client1 : j'aurais voulu savoir le/ les arrhes qu'il fallait verser avant  
 Agent1 : vous désirez savoir s'il vous plaît ? le ? j'ai pas compris, excusez-moi/  
 Client2 : le/ les arrhes, le e l/ l'avance qu'il faut faire  
 Agent2 : ah l'agence ? e(m)  
 Client3 : l'avance pardon  
 Agent3 : ah l'avance ok, »

Le client sait, au vu des différents échanges précédents, que l'agent ne maîtrise pas le français. Il commence malgré tout à se répéter comme dans le cas d'une incompréhension liée à l'audition. L'agent précisant qu'il n'a pas compris (et non pas mal entendu), le client suppose une incompréhension sur le terme 'arrhes' et cherche à le décrire brièvement. S'ensuit une deuxième incompréhension mais celle-ci uniquement auditive au vu de la réponse de l'agent2 qui répète ce qu'il a compris. La construction du sens se termine par l'approbation de l'agent sur la représentation du sens du terme.

Ce même exemple illustre également l'importance du contexte dans le dialogue et peut se retrouver dans la vie de tous les jours lorsque, pour ce même appel téléphonique, nous jugeons l'identité de l'appelant importante (pour laisser un message à un collègue par exemple). Nous lui demanderons, parfois bien plus tard dans le dialogue, de préciser à nouveau son nom.

### 2.3.3. Les attentes : plus proche de la réalité ?

Cette étude nous a permis de mettre en relief certaines irrégularités liées à la maîtrise de la langue, irrégularités qui démontrent la difficulté de modéliser un énoncé mais dont nous devons tenir compte dans un modèle multi-utilisateur. Suite à ces irrégularités, nous avons relevé plusieurs réactions possibles face à un énoncé (**Tableau 12**).

Enoncé	Correct	Incorrect
Compris	Compris	Enoncé corrigé
		Fragment ignoré
Incompris	Incompris	Incompris

**Tableau 12** : Réactions de l'allocutaire face à un énoncé

Ainsi, un énoncé correct (cf. l'exemple des arrhes précédent) peut être compris ou incompris par l'allocutaire. Un énoncé incorrect (cf. excursion *mirées*) peut aussi être compris ou incompris. Dans le cas où il est compris, l'énoncé peut être corrigé (c'est le cas de combien *des* étoiles corrigé en combien d'étoiles), ou le fragment ignoré (c'est le cas avec *mirées*).

La réaction que nous avons détectée semble provenir du fait que l'énoncé incorrect répond plus ou moins aux attentes de l'allocutaire. Dès lors, ce dernier corrige sa représentation de l'énoncé, ignore le fragment incorrect ou engage un sous-dialogue de résolution de son incompréhension (que celle-ci soit acoustique ou linguistique). Dans une approche par concepts comme celle de [Rouillard 00], nous pouvons imaginer des concepts principaux et des concepts secondaires. En effet, nous pensons que la réaction de l'allocutaire dépend principalement de sa compréhension du concept principal (permettant de déterminer l'acte), l'incompréhension de concepts secondaires n'affectant pas la réponse aux attentes de celui-ci. Ainsi, l'incompréhension d'un concept principal apporte une rupture au dialogue et aboutit à un sous-dialogue tandis que celle de concept secondaire n'affecte pas la compréhension de l'idée principale liée à l'énoncé (malgré un fragment incompréhensible, l'énoncé semble répondre aux attentes de l'allocutaire).

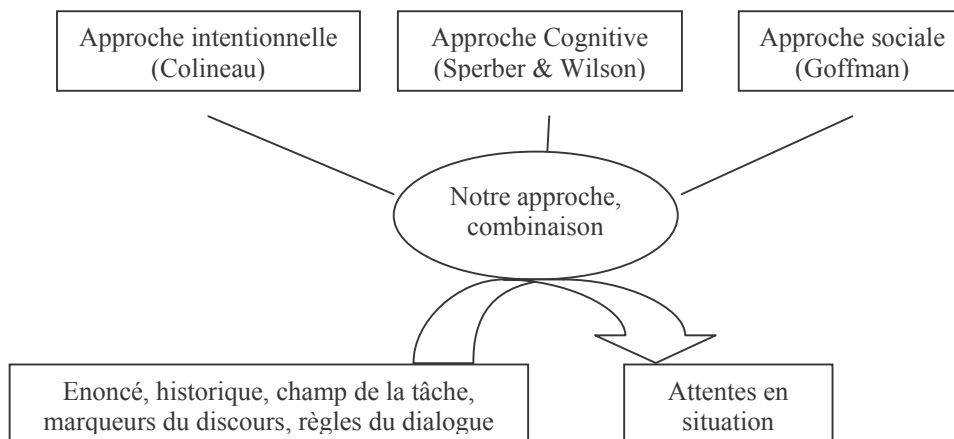
Cette étude justifie, selon nous, la modélisation au niveau pragmatique de l'énoncé, c'est-à-dire notre notation en actes de dialogue. C'est, en effet, à ce niveau que nous pourrions le mieux appréhender ces irrégularités et leurs effets. Un long travail d'annotation du corpus de 31 dialogues en actes, buts, stratégies et échanges a donc été effectué. Cette annotation nous a permis, dans un premier temps, d'estimer les règles de prédiction (ex :  $F^{\text{fs}}_A(p) \rightarrow F^{\text{s}}_B(p) \mid F^{\text{fs}}_B(p)$  ) puis, dans un deuxième temps, par une étude statistique, de valider cette estimation et de l'affiner en termes de probabilités associées aux attentes.

### 3. Application

#### 3.1. Combinaison des 3 approches

##### 3.1.1. Notre approche

Dans cette thèse, nous nous attachons à définir et catégoriser toutes les attentes possibles dans le dialogue puis nous cherchons à les modéliser pour aboutir à un algorithme de gestion efficace. Pour cela nous prenons en compte l'historique du dialogue, le champ de la tâche, les marqueurs de discours (thèse de N. Colineau [Colineau 97]) et les règles du dialogue. Mais nous nous inspirons aussi de la psychologie sociale avec la théorie des faces de Goffman [Goffman 73, 74] et de la psychologie cognitive avec les travaux de Sperber & Wilson [Sperber 86] sur la pertinence (**Figure 17**). Il s'agira également, dans le cas d'un dialogue oral, de gérer efficacement les incertitudes en provenance du système de reconnaissance.



**Figure 17** : Ebauche du modèle des attentes

Notre démarche est donc la synthèse de trois approches. L'approche intentionnelle, dirigée par les buts du locuteur, sera fondée sur les travaux de Colineau sur les marqueurs du discours (sémantiques, syntaxiques, etc.) et sur les actes de langage. Si catégoriser un acte actionnel, communicationnel, assuré ou délégué semble pertinent, tenir compte du type d'acte permet de prendre le contexte en considération en précisant où en est le dialogue (requête, confirmation, incidence, politesse, etc.). La structure des enchaînements des actes permet de prédire l'ensemble des actes pouvant intervenir en réponse à l'acte énoncé. Cette approche cherchera à définir les intentions communicatives de chacun. Mais l'étude des actes de langage ne suffit pas toujours. Nous adjoindrons alors à cette approche une

approche plus cognitive fondée sur l'idée d'un environnement cognitif qui évolue au cours du dialogue. Les attentes y sont davantage les effets de l'offre de l'autre locuteur. Nous tiendrons ainsi compte du fait que le but de chaque interlocuteur est d'avoir un discours pertinent donc de minimiser le coût (en terme de charge cognitive) du traitement cognitif de l'autre. Enfin, l'homme est fortement conditionné par la société. Une approche sociale semble alors importante pour comprendre certains énoncés fortement ritualisés où chaque locuteur doit se demander ce qu'il gagne et ce qu'il perd à donner une certaine réponse. Nous essayerons ainsi de tenir compte des raisonnements normalisés (par le milieu socioculturel), des rôles de chacun, de la hiérarchie qui donnera des informations de dominance de l'un sur l'autre et surtout des interactions par intercompréhension.

Nous pourrions nous intéresser à d'autres approches (émotionnelle, conative, etc.) mais elles n'interviennent que très peu dans un dialogue informatif.

### **3.1.2. Des exemples**

Dans cette partie, nous observerons des exemples d'analyse, à travers notre modèle, de dialogues issus du corpus de NESPOLE!. Nous montrerons les attentes de l'allocutaire B en rapport à ce qu'a dit le locuteur A. Nous découpons les attentes selon trois contextes (intentionnel, cognitif et social). P représentera le contenu propositionnel énoncé.

Nous étudierons en particulier les attentes face à une demande qui peut être de trois ordres : information, confirmation ou précision. La demande de précision est un cas particulier de la demande d'information dans le sens qu'elle éloigne le dialogue du but initial afin de répondre à un sous-but qui permettra de résoudre le but plus général.

Pour ces exemples, nous noterons s(p) pour savoir(p), qui correspond à connaître une information à propos de p, et c(p) pour croire(p), qui correspond à croire une information sur p.

### Demande d'information

Contexte intentionnel :  $\text{But}_A = s_A(p)$ , A veut connaître la réponse à sa question.

Contexte cognitif :  $\neg s_A(p)$ , A ne connaît pas la réponse à sa question.

Contexte social : la situation agent-client peut aider à coopérer.

Dans ce contexte, une demande d'information pourra être représentée ainsi :

$\frac{F_A^{fs}(p)}{\neg s_A(p) \wedge c_A(\text{coopere}_B)}$	$\Rightarrow$	$\frac{F_B^s(p)}{s_B(p)}$	B donne la réponse s'il sait
	$\Rightarrow$	$\frac{F_B^{fs}(p)}{\neg s_B(p) \wedge \text{coopere}_B}$	B demande des précisions sur P s'il ne sait pas et qu'il coopère
	$\Rightarrow$	$\frac{\neg F_B^s(p)}{\neg \text{coopere}_B}$	B ne donne pas de réponse s'il ne coopère pas

#### Exemple :

A la demande d'information du client : « j'aurais voulu savoir le/ les arrhes qu'il fallait verser avant », l'agent demande des précisions liées à son incompréhension : « vous désirez savoir s'il vous plaît ? le ? j'ai pas compris, excusez-moi/ ».

Le but n'est pas atteint mais donnera lieu à un sous-but qui portera sur la demande de précision.



### **Demande de confirmation**

Contexte intentionnel :  $\text{But}_A = s_A(p)$ , A veut connaître la réponse à sa question.

Contexte cognitif :  $c_A(p)$ , A croit connaître la réponse (ex : situation d'examen ou demande de confirmation).

Contexte social : la situation agent-client (ou la situation d'examen) peut aider à coopérer.

Dans ce contexte, une demande de confirmation pourra être représentée ainsi :

$F_A^{\text{fd}}(p)$	$\Rightarrow$	$F_B^{\text{d}}(p)$	B donne la réponse s'il sait
$c_A(p) \wedge c_A(\text{coopere}_B)$		$s_B(p)$	
	$\Rightarrow$	$F_B^{\text{fs}}(p)$	B demande des précisions sur P s'il ne sait pas et qu'il coopère
		$\neg s_B(p) \wedge \text{coopere}_B$	
	$\Rightarrow$	$\neg F_B^{\text{d}}(p)$	B ne donne pas de réponse s'il ne coopère pas
		$\neg \text{coopere}_B$	

Exemple :

A la demande de confirmation du client : « c'est e le centre de tourisme APT ? », l'agent répond par la positive car il sait : « oui, c'est bien là oui ».

Le but est atteint, A possède la réponse à sa question.

### Demande de précision

Contexte intentionnel :  $But_A = s_A(p)$ , A veut connaître la réponse à sa question.

Contexte cognitif :  $\neg s_A(p)$ , A ne connaît pas la réponse à sa question.

Contexte social : la situation agent-client peut aider à coopérer.

Dans ce contexte, une demande de précision pourra être représentée ainsi :

$F^{\text{fs}}_A(p)$	$F^{\text{s}}_B(p)$	B donne la réponse
$\neg s_A(p) \wedge c_A(\text{coopere}_B)$	$\Rightarrow$ $s_B(p)$	s'il sait
	$F^{\text{fs}}_B(p)$	B demande des précisions sur P
$\Rightarrow$	$\neg s_B(p) \wedge \text{coopere}_B$	s'il ne sait pas et qu'il coopère
	$F^{\text{fs}}_B(p)$	B demande une clarification sur P
$\Rightarrow$	$\neg \text{compris}_B(p) \wedge \text{coopere}_B$	s'il n'a pas compris et qu'il coopère
	$\neg F^{\text{s}}_B(p)$	B ne donne pas de réponse
$\Rightarrow$	$\neg \text{coopere}_B$	s'il ne coopère pas

### Exemple :

A la demande de précision du client : « pardon vous avez dit Grenoble ? », l'agent répond par la positive car il sait : « oui ». Le but est atteint, le dialogue va pouvoir continuer avec le but (ou sous-but) du niveau supérieur.

C'est ce que nous pouvons voir avec le dialogue suivant en situation coopérative :

Enoncé	But	Acte	Attentes
A: Habitez-vous à Grenoble ?	$?s_A(p)$	$F^{\text{fs}}_A(p)[\neg s_A(p)]$	$F^{\text{s}}_B(p)[s_B(p)] \vee F^{\text{fs}}_B(p)[\neg s_B(p)]$ $\vee F^{\text{fs}}_B(p)[\neg \text{compris}_B(p)]$
B: Pardon, vous avez dit Grenoble ?	$?s_B(p')$	$F^{\text{fs}}_B(p')[\neg \text{compris}_B(p')]$	$F^{\text{s}}_A(p')[s_A(p')] \vee F^{\text{fs}}_A(p')[\neg s_A(p')]$
A: Oui	$+s_B(p')$	$F^{\text{s}}_A(p')[s_A(p')]$	$F^{\text{s}}_B(p)[s_B(p)] \vee F^{\text{fs}}_B(p)[\neg s_B(p)]$
B: Non, j'habite dans la banlieue	$+s_A(p),$ $++s_B(p')$	$F^{\text{s}}_B(p)[s_B(p)]$	

Ainsi peut-on voir que l'acte suivant devrait appartenir aux attentes et permet d'affiner le contexte.

### **3.2. Mise en œuvre**

Représenter les attentes consiste, notamment, selon notre approche, en une représentation des différents contextes (cognitif, intentionnel et social). Pour cela, nous avons utilisé une approche corpus dont nous présenterons une analyse plus loin. Cette approche nous a permis de valider une notation des attentes en termes d'actes et de règles d'association avec probabilité d'apparition. Ainsi, nous aurons des règles de la forme  $F_{A}^{fs}(p) \rightarrow F_{C}^s(p) [93\%] \mid F_{C}^{fs}(p) [5\%]$ .

Les règles représenteront les aspects non seulement intentionnels mais aussi cognitifs (par exemple, à une demande d'information, l'allocutaire répondra s'il possède la réponse ou demandera des précisions s'il ne la possède pas). Elle ressemblent à des règles de logique modale classique que nous pourrions traduire en règles de logique modale normale par les travaux de [Gasquet & Herzig 96]. En revanche, les probabilités d'apparition de chaque règle permettent de mettre en jeu le contexte social. Ainsi, à partir d'une annotation fine d'un corpus en termes d'actes de langages, nous pouvons proposer ces règles et leurs probabilités d'apparition. Liées à une gestion des attentes qui suit celle des buts, c'est-à-dire une gestion de pile d'attentes, ces règles nous permettent de prendre en compte non seulement l'attente la plus probable dans un certain contexte, mais aussi les autres.

Le dialogue s'organise ainsi en fonction de la tâche à accomplir, et aussi en fonction de la rationalité des interactants, elle-même dépendant de contraintes psychologiques (processus cognitif de chacun) et sociales (conventions, rôles, etc.).

Cette étude de corpus réels présentés dans le chapitre « approche corpus » nous a permis dans un premier temps d'estimer les règles d'association entre actes et attentes. Nous avons donc pu proposer des règles de la forme  $F_{A}^f(p) \rightarrow F_{C}(p)$ . Dans le but d'une modélisation plus fine des rôles et du contexte, nous avons, ensuite, affiné ces règles en leur adjoignant des probabilités d'apparition liées à leur fréquence d'occurrence dans les dialogues du corpus. Pour cela, nous avons dû analyser ces même dialogues, transcrits puis annotés.

#### **3.2.1. Notation**

Une première étude des dialogues enregistrés puis transcrits dans le cadre du projet NESPOLE! [Burger & al. 01], [Besacier & al. 01] nous a permis de montrer certaines irrégularités liées à la maîtrise de la langue [Fouquet 01]. Nous y avons montré que ces irrégularités affectent plus ou moins la suite du dialogue. Nous nous sommes ensuite penchés sur les raisons de cette gêne plus ou moins importante. Nous pensons, comme nous l'avons proposé dans [Fouquet 02], que l'énoncé répond

néanmoins plus ou moins aux attentes de l'allocutaire. En effet, si la réponse du locuteur, même erronée, correspond à ce qu'attendait l'allocutaire, celui-ci n'a pas de raisons d'alourdir le dialogue en interrompant le locuteur sur une partie qui ne l'intéresse pas. Ayant obtenu son information, il laissera le dialogue se poursuivre normalement. En revanche, si cette réponse est trop erronée, l'allocutaire n'a pas ce qu'il attendait. Selon le principe d'engagement, il souhaite sa réponse. Dès lors, il engagera un sous-dialogue afin d'obtenir son renseignement. De même, si le locuteur est trop incompréhensible, les attentes de l'allocutaire ne sont pas satisfaites et celui-ci engagera un sous-dialogue de résolution de cette incompréhension.

Cette raison nous amène à envisager en premier lieu une approche par concepts [Rouillard 00]. Nous y distinguerons en revanche des concepts principaux et des concepts secondaires. Les premiers déterminent, selon nous, l'acte. Ils sont primordiaux dans la compréhension globale de l'énoncé et apportent une rupture et une sous-dialogue en cas d'incompréhension. Les seconds, moins importants, n'affectent pas la réponse aux attentes. En effet, malgré un fragment incompréhensible, l'énoncé semble répondre aux attentes de l'allocutaire.

En revanche, l'idée des concepts étant trop liée à la tâche, nous préférons modéliser le dialogue au niveau pragmatique, ce qui nous amène à utiliser une notation pragmatique de la forme  $F(p)$  où  $F$  est la force illocutoire et  $p$  le contenu propositionnel. Nous aurons donc des attentes de la forme  $F_A^{fs}(p) \rightarrow F_C^s(p)$ . C'est pour nous la seule manière de rester générique en restant hors de la tâche et en s'affranchissant davantage du niveau de maîtrise de la langue.

### **3.2.2. Estimation de règles**

Cette première étude nous a donc permis, outre de choisir une notation de 'haut niveau' (la notation pragmatique), d'estimer des règles d'associations entre actes et attentes. Ainsi, nous avons constaté des règles de la forme  $F_A^{fs}(p) \rightarrow F_C^s(p) \mid F_C^{fs}(p)$  en fonction du contexte. C'est que nous pouvons voir ci-dessous à travers deux actes, la demande d'information et celle de confirmation. Dans ces deux cas, le contexte intentionnel est le même (le but du locuteur  $A$  est de connaître la réponse à sa question  $S_A(p)$ ). De même, le contexte social est le même car nous restons dans le contexte du projet NESPOLE! mettant en jeu un agent et un client pour du renseignement touristique, ce qui aide à la coopération des interlocuteurs. En revanche, c'est sur le contexte cognitif que sera la différence. En effet, dans une demande d'information, le locuteur ne connaît pas la réponse ( $\neg S_A(p)$ ), tandis que, dans une demande de confirmation, il croit connaître la réponse ( $C_A(p)$ ). Ainsi, on pourra les représenter respectivement par les deux règles suivantes :

$F_{A}^{fs}(p)$	$\Rightarrow$	$F_{B}^{s}(p)$	B donne la réponse
$\neg S_{A}(p) \wedge C_{A}(Coopere_{B})$		$S_{B}(p)$	s'il sait
	$\Rightarrow$	$F_{B}^{fs}(p)$	B demande des précisions sur P
		$\neg S_{B}(p) \wedge Coopere_{B}$	s'il ne sait pas et qu'il coopère
	$\Rightarrow$	$\neg F_{B}^{s}(p)$	B ne donne pas de réponse
		$\neg Coopere_{B}$	s'il ne coopère pas

Exemple : A la demande d'information du client : « j'aurais voulu savoir le/ les arrhes qu'il fallait verser avant », l'agent demande des précisions liées à son incompréhension : « vous désirez savoir s'il vous plaît ? le ? j'ai pas compris, excusez-moi/ ». Le but n'est pas atteint mais donnera lieu à un sous-but qui portera sur la demande de précision.

$F_{A}^{fd}(p)$	$\Rightarrow$	$F_{B}^{d}(p)$	B donne la réponse
$C_{A}(p) \wedge C_{A}(Coop_{B})$		$S_{B}(p)$	s'il sait
	$\Rightarrow$	$F_{B}^{fs}(p)$	B demande des précisions sur P
		$\neg S_{B}(p) \wedge Coop_{B}$	s'il ne sait pas et qu'il coopère
	$\Rightarrow$	$\neg F_{B}^{d}(p)$	B ne donne pas de réponse
		$\neg Coop_{B}$	s'il ne coopère pas

Exemple : A la demande de confirmation du client : « c'est e le centre de tourisme APT ? », l'agent répond par la positive car il sait : « oui, c'est bien là oui ». Le but est atteint, A possède la réponse à sa question.

Enoncé	But	Acte	Attentes
C: j'aurais voulu savoir le/ les arrhes qu'il fallait verser avant	?S <sub>C</sub> (p)	F <sup>fs</sup> <sub>C</sub> (p)[¬ S <sub>C</sub> (p)]	F <sup>s</sup> <sub>A</sub> (p)[S <sub>A</sub> (p)] ∨ F <sup>fs</sup> <sub>A</sub> (p')[¬ S <sub>A</sub> (p)]
A: vous désirez savoir s'il vous plaît ? le ? j'ai pas compris, excusez-moi/	?S <sub>A</sub> (p')	F <sup>fs</sup> <sub>A</sub> (p')[¬ S <sub>A</sub> (p)]	F <sup>s</sup> <sub>C</sub> (p')[S <sub>C</sub> (p')] ∨ F <sup>fs</sup> <sub>C</sub> (p'')[¬ S <sub>C</sub> (p')]
C: le/ les arrhes, le e / l'avance qu'il faut faire	+S <sub>A</sub> (p')	F <sup>s</sup> <sub>C</sub> (p')[S <sub>C</sub> (p')]	F <sup>s</sup> <sub>A</sub> (p)[S <sub>A</sub> (p)] ∨ F <sup>fs</sup> <sub>A</sub> (p')[¬ S <sub>A</sub> (p)]
A: ah l'agence ? e(m)	?S <sub>A</sub> (p')	F <sup>fs</sup> <sub>A</sub> (p')[¬ S <sub>A</sub> (p)]	F <sup>s</sup> <sub>C</sub> (p')[S <sub>C</sub> (p')] ∨ F <sup>fs</sup> <sub>C</sub> (p'')[¬ S <sub>C</sub> (p')]
C: l'avance pardon	+S <sub>A</sub> (p')	F <sup>s</sup> <sub>C</sub> (p')[S <sub>C</sub> (p')]	F <sup>s</sup> <sub>A</sub> (p)[S <sub>A</sub> (p)] ∨ F <sup>fs</sup> <sub>A</sub> (p')[¬ S <sub>A</sub> (p)]
A: ah l'avance ok,	++S <sub>A</sub> (p'), ++ S <sub>C</sub> (p)	F <sup>s</sup> <sub>A</sub> (p)[S <sub>A</sub> (p)]	

Tableau 13 : Exemple de dialogue avec gestion des attentes

### 3.2.3. Probabilités

Selon nous, les règles représentent les aspects intentionnels mais aussi cognitifs (par exemple, à une demande d'information, l'allocutaire répondra s'il possède la réponse ou demandera des précisions s'il ne la possède pas). En revanche, les probabilités d'apparition de chaque règle permettent de mettre en jeu le contexte social. En effet, selon nous, la fréquence d'apparition des différents types de réponse dépend des rôles de chacun. Ainsi, dans la situation client/agent, l'agent pose davantage de demandes de précisions (sur des paramètres inhérents à la tâche) que le client. De même, dans la situation maître/élève, le maître pose davantage de demandes et l'élève de réponses.

L'introduction de probabilités dans ces règles permet de définir un ordre de priorité dans les attentes du locuteur. Ainsi, lorsqu'il pose une demande d'information, il espère d'abord une réponse. Il s'attend donc d'abord à celle-ci, mais s'attend aussi fortement à une demande de précision sur sa question en imaginant que l'interlocuteur peut ne pas avoir de réponse à donner par manque d'information. Il peut aussi s'attendre à un refus de réponse dans le cas d'un interlocuteur peu coopératif, une requête de répétition dans le cas d'un interlocuteur qui aurait mal entendu, une requête de reformulation dans le cas d'un interlocuteur qui aurait mal compris, etc.

Une règle tiendra donc compte des connaissances des interlocuteurs mais aussi des rôles et des stratégies employées. Elle aura la forme :

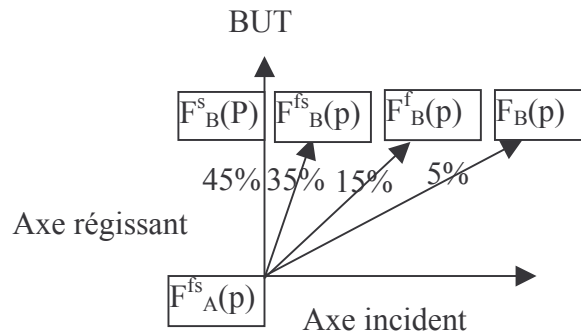
$$\frac{F_A(p)}{S_A(q)} \Rightarrow \frac{F'_B(p')}{S_B(q') \wedge \text{Str}'_B} \quad [x' \%]$$

$$\Rightarrow \frac{F''_B(p'')}{S_B(q'') \wedge \text{Str}''_B} \quad [x'' \%]$$

$$\Rightarrow \frac{F'''_B(p''')}{S_B(q''') \wedge \text{Str}'''_B} \quad [x''' \%]$$

A pose un acte  $F_A(p)$  en fonction d'un certain nombre de connaissances  $S_A(q)$ . Ses attentes seront par exemple à  $x'\%$  un acte de B  $F'_B(p')$  en fonction des connaissances de B  $S_B(q')$  et des sa stratégie  $\text{Str}'_B$ .

Dans un dialogue dirigé par les buts, nous pensons que l'acte le plus probable à la suite d'un acte est l'acte qui va tenter d'atteindre le but. Ainsi, nous pensons que les probabilités décrivent l'écart d'incidence par rapport à l'axe régissant de Luzzati. Par exemple (**Figure 18**) :



**Figure 18** : Les probabilités comme écart d'incidence

Ainsi, lorsque A pose un acte  $F_A^{fs}(p)$ , sa première attente sera celle qui dirige le plus le dialogue vers le but. Ensuite plus la probabilité d'apparition est faible, plus l'incidence est élevée. La deuxième attente s'écarte du but afin pour mieux s'en rapprocher plus tard. Et ainsi de suite jusqu'à la dernière attente, la moins probable. Si celle-ci est produite, alors l'incidence sera telle qu'elle peut remettre en cause le dialogue.

## **Conclusions théoriques**

Dans le cadre de la communication, le modèle du dialogue est un élément indispensable. S'il se résume à une approche par mots-clés (remplissage de blancs), il semble plus libre sur la syntaxe, mais contraint l'utilisateur à des requêtes plutôt simples sur des bases de données. Cela peut suffire dans certains cas (réservation ferroviaire ou recherche par mots-clés). L'élaboration d'une couche de dialogue plus complexe permet d'autres interactions plus riches et plus souples. Pour la traduction, la sémantique de l'énoncé est primordiale afin de ne pas générer d'incompréhension. Une analyse par mots-clés est insuffisante pour rendre compte du sens de l'énoncé.

En dialogue humain (éventuellement médiatisé par la machine) comme en dialogue homme-machine, nous constatons que, sur les différents modèles étudiés, peu sont utilisables dans le contexte de notre sujet de recherche. De plus, ces modèles semblent insuffisants seuls. C'est donc par composition d'un modèle intentionnel avec d'autres travaux issus des psychologies cognitive et sociale que nous avons tenté d'aborder au mieux les attentes en dialogue.

Notre modèle semble refléter ce qui se passe en communication verbale. Il s'intéresse non plus seulement au système mais surtout au locuteur et à ses attentes. Les attentes du locuteur sont vues comme les résultats espérés de ses actes. Elles représentent son état mental contenant ses buts, ses engagements, ses connaissances, des règles de conversation et les rôles des participants. Elle résultent donc d'une combinaison de différentes disciplines.

Nous proposons des règles pour les représenter. Ces règles, proches de la logique modale, tiennent compte des actes, des rôles, des stratégies, de la coopération des intervenants et du contexte cognitif. Elle présupposent des dialogues finalisés dans lesquels on cherche à atteindre un but.

La représentation en **actes** de langage permet d'aborder les aspects pragmatiques du dialogue et de prendre ainsi en compte le contexte d'énonciation. La **gestion des intentions** se fait par le principe des actes posant des buts et par le suivi de ces buts jusqu'à leur satisfaction. **La pertinence** est représentée par la coopération. La **gestion des aspects sociaux** se fait par le rôle du locuteur et les probabilités associées. Enfin, la **gestion des implicatures et des présuppositions** est issue de la gestion des croyances et connaissances remises en jour en permanence.

Notre modèle utilise actuellement des probabilités. Nous allons donc aborder le moyen de les définir, c'est-à-dire l'approche à base de corpus adoptée.



## **Chapitre II : Approche à base de corpus**

Dans le chapitre précédent présentant la partie théorique de notre approche, nous avons introduit des probabilités dans notre modélisation. Pour les déterminer, une approche à base de corpus permet de valider cette notion dans un contexte de dialogue réel humain ou homme-machine. Notre objectif est donc l'étude des attentes et leur modélisation en dialogue humain ainsi qu'en dialogue homme-machine.

Les objectifs du projet Nespole! de traduction de parole pour du e-commerce sont de pouvoir converser avec un agent de tourisme à l'étranger, chacun dans sa langue maternelle. Modéliser le dialogue permet de restreindre l'ensemble des

solutions de reconnaissance ainsi que de gérer correctement les ambiguïtés de traduction. Pour nous, cela nécessite une bonne gestion des attentes des locuteurs.

Les objectifs du projet PVE portent sur l'étude exploratoire d'un assistant électronique en entreprise. Une bonne gestion du dialogue est nécessaire pour traiter les tâches que l'assistant doit réaliser. Cette gestion du dialogue repose sur l'identification des actes posés par les interlocuteurs. Une bonne modélisation des attentes de chacun est donc encore nécessaire.

Nous avons présenté précédemment la méthodologie de conception incrémentale adoptée au long de ces différents projets et de notre thèse. Plusieurs cycles nécessitent des corpus. Ainsi, après avoir défini le modèle, l'approche à base de corpus a permis de définir le vocabulaire de l'application et ses contraintes linguistiques. Le corpus fournit des données objectives pour le système à concevoir et évaluer. Il permet de progresser dans un système en spirale en évaluant si le cahier des charges est cohérent et répond bien au besoin.

Le rôle du *corpus-pilote* est d'étudier les usages et de cibler correctement les besoins. Puis dans l'étape de conception, il faut évaluer ce que sera le système final : pour cela on a recours à des *corpus-simulés*, enregistrés par exemple avec des techniques de magicien d'Oz (un compère simulant le système de dialogue). Dans le cycle de développement du système, les versions successives sont évaluées à partir d'enregistrements de dialogues avec des usagers plausibles ou réels. L'évaluation prend en compte différentes données sociales et en particulier celles qui sont relatives aux usages possibles du système et de systèmes concurrents. Les évaluations portent sur le comportement technique des systèmes et sur les appréciations des usagers (confort, utilisabilité ...). On utilise parfois des corpus-tests préenregistrés contenant certaines difficultés qu'un système devrait surmonter. Comme il ne pourra s'agir de dialogues réels, qui par définition ne peuvent être préenregistrés, seuls certains aspects du dialogue oral peuvent faire l'objet de tests à partir de corpus de ce type, par exemple la compréhension des énoncés du dialogue.

Dans le cas qui nous occupe, l'étude d'un corpus permet d'aborder les particularités liées au dialogue, celles liées au dialogue oral et celles liées au dialogue dans ce cadre (simulation de la situation future). Nous avons ainsi élaboré trois collectes. La première concerne le corpus-pilote pour du dialogue homme-homme dans le cadre de la traduction de parole (Nespole!). La deuxième collecte concerne le corpus-pilote pour du dialogue homme-machine dans le cadre du remplacement de boîte vocale par assistant virtuel (PVE). Enfin, la troisième collecte concerne du corpus-simulé recueilli par technique de magicien d'Oz décrit plus loin.

## 1. Corpus-pilote homme-homme pour Homme-Homme : Nespole!

Durant l'été 2000, nous avons collecté 31 dialogues homme-homme en français. Nous les avons transcrits ensuite pendant l'automne 2000 avant d'en faire une annotation en actes, buts illocutoires, stratégie et structure, puis une analyse stochastique. Nous décrivons dans ce chapitre le protocole suivi et le corpus obtenu.

### 1.1. Protocole Nespole!

Dans le cadre de cette collecte, nous avons adopté la démarche suivante. La première étape a consisté en la mise en place d'une **plate-forme** de test. Cette plate-forme intègre la connexion à l'agence italienne à travers Internet. Pour cela, nous avons utilisé NetMeeting<sup>TM</sup>. Cet outil nous a permis en outre la mise en place d'un tableau partagé pour l'affichage des pages Internet. L'agent peut ainsi montrer des pages Internet au client. Enfin, pour l'acquisition du corpus, la plate-forme utilise TotalRecorder pour enregistrer en stéréo ce que chacun dit. Nous avons enregistré sur une voie ce que dit le client et sur l'autre ce que dit l'agent. La deuxième étape a consisté en la **traduction du site web** italien APT (de l'anglais vers le français). La troisième étape fut l'**élaboration de 6 scénarii** qui ont servi à cadrer les dialogues à recueillir. Dans ceux-ci, un client appelle une agence de tourisme afin d'organiser ses vacances dans le Trentin (région du Nord-Est italien). Lui sont alors notifiées les différentes activités qui l'intéressent et divers renseignements à remplir comme ses dates (ou sa période de vacances), son mode de transport, le nombre de membres de la famille, etc. S'ensuivent la **diffusion** de la demande à participation, le **recueil** des participations possibles, et la **convocation** jusqu'au jour de l'**enregistrement**. Il est à noter que l'agent est italien mais parle français pour l'occasion. Pour cette phase, on commence par un accueil du sujet, suivie d'une explication du projet et de la tâche qui lui sera demandée. Puis on fait quelques tests techniques de calibrage du micro et de fonctionnement de la plate-forme de test (des tests donc sur l'enregistrement avec TotalRecorder et sur la transmission avec Netmeeting). On enregistre alors le dialogue qui durera en 12 et 30 minutes. Lorsque les **31 dialogues** sont enregistrés, on commence une longue phase de **transcription** consistant en la retranscription par écrit des dialogues oraux. Nous avons effectué cette transcription, sous traitement de texte, avec Selma El-Moundi, Solange et Coralie Hollard. Nous l'avons ensuite vérifiée et validée. Suite à cela, nous avons effectué l'**élaboration d'une grille d'analyse** avec Anne-Claire Descalle et Solange Hollard. Cette grille met en relief le rôle des participants à travers un identificateur de locuteur (A pour Agent, C pour

Client). Elle contient les tours de parole avec un numéro d'intervention puis l'énoncé prononcé par le locuteur et l'acte de dialogue correspondant, suivi du contenu propositionnel, du but illocutoire et de la stratégie employée.

Enfin, la structure des dialogues a été représentée à travers les échanges. Une règle assez simple est que lorsqu'un but est posé, un échange débute, lorsqu'il est satisfait, l'échange se termine. Différents choix ont dû être faits quant à l'annotation avec Anne-Claire Descalle. En particulier, suite aux nombreux phatiques (« mm ») qui coupent l'énoncé en deux et au temps de transfert du signal (environ 1 seconde en moyenne), nous avons opté pour une reconstruction des énoncés.

Ensuite, pour les actes, l'ouverture (« bonjour ») a été annotée  $F^p$  car elle offre la possibilité de continuer ou non, le remerciement (« merci ») a également été annoté  $F^p$  car s'il clôt un échange, il offre la possibilité d'en ouvrir un autre. En revanche la clôture (« au revoir ») a été annotée  $F$  car elle incite à clore le dialogue et raccrocher. Notons que dans le cadre de PVE, le rôle de la machine étant mieux défini, la clôture de celle-ci (« au revoir » côté machine) a été annoté  $F^f$  car si elle fait clore le dialogue, elle laisse toujours la dernière initiative au client.

D'autres actions ont nécessité des choix. C'est le cas des demandes d'envoi ou d'attente par exemple. Une demande d'envoi ( $F^f$ (Pouvez-vous m'envoyer une brochure)) correspond à un  $F^{fs}$  mais avec une idée de réalisation d'action. Le but se réalise par le  $F^s$  suivant qui sera annoté comme  $F$  ( $F$ (je vous l'envoie)) même si le marqueur d'action n'est pas présent ( $F$ (oui d'accord)).

Pour la mise en attente ( $F^f$ (attendez un instant s'il vous plait)), le même choix a été fait. Mais l'interlocuteur attend parfois sans rien dire (ni « d'accord »,  $F$  non marqué linguistiquement, ni « j'attends »,  $F$  marqué linguistiquement).

A l'opposé, certaines énumérations peuvent prendre plusieurs tours de parole. Le but est pour nous atteint à la première réponse. Il est ensuite satisfait après le dernier tour de parole en réponse à ce but.

Enfin, dans le cas d'une répétition ou reformulation de l'expression d'un but, le premier but a été considéré comme abandonné. Munis de cette grille d'analyse et des ces choix, Anne-Claire Descalle et moi avons effectué une **annotation** de ces 31 dialogues. Ces annotations ont permis une **analyse** fine qui a servi à valider notre notion d'attentes à travers une approche stochastique que nous aborderons plus loin.

## 1.2. Description du corpus NESPOLE!

La collecte dans le cadre du projet Nespole! a permis d'enregistrer 191 dialogues monolingues dans différentes langues dont 31 en français. Ces 31 dialogues ont été transcrits puis annotés. Nous avons relevé 4633 actes répartis comme indiqué dans la description du corpus ci-après (**Tableau 14**) :

Rôle	F	F <sup>d</sup>	F <sup>f</sup>	F <sup>fd</sup>	F <sup>fs</sup>	F <sup>p</sup>	F <sup>s</sup>	Total
Agent	80	7	69	2	269	165	1672	2264
Client	66	2	23	4	442	115	1717	2369

**Tableau 14** : Répartition des 4633 actes du corpus Nespole! en fonction des rôles.

Un exemple d'annotation peut être observé ci-dessous (**Tableau 15**) :

C/A	Tr	Enoncé	Acte	Contenu propositionnel (P)	But	Str	Echange
C	1	Allô bonjour	Cfp(P)	salutations			Ouverture
A	2	Bonjour Monsieur	Afp(P)	salutations			
C	3	c'est le centre de tourisme APT	Cffd(P)	demande de confirmation	?Afd(P0)	D	Déb0 = demande confirmation
	4	bonjour	Cfp(P)	salutations			
	5	j'appelle	Cfs(P)	1 <sup>er</sup> essai d'exposer le motif de l'appel		D	
A	6	oui c'est bien là oui	Afd(P)	confirmation par A du lieu (APT)	+Afd(P0)	R	
C	7	c'est bien là d'accord	Cfs(P)	Phatique Ø	++Afd(P0)	R	Fin0 = confirmation.
	8	j'appelle pour avoir des renseignements pour organiser une semaine au ski au mois de décembre	Cffs(P)	Requête principale : C expose le motif de son appel	?Cs(P1)	D	Déb1 = demande rsg. pour organiser ses vacances.
A	9	oui	Afs(P)	réaction positive signe de compréhension du but de C + Possibilité de répondre à la question		R	
C	10	et donc je pense partir avec mes enfants et ma femme et donc	Cfs(P)	Compléments d'information de C.		D	
	11	j'aimerais avoir des renseignements sur la meilleure ville	Cffs(P)	Demande ; la meilleure ville.	- Cs(P1) ^ ?Cs(P2)	D	Déb2= demande de renseignements.
	12	quel est le meilleur endroit ou aller	Cffs(P)	demande d'infos/conseils ; meilleur endroit où aller.	- Cs(P2) ^ ?Cs(P3)	D	Déb3 = quel est le meilleur endroit ?
	13	et je souhaite	Cffs(P)	début de demande	-Cs(P3)	D	
A	14	Oui	Afs(P)	Phatique Ø		R	
	15	vous connaissez déjà notre région ou c'est la première fois que vous venez ici en Trentino	Affs(P)	demande ; connaissance de la région.	?As(P4)	K	Début4 = demande si C connaît la région.
C	16	C'est la première fois que je que je viens	Cfs(P)	réponse ; non	+As(P4)	R	
A	17	Ah oui	Afs(P)	Phatique Ø		R	

**Tableau 15** : Exemple d'annotation, corpus pilote pour du dialogue humain (Nespole!)

## 2. Corpus-pilote homme-homme pour Homme-Machine : PVE

Afin de définir les spécifications du modèle de dialogue, les tâches habituelles et le vocabulaire de l'application, nous avons enregistré durant le printemps 2002 plus de 800 dialogues réels entre des secrétaires et leurs interlocuteurs. Ces dialogues homme-homme collectés mettaient en relation des secrétaires et leurs appelants sur le canal téléphonique. Il s'agit de dialogues oraux entre humains. Les 44 dialogues les plus représentatifs des tâches les plus fréquentes d'un assistant ont été sélectionnés puis transcrits pour analyse. Leur annotation en actes et structures a permis d'élaborer une plate-forme de magicien d'Oz pour le recueil du corpus-simulé de 86 dialogues homme-machine (annotés *in vivo* en actes) dont nous présenterons la mise en œuvre plus loin.

### 2.1. Protocole PVE

Pour cette collecte, la démarche suivante a été suivie. La première étape, après une étude des **usages**, a consisté en un **démarchage** auprès de secrétaires. Après les accords nécessaires à l'écoute sur ligne téléphonique, nous avons mis en place un **service d'écoute** et d'enregistrement sur les lignes téléphoniques des secrétaires acceptant de se prêter à l'expérimentation. Nous avons ensuite **enregistré** plus de 800 dialogues réels dans ces conditions. Après un **classement** par tâches et une **sélection** de dialogues les plus pertinents pour notre projet, nous avons dégagé les trois tâches les plus fréquentes : joindre une personne, prendre un rendez-vous et réserver une salle. Nous avons ensuite ajouté 3 autres tâches considérées comme importantes pour un assistant virtuel : gérer un agenda partagé, recevoir une information, et envoyer un document. Nous avons ainsi obtenu six tâches représentatives d'un assistant :

1. [STAND+RDIR] redirection d'appel vers un membre de l'entreprise
2. [RDV] prise de rendez-vous avec un membre de l'entreprise
3. [RESA] réservation de salle
4. [RSG-AGD] consultation ou modification d'un agenda partagé
5. [ENV-DOC] envoi de document à un membre de l'entreprise ou à un extérieur.
6. [COM-INFO] communication d'une information par la machine

Nous avons sélectionné 44 dialogues (d'une durée de 2 à 7 minutes environ) considérés comme les plus représentatifs de ces tâches. Après **élaboration d'une**

**norme de transcription** suivant celle de Nespole!, l'étape suivante a consisté en la **transcription** (cf. [Descalle 02]) des 44 dialogues choisis. Après cette transcription, nous avons élaboré une **grille d'annotation** qui a permis l'**annotation** de ces transcriptions en termes de buts illocutoires, stratégies, actes et échanges.

Pour ces derniers, Anne-Claire Descalle a d'abord élaboré une structure d'échanges générique pour chaque dialogue, puis une structure d'échanges générique à tous les dialogues. Pour cela, elle a suivi un modèle d'analyse de dialogue structural (correspondant à l'école de Genève [Roulet & Moeschler 85]) avec une composante dynamique (à la suite de [Luzzati 95] & [Bilange 92]).

À partir de ces transcriptions, nous avons par ailleurs défini le **vocabulaire** utilisé dans le corpus homme-homme ainsi que des **classes** de mots propres à l'application. Puis une analyse de ces transcriptions a été faite alors en vue de développer le modèle de dialogue et le premier magicien d'Oz. Nous avons ainsi déterminé 122 **énoncés typiques** dans le corpus homme-homme.

Ces énoncés, contenant des variantes et des paramètres, seront ceux utilisés par le système dans sa première version. Les variantes, déduites de l'analyse, ont été introduites dans le but de simuler du dialogue plus naturel. Les énoncés de réponse contiennent en outre un certain nombre de paramètres propres à la tâche comme les nom et patronyme de l'utilisateur ou des membres de l'entreprise, les éléments de date (jour de la semaine, jour du mois, mois, année) ou de moment (heure, minute, moment de la journée). Un exemple d'énoncé peut être observé ci-dessous :

```
<DPRESET id="omni_rsg_dispo" type="R" act="fs" name=" [sexe] [nom] (est/n'est pas) disponible (ce jour-là)."/>
```

id représente l'étiquette d'identifiant pour repérer chaque énoncé, type représente la stratégie associée (ici réactif), act est l'acte de langage correspondant à l'énoncé et name contient l'énoncé sous une forme condensée. Ainsi cet énoncé pourra se décliner en quatre variantes (**Tableau 16**) :

[sexe] [nom] est disponible.	[sexe] [nom] est disponible ce jour-là.
[sexe] [nom] n'est pas disponible.	[sexe] [nom] n'est pas disponible ce jour-là.

**Tableau 16** : Variantes et paramètres pour le magicien d'Oz (PVE)

[sexe] et [nom] sont deux paramètres qu'il conviendra d'instancier au moment de la réponse. Ainsi, au moment de la réponse « [sexe] [nom] est disponible ce jour-là » pourra se remplacer par « monsieur dupond est disponible ce jour-là ».

## 2.2. Description du corpus-pilote PVE

Nous avons repris les annotations pour représenter les 44 dialogues en XML sous la forme décrite ci-après. En premier lieu, l'entête reprend diverses définitions comme celle du locuteur :

```
<DSPEAKERS>
  <DSPEAKER id="A" name="Vocalisa" type="Agent" />
  <DSPEAKER id="C" name="User" type="Client" />
</DSPEAKERS>
```

Définition des actes de langage utilisés :

```
<DACTS>
  <DACT id="f" name="Faire" />
  <DACT id="ff" name="Faire Faire" />
  <DACT id="fp" name="Faire Pouvoir" />
  <DACT id="fd" name="Faire Devoir" />
  <DACT id="fs" name="Faire Savoir" />
  <DACT id="ffs" name="Faire Faire Savoir" />
  <DACT id="ffd" name="Faire Faire Devoir" />
  <DACT id="ffp" name="Faire Faire Pouvoir" />
</DACTS>
```

Définition des types de tâches relevées :

```
<DTASKS>
  <DTASK id="0" name="null changement de tâche" />
  <DTASK id="1" name="standard et redirection" />
  <DTASK id="2" name="rendez-vous" />
  <DTASK id="3" name="réservation de salle et de matériel" />
  <DTASK id="4" name="communication d'une information" />
  <DTASK id="5" name="renseignement sur l'agenda" />
  <DTASK id="6" name="envoi de document" />
</DTASKS>
```

Ensuite se déroule un dialogue de la forme suivante :

```
<DIALOG type="training" ref="1">
  <COMMENT> c4d55 : RDV--> réussite. ACTE BUT STR ECHANGE
</COMMENT>
  <TASK id="2">
    <UTTERANCE speaker="A" act="fp">
      <TRS> CELAIO bonjour </TRS>
    </UTTERANCE>
    [...]
    <UTTERANCE speaker="A" act="f">
      <TRS> au revoir </TRS>
    </UTTERANCE>
  </TASK>
</DIALOG>
```

Nous pouvons alors voir un exemple de dialogue complet (passé par une XSL) :



## Dialogue numéro 1 Tâche : 2

Locuteur	Acte	Énoncé
A	fp	CELAIO bonjour
C	fp	oui bonjour euh
C	ffs	je voudrais savoir en fait euh euh si il faut prendre rendez-vous euh
C	fp	pour en fait euh parce que j' aimerais poser des questions sur l' orientation
A	fs	oui c'est mieux oui effectivement
C	fs	euh
A	ffs	vous êtes étudiante où là
C	fs	euh je suis étudiante en deuxième année de DEUG MIASS
A	fs	en deuxième année de DEUG MIASS oui
C	fd	voilà donc euh
A	fs	mmm
C	fs	j' aimerais avoir certaines précisions sur des des différents débouchés en fait
A	fs	bien sûr oui alors euh
A	ffs	quand est-ce que vous voulez venir demain
C	fs	euh ben euh demain matin si vous voulez
A	fs	alors je regarde l' agenda hein du conseiller
C	fs	oui
A	fs	je vais regarder tout de suite
A	ffd	demain c'est vendredi dix-sept hein
C	fd	voilà oui
A	fs	alors euh il a un rendez-vous à neuf heures
A	fp	vous pouvez venir à dix heures si vous voulez
C	fs	d'accord mais y a pas de problèmes
A	fs	d'accord
A	ffs	votre nom c'est
C	fs	NOM_PERSONNE
A	ffs	DEBUT_EPELATION_NOM_PERSONNE
C	fs	FIN_EPELATION_NOM_PERSONNE
A	fs	d'accord donc euh MIASS
A	ffs	éventuellement un numéro de téléphone si y a besoin
C	fs	euh oui zéro quatre
A	fs	mmm
C	fs	soixante seize
A	fs	mmm
C	fs	quatre-vingt cinq
A	fs	oui
C	fs	zéro deux
A	fs	mmm mmm
C	fs	quarante deux
A	f	d'accord c'est noté
A	fp	alors demain matin à dix heures
C	fs	d'accord
C	fp	ben merci beaucoup
A	fp	de rien je vous en prie
C	fp	au revoir
A	f	au revoir

Le corpus-pilote du projet PVE est composé de 2276 actes répartis comme suit, en fonction des tâches relevées précédemment (**Tableau 17**) :

Tâche	Rôle	F	F <sup>d</sup>	F <sup>f</sup>	F <sup>fd</sup>	F <sup>fs</sup>	F <sup>p</sup>	F <sup>s</sup>	Total	Total
<b>Stand RDIR</b>	Agent	5	6	3	4	11	11	62	102	206
	Client	5	8	5	6	9	18	53	104	
<b>RDV</b>	Agent	9	14	9	8	26	57	132	255	509
	Client	11	12	3	13	22	41	152	254	
<b>RESA Salle,matériel</b>	Agent	9	18	4	12	21	20	84	168	350
	Client	5	13	1	16	26	31	90	182	
<b>RSG AGD</b>	Agent	0	5	2	1	4	2	20	34	73
	Client	3	1	1	2	4	11	17	39	
<b>Total</b>	Agent	23	43	18	25	62	90	298	559	1138
	Client	24	34	10	37	61	101	312	579	

**Tableau 17** : Répartition des 1138 actes du corpus-pilote PVE en fonction des tâches et du rôle.

Nous verrons plus loin l'utilisation que nous avons faite des annotations du corpus pour notre approche stochastique. Ce corpus nous a permis d'élaborer une plate-forme de type magicien d'Oz afin de recueillir du corpus simulé de dialogue homme-machine. Selon la méthodologie employée, nous avons d'abord élaboré une première version que nous allons présenter à présent.

### 3. Corpus-simulé homme-machine version 0 : Pré-tests PVE

Comme nous l'avons vu dans le chapitre précédent, les transcriptions nous ont servi à élaborer un corpus adapté aux tâches auxquelles le système doit pouvoir répondre. Ce corpus étant fait de dialogues humains, il ne peut être suffisant pour l'adapter à du dialogue homme-machine. Nous avons donc collecté de nouveaux dialogues oraux par le biais d'un magicien d'Oz afin d'avoir du dialogue homme-machine. La démarche fut la suivante.

#### 3.1. Protocole corpus-simulé PVE version 0, pré-tests

Nous avons commencé par l'élaboration d'une **plate-forme** en vue d'une expérimentation de type Magicien d'Oz (version 0) pour recueillir du corpus homme-machine adapté à la tâche à réaliser et valider certaines hypothèses avant la modélisation effective. La plate-forme sera décrite plus en détail dans le chapitre « Magicien d'Oz ». Notons seulement que la plate-forme contient un système de reconnaissance vocale en entrée et de synthèse de parole en sortie, de sorte que la modalité d'interaction avec l'utilisateur est essentiellement vocale, les seules autres actions consistant à se mettre en relation avec le service d'assistant, décrocher, raccrocher et éventuellement appuyer pour parler. L'expérimentation a alors consisté à faire passer des ergonomes devant le système en tant que client et en tant que magicien simulant l'agent. Ceux-ci connaissaient donc la visée magicien d'Oz de l'expérimentation. Le sujet avait alors le rôle d'utilisateur puis de magicien 'contrôleur de dialogue', le magicien 'annotateur' étant géré par moi-même. Les dialogues étaient **recueillis** sous forme orale et écrite (à travers la reconnaissance vocale) et **annotés** *in vivo* en actes. Des questions sur les remarques, commentaires et améliorations à apporter ont été posées aux utilisateurs et aux expérimentateurs magiciens. Ces retours ont permis une amélioration significative de l'ergonomie de l'interface. En outre, les dialogues recueillis ont permis de déterminer le **vocabulaire** utilisé dans le corpus homme/machine et les **énoncés typiques** dans le corpus homme/machine. Enfin, cette expérimentation a permis de valider nos premières hypothèses. La qualité du dialogue dépendant de la qualité des hypothèses de reconnaissance vocale, une **transcription** a été faite sur les énoncés réellement prononcés.

### 3.2. Description du corpus-simulé PVE version 0, prétests

Nous avons repris les annotations et les transcriptions pour représenter les 14 dialogues obtenus en XML sous la forme décrite ci-après. En premier lieu, l'entête reprend diverses définitions comme celle du locuteur ou des actes comme nous l'avons vu au chapitre précédent. L'entête de définition des tâches a été modifiée pour obtenir une entête de la forme :

```
<DTASKS>
  <DTASK id="ouverture" name="null, changement de tâche" />
  <DTASK id="stand" name="standard et redirection" />
  <DTASK id="rdv" name="rendez-vous" />
  <DTASK id="resa" name="réservation de salle et de
matériel"/>
  <DTASK id="comminfo" name="communication d'une information"
/>
  <DTASK id="rsgagd" name="renseignement sur l'agenda" />
  <DTASK id="envdoc" name="envoi de document" />
</DTASKS>
```

Le dialogue se déroule ensuite de la façon suivante :

```
<DIALOG type="pretest" ref="2">
  <TASK id="ouverture">
    <UTTERANCE speaker="A" act="fp">
      <PRESET id="omni_bonjour2"/>
    </UTTERANCE>
  </TASK>
  <TASK id="resa">
    <UTTERANCE speaker="C" act="ffs">
      <SIGNAL> 1597.wav </SIGNAL>
      <ASR> j' aimerais réserver une salle </ASR>
      <TRS> j' aimerais réserver une salle </TRS>
      <ANALYZE>
        <PART act="ffs"> j' aimerais réserver
une salle </PART>
      </ANALYZE>
    </UTTERANCE>
    <UTTERANCE speaker="A" act="ffs">
      <PRESET id="resa_materiel"/>
    </UTTERANCE>
  [...]
  <UTTERANCE speaker="A" act="ff">
    <PRESET id="omni_au_revoir"/>
  </UTTERANCE>
  <UTTERANCE speaker="C" act="f">
    <SIGNAL> 1604.wav </SIGNAL>
    <ASR> revoir </ASR>
    <TRS> au revoir </TRS>
    <ANALYZE>
      <PART act="f"> au revoir </PART>
    </ANALYZE>
  </UTTERANCE>
</TASK>
</DIALOG>
```

Nous pouvons voir dans cet exemple le rajout d'une tâche d'ouverture. En effet, à l'instar de Luzzati ou Bilange, nous pensons que cette phase est indépendante de la tâche qui suit. Cette phase, primordiale pour bien débiter un dialogue homme-machine, sera étudiée à part par la suite.

Un exemple de dialogue complet peut être observé ici :

**Dialogue numéro 2**

**Tâche : ouverture**

Locuteur	Acte	Énoncé	Signal
A	fp	Bonjour, ici la secrétaire virtuelle Vocalisa. Quelle est votre demande ?	

**Tâche : resa**

Locuteur	Acte	Énoncé	Signal
C	ffs	j' aimerais réserver une salle j' aimerais réserver une salle	1597.wav
A	ffs	Avez-vous besoin du matériel de projection ?	
C	fs	oui j' aimerais bien un message projection oui j' aimerais bien un matériel de projection	1598.wav
A	ffs	Ce serait pour quelle date précisément ?	
C	fs	ce serait pour lundi prochain à seize heures ce serait pour lundi prochain à seize heures	1599.wav
A	ffs	C'est à quel nom ?	
C	fs	est-ce que monde dominique blanc de euh c'est au nom de dominique blanc	1600.wav
A	ffs	Voulez-vous la salle de vingt-cinq places ou celle de trente places ?	
C	fs	j' aimerais salle de trente places s' il vous plaît j' aimerais celle de trente places s' il vous plaît	1601.wav
A	ffd	Nous disons donc lundi 17 à 16 heures salle B2 ?	
C	fd	c'est parfait c'est parfait	1602.wav
A	ff	Pouvez-vous m'envoyer une confirmation par fax ou par courriel, s'il vous plaît ?	
C	f	ah problème merci au revoir pas de problème merci au revoir	1603.wav
A	fs	okay	
A	ff	Au revoir mademoiselle	
C	f	revoir au revoir	1604.wav

Chaque énoncé est représenté sous sa forme reconnue par le système de reconnaissance vocale puis sous sa forme retranscrite (en rouge).

Le corpus de prétests est composé de 722 actes répartis ainsi (**Tableau 18**) :

Tâche	Rôle	F	F <sup>d</sup>	F <sup>f</sup>	F <sup>fd</sup>	F <sup>fs</sup>	F <sup>p</sup>	F <sup>s</sup>	Total	Total
<b>Stand RDIR</b>	Agent	4	0	6	0	8	1	8	27	70
	Client	5	0	2	0	12	5	19	43	
<b>RDV</b>	Agent	3	0	1	2	8	0	2	16	36
	Client	0	2	0	0	3	2	13	20	
<b>RESA</b> <b>Salle,matériel</b>	Agent	5	0	3	5	15	3	3	32	72
	Client	3	5	3	0	6	3	20	40	
<b>RSG AGD</b>	Agent	10	1	7	1	6	0	6	31	71
	Client	2	1	2	1	15	7	12	40	
<b>ENVDOC</b>	Agent	5	0	5	0	9	1	5	25	58
	Client	5	0	8	0	3	5	12	33	
<b>Comminfo</b>	Agent	1	0	3	0	5	2	6	17	29
	Client	3	0	0	0	0	0	9	12	
<b>Ouverture</b>	Agent	0	0	0	0	13	12	0	25	25
	Client	0	0	0	0	0	0	0	0	
<b>Total</b>	Agent	28	1	25	8	64	17	30	173	361
	Client	18	8	15	1	39	22	85	188	

**Tableau 18** : Répartition des 361 actes du corpus de prétests PVE en fonction des tâches et du rôle.

Cette phase de prétests nous a donc permis d'alimenter notre plate-forme d'expérimentation en données homme-machine et de l'adapter à de tels types de dialogue. Les dialogues ont été rajoutés au premier corpus afin d'obtenir un corpus final de 1110 phrases composées d'un vocabulaire de 1119 mots. Une expérimentation réelle a été ensuite mise en place tenant compte des améliorations à apporter tant au niveau de l'interface qu'au niveau des énoncés à prévoir.

#### 4. Corpus simulé homme-machine version 1 : TestsPVE

La phase précédente a constitué, dans notre démarche, une version 0 du système final. Son analyse a permis la mise en place de la version 1 et de l'expérimentation associée. Le but de cette expérimentation est de recueillir du corpus homme-machine adapté à la tâche à réaliser, et de valider certaines hypothèses avant la modélisation effective.

##### 4.1. Protocole TestsPVE

Pour cette expérimentation, nous avons mis en place **une plate-forme** de type magicien d'Oz à partir de la précédente. Nous avons ensuite élaboré un **questionnaire** concernant la qualité perçue du système. Après la **diffusion** de la demande à participation, le **recueil** des participations possibles, la **convocation** de ces participants, le **passage de l'expérimentation** a lieu. Les six tâches retenues ont été réparties en deux situations d'expérimentation :

Situation a) :	<ul style="list-style-type: none"> <li>✓ La prise de rendez-vous, notée RDV ;</li> <li>✓ La réservation de salle et de matériel, notée RESA ;</li> <li>✓ L'envoi de documents, notée ENV-doc</li> </ul>
Situation b) :	<ul style="list-style-type: none"> <li>✓ La tenue du standard et la redirection d'appel, notée STAND+RDIR</li> <li>✓ La tenue d'agenda, notée RSG-agd</li> <li>✓ La communication d'une information, notée COM-info</li> </ul>

**Tableau 19** : Les six tâches retenues dans l'expérimentation en magicien d'Oz (PVE)

Concernant la **tâche** à accomplir, les sujets sont invités à interagir avec le prototype d'assistant de communication, en jouant le rôle d'un utilisateur, désireux de faire accomplir à la machine différentes tâches, précisées dans les consignes. Ces tâches correspondent aux situations génériques définies dans les étapes précédentes. Il y a 24 **sujets** âgés de 23 à 56 ans, 7 femmes et 17 hommes. Tous avaient déjà travaillé sur ordinateur.

Le dispositif comprend deux salles. Une salle « sujet » contient une machine destinée au sujet, et une machine dédiée au module d'actes de dialogue (le « serveur

d'actes ») que nous décrirons dans le chapitre « Magicien d'Oz ». Un expérimentateur accueille le sujet dans cette salle, et reste présent pendant toute la passation, veillant au bon déroulement de la séance et contrôlant le serveur d'actes. Une salle « magicien » contient une machine, gérée par deux expérimentateurs.

Toute la séance se déroule oralement. Le sujet entend l' « assistant de communication » grâce à des écouteurs, et répond en parlant dans un micro. Il « décroche » et « raccroche » en cliquant sur l'écran de sa machine. Les prises de parole sont réglées par un dispositif « push to talk ».

Le déroulement des séances est le suivant. Dans la salle « sujet », le sujet est accueilli. La tâche lui est présentée ainsi que les consignes (situation a ou b). Ensuite le matériel est essayé, le micro est calibré, les boutons « décrocher » « raccrocher » et le système « push to talk » sont testés. Outre les explications orales de l'expérimentateur, le sujet dispose d'une feuille de consignes, qui comprend également un emplacement destiné aux notes prises pendant la séance. Il peut donc donner ses commentaires sur les scénarii prévus et sur le système.

La tâche est alors réalisée. Le sujet doit par trois fois (une fois pour chaque tâche à effectuer) se connecter au système Vocalisa, dialoguer avec le système, puis se déconnecter quand le dialogue est considéré terminé.

Suite à ces trois 'appels', l'entretien commence. Pour celui-ci, l'un des expérimentateurs « magiciens » vient recueillir les impressions générales du sujet et l'interroge notamment sur les réponses du système, l'enchaînement du dialogue, le retour d'information (*feedback*), le guidage, les points positifs et négatifs de l'ensemble et les améliorations souhaitées.

Le questionnaire et les résultats seront présentés en annexe. Lorsque les 24 sujets sont passés, nous avons procédé au **dépouillement** du questionnaire et des résultats.

#### **4.2. Corpus simulé PVE version 1, tests**

Les 72 dialogues obtenus ont été traités comme les précédents. Ils ont été mis au format XML. Les annotations *in vivo* en actes de dialogue ont été ensuite vérifiées. Les énoncés ont été transcrits *in vitro* afin de tenir compte des erreurs de reconnaissance. Afin d'uniformiser et ainsi pouvoir comparer, les entêtes sont les mêmes que pour la phase de pré-tests. Un dialogue de test a la forme XML suivante :



```

<DIALOG type="test" ref="2">
  <TASK id="ouverture">
    <UTTERANCE speaker="A" act="fp">
      <PRESET id="omni_bonjour"/>
    </UTTERANCE>
  </TASK>
  <TASK id="resa">
    <UTTERANCE speaker="C" act="ffs">
      <SIGNAL> 1758.wav </SIGNAL>
      <ASR> c' est bonjour ici dominique blanc je
suis vraiment doctoral je dois soutenir ma
thèse avant la fin du mois que j' ai je
souhaite réserver une salle équipee d' un
rétro-projecteur pour le vingt-deux personne
</ASR>
      <TRS> bonjour ici dominique blanc je fini
mon doctorat je dois finir ma thèse avant la
fin du mois et je souhaite réserver une
salle équipéee d' un rétro-projecteur pour
une vingtaine de personnes </TRS>
      <ANALYZE>
        <PART act="fp"> bonjour </PART>
        <PART act="fs"> ici dominique
blanc</PART>
        <PART act="fs"> je fini mon
doctorat</PART>
        <PART act="fs"> je dois finir ma thèse
avant la fin du mois </PART>
        <PART act="ffs"> et je souhaite
réserver une salle équipéee d' un
rétro-projecteur pour une vingtaine de
personnes </PART>
      </ANALYZE>
    </UTTERANCE>
    <UTTERANCE speaker="A" act="ffs">
      <PRESET id="resa_date"/>
    </UTTERANCE>
    [...]
    <UTTERANCE speaker="C" act="ff">
      <SIGNAL> 1763.wav </SIGNAL>
      <ASR> très bien au revoir </ASR>
      <TRS> très bien au revoir </TRS>
      <ANALYZE>
        <PART act="fd"> très bien </PART>
        <PART act="ff"> au revoir </PART>
      </ANALYZE>
    </UTTERANCE>
    <UTTERANCE speaker="A" act="ffd">
      <PRESET id="resa_resume2"/>
    </UTTERANCE>
    <UTTERANCE speaker="A" act="f">
      <PRESET id="omni_bonjour"/>
    </UTTERANCE>
  </TASK>
</DIALOG>

```

Nous pouvons alors voir un exemple de dialogue de test complet comprenant l'annotation des tâches effectuées, des tours de parole, du locuteur pour chaque tour, de son acte associé à l'hypothèse de reconnaissance, l'énoncé transcrit et le fichier de signal correspondant.

### Dialogue numéro 2

Tâche : ouverture

Locuteur	Acte	Enoncé	Signal
A	fp	Bonjour, ici l'assistante virtuelle Vocalisa. Quel est le motif de votre appel, s'il vous plaît ?	

Tâche : resa

Locuteur	Acte	Enoncé	Signal
C	ffs	c' est bonjour ici dominique blanc je suis vraiment doctoral je dois soutenir ma thèse avant la fin du mois que j' ai je souhaite réserver une salle équipée d' un rétro-projecteur pour le vingt-deux personne bonjour ici dominique blanc je fini mon doctorat je dois finir ma thèse avant la fin du mois et je souhaite réserver une salle équipée d' un rétro-projecteur pour une vingtaine de personnes	1758.wav
A	ffs	Ce serait pour quelle date ?	
C	fs	euh je n' est pas de les dates absolument précisé disant la semaine prochaine euh je n' est pas de date absolument précise disons la semaine prochaine	1759.wav
A	ffs	Voulez-vous la salle de vingt-cinq places ou celle de trente places ?	
C	fs	je préfère ai la salle de trente places je préférerais la salle de trente places	1760.wav
A	fs	J'aurais une salle à ce moment là mais sans rétro-projecteur, sinon la salle avec rétro-projecteur sera libre lundi 31 à 10 heures	
C	ffs	oui mais faut absolument un rétro-projecteur trente donc euh je te dire compris les dates trente me tromper de oui il me faut absolument un rétro-projecteur donc euh je n'ai pas bien compris la date le trente ou le trente et un	1761.wav
A	fs	lundi 31 à 10 heures	
C	fs	d' accord très bien d' accord très bien	1762.wav
A	ffd	Nous disons donc lundi 31 à 10 heures salleB2	
C	ff	très bien au revoir très bien au revoir	1763.wav
A	ffd	Nous disons donc lundi 31 à 10 heures salleB2	
A	f	Au revoir madame	

L'expérimentation en magicien d'Oz nous a permis de collecter 86 dialogues homme-machine mettant en situation un client et l'assistant virtuel d'une entreprise. Le corpus est composé de six tâches plus une septième artificiellement rajoutée à l'expérimentation. Cette septième tâche concerne en fait l'ouverture du dialogue (donc la première phrase du dialogue) et la phase de transition d'une tâche à l'autre dans le même dialogue (« Que puis-je faire d'autre pour vous ? »). Elle sera notée Ouverture.

Les dialogues sont constitués de 1626 actes de dialogue. Nous pouvons voir la répartition de ces actes en fonction des tâches et des rôles sur le **Tableau 20** :

Tâche	Rôle	F	F <sup>d</sup>	F <sup>f</sup>	F <sup>fd</sup>	F <sup>fs</sup>	F <sup>p</sup>	F <sup>s</sup>	Total	Total
Stand RDIR	Agent	17	0	23	4	54	0	48	146	365
	Client	14	4	22	0	45	23	111	219	
RDV	Agent	6	0	9	8	31	0	11	65	156
	Client	10	8	3	0	14	7	49	91	
RESA Salle,matériel	Agent	7	0	16	13	41	4	21	102	241
	Client	14	9	4	0	26	15	71	139	
RSG AGD	Agent	22	3	8	9	29	0	26	97	237
	Client	7	8	23	1	26	11	64	140	
ENVDOC	Agent	31	0	23	0	20	0	24	98	278
	Client	27	0	42	0	24	22	65	180	
Comminfo	Agent	9	0	10	5	19	12	39	94	207
	Client	18	8	13	0	15	11	48	113	
Ouverture	Agent	0	0	1	0	72	60	2	135	142
	Client	0	0	0	0	3	2	2	7	
Total	Agent	92	3	90	39	266	76	171	737	1626
	Client	90	37	107	1	153	91	410	889	

**Tableau 20** : Répartition des 1626 actes du corpus de tests PVE en fonction des tâches et rôles.

## Conclusion sur les corpus

Le corpus de dialogues humains NESPOLE! est composé de 31 dialogues et 4633 actes de dialogues qui portent sur la tâche de renseignement touristique. Pour l'analyse du corpus nous avons suivi la méthodologie présentée précédemment pour le dialogue homme-homme. Une phase de collecte de dialogues oraux a été suivie de la transcription des dialogues puis de leur annotation.

Le corpus de dialogues PVE se répartit en 7 tâches dont une artificielle nommée 'ouverture'. La distribution des dialogues en fonction des tâches est la suivante (**Tableau 21**) :

Tâches	Corpus-pilote	Pré-tests	Tests
Redirection, Standard	5	4	17
Rendez-vous	11	2	10
Réservation de salle	10	3	12
Communication d'information	0	2	12
Gestion d'un agenda partagé	2	4	16
Envoi de document	0	3	20
Ouverture	-	13	72

**Tableau 21** : Description du corpus PVE en fonction des tâches.

Pour l'analyse du corpus, nous avons suivi la méthodologie incrémentale présentée précédemment pour le dialogue homme-machine. Une phase de collecte de dialogues oraux a été suivie de la transcription des dialogues, puis de leur annotation. Le corpus-pilote ainsi créé a permis le développement d'un magicien d'Oz dans une première version. Une nouvelle collecte, via le système en magicien d'Oz, a été transcrite puis annotée afin de créer le corpus de pré-tests qui a permis d'ajuster le système en magicien d'Oz à du dialogue homme-machine. La troisième campagne de collecte s'est ensuite faite sur la plate-forme à base de magicien d'Oz dans sa version finale. Les dialogues obtenus ont été transcrits puis annotés afin d'en faire une analyse qui permettra le développement du système final.

Nous allons maintenant décrire le système à base de magicien d'Oz que nous avons développé pour faire de la collecte de dialogues homme-machine.

## **C. Expérimentation et tests**



## **Chapitre I : Plate-forme d'expérimentation (Magicien d'oz)**

Souhaitant concevoir des systèmes interactifs de dialogue homme-machine, nous manquons de corpus (homme-machine) en langue naturelle. Or, l'étude de la communication homme-machine ne peut se fonder uniquement sur le modèle de communication humaine. L'interaction homme-machine fait apparaître des phénomènes différents de l'interaction humaine. On se trouve ainsi en présence du problème circulaire relevé par [Siroux et al. 89], de disposer de données pour réaliser un système qui n'a pas encore produit ces données. « *La conception d'un système exige que soit identifié le langage de ses utilisateurs, et cette identification exige l'existence du système. Pour sortir de cette contradiction, la solution fréquemment adoptée est celle d'expérimentation "Magicien d'Oz".* [Polity et al. 90] ».

## **1. Magicien d'Oz**

### **1.1. Magicien d'Oz pour du dialogue**

Dans une expérience de type Magicien d'Oz, un compère humain simule les comportements de la machine à l'insu de l'utilisateur. Ce dernier pense donc dialoguer avec une machine, alors qu'il dialogue avec un compère. Cette technique entraîne de nombreuses contraintes : effort d'organisation, de préparation du matériel, choix de sujets motivés pour se soumettre à l'expérience, prise de rendez-vous, déplacement sur le lieu de l'enregistrement, consignes plus ou moins précises à respecter pour le compère, etc.

Ces simulations sont donc délicates et coûteuses à élaborer. Les corpus de données ainsi recueillis sont souvent peu volumineux, et il n'est pas sûr que les dialogues obtenus dans ces conditions ne soient pas biaisés. Il existe, en effet, des comportements langagiers induits par la machine : similitudes, enrichissements, appauvrissements, etc. Or, avec un magicien d'Oz, ces phénomènes sont parfois engendrés par les propres comportements du compère.

Des travaux comme ceux de [Rouillard 00] ou [Lemeunier 00] tentent de se passer de ce type d'expérimentation au profit d'une application réelle enrichie au fur et à mesure des tests. Dans le cadre du projet PVE, [Nguyen 03a] propose également une application réelle enrichie au fur et à mesure par l'adjonction de nouvelles données (nouvelles tâches, nouveau type d'énoncé, etc.).

Pour le recueil, des données nous avons opté pour une solution de type magicien d'Oz selon la méthodologie présentée au premier chapitre : collecte de corpus homme-homme, transcription, annotation et analyse des dialogues humains obtenus. Nous avons ensuite élaboré une première plate-forme de type magicien d'Oz simulant le système. Nous avons recueilli un premier corpus de dialogues homme-machine dans le cadre de l'application visée. Ce premier corpus a été analysé afin d'améliorer la qualité de la simulation et l'ergonomie de l'utilisation de la plate-forme. La deuxième version de la plate-forme a permis une expérimentation de recueil de dialogues homme-machine qui ont été analysés en vue du système final.

Nous verrons donc plus en détails les raisons pour lesquelles nous avons opté pour l'utilisation de magicien d'Oz. Nous nous attacherons à leur description. Nous aborderons enfin l'apport de la notion d'attentes comme aide à l'humain avant d'être aide au système.



## 1.2. Magicien d'Oz pour du dialogue oral

Les interfaces utilisant le canal de parole sont souvent plus adaptées que les interfaces uniquement graphiques. Elles semblent plus naturelles et permettent leur accès à certains individus non-voyants, handicapés, utilisant leurs yeux ou leurs mains pour d'autres tâches, etc.

La plupart des systèmes de dialogue oral considèrent que l'utilisateur a déjà appris le vocabulaire et la forme de dialogue (grammaire) possibles à chaque étape du dialogue. Si de telles considérations sont concevables pour des utilisateurs experts, elles le sont beaucoup moins pour le grand public, même sur des tâches de complexité moyenne. Le problème de la conception de systèmes de dialogue oral est loin d'être évident. La formulation est très variable d'une personne à l'autre (différents mots et structures pour un même sens) et pour une même personne. De plus, la langue évolue, créant par exemple des mots comme « courriel », mots que le système de dialogue oral devra éventuellement être capable de comprendre. Enfin, l'utilisateur répondra différemment selon la structure du dialogue et son historique (en particulier l'énoncé précédent).

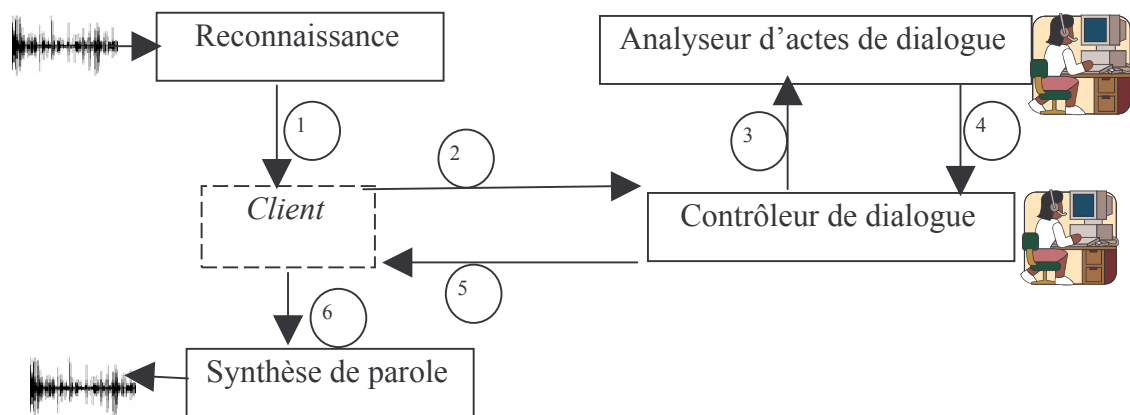
Une solution consiste à développer des systèmes de dialogue oral pour des tâches restreintes. L'étude de dialogues humains étant insuffisante pour créer un système dans lequel une machine remplacerait l'un des deux interlocuteurs, nous avons développé un premier magicien d'Oz simulant le comportement du système (un assistant en entreprise). Ainsi, nous avons élaboré un corpus de dialogue oral homme-machine. L'analyse de ce corpus a permis de montrer certaines différences entre le dialogue oral humain et le dialogue oral homme-machine.

Nous présenterons d'abord l'architecture générale de la plate-forme de test considérant le magicien d'Oz comme une solution acceptable pour capturer des dialogues oraux homme-machine et tester la première version du système (les phrases servant de réponse). Nous aborderons ensuite les modules de reconnaissance vocale, interprétation en actes, contrôle de dialogue et synthèse de parole.

Les sujets appellent l'assistant virtuel, actuellement *via* micro-casque, pour les tâches suivantes : mise en relation avec une personne, prise de rendez-vous, réservation de salle, gestion d'un agenda partagé, envoi de documents ou réception d'informations. Le magicien d'Oz répond à ces tâches par des actions appropriées. Il interprète l'acte de dialogue du locuteur en fonction de la situation, contrôle les différentes étapes qui permettent de conduire au but dialogique de l'utilisateur, et répond par l'action la plus appropriée (la réponse correspondant à la situation).

## 2. Architecture générale du système

Le système de dialogue est composé de quatre parties dont la reconnaissance vocale en entrée et la synthèse de parole en sortie. Les deux autres sont simulés dans notre approche par des magiciens d'Oz. L'un simule la partie interprétation en actes par l'annotation en actes de dialogue des hypothèses de reconnaissance. L'autre s'occupe du contrôle de dialogue. Il reçoit les hypothèses de reconnaissance et les actes de dialogues associés, choisit la réponse la plus appropriée, et la renvoie à la synthèse de parole du client. La **Figure 19** montre l'architecture du système de dialogue dans lequel chacune des quatre parties est implantée comme un serveur qui communique avec les autres parties. Chaque partie sera décrite dans ce chapitre.



**Figure 19** : Architecture du système de dialogue.

Le scénario est le suivant :

1. Le client reçoit l'**hypothèse** de reconnaissance d'un serveur de reconnaissance local.
2. Le client envoie cette **hypothèse** au contrôleur de dialogue.
3. Le contrôleur de dialogue envoie cette **hypothèse** à l'analyseur d'actes de dialogue.
4. L'analyseur d'actes envoie en retour l'**acte de dialogue** correspondant à l'hypothèse reçue.
5. Le contrôleur de dialogue choisit la **réponse** la plus appropriée et l'envoie au client.
6. Le client envoie la **réponse** reçue au serveur local de synthèse

Sur l'architecture plus détaillée que nous pouvons voir en **Figure 20**, les modules sont séparés en fonction de leur utilisation :

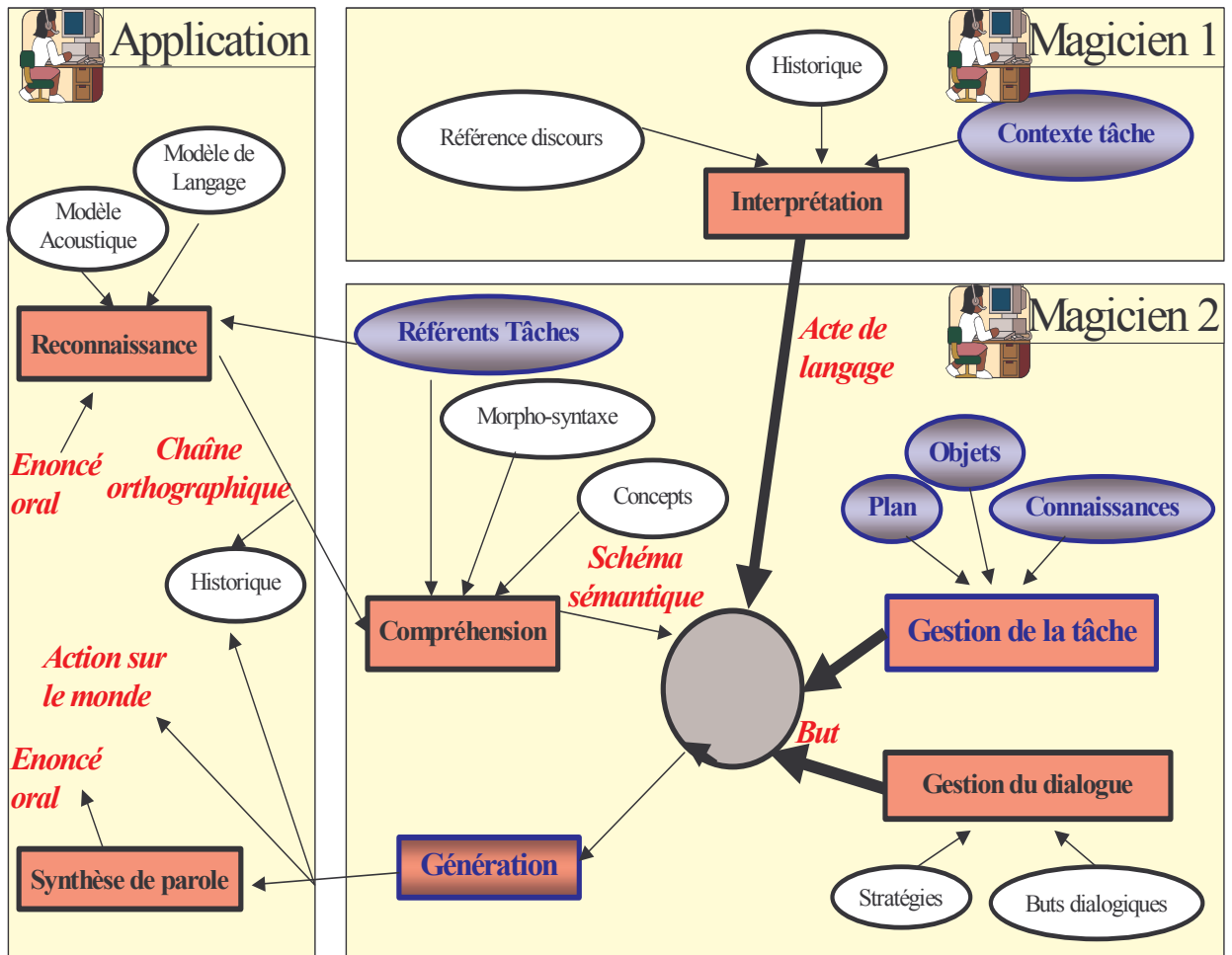


Figure 20 : Architecture détaillée, système de dialogue à base de magiciens d'Oz

Les modules de reconnaissance et de synthèse sont utilisés du côté de l'application. Le module d'interprétation est réalisé par le premier magicien d'Oz dont le rôle est d'identifier l'acte de dialogue contenu dans l'énoncé. Dans l'idéal, il devrait identifier les actes présents dans l'énoncé mais le temps de traitement étant rédhibitoire, nous avons choisi de n'identifier que l'acte indirect qui sera important pour la suite. Les autres modules sont intégrés système magicien d'Oz principal. Les items qui dépendent de la tâche ont été isolés. Ce sont ceux en dégradé sur notre schéma. Ainsi chaque module a un certain nombre de paramètres, dont certains dépendant de la tâche. Nous allons donc développer maintenant le côté application avec les modules de reconnaissance et de synthèse ainsi que les deux magiciens d'Oz qui traitent l'interprétation et le contrôle du dialogue.

### 3. Reconnaissance de parole, un modèle de langage spécifique.

La reconnaissance automatique de la parole revient à effectuer de manière automatique la « *transcription d'un signal vocal (en entrée) en suite de mots (en sortie)* » [Gauvain 00]. Les premières méthodes suivaient les techniques de reconnaissance de formes en se fondant sur les connaissances. Très vite ont suivi les approches fondées sur la reconnaissance de mots isolés puis sur les méthodes stochastiques à base de modèles de Markov cachés. C'est dans cette continuité que la reconnaissance de parole continue apparaît, avec dans un premier temps des systèmes mono-locuteurs nécessitant un apprentissage préalable à l'utilisation puis des systèmes multilocuteur ne nécessitant pas d'apprentissage préalable. Dans ces nouvelles approches, le système a deux composantes principales, un modèle acoustique appris sur un grand nombre de locuteurs et permettant le décodage acoustico-phonétique de l'énoncé oral, et un modèle de langage appris sur un grand nombre de textes permettant de fixer des probabilités de succession de mots et ainsi de définir la limite des mots et leur reconnaissance.

Le système de reconnaissance vocale en parole continue française utilise la **boîte à outils** Janus-III [Woszczyzna & al. 93] de CMU. Le modèle acoustique dépendant du contexte issu de [Besacier & al. 01] a été appris sur un corpus, extrait de Bref80 [Lamel & al. 91], de 12 heures de parole continue prononcée par 72 locuteurs. Le module de reconnaissance vocale a été adapté aux tâches choisies en créant un modèle de langage spécifique à ces tâches.

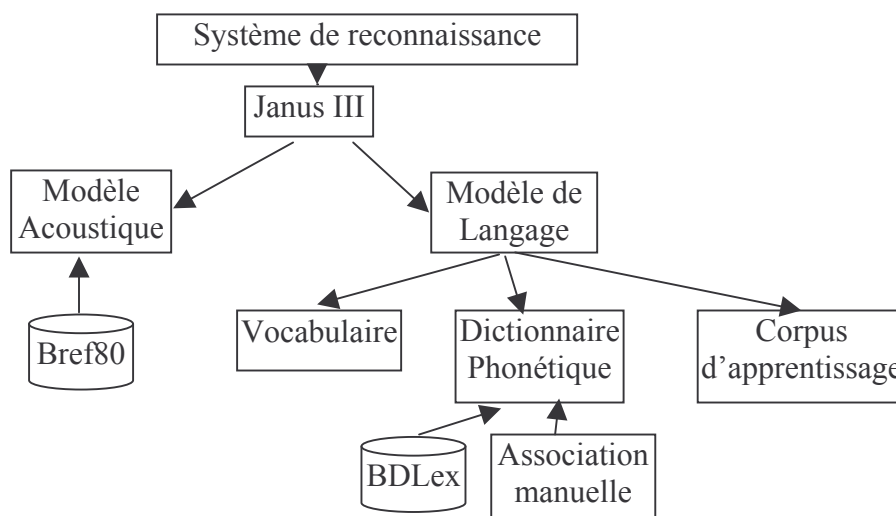


Figure 21 : Architecture du système de reconnaissance vocale

### **3.1. Corpus & vocabulaire pour l'apprentissage d'un modèle de langage.**

Utiliser la reconnaissance et la synthèse de parole permet de simuler le système tel qu'il pourra être, c'est à dire dépendant des résultats de reconnaissance. Aucun système général n'étant satisfaisant dans le cadre du dialogue, nous avons élaboré un modèle de langage plus spécifique à notre application. Nous avons suivi la méthodologie suivante :

- ✓ collecte d'un corpus de 801 dialogues homme-homme dans le cadre de l'application ciblée
- ✓ transcription de 44 dialogues de ce corpus
- ✓ analyse de ce corpus en terme de vocabulaire
- ✓ par ce vocabulaire, établissement d'un dictionnaire phonétique spécifique à l'application
- ✓ apprentissage d'un modèle de langage avec ce vocabulaire et ce corpus

Puis :

- ✓ collecte d'un corpus de 14 dialogues homme-machine dans le cadre de l'application ciblée (prétests ; première version du système à base de magicien d'Oz)
- ✓ transcription de ce corpus
- ✓ analyse de ce corpus en terme de vocabulaire
- ✓ par ce vocabulaire, enrichissement du dictionnaire phonétique spécifique à l'application
- ✓ concaténation du vocabulaire et du corpus avec ceux de la première phase
- ✓ apprentissage d'un modèle de langage avec le vocabulaire (1119 mots) et le corpus (1110 énoncés) résultant

Pour l'élaboration et l'enrichissement du dictionnaire phonétique, une partie du vocabulaire est présente dans le dictionnaire BDLex associant les mots et leur forme phonétique. Cette partie a été extraite de BDLex. Le vocabulaire absent de BDLex (environ 500 formes phonétiques différentes) a été écrit manuellement (ex : labo, login, ok, okay, excusé, e-mail, embêté, ennuyé, visioconférences, verras, trompée, transmette, réceptionniste, aïe, etc.). Nous avons placé en annexe la liste des mots rajoutés et leur forme phonétique associée.

### **3.2. Apprentissage d'un modèle de langage.**

Comme tous les systèmes de reconnaissance de parole récents, Janus utilise des modèles de langage. El-Bèze [El-Bèze 93] définit un modèle de langage comme la composante du système de reconnaissance qui se charge d'introduire les contraintes imposées par la syntaxe de la langue. Les modèles de langage, actuellement dans les systèmes de reconnaissance à grand vocabulaire les plus performants, sont fondés sur une approche probabiliste, compatible en cela avec les autres composantes du système de reconnaissance. Tandis que le module de reconnaissance évalue les probabilités des phrases par rapport aux observations acoustiques, le modèle de langage détermine à partir de corpus textuels les probabilités d'apparition des mots ainsi que celles de suites de mots.

Ces modèles de langage probabilistes reposent le plus souvent sur un paradigme empirique : une bonne estimation de la probabilité d'un événement linguistique peut-être obtenue en observant cet événement sur un corpus de texte de taille suffisante. Les nécessités induites par le processus de reconnaissance, conjuguées avec les capacités actuelles des analyseurs morpho-syntaxiques reposant sur une analyse structurelle de la phrase (capacités encore réduites sur des corpus généraux de grande taille), font que les modèles de langage ne prennent en compte que les contraintes locales de la syntaxe, par des modèles dits *n*-grammes ([Rosenfeld 94] ; *n* étant généralement inférieur ou égal à 5), où l'on estime la probabilité d'une phrase à partir des probabilités conditionnelles d'apparition d'un mot ou d'une classe de mots, étant donnés les *n*-1 mots ou classes de mots précédents. Cette approche est particulièrement intéressante pour son efficacité et sa robustesse, mais est limitée à la modélisation des structures linguistiques locales.

La reconnaissance se fonde alors sur une comparaison entre une forme apprise et une forme présentée. Si l'on ne tient pas compte du contexte, les possibilités « explosent » et la reconnaissance est difficile. Utiliser le contexte permet de lever de nombreuses ambiguïtés. Cependant, certaines applications nécessitent une reconnaissance en temps réel. L'information future n'est donc pas présente. Nous ne pouvons utiliser que sur l'historique, les informations passées et présentes.

Actuellement, des modèles bigrammes et trigrammes de mots sont utilisés couramment dans les systèmes de reconnaissance de la parole continue. Certains leur préfèrent des modèles pentagrammes, plus performants mais plus lourds à stocker. En fait, le choix se fera en fonction de l'application. Les modèles bigrammes et trigrammes de mots, estimés sur de très grands textes, fournissent, selon El-Bèze,

une bonne contrainte syntaxique locale, au prix de modèles de très grande taille. Pour des langues comportant des flexions comme le français ou l'allemand, l'utilisation de modèles de n-grammes de mots nécessitent des tailles de corpus et de modèles plus importantes que pour l'anglais. Différents modèles de n-grammes, de classes et de mots ont été développés, permettant de réduire le nombre de paramètres à estimer et la taille du corpus d'apprentissage et la taille du modèle.

Pour regrouper les mots dans des classes, nous pouvons utiliser des méthodes manuelles, plus lisibles, mais nécessitant un expert. Elles sont utiles pour regrouper les nombres, les dates ou les jours, et ainsi alléger le corpus d'apprentissage. *A contrario*, nous pouvons employer des méthodes statistiques (par minimisation de perplexité ou par mesure de similarité en analyse sémantique latente), mais ces méthodes nécessitent un corpus de taille importante. Cela étant plutôt rarement le cas en dialogue oral, une approche par expert adaptée à des petites tâches a été adoptée.

Une approche mixte a été proposée dans [Fouquet 99] combinant analyse linguistique et analyse statistique. Cette approche propose l'extension de petit corpus par des règles permettant de générer des énoncés non prononcés mais tout à fait possibles. Cette approche nécessite alors un expert afin d'élaborer les règles de possibilités qui permettent d'étendre « *Je veux* » et « *Je voudrais réserver* » à « *Je veux* », « *Je voudrais réserver* » mais aussi « ***Je veux réserver*** » et « ***Je voudrais*** » qui sont des formes tout à fait acceptables en dialogue oral. Deux apprentissages stochastiques sont ensuite réalisés sur le premier corpus et sur son extension afin de donner un poids plus important aux énoncés effectivement trouvés. Le modèle mixte propose une simple composition des deux apprentissages afin de rehausser les énoncés possibles mais garder le même écart avec les énoncés apparus en corpus.

Le vocabulaire issu des transcriptions a servi de base à l'apprentissage de différents modèles de langage afin de définir le meilleur modèle pour notre application. Nous avons combiné différents paramètres pour nos tests : le **vocabulaire**, le **corpus d'apprentissage**, la **boîte à outils**, la prise en compte du **mot inconnu** et l'utilisations de modèles à **classes**.

Les premiers tests combinent le corpus d'apprentissage du journal **Le Monde (1997-2001)** avec le vocabulaire de notre application (1119 mots) enrichi des X mots les plus fréquents du Monde où X varie de 0 à 25000. Considérant l'amélioration du web comme corpus [Vaufreydaz 96], nous avons testé également des modèles appris sur un corpus collecté sur **Internet**, WebFr4 [Vaufreydaz 02]. Nous avons enfin testé les modèles appris sur le **corpus spécifique** (les transcriptions des dialogues).

Ensuite, nous avons testé deux boîtes à outils: **Janus III** de CMU (Clarkson & Rosenfeld 97) (<[http://www.speech.cs.cmu.edu/SLM\\_info.html](http://www.speech.cs.cmu.edu/SLM_info.html)>)et **SRI-LM** Language Modeling Toolkit (<http://www.speech.sri.com/projects/srilm/>).

Enfin nous avons étudié dans notre contexte la prise en compte du mot inconnu et l'utilisation de modèles à classes.

### **3.3. Evaluation des modèles de langage.**

Nous avons comparé 8 modèles afin de conserver le plus adapté à un système de dialogue : notre corpus spécifique et la boîte à outils SRI-LM. Pour évaluer les performances du système de reconnaissance vocale, nous avons essayé avec celui-ci de reconnaître 63 phrases prononcées pendant un premier beta-test du système.

Ces phrases sont issues de situations de dialogues réelles dans lesquelles les utilisateurs avaient à parler, avec des requêtes spécifiques définies par avance, au système simulé par magicien d'Oz. Les meilleures performances ont été obtenues avec notre corpus spécifique et la boîte à outils SRI-LM. Les autres solutions sont trop mauvaises pour les considérer acceptables. Le Monde est trop loin des spécificités du dialogue. Le web ignore les spécificités des phrases courtes trouvées dans nos dialogues oraux. Le taux d'erreur de mot obtenu dans nos meilleures conditions est de 39% ce qui semble acceptable pour les magiciens qui n'ont demandé que trois fois à l'utilisateur de répéter dans les six dialogues ayant servi au test. Il ressort de cette analyse différentes conclusions :

- Un **vocabulaire** de 40 000 mots est inutile en dialogue oral car cela fait trop de mots inutilisés. Un vocabulaire de 10 000 mots semble plus adapté à un dialogue oral en général, mais c'est avec le vocabulaire de l'application uniquement que les résultats ont été les meilleurs. Cette configuration peut omettre des mots fréquemment dits à l'oral ce qui nécessite donc de bien élaborer le corpus.

- Concernant le **corpus d'apprentissage**, Internet et Le Monde donnent les même résultats. Le taux d'erreur de reconnaissance de mots est supérieur à 55%. Le corpus de l'application offre un taux d'erreur de reconnaissance de mot bien plus intéressant car inférieur à 39%.

- La **boîte à outils** SRI-LM offre de meilleurs résultats que Janus III.

- Etant donné que notre corpus d'apprentissage est plutôt petit (1110 phrases), l'introduction du **mot inconnu** permet de ne pas pénaliser les mots autour d'un mot non reconnu.



- Notre corpus d'apprentissage étant relativement petit, certains mots, totalement interchangeables, comme les prénoms, les noms ou les lieux peuvent être absents du corpus alors qu'ils pourraient être prononcés. Nous pouvons retrouver dans notre corpus « je suis Dominique Blanc » et « mon nom est Jean Dupond » sans avoir d'occurrence de « je suis Jean Dupond », pourtant plausible. Nous avons donc utilisé des modèles à base de **classes**. Ces classes ont été obtenues manuellement à partir des transcriptions. Nous avons retenu les nom, prénom, genre, date, jour, mois, moment, lieu, pays et fermeture (au\_revoir, adios, à\_demain, à\_tout\_de\_suite, etc.). Nous avons opté pour des classes génériques, c'est-à-dire indépendantes du système (et de la tâche). Le modèle créé devrait ainsi rester valable pour différents systèmes de dialogue. L'adjonction de la notion de classe a montré une nette amélioration en termes de reconnaissance pour les raisons énoncées ci-dessus.

### 3.4. Au niveau de l'ergonomie de l'interface

Au niveau de l'interface, la prise en compte de reconnaissance vocale nécessite quelques choix ergonomiques. Nous avons testé trois solutions :

- **Main libre** à base de seuil d'énergie de la voix. L'enregistrement démarre au moment où l'on parle, se termine peu après qu'on a fini de parler. Pour la gestion de prise de parole commençant par une plosive, la gestion d'une mémoire tampon est nécessaire, mais force le système à enregistrer en permanence le signal émis. Ainsi, il est possible de récupérer le signal, non pas à partir du moment précis où le locuteur commence à s'exprimer, mais un peu en amont, afin de ne pas risquer de manquer le début de la phrase. Cette solution est idéale puisqu'elle libère l'utilisateur d'une contrainte forte. En revanche, elle nécessite un calibrage important afin d'éviter que trop de faux départs ou qu'une non-détection ne viennent perturber le dialogue.

- « **Push-to-talk** » faisant intervenir l'utilisateur. Cette méthode se rapproche du talkie-walkie pour lequel l'utilisateur appuie sur un bouton pour parler. Ici, il clique sur un bouton pour parler, le maintient enfoncé durant son énoncé et le relâche quand il a terminé. Moins bonne que la précédente, cette méthode nécessite un effort cognitif de la part de l'utilisateur, et de nombreux oublis ont pu être observés.

- « **Push-to-talk, Push-to-stop** ». Pour cette méthode, l'utilisateur clique pour parler et clique à nouveau pour arrêter l'enregistrement et démarrer la reconnaissance. Cette méthode nécessite un effort cognitif important. Cet effort est trop important selon nous pour des sujets.

Nous avons opté pour le « push-to-talk » en raison du trop grand nombre de faux départs constatés dans notre méthode « main libre ».

## 4. Synthèse de parole française

La dernière partie de notre système de dialogue oral consiste en une synthèse de parole (en anglais Text-To-Speech d'où l'acronyme TTS) qui produit un signal audio à partir du texte.

La synthèse a été développée en tant que serveur qui reçoit le texte de l'utilisateur et répond en créant le fichier son correspondant et en le jouant. Ce serveur a été intégré au serveur de reconnaissance vocale parce qu'ils utilisaient tous les deux le son et que la reconnaissance vocale doit être stoppée pendant le processus de synthèse.

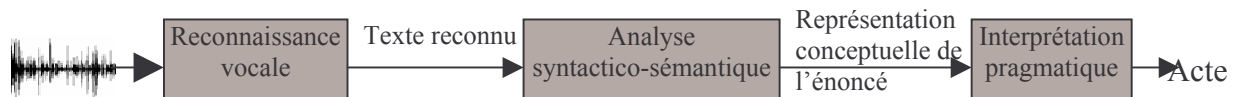
Afin de minimiser le temps d'attente de l'utilisateur, la synthèse doit se faire le plus vite possible. Le système Mbrola TTS [Baghein & al. 00] de la faculté Polytechnique de Mons (<<http://tcts.fpms.ac.be/synthesis/euler>>) nous semble suffisamment rapide. Il nécessite environ 30% du temps de la phrase pour produire le fichier son correspondant. La qualité de synthèse vocale nous a semblé juste acceptable pour notre application.

Les utilisateurs ont critiqué la qualité de la synthèse vocale. Ils dénonçaient la prosodie de cette voix synthétique, monocorde et trop « mécanique » à leur goût. Cependant, à la suite de [Rouillard 00], nous pensons que « *le fait que la machine parle avec une voix mécanique n'est pas nécessairement un défaut du système. En effet, si l'on observe le DHM d'un point de vue cognitif, cette voix de robot rappelle continuellement à l'utilisateur que c'est à une machine qu'il a affaire. Le fait que l'ordinateur parle de cette manière permet à l'utilisateur de bien prendre conscience du statut et du rôle de son interlocuteur.* » [Rouillard 00]

Nous pensons aussi qu'une voix synthétique qui se rapproche de la voix humaine peut inciter l'utilisateur à attribuer à la machine un niveau de compréhension dialogique en rapport avec son élocution. La personne qui entend une machine parler presque comme un humain, risque d'attendre d'elle des capacités de compréhension proches de celles d'un humain. On tolérera donc moins d'erreurs de la part d'une machine qui parle convenablement, puisque l'on s'imagine que sa façon de s'exprimer est étroitement liée à ses (hautes) capacités calculatoires.

## 5. Magicien d'Oz et collecte (aide à l'annotation)

Afin d'être le plus générique possible, c'est à dire le plus indépendant possible par rapport à la tâche, le modèle de dialogue est fondé sur la théorie des actes de langage introduite par Austin [Austin 62] & Searle [Searle 72]. Un acte est vu comme la manière de communiquer le contenu propositionnel. Utiliser les actes de dialogue contenus dans l'énoncé est une source de connaissance importante pour l'interprétation de cet énoncé. Nous considérons qu'un dialogue peut être décrit par un ensemble limité mais ouvert d'actes de dialogue. Ainsi, considérant que la compréhension du texte en actes est correctement faite, travailler à un niveau pragmatique nous permet d'aborder des combinaisons d'actes et des stratégies de dialogue afin de trouver la réponse la plus adéquate.



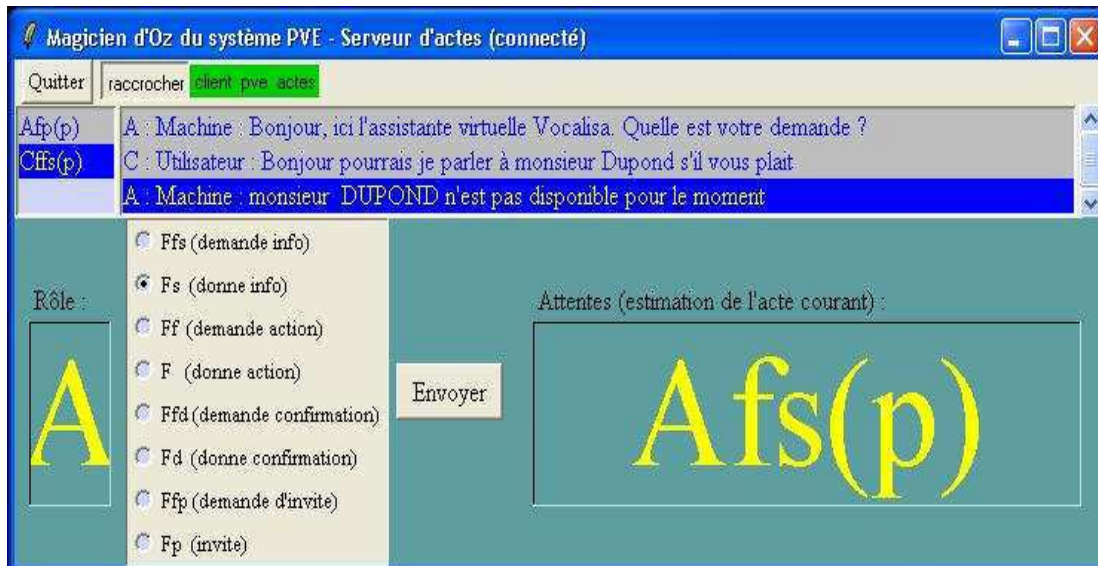
**Figure 22** : Du signal de parole aux actes (et attentes)

Comme nous pouvons le voir en **Figure 22**, la compréhension de parole revient à essayer de saisir le sens global de l'énoncé. Selon [Pierrel 87], elle est opposable à la reconnaissance de parole en ce sens que cette dernière cherche à reconstituer le message élément par élément. C'est la partie la plus difficile à l'heure actuelle. Mais c'est également la partie la plus importante si l'on désire obtenir des dialogues plus naturels et aisés avec la machine. Pour nous, cela revient à analyser l'énoncé sous un angle pragmatique qui nous permet de représenter le sens global de l'énoncé.

L'analyse en actes revient à associer l'énoncé ou sa représentation conceptuelle à un acte de dialogue. C'est un point crucial du projet car de la correcte identification de ces actes dépend tout le reste du modèle. Nous avons trouvé quelques travaux sur l'automatisation de cette identification, comme la thèse de Colineau [Colineau 97] qui utilise des réseaux de neurones. Nous pouvons proposer également une approche fondée sur l'analyse sémantique latente (LSA) à partir des actes identifiés et d'une analyse morpho-conceptuelle des énoncés associés. En revanche, il n'existe pas à notre connaissance d'applications génériques.

C'est pourquoi une plate-forme de type magicien d'Oz a été développée pour annoter l'hypothèse de reconnaissance vocale en actes de dialogue. Le magicien reçoit la phrase reconnue par le système. Il peut garder en mémoire des marqueurs du

discours de Colineau comme « je veux », « je voudrais », etc. Il choisit, dans une liste d'actes, l'acte le plus approprié. Pour faire ce choix, il est aussi aidé par une présentation des attentes, introduites dans [Fouquet 01], déduites de l'acte précédent [Fouquet 02]. La **première attente**, c'est-à-dire celle qui maximise la probabilité de prédiction, est **présélectionnée**. Lorsque le magicien envoie l'acte correspondant à la phrase, les attentes sont **mises à jour** de sorte que les probabilités sont apprises dans le contexte de l'application.



**Figure 23** : Interface du Magicien d'Oz pour l'annotation en acte de dialogue

Dans l'exemple (**Figure 23**), le compère magicien reçoit une première phrase de demande d'information : « Bonjour, pourrais-je parler à monsieur Dupond s'il vous plaît ? ». Il l'annote  $F_{C}^{fs}(p)$  qui représente la demande d'information du client. L'attente la plus probable pour cet acte  $F_{C}^{fs}(p)$  est une réponse de l'ordinateur  $F_{A}^{s}(p)$  avec une probabilité de 77%. Lorsque le magicien recevra la phrase suivante (« Monsieur Dupond n'est pas disponible pour le moment ») du système,  $F_{A}^{s}(p)$  sera présélectionné de sorte qu'il lui suffit de confirmer si cet acte est le bon. Dans ce cas, la probabilité précédente est incrémentée en prenant en compte cette nouvelle occurrence. Dans le cas contraire (par exemple l'attente suivant,  $F_{A}^{fs}(p)$ , correspondant ici à une demande de précision avec la phrase « avez-vous essayé son poste direct ? », le magicien sélectionne l'acte  $F_{A}^{fs}(p)$  et le confirme pour l'envoyer au système et les probabilités seront mises à jour avec cette nouvelle occurrence.

Cette aide à l'annotation par la prédiction de l'acte suivant est un sous-produit intéressant de notre théorie des attentes. Le prochain chapitre sera consacré à l'évaluation de cette prédiction d'acte de dialogue en utilisant la première attente.

## 6. Magicien d'Oz et contrôleur de dialogue

Pour l'élaboration de notre magicien simulant le contrôleur de dialogue, nous avons suivi la méthodologie de conception en spirale. Ainsi, nous avons commencé un premier système frustré que nous avons testé et amélioré plusieurs fois afin d'obtenir une version stable permettant d'utiliser le système en conditions réelles.

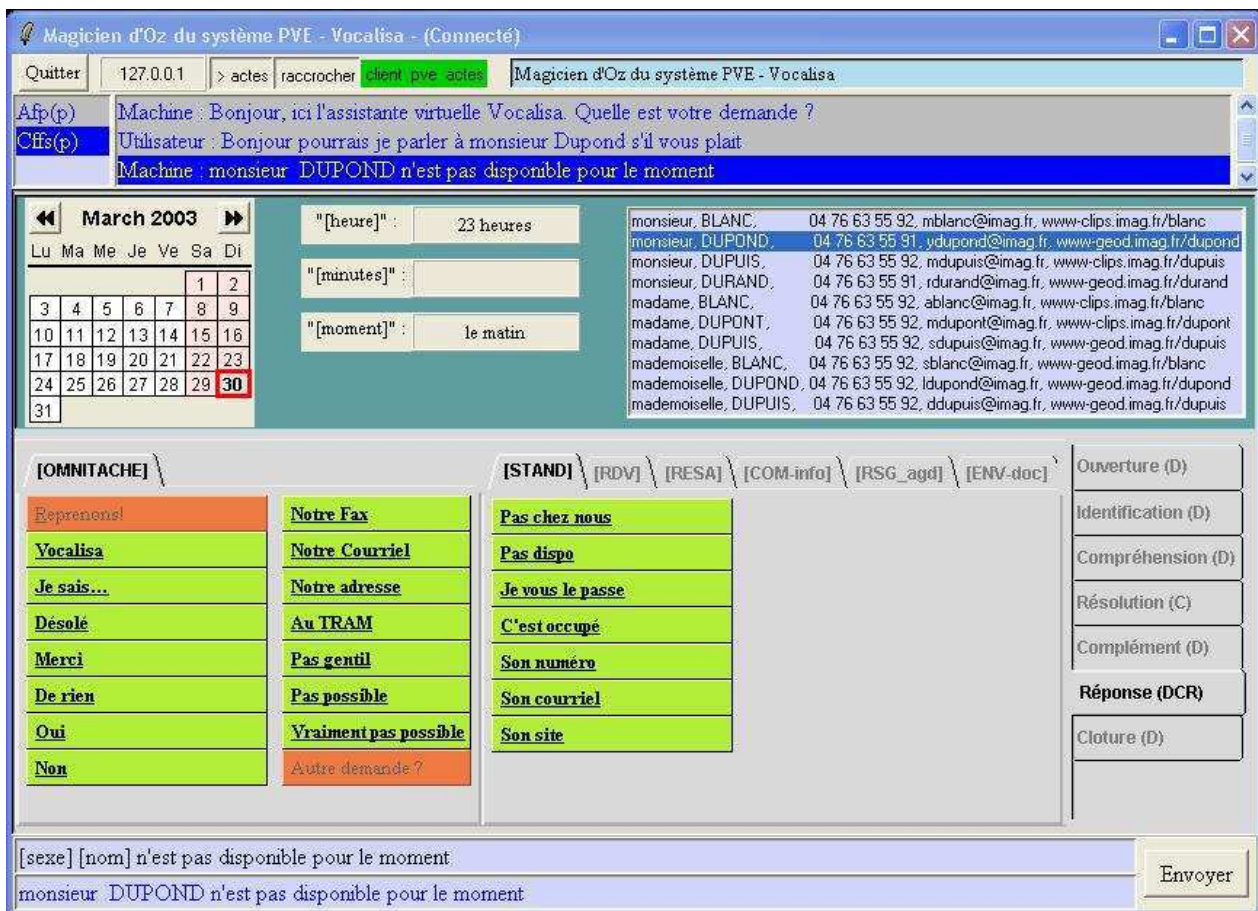
### 6.1. Plate-forme mettant en jeu un magicien d'Oz

Notre point de départ réside dans l'analyse du corpus collecté en conditions réelles. Cette analyse nous a permis de déterminer 122 phrases que le système devra être en mesure de prononcer et des paramètres liés à la tâche comme les coordonnées du personnel de l'entreprise, les indications relatives à la date (jour, mois, année, jour de la semaine). Suite à cette analyse et selon le développement proposé, nous avons élaboré plusieurs versions de magiciens. Nous pouvons voir l'évolution de la plate-forme dans le tableau suivant :

V.	Ajout	Particularités	Commentaires sur la plate-forme
1	<b>Indépendante</b> de la tâche	Fichier de configuration externe pour les énoncés et les paramètres	Indépendante donc adaptable à d'autres tâches mais peu utilisable en temps réel car demande de gros efforts d'apprentissage au magicien. Impossible à utiliser dès lors que le nombre de phrases prononçables est important
2	Dépendante de la tâche	<b>Onglets pour séparer les tâches</b> <b>Paramètres</b> présentés de façon plus <b>ergonomique</b> (e.g. : calendrier avec un clic pour la date au lieu de joursemaine, jourmois, mois, année)	Pas assez efficace
3	Stratégies employées	<b>Onglets pour séparer les stratégies</b>	Encore trop d'étiquettes affichées
4	Structure générique	<b>Onglets pour séparer les phases</b>	Nombre de manipulations important pour obtenir une étiquette
5	Stratégie	<del>Onglets</del> → <b>couleurs</b>	Encore beaucoup d'étiquettes
6	Attentes	Etiquettes correspondants aux <b>attentes</b> de l'utilisateur soulignées, les autres grisées	Moins d'étiquettes à regarder, efficacité accrue par la mise en relief des énoncés répondant aux attentes
7	Reconnaissance de parole	<i>Ad hoc</i>	Non parfait mais le mieux
8	Synthèse de parole	Intégré au serveur de reconnaissance pour éviter de reconnaître la synthèse	Voix synthétique acceptable mais loin d'être parfaite.

**Tableau 22** : Evolution de la plate-forme magicien d'Oz

Selon l’hypothèse que les dialogues homme-machine sont différents de ceux entre humains particulièrement en dialogue oral où les conventions sociales sont omniprésentes, nous avons développé un magicien d’Oz afin de simuler complètement le modèle de dialogue. Le compère reçoit l’hypothèse de reconnaissance vocale de l’utilisateur (afin de tenir compte des erreurs de reconnaissance). Il reçoit également l’acte de dialogue, correspondant à cette hypothèse, du premier magicien (**Figure 23**) et l’attente la plus probable associée. Il a alors le rôle de contrôleur de dialogue, contrôleur de tâche et générateur de réponse en choisissant les paramètres adéquats dans l’interface de la tâche et une réponse appropriée dans la liste des 122 phrases obtenues à partir des transcriptions manuelles des dialogues réels (**Figure 24**).



**Figure 24** : Interface du Magicien d’Oz pour le contrôleur de dialogue

Lorsqu’il a choisi la meilleure réponse et ses paramètres associés, il l’envoie à la synthèse du client. Lorsque l’utilisateur dit « Bonjour, pourrais-je parler à monsieur Dupond s’il vous plaît ? », le magicien peut sélectionner monsieur Dupond dans l’annuaire et répondre « Monsieur Dupond n’est pas disponible pour le

moment » ou « Je vous passe monsieur Dupond » en sélectionnant l'étiquette voulue puis la phrase correspondante dans le menu lié à cette étiquette.

Le système doit également gérer le retour visuel de chaque opération, faciliter la conscience collective en affichant les machines interconnectables, et gérer les aspects réseau (connexion, déconnexion et protocole de transfert d'information).

Une phase de configuration est donc nécessaire en début d'expérimentation. Celle-ci ne nécessite pas la présence du magicien ; l'expérimentateur suffit. La configuration effectuée, le système offre différents composants que nous allons décrire maintenant avec les choix ergonomiques que nous avons faits.

## **6.2. Composants de la plate-forme et choix ergonomiques associés**

### **6.2.1. Paramètres**

Le magicien simule une secrétaire virtuelle capable de résoudre différentes tâches. Il est d'abord nécessaire de lui présenter les différents paramètres des tâches qui lui incombent. Ainsi, un calendrier lui est proposé plutôt qu'une liste de jours. En effet, le calendrier permet au magicien de s'affranchir de la cohérence d'une date proposée. En un clic, il sélectionne quatre paramètres (le jour de la semaine, le jour du mois, le mois et l'année).

De même, un annuaire du personnel de l'entreprise lui est présenté afin de sélectionner en un clic cinq paramètres (le sexe, le nom, le numéro de téléphone, l'adresse mail et l'adresse Internet).

Enfin, restent quelques paramètres, dont l'heure, qui n'ont pu être réglés autrement qu'individuellement, étant particuliers dans leur énonciation (midi, et quart, moins le quart, etc.). En outre, nous ne voulons pas proposer toutes les heures ni toutes les minutes. Un rendez-vous à 22h43 ne nous semble pas pertinent dans le cadre de l'entreprise.

Pour une efficacité accrue, nous proposons au magicien des valeurs par défaut aux différents paramètres. Ainsi, le jour en cours sera présélectionné ainsi que l'heure suivante et un membre du personnel.

La partie paramètres des tâches étant présentée, le magicien doit désormais pouvoir répondre à son interlocuteur. Pour cela nous lui proposons des phrases avec paramètres. Etant donnée la complexité liée au nombre de réponses possibles (122

phrases à paramètres et variantes), il est nécessaire de les organiser pour faciliter le travail du magicien.

#### **6.2.2. Système pluritâche**

Nous avons dans un premier temps choisi de classer les phrases par tâche avec une catégorie omnitâche correspondant à toutes les phrases valables pour toutes les tâches. Ainsi, en début de dialogue, le magicien sélectionne l'onglet correspondant à la tâche en cours et ne voit que les phrases de cette tâche (plus celles omnitâches), réduisant ainsi plus ou moins par six (car six tâches ont été sélectionnées pour nos tests) le nombre de phrases présentées.

#### **6.2.3. Énoncés synthétiques**

Dans un deuxième temps, nous avons réduit l'effort cognitif du magicien en lui présentant uniquement des étiquettes simplifiées des phrases avec dans le menu associé les phrases correspondantes. Pour un rendez-vous, une étiquette « résumé » masque les quatre phrases associées dont « nous disons donc [joursemaine] [jourmois] à [heure] [minutes] salle B2 avec [sexe] [nom] ». Ainsi, plutôt que de relire toutes les phrases, le magicien choisit d'abord l'étiquette correspondant à ce qu'il souhaite dire, puis sélectionne la phrase plus précise qu'il souhaite proposer.

#### **6.2.4. Architecture des dialogues**

Dans un souci de respecter le déroulement logique du dialogue, l'analyse du corpus homme homme a permis d'élaborer un diagramme de l'architecture de chaque type dialogue ([Descalle 02]). Nous avons suivi ce diagramme en classant chacun de ces 122 énoncés en fonction des tâches et des phases dans lesquelles ils sont susceptibles d'intervenir, mettant chaque phase soulignée en onglets.

Le magicien choisit donc la tâche en début de dialogue mais peut en changer en cours de dialogue. Il suit ensuite le diagramme en sélectionnant la phase (par exemple introduction, compréhension du but ou réponse) et choisit dans les énoncés de la phase sélectionnée.

#### **6.2.5. Les attentes comme aide au contrôle**

Le magicien utilise la première attente déduite de l'acte reçu. Celle-ci est mise en relief (l'étiquette correspondante est soulignée) afin de faciliter le choix. S'il veut répondre à la première attente du locuteur, il lui suffira de choisir sa réponse parmi celles soulignées. Le magicien peut également choisir une autre réponse. Ainsi, les attentes sont d'abord une aide à l'humain avant d'être une aide au système.



#### 6.2.6. Stratégies

Le système de contrôle du dialogue par magicien d'Oz utilise les trois stratégies les plus représentées parmi celles décrites dans [Caelen 92].

La stratégie **directive**, guidée par les buts de la machine, intervient dans les phases d'ouverture, d'introduction et de clôture, c'est à dire lors de l'initiative de la machine. Tant qu'il cherche à connaître le but du client, l'agent est directif en posant des questions notamment des demandes de précision.

La stratégie **réactive**, dirigée par les données (les énoncés du client), fait l'hypothèse que le client connaît précisément son but et les moyens d'y parvenir. L'agent répond donc simplement au client sans poser de questions.

La stratégie **coopérative**, guidée par les buts de l'utilisateur, suppose que le client ne maîtrise pas complètement le moyen d'accéder à son but, soit au niveau de la tâche, soit au niveau de l'interface elle-même. La machine tente d'inférer les buts de l'utilisateur.

La dernière aide consiste donc en la coloration de coopération. En plus des réponses, la stratégie correspondante (directive, réactive ou coopérative) est montrée avec un code de couleur de sorte que le magicien puisse choisir sa stratégie lorsqu'il répond. Il peut donc être directif puis réactif ou coopératif en choisissant la phrase par sa couleur lorsqu'il a le choix de différentes stratégies. Pour l'exemple précédent, choisir « Monsieur Dupond n'est pas disponible pour le moment » ou « Je vous passe monsieur Dupond » correspond à une stratégie réactive tandis que « Avez-vous essayé son poste direct ? » semble être davantage directif.

#### 6.6.7. Enoncé final

Pour finir, il était nécessaire de présenter la phrase à envoyer en réponse. Il nous a semblé important de la présenter sous la forme originale (avec les paramètres) mais aussi sous la forme finie prête à envoyer (avec les valeurs des paramètres). La première forme permet de savoir quels paramètres sont pertinents tandis que la deuxième permet de mieux se rendre compte de ce qui sera envoyé et de vérifier que ce sont bien ces valeurs que nous voulons pour les paramètres.

## **Conclusion sur le système**

Nous avons présenté dans ce chapitre les différents modules (dont deux simulés par magicien d'Oz) développés pour l'élaboration d'un corpus de dialogues oraux homme-machine et l'annotation de parole. Pour ces deux magiciens, nous avons montré que la théorie des attentes semble être une perspective intéressante pour améliorer la compréhension en actes de dialogues mais aussi le contrôle du dialogue en prédisant l'acte de dialogue le plus attendu et les autres moins attendus mais possibles.

Le magicien « annotateur d'actes » précédemment décrit a permis de simuler le typage de l'acte (et ainsi une partie de l'analyse pragmatique) de l'énoncé reconnu. En outre, proposant l'attente la plus probable comme acte présélectionné, l'annotateur a enrichi les statistiques d'enchaînement des actes et adapté les probabilités d'attentes aux tâches.

Cette analyse fournie au magicien « contrôleur de dialogue » permet de mettre en relief les réponses correspondant à l'attente la plus probable. Si le client pose un faire-faire-savoir, seront mises en valeur les phrases du magicien correspondant à un faire savoir.

Les modules présentés nous ont donc servi à enregistrer 12 dialogues pour chacune des 6 tâches retenues considérant les tâches les plus communes pour des secrétaires. L'analyse de ces 72 dialogues nous permettra de tester et de mettre à jour le système de dialogue. Elle servira ensuite à l'automatisation de l'interprétation pour une tâche spécifique et un contrôleur de dialogue générique utilisant tous deux les attentes en dialogue.

L'expérimentation en magicien d'Oz a permis le calcul des attentes pour modéliser le dialogue avec toutes ces données : opportuniste pour être dynamique et flexible. Nous allons à présent aborder cette modélisation stochastique proposée.

## **Chapitre II : Modélisation stochastique**

Nous nous intéressons dans cette partie au dialogue oral humain et homme-machine. Nous nous pencherons particulièrement sur l'apport de la modélisation stochastique des attentes des interlocuteurs dans un système de dialogue oral.

Nous présentons d'abord la partie stochastique de notre approche avec la gestion des rôles des interlocuteurs. Nous nous appuyons ensuite sur l'étude (publiée dans [Fouquet 00]) du corpus de renseignement touristique [Besacier et al. 01] pour proposer une première estimation de ce que seront pour nous les attentes en situation. Nous proposerons, par une analyse plus approfondie de ce même corpus, une validation de la première estimation et un affinage en terme de probabilités associées

à chaque attente. Enfin, nous comparerons cette nouvelle approche avec une approche, uniquement stochastique, fondée sur la prédiction d'actes de parole.

## **1. L'importance des rôles en dialogue oral finalisé**

Pour la modélisation des attentes, nous avons choisi d'utiliser les actes de langage et de leur adjoindre le rôle du locuteur. En effet, nous pensons que ce rôle a une grande importance dans l'évolution du dialogue. L'agent touristique et l'assistant virtuel ne dialogueront pas comme leur client. Pour montrer cette importance, nous avons analysé un corpus de pré-tests du système PVE. Dans ce système, un magicien d'Oz simule une secrétaire virtuelle. L'analyse en terme d'actes et attentes associées montre un déséquilibre important entre les interlocuteurs. En effet, l'acte du locuteur correspond très souvent (à 89 contre 8) à la première attente du système, à savoir que lorsque le système pose une question, le sujet y répond. A l'inverse, l'acte du système correspond davantage (à 19 contre 7) à la deuxième attente du locuteur à savoir que lorsque le locuteur pose une question, le système demande davantage de précisions. Le système adoptant le plus souvent une stratégie directive, il pose davantage de questions alors que le locuteur, plus réactif y répond. En fait, pour une demande d'information, le sujet donne l'information tandis que le système donne une réponse implicite d'acceptation en demandant des précisions.

Ex : C : Je voudrais joindre monsieur Durand. ( $F^{\text{fs}}_{\text{C}}(\text{p})$ )

A : C'est personnel ou professionnel ? ( $F^{\text{fs}}_{\text{A}}(\text{p})$ )

En dialogue finalisé et plutôt dirigé, la succession d'actes sera donc de la forme questions/réponses. Cela peut s'expliquer par le fait que le système connaît la marche à suivre pour résoudre le but du client. Il remplit petit à petit les blancs en posant des questions précises. Le client est davantage réactif.

## 2. Méthodologie pour la prédiction des actes vs attentes

Notre but, ici, sera la comparaison de deux approches possibles, considérant chacune le niveau pragmatique de l'énoncé en cours. Nous nous appuierons donc sur les actes en situation. Nous nous inspirerons, dans un premier temps, de l'analyse conversationnelle présentée par [Bilange 92]. Ainsi, nous prendrons en considération les ruptures et réparations ou les demandes de précision, soulevées notamment dans notre corpus. Nous proposerons alors une règle de la forme  $F^{\text{fs}}_A(p) \rightarrow F^{\text{s}}_B(p) \mid F^{\text{fs}}_B(q)$  ; lorsque l'acte  $F^{\text{fs}}_A(p)$  apparaît, les attentes du locuteur A sont  $F^{\text{s}}_B(p)$  ou  $F^{\text{fs}}_B(q)$ . Nous chercherons dès lors à comparer les deux approches. La première, plus simpliste, permet de prédire l'acte suivant à partir de l'acte en cours. La deuxième, plus fine, se fonde sur les attentes et leur gestion pour déterminer l'acte suivant.

Cette approche considère les huit actes de langage suivant (**Tableau 23**) :

F : action	F <sup>f</sup> : demande d'action
F <sup>s</sup> : information	F <sup>fs</sup> : demande d'information
F <sup>d</sup> : engagement	F <sup>fd</sup> : demande d'engagement
F <sup>p</sup> : possibilité, invite	F <sup>fp</sup> : demande de possibilité

**Tableau 23** : Liste des actes utilisés pour les tests

### 2.1. Prédiction des actes

Dans cette approche, nous adoptons une démarche stochastique pour prédire l'acte subséquent le plus probable à partir d'un historique plus ou moins grand. Nous utilisons les probabilités d'actes de dialogue n-grammes pour calculer l'acte suivant le plus probable, ce qui nous donne par la formule de probabilité conditionnelle de Bayes :

$$P(A_{n+1} \mid A_1, \dots, A_n) = P(A_1, \dots, A_n, A_{n+1}) / P(A_1, \dots, A_n)$$

Les probabilités seront alors estimées à partir de techniques de fréquence relative. Pour prédire l' $i^{\text{ème}}$  acte de dialogue  $A_i$ , les  $n-1$  actes de dialogues précédents permettent de déterminer l'acte de dialogue le plus probable par la formule :

$$A_i = \operatorname{argmax}_a P(A \mid A_{i-1}, A_{i-2}, \dots, A_{i-n+1})$$

Notre corpus est composé de dialogues réels et comme dans de nombreuses situations réelles, il n'est pas possible de collecter un très grand nombre de données pour estimer correctement les statistiques. Il n'est donc pas raisonnable d'utiliser les techniques classiques. Nous devons trouver une solution pour deux problèmes importants. D'une part, le modèle de dialogue fondé sur les séquences de n-grammes actes de dialogues ne peut pas être utilisé dans le cas d'une entrée non attendue. D'autre part, les modèles à base de n-grammes peuvent prédire plusieurs actes avec la même probabilité dans le cas d'un manque de données d'apprentissage. Le traitement de ces deux problèmes consiste, dans notre expérimentation, à utiliser le modèle (n-1)-gramme et récursivement.

Dans un premier temps, nous avons appliqué cette méthode sur le corpus PVE obtenu par magicien d'Oz en considérant un acte par tour de parole. Nous avons testé plusieurs tailles d'historiques et deux cas, inclusion de la tâche d'ouverture dans les statistiques ou non. L'historique 1 correspond aux unigrammes, c'est-à-dire à la répartition statistique dans le corpus indépendamment de tout acte précédent. L'historique 2 correspond aux bigrammes et s'appuie donc sur l'acte précédent pour la prédiction de l'acte courant. De même, l'historique 6 correspond à la prédiction d'un acte à partir des 5 précédents.

Par ailleurs, nous avons testé deux cas. Le premier correspond à un taux de prédiction global indépendamment de la tâche. Le deuxième cas supprime des statistiques la première tâche d'ouverture car celle-ci est prédictible à 100% (l'agent prend toujours la parole de la même manière, avec le même acte). Les résultats de cette prédiction sont décrit ci-après (**Tableau 24**) :

	7 tâches		6 tâches	
	Agent	Client	Agent	Client
Nb actes	163	151	141	151
n=1	24,5%	33,1%	28,4%	33,1%
n=2	39,3%	60,3%	33,3%	62,9%
n=3	38,7%	60,3%	33,3%	61,6%
n=4	39,3%	59,6%	34,75%	60,3%
n=5	37,4%	53,6%	31,9%	56,3%
n=6	36,8%	53,0%	31,2%	55,6%

**Tableau 24** : Taux de prédiction d'acte par la méthode statistique dans PVE (un acte par tour)

Nous pouvons observer une meilleure prédiction de l'acte du client. Cela est dû au fait que le dialogue est plutôt dirigé par l'agent indépendamment de la tâche (l'agent sait quels paramètres il attend et pose directement les questions sur ces paramètres). Ces résultats sont un peu moins fiables que ceux de VerbMobil [Reithinger & Maier 95] qui obtient 44,2% pour  $n=1$ , 66,5% pour  $n=2$ , 81,5% pour  $n=3$ . Cependant, leur corpus est mono-tâche et contient un grand nombre de données (environ 7200 actes de dialogue). En outre, nos résultats corroborent ceux de VerbMobil à savoir qu'un grand historique est inutile, le taux le meilleur étant souvent trouvé pour  $n=2$  c'est à dire pour les bigrammes.

Dans un deuxième temps, nous avons réannoté le corpus pour découper chaque tour de parole en un certain nombre d'actes, et ainsi se rapprocher de la réalité. Les calculs sont identiques aux précédents mais reflètent davantage la structure du dialogue.

Nous pourrions, au final, adopter l'approche de génération d'acte par règle [Imberdis 95], par réseaux neuronaux ([Gouiaa et al. 99] à partir de l'analyse de [Colineau 94] du corpus de [Ozkan 94]) ou stochastique [Maynard, Lefèvre 02].

## 2.2. Génération et gestion des attentes

Dans les systèmes de dialogue oral, la compréhension des attentes du locuteur est nécessaire en sus de la sémantique et de la structure syntaxique de l'énoncé. La reconnaissance d'attentes n'est pas un problème aisé. Il existe de nombreuses manières de laisser transparaître une attente qui peut résulter de différentes formulations voire ne pas être marquée linguistiquement. De plus, la classification de l'attente du locuteur dépend d'un cadre fondamental donc il existe de nombreuses classifications selon les approches adoptées.

Notre modèle est dérivé de la théorie des actes de langage et plus précisément du modèle de prédiction d'actes présenté ci-dessus. Il utilise également les statistiques n-grammes. Les attentes sont un ensemble de réponses que l'utilisateur est susceptible d'attendre lorsqu'il formule un énoncé. Pour un énoncé de demande, il pourra s'attendre à une réponse, des précisions ou *a contrario* une demande de précision. La notation est celle présentée auparavant.

L'identification du locuteur et le contenu propositionnel sont ajoutés à cette notation afin d'obtenir des actes de dialogue de la forme :  $F_{U}^{fs}(p)$ .  $F^{fs}$  est l'acte de langage.  $U$  représente l'identifiant (donc le rôle) de l'utilisateur (ou  $C$ , client).  $M$  représente la machine (ou  $A$ , agent ou assistant).  $(p)$  est la représentation logico-sémantique du contenu propositionnel non développé pour le moment dans notre

système. Inclure l'identification du locuteur permet de tenir compte des rôles des interactants et d'aborder les attentes vues du côté locuteur ce qui correspond à une nouvelle direction considérant les attentes indépendamment de la tâche.

Considérant que l'approche, trop simpliste, de prédiction d'acte suivant n'est pas assez pertinente en dialogue, nous avons affiné ce modèle en lui ajoutant la notion d'attentes. En effet, les actes comme  $F(p)$ ,  $F^s(p)$ ,  $F^d(p)$  ou  $F^p(p)$  n'ont pas la même importance que ceux qui posent des buts comme  $F^f(p)$ ,  $F^{fs}(p)$ ,  $F^{fd}(p)$  ou  $F^{fp}(p)$ . Ces derniers ont, pour nous, un lien très fort avec l'acte fourni en réponse. Lorsque je demande une information, je souhaite que l'allocutaire me fournisse la réponse. Inversement, lorsque je donne une information, je n'ai pas d'attentes vis-à-vis de mon allocutaire mais je réponds davantage aux siennes. Nous formulons donc l'hypothèse que lorsque l'acte courant pose un but (donc des attentes) à savoir  $F^f(p)$ ,  $F^{fs}(p)$ ,  $F^{fd}(p)$  ou  $F^{fp}(p)$ , l'acte (immédiat pour les bi-grammes) de l'allocutaire doit être, dans un dialogue coopératif, un acte parmi les attentes du locuteur (respectant ainsi les maximes de Grice).

De plus, l'attente est orientée vers l'allocutaire. L'utilisateur pose des attentes vis-à-vis de l'allocutaire. Nous nous intéressons donc à l'acte fourni en réponse. Les n-grammes dont les éléments portent sur le même locuteur ne nous intéressent pas. Nous ne retiendrons donc que les changements de locuteur (pour les bi-grammes  $\{F_A(p), F_B(p)\}$  et  $\{F_B(p), F_A(p)\}$ ) et les actes dont l'historique contient un acte posant des buts. Ainsi, nous pouvons généraliser pour une taille d'historique quelconque. **Dans l'historique du dialogue, seuls nous intéressent les énoncés où l'interlocuteur pose une attente : les cas de demande d'action, d'information ou d'engagement.** Nous ne cherchons à prédire que les actes ayant dans leur historique un tel type d'énoncé.

A partir de cette génération d'attentes, nous pouvons adopter un mécanisme de gestion des attentes qui suit celui des buts proposé par [Caelen 92]. Ainsi, chaque acte qui pose un but, pose ses attentes associées dans une pile. Par exemple, lorsque je propose un  $F^{fs}_A(p)$ , mes attentes sont, si je suis un agent  $\{F^{fs}_B(p), F^{fs}_B(p), F^p_B(p), F^f_B(p), F^{fd}_B(p)\}$ . Ces attentes sont empilées dans la pile d'attentes. Lorsque le but est satisfait ou abandonné, les attentes sont dépilées. Ces attentes seront donc dépilées lorsqu'un des actes attendus est proposé (avec l'exemple précédent,  $F^{fs}_B(p)$  permet de dépiler simplement les attentes). Ainsi, les  $F(p)$ ,  $F^s(p)$ ,  $F^d(p)$  et  $F^p(p)$  ne génèrent-ils pas d'attentes mais permettent davantage de les dépiler.

Le but est un état que désire atteindre l'utilisateur. Ce peut être un état mental (ex : connaître une information ou acquérir un savoir-faire) ou situationnel (ex : un



fax envoyé). Un but peut se représenter à l'aide d'un prédicat  $b$  (ex  $b = s_A(\text{heure})$ ). Le but a plusieurs états. Un nouveau but ( $?b$ ) est un but qui vient d'être posé par l'expression du locuteur. Ce but est atteint ( $+b$  ou  $\dagger b$  selon Caelen) lorsque l'état de la situation rend le prédicat vrai ( $s_A(\text{heure})$ ). Il est ensuite satisfait ( $++b$  ou  $\ddagger b$  selon Caelen) lorsque le locuteur manifeste son accord explicitement (merci) ou implicitement sur l'atteinte du but. Un but peut aussi être mis en attente ( $-b$ ) lorsque l'un des interlocuteurs résout temporairement un autre problème (un autre but est posé). Il peut être réparé ou déplacé ( $b'$ ) lorsqu'à la suite d'une incompréhension ou d'un compromis, il est modifié. Il peut aboutir à un sous-but ( $sb$ ) lorsque le problème est décomposé en sous-problèmes. Il peut enfin être abandonné ( $@b$ ) à la suite d'un échec et d'un souhait d'abandon de la part du locuteur.

Nous pouvons voir un exemple de règles de gestion (volontairement simplifié) ci-dessous ( $?$  pour poser un but,  $+$  pour l'atteindre,  $++$  pour le satisfaire et  $@$  pour l'abandonner ;  $s$  pour un sous but) :

$$\text{But} = \emptyset \wedge \text{Acte} = F_{Y}^{\text{fs}}(p) \Rightarrow ?S_X(p) \wedge \text{Empile}(F_X^s(p) \vee F_X^{\text{fs}}(q))$$

$$\text{But} = S_X(p) \wedge \text{Acte} = F_Y^s(p) \Rightarrow +S_X(p) \wedge \text{Empile}(F_X^s(+S_X(p)) \vee (F_X^{\text{fs}}(q) \wedge S_X(q) \neq S_X(p)))$$

$$\text{But} = S_X(p) \wedge \text{Acte} = F_Y^{\text{fs}}(q) \wedge S_Y(q) = sS_X(p) \Rightarrow -S_X(p) \wedge ?S_Y(q) \wedge \text{Empile}(F_X^s(q) \vee F_X^{\text{fs}}(q))$$

$$\text{But} = S_X(p) \wedge \text{Acte} = F_Y^{\text{fs}}(q) \wedge S_Y(q) \neq sS_X(p) \Rightarrow @S_X(p) \wedge ?S_Y(q) \wedge \text{Dépile} \wedge \text{Empile}(F_X^s(q) \vee F_X^{\text{fs}}(q))$$

$$\text{But} = S_X(p) \wedge \text{Acte} = F_Y^s(q) \wedge q = +S_X(p) \Rightarrow ++S_X(p) \wedge \text{Dépile}$$

$$\text{But} = S_X(p) \wedge \text{Acte} = F_Y^{\text{fs}}(q) \wedge +S_X(p) \wedge S_X(q) \neq S_X(p) \Rightarrow ++S_X(p) \wedge ?S_X(q) \wedge \text{Dépile} \wedge \text{Empile}(F_Y^s(q) \vee F_Y^{\text{fs}}(q))$$

Empiler un but ( $?$ ) reviendra donc à empiler ses attentes associées, et inversement, le dépiler ( $++$  ou  $@$ ) reviendra à dépiler les attentes de leur satisfaction. Atteindre un but permettra d'empiler les attentes liées à la satisfaction du but.

Selon le mécanisme de gestion des buts et à partir de notre analyse du corpus, nous pouvons formuler des règles de la forme  $F_A^{\text{fs}}(p) \rightarrow F_B^s(p)(85,96\%) \mid F_B^{\text{fs}}(q)(10,53\%) \mid F_B^{\text{fd}}(q)(3,51\%)$  ; lorsque l'acte  $F_A^{\text{fs}}(p)$  apparaît, les attentes du locuteur A sont  $F_B^s(p)$  à 85,96% ou  $F_B^{\text{fs}}(q)$  à 10,53% ou encore  $F_B^{\text{fd}}(q)$  à 3,51%. Ainsi, nous n'empilerons pas seulement les attentes mais aussi leur probabilité d'apparition. Cette pondération (qui reflète les rôles des participants) permet de tenir compte de l'aspect social du dialogue. Dès lors nous pourrons proposer un modèle opportuniste qui, considérant l'acte généré, propose les attentes associées. Nous pouvons ainsi prédire les attentes les plus probables comme les autres (moins probables) ce qui nous permet d'éviter de tomber dans l'impasse des non-attendus.

### 3. Description statistique des corpus

Nous décrivons dans cette partie la répartition statistique des actes et leur cooccurrence dans les corpus étudiés.

#### 3.1. Corpus de dialogues homme-homme NESPOLE!

Le corpus de NESPOLE! décrit précédemment a été découpé en deux parties pour nos tests de prédiction. La partie ‘Apprentissage’ sert à entraîner le système à travers une analyse statistique des cooccurrences qui définit les probabilités. Elle est constituée de 4454 actes de dialogue répartis comme suit (**Tableau 25**) :

Rôle	F	F <sup>d</sup>	F <sup>f</sup>	F <sup>fd</sup>	F <sup>fs</sup>	F <sup>p</sup>	F <sup>s</sup>	Total
Agent	77	1	68	0	263	159	1615	2183
Client	65	0	22	1	429	110	1644	2271

**Tableau 25** : Répartition par rôle des 4454 actes du corpus d’apprentissage de Nespole!.

Le corpus de pré-tests a servi à valider l’annotation et n’a pas été utilisé pour l’apprentissage ni pour l’évaluation. Il est cependant composé de 74 actes de dialogues (**Tableau 26**) :

Rôle	F	F <sup>d</sup>	F <sup>f</sup>	F <sup>fd</sup>	F <sup>fs</sup>	F <sup>p</sup>	F <sup>s</sup>	Total
Agent	0	3	3	0	5	2	29	42
Client	4	3	0	1	12	1	11	32

**Tableau 26** : Répartition par rôle des 74 actes du corpus de pré-tests de Nespole!.

La partie ‘Tests’ sert à l’évaluation de la prédiction. Elle est constituée de 189 actes de dialogue répartis comme sur le **Tableau 27** ci-dessous :

Rôle	F	F <sup>d</sup>	F <sup>f</sup>	F <sup>fd</sup>	F <sup>fs</sup>	F <sup>p</sup>	F <sup>s</sup>	Total
Agent	3	6	1	2	6	6	67	91
Client	1	2	1	3	13	5	73	98

**Tableau 27** : Répartition par rôle des 189 actes du corpus de test de Nespole!.

### 3.2. Corpus de dialogues homme-machine PVE

Le corpus de PVE précédemment décrit est divisé en corpus-pilote, prétests et tests. Pour nos tests sur du dialogue homme machine, nous avons regroupé les prétests et 51 tests pour obtenir un corpus d'apprentissage de 1544 actes (**Tableau 28**). Les 21 tests restants ont servi de corpus d'évaluation (**Tableau 29**).

Tâche	Rôle	F	F <sup>d</sup>	F <sup>t</sup>	F <sup>fd</sup>	F <sup>fs</sup>	F <sup>p</sup>	F <sup>s</sup>	Total	Total
Stand RDIR	Agent	17	0	23	2	46	1	41	130	329
	Client	13	2	19	0	46	22	97	199	
RDV	Agent	7	0	6	8	32	0	10	63	150
	Client	5	7	6	0	16	8	49	87	
RESA Salle,matériel	Agent	11	0	15	15	47	5	20	113	260
	Client	12	12	7	0	29	15	72	147	
RSG AGD	Agent	23	4	15	8	26	0	25	101	241
	Client	8	7	20	2	34	17	52	140	
ENVDOC	Agent	24	0	23	0	24	1	22	94	252
	Client	21	0	38	0	22	21	56	158	
Comminfo	Agent	7	0	13	3	22	11	34	90	185
	Client	16	6	7	0	11	8	47	95	
Ouverture	Agent	0	0	1	0	63	54	2	120	127
	Client	0	0	0	0	3	2	2	7	
Total	Agent	89	4	96	36	260	72	154	711	1544
	Client	75	34	93	2	161	93	375	833	

**Tableau 28** : Répartition en tâches et rôles des 1544 actes du corpus d'apprentissage stochastique PVE.

Tâche	Rôle	F	F <sup>d</sup>	F <sup>t</sup>	F <sup>fd</sup>	F <sup>fs</sup>	F <sup>p</sup>	F <sup>s</sup>	Total	Total
Stand RDIR	Agent	4	0	6	2	16	0	15	43	106
	Client	6	2	5	0	11	6	33	63	
RDV	Agent	2	0	4	2	7	0	3	18	42
	Client	5	3	1	0	1	1	13	24	
RESA Salle,matériel	Agent	1	0	4	3	9	0	4	21	53
	Client	5	2	0	0	3	3	19	32	
RSG AGD	Agent	9	0	0	2	9	0	7	27	67
	Client	1	2	5	0	7	1	24	40	
ENVDOC	Agent	12	0	5	0	5	0	7	29	84
	Client	11	0	12	0	5	6	21	55	
Comminfo	Agent	3	0	0	2	2	3	11	21	51
	Client	5	2	6	0	4	3	10	30	
Ouverture	Agent	0	0	0	0	22	18	0	40	40
	Client	0	0	0	0	0	0	0	0	
Total	Agent	31	0	19	11	70	21	47	199	443
	Client	33	11	29	0	31	20	120	244	

**Tableau 29** : Répartition en tâches et rôles des 443 actes du corpus de test stochastique PVE.

## 4. Tests en dialogue homme-homme : Sur le corpus NESPOLE!

### 4.1. Prédiction d'actes

Pour l'approche par prédiction d'actes, nous avons regroupé les 30 premiers dialogues du corpus Nespole! collecté. Ces dialogues nous ont permis de définir la fréquence de cooccurrence de chaque acte. Nous avons obtenu 4315 actes dont 3582 actes en réponse (c'est-à-dire actes formulés par l'allocutaire en réponse au locuteur). Nous ne présenterons pas ici la matrice de cooccurrence des bi-grammes {Acte-en-cours, Acte-suivant} brute (la matrice fréquentielle est présentée en **Tableau 32**. Cette matrice nous fournira la probabilité de cooccurrence  $P(\text{Acte-courant}, \text{Acte-suivant})$  mais aussi la probabilité d'apparition des actes,  $P(\text{Acte-courant})$ , de l'agent (**Tableau 30**) et du client (**Tableau 31**).

Acte Agent	$F_A(p)$	$F^d_A(p)$	$F^f_A(p)$	$F^{fd}_A(p)$	$F^{fs}_A(p)$	$F^p_A(p)$	$F^s_A(p)$	Total
Fréquence (%)	2,02	1,67	1,60	2,25	3,82	3,57	34,04	48,97

**Tableau 30** : Fréquence d'apparition des actes de l'agent dans les 30 dialogues

Acte Client	$F_C(p)$	$F^d_C(p)$	$F^f_C(p)$	$F^{fd}_C(p)$	$F^{fp}_C(p)$	$F^{fs}_C(p)$	$F^p_C(p)$	$F^s_C(p)$	Total
Fréquence (%)	1,53	2,90	0,58	1,09	0,02	7,53	2,25	33,74	49,64

**Tableau 31** : Fréquence d'apparition des actes du client dans les 30 dialogues

Nous pouvons constater que l'agent prend la parole dans 48,97% des tours et le client dans 49,64%, le reste étant le début (0,72%) et la fin (0,67%) du dialogue. De plus, ces fréquences sont plus ou moins similaires. En effet, l'acte  $F^{fp}_A(p)$  n'apparaît jamais dans notre corpus et les demandes d'information  $F^{fs}(p)$  du client sont *grosso modo* deux fois supérieures en nombre à celles de l'agent. De même, l'agent semble proposer davantage d'invites  $F^p(p)$ . Enfin, les  $F^s(p)$  semblent équivalents, mais à cause des rôles de chacun, la proportion de phatiques (actes à contenu propositionnel nul) influe sur la distribution des actes. Ainsi, l'agent propose-t-il beaucoup d'actes  $F^s(p)$  à contenu informationnel tandis que le client fournit davantage de  $F^s(p)$  phatiques, c'est-à-dire dont le contenu propositionnel est nul mais qui ont pour fonction de maintenir la communication (en montrant qu'il écoute ou qu'il a compris ce qui est dit). Cela nous montre bien que les rôles ne sont pas interchangeables et qu'il est important de les considérer dans le dialogue.

Par la formule de Bayes, nous obtenons la probabilité d'apparition de chaque acte en contexte,  $P(\text{Acte-suivant} \mid \text{Acte-courant})$ . Nous pouvons le voir sur le **Tableau 32** présentant la fréquence d'apparition de chaque Acte-suivant (en colonne) en rapport à l'acte énoncé (en ligne) :

Acte\suivant	$F_A(p)$	$F_A^d(p)$	$F_A^f(p)$	$F_A^{fd}(p)$	$F_A^{fs}(p)$	$F_A^p(p)$	$F_A^s(p)$	$F_C(p)$	$F_C^d(p)$	$F_C^f(p)$	$F_C^{fd}(p)$	$F_C^{fp}(p)$	$F_C^{fs}(p)$	$F_C^p(p)$	$F_C^s(p)$	debut	fin
$F_A(p)$	0	0	9,20	1,15	1,15	1,15	9,20	2,30	1,15	2,30	0,00	0	8,05	0	52,87	0	11,49
$F_A^d(p)$	1,39	0	0	0	1,39	0	16,67	0	0	1,39	2,78	0	8,33	4,17	63,89	0	0
$F_A^f(p)$	2,90	0	0	7,25	1,45	1,45	18,84	52,17	0	0	1,45	0	0,00	0	14,49	0	0
$F_A^{fd}(p)$	0	0	0	0	0	0	3,09	1,03	85,57	0	0	0	3,09	0	7,22	0	0
$F_A^{fs}(p)$	0	0	0	0	3,03	3,03	3,64	0	0,61	0,61	0	0	4,24	1,21	83,64	0	0
$F_A^p(p)$	0,65	0	0	0,65	1,30	0,65	7,14	9,09	1,30	0,65	1,30	0	5,84	28,57	42,21	0	0,65
$F_A^s(p)$	1,63	0,20	1,29	1,23	3,34	1,97	6,06	0,54	2,18	0,82	1,84	0,07	10,48	1,02	67,26	0	0,07
$F_C(p)$	9,09	0	7,58	15,15	1,52	4,55	31,82	0	1,52	0	0	0	0	1,52	4,55	0	22,73
$F_C^d(p)$	4,00	0	2,40	4,80	11,20	2,40	60,00	0	0	0	0,80	0	1,60	1,60	11,20	0	0
$F_C^f(p)$	68,00	0	0	0	8,00	0	16,00	0	0	0	0	0	0	0	8,00	0	0
$F_C^{fd}(p)$	0	89,36	0	0	0	2,13	2,13	0	0	0	0	0	2,13	2,13	0	0	2,13
$F_C^{fp}(p)$	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0
$F_C^{fs}(p)$	0,31	0,31	0,31	3,38	6,77	2,15	76,92	0	0	0	0,62	0	6,46	0,92	1,85	0	0
$F_C^p(p)$	11,34	2,06	0	2,06	1,03	35,05	21,65	0	0	0	2,06	0	13,40	2,06	7,22	0	2,06
$F_C^s(p)$	1,30	1,65	2,27	2,95	4,53	3,43	65,52	0,34	0,34	0,55	0,62	0	6,80	1,24	8,45	0	0
debut	0	0	0	0	0	61,29	3,23	0	0	0	3,23	0	6,45	19,35	3,23	3,23	0
fin	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0

**Tableau 32** : Pourcentages d'actes suivant par rapport à chaque acte (Nespole!)

Ce que nous avons vu pour les unigrammes (répartition statistique des actes indépendamment des précédents) et les bigrammes (fréquence de cooccurrence des actes) se généralise ensuite pour n quelconque. Dès lors, nous pouvons proposer un modèle simple de prédiction d'acte à partir de l'acte énoncé. En effet, à partir de cette matrice, nous pouvons définir l'acte suivant le plus probable (celui qui maximise la probabilité). L'acte prédit est donc celui qui a la plus haute probabilité :

$$\text{Acte-prédit} = \operatorname{argmax}_{\text{acte} \in \{\text{Ens. Actes}\}} P(\text{Acte} \mid \text{Acte}_{i-1}, \text{Acte}_{i-2}, \dots, \text{Acte}_{i-n+1}),$$

soit pour  $n=2$ ,  $\operatorname{argmax}_{\text{acte} \in \{\text{Ens. Actes}\}} P(\text{Acte-suivant} \mid \text{Acte-courant})$ . Dans ce cas, la prédiction sera  $F_C^s(p)$  si l'on a  $F_A^{fs}(p)$  car cet acte obtient 83,64% de probabilité du succéder à l'acte courant.

#### 4.2. Génération et gestion des attentes

Pour l’approche fondée sur la théorie des attentes, nous nous intéressons en particulier à l’acte fourni en réponse. Nous érudons les n-grammes dont les éléments portent sur le même locuteur. De plus, seuls les actes délégués, c’est-à-dire posant un but, posent selon nous des attentes. Nous retiendrons ainsi les changements de locuteur (pour les bi-grammes  $\{F_A(p), F_B(p)\}$  et  $\{F_B(p), F_A(p)\}$  par exemple) et les n-grammes dont l’historique contient un acte délégué (**Tableau 33**).

Acte\suisvant	$F_A(p)$	$F^d_A(p)$	$F^f_A(p)$	$F^{fd}_A(p)$	$F^{fs}_A(p)$	$F^p_A(p)$	$F^s_A(p)$	$F_C(p)$	$F^d_C(p)$	$F^f_C(p)$	$F^{fd}_C(p)$	$F^{fp}_C(p)$	$F^{fs}_C(p)$	$F^p_C(p)$	$F^s_C(p)$
$F^f_A(p)$								76,60	0,00	0,00	2,13	0,00	0,00	0,00	21,28
$F^{fd}_A(p)$								1,06	88,30	0,00	0,00	0,00	3,19	0,00	7,45
$F^{fs}_A(p)$								0,00	0,67	0,67	0,00	0,00	4,70	1,34	92,62
$F^f_C(p)$	73,91	0,00	0,00	0,00	8,70	0,00	17,39								
$F^{fd}_C(p)$	0,00	95,45	0,00	0,00	0,00	2,27	2,27								
$F^{fp}_C(p)$															
$F^{fs}_C(p)$	0,34	0,34	0,34	3,75	7,51	2,39	85,32								
debut	0,00	0,00	0,00	0,00	0,00	63,33	3,33	0,00	0,00	0,00	3,33	0,00	6,67	20,00	3,33

**Tableau 33** : Pourcentages d’attentes par rapport à chaque acte posant un but (Nespole!)

Selon le mécanisme de gestion des buts et à partir de notre analyse du corpus, nous pouvons formuler des règles de la forme  $F^{fs}_A(p) \rightarrow F^s_B(p)(85,96\%) \mid F^{fs}_B(q)(10,53\%) \mid F^{fd}_B(q)(3,51\%)$  ; lorsque l’acte  $F^{fs}_A(p)$  apparaît, les attentes du locuteur A sont  $F^s_B(p)$  à 85,96% ou  $F^{fs}_B(q)$  à 10,53% ou encore  $F^{fd}_B(q)$  à 3,51%. Ainsi, nous n’empilerons pas seulement les attentes mais aussi leur probabilité d’apparition. Cette pondération (qui reflète les rôles des participants) permet de tenir compte de l’aspect social du dialogue.

Dès lors nous pourrions proposer un modèle opportuniste qui, considérant l’acte généré, propose les attentes associées. Grâce à ce modèle, nous pouvons prédire les attentes les plus probables comme les autres (moins probables), ce qui nous permet d’éviter de tomber dans l’impasse des non-attendus.

### 4.3. Test : Comparaison des deux approches

Nous pouvons comparer les deux approches proposées. La première est fondée uniquement sur la prédiction stochastique de l'acte suivant. La seconde, qui nous intéresse par son côté plus 'intelligent', reprend une certaine prédiction d'acte (seulement ceux en réponse) et lui adjoint une gestion 'intelligente' des attentes.

Nous avons émis l'hypothèse que le deuxième système devrait offrir de meilleures performances que le premier, considérant que la compréhension est plus fine. Pour tester cette hypothèse, nous avons analysé le 31<sup>ème</sup> dialogue (considéré donc comme dialogue de test) selon le même protocole que précédemment. La matrice de cooccurrence de bi-grammes {Acte-courant, Acte-suivant} est la suivante (**Tableau 34**) :

Acte/suivant	F <sub>A</sub> (p)	F <sup>d</sup> <sub>A</sub> (p)	F <sup>f</sup> <sub>A</sub> (p)	F <sup>fd</sup> <sub>A</sub> (p)	F <sup>fs</sup> <sub>A</sub> (p)	F <sup>p</sup> <sub>A</sub> (p)	F <sup>s</sup> <sub>A</sub> (p)	F <sub>C</sub> (p)	F <sup>d</sup> <sub>C</sub> (p)	F <sup>fd</sup> <sub>C</sub> (p)	F <sup>fs</sup> <sub>C</sub> (p)	F <sup>p</sup> <sub>C</sub> (p)	F <sup>s</sup> <sub>C</sub> (p)	A	C	A+C
F <sub>A</sub> (p)													1	1		1
F <sup>d</sup> <sub>A</sub> (p)							1					1	1	3		3
F <sup>f</sup> <sub>A</sub> (p)								2						2		2
F <sup>fd</sup> <sub>A</sub> (p)									2					2		2
F <sup>fs</sup> <sub>A</sub> (p)													4	4		4
F <sup>p</sup> <sub>A</sub> (p)												3		3		3
F <sup>s</sup> <sub>A</sub> (p)	1		1		1	1	2				2		14	22		22
F <sub>C</sub> (p)				2											2	2
F <sup>d</sup> <sub>C</sub> (p)							1						1	2		2
F <sup>fd</sup> <sub>C</sub> (p)		2				1								3		3
F <sup>fs</sup> <sub>C</sub> (p)							4							4		4
F <sup>p</sup> <sub>C</sub> (p)		1				1	1							3		3
F <sup>s</sup> <sub>C</sub> (p)			1		3		13			2	1	1	2	23		23
debut										1						1
														37	37	75

**Tableau 34** : Matrice de cooccurrence des bigrammes {Acte-courant, Acte-suivant} dans le 31<sup>ème</sup> dialogue de notre corpus (Nespole!)

Nous pouvons donc comparer l'acte effectivement produit après chaque acte et celui prédit. Si nous sélectionnons l'acte suivant le plus probable, nous obtenons 45 actes correctement prédits contre 30 qui peuvent être considérés comme des erreurs (en fait, ils sont seulement moins probables). Le taux d'erreur est donc de 40% ce qui semble montrer que l'approche par prédiction d'acte n'est peut-être pas la plus pertinente et qu'elle mériterait d'être affinée.

En reprenant notre 31<sup>ième</sup> dialogue, nous pouvons tester notre approche par gestion d'attentes de la même manière que précédemment. Nous ne conserverons donc que les actes qui posent des attentes comme les  $F^f(p)$ , les  $F^{fd}(p)$ ,  $F^{fp}(p)$  ou  $F^{fs}(p)$ . L'attente respective la plus probable est donc  $F(p)$ ,  $F^d(p)$ ,  $F^p(p)$  et  $F^s(p)$ . Nous obtenons alors 14 actes correctement prédits contre 1 qui n'est pas le plus probable. En termes de taux d'erreur, nous obtenons donc 6,67%, ce qui représente bien moins que les 40% démontrés ci-dessus.

Après affinage de l'annotation reflétant davantage les actes indirects des participants, nous avons refait les tests pour un historique  $n$  allant de 1 à 8. Nous avons alors séparé la prédiction des actes de l'agent de la prédiction de ceux du client (l'utilisateur du système). Les résultats sont sur le **Tableau 35** ci après :

Historique	Actes				Attentes			
	Agent		Utilisateur		Agent		Utilisateur	
	Taux	Nb	Taux	Nb	Taux	Nb	Taux	Nb
n=1	80,22	91	76,53	98	80,22	91	76,53	98
n=2	80,22	91	80,61	98	93,75	16	100,00	9
n=3	80,22	91	82,65	98	82,61	23	100,00	10
n=4	78,02	91	80,61	98	75,86	29	100,00	13
n=5	78,02	91	79,59	98	77,14	35	100,00	15
n=6	75,82	91	77,55	98	73,17	41	100,00	18
n=7	73,63	91	78,57	98	73,91	46	100,00	20
n=8	72,53	91	77,55	98	70,59	51	87,50	24

**Tableau 35** : Taux (%) de prédictions d'acte et de prédiction d'attentes et nombre à prédire (Nespole !).

Une amélioration significative a été apportée quant à la prédiction des actes. Celle-ci reste néanmoins en deçà de la prédiction d'attentes, quelle que soit la taille d'historique. Du fait que les attentes sont vues comme une sous-partie (une restriction) des actes, il y a beaucoup moins d'attentes à prédire, mais elles sont bien prédites. En outre, un historique de taille 2 est suffisant pour prédire actes ou attentes. Nous ne constatons pas d'amélioration significative au delà.



Ces résultats sont encourageants à double titre. En effet, en essayant d'interpréter nos résultats, nous trouvons les raisons de ces taux si intéressants. Il nous semble ainsi plus logique de ne chercher à prédire que les attentes fondées sur les actes en réponse. Ce sont bien ces derniers qui offrent des attentes. Prenons l'exemple d'une demande d'information. Cet acte demande une réponse. Les attentes associées sont donc plutôt prévisibles (la réponse à cette demande). En revanche, dans le cas d'une simple réponse, l'acte suivant peut être aussi bien une question de l'un des interlocuteurs qu'un apport d'information de l'un des interlocuteurs. Dès lors, toute prévision n'apporte que peu de précision. C'est pourquoi nous ne conserverons, dans notre approche, que les actes en réponse ou plutôt l'association acte courant / acte en réponse. Liés à une bonne gestion des attentes, ces actes nous permettent donc de fournir un modèle non seulement plus générique mais aussi plus 'intelligent' c'est-à-dire plus proche du fonctionnement humain.

Cette deuxième méthode, associée au mécanisme de gestion des attentes (qui reflète de manière plus 'intelligente' le déroulement du dialogue) devrait donc offrir des résultats beaucoup plus pertinents (l'efficacité de la gestion des attentes étant plus ou moins liée à celle de la gestion des buts). En outre, nous n'avons conservé, pour les besoins de l'expérience, que le premier acte, le plus probable. Or, nous pensons qu'il est souhaitable de conserver les autres afin de mieux gérer les attentes puisque le plus probable est pris en compte mais les suivants ne sont pas écartés. Dès lors, le taux d'erreur devrait diminuer encore.

### **Conclusion en dialogue homme-homme**

Dans cette partie, nous nous sommes intéressés aux dialogues entre humains coopératifs de renseignement touristique avec des rôles non équivalents. Nous avons montré que, dans ce cadre, l'acte suivant correspond le plus souvent à l'attente la plus probable. Dans le cadre du dialogue homme-machine, c'est, si possible, cet acte qui serait choisi, engendrant parfois quelques erreurs. Associé à un mécanisme de gestion des attentes, nous pourrions, en revanche, disposer d'une plus grande flexibilité dans la compréhension de l'énoncé. Dès lors, les actes moins probables seraient envisageables. Le premier acte n'étant pas toujours possible (nous ne comprenons pas tout), le deuxième pourrait être sélectionné. C'est ce que nous pourrions voir aussi en suivi de dialogue dans le cadre du dialogue entre humains médiatisé par la machine. Dans ce cadre, nous pourrions proposer toutes les attentes avec leur probabilité d'apparition, et même affiner rétroactivement le contexte (intentionnel, cognitif et social) en fonction de l'acte effectivement choisi. Cette approche nous permettrait alors de modéliser le dialogue de manière plus cognitive.

## **5. Tests en dialogue homme-machine : Dans le corpus PVE**

Nous avons montré dans la partie précédente l'intérêt de prendre en compte les attentes en dialogue oral humain sur un corpus de renseignement touristique [Besacier et al., 2001]. Nous montrerons ici l'intérêt de ces attentes en dialogue homme-machine. Pour cette partie, nous avons suivi la méthodologie présentée précédemment. Les dialogues sont issus de la campagne d'enregistrement en magicien d'Oz dans le cadre du projet PVE. Le corpus obtenu a été annoté en direct, c'est à dire lors de l'enregistrement. Ces annotations en actes ont subi plusieurs passes de vérification et mise à jour afin de refléter au mieux les actes indirects des participants et de gommer les erreurs d'interprétation dues aux phénomènes de direct et aux erreurs en amont. A partir de ces annotations, des tests ont été menés pour évaluer les deux approches présentées précédemment : prédiction d'actes et prédiction d'attentes.

### **5.1. Statistiques de succession**

Nous avons observé, dans la partie « Prédiction d'actes », que l'approche statistique, fondée sur les n-grammes incluant deux niveaux de recuit (backoffs) apporte les meilleurs résultats pour  $n = 2$ . Nous présenterons donc dans cette partie les statistiques fondées sur les bigrammes. Considérée comme suffisante, elle permettra une comparaison précise pour les attentes. Nous généraliserons ensuite, dans la partie comparaison, les résultats avec un historique variant de 1 à 8.

Pour l'approche par prédiction d'actes, les 65 dialogues (comprenant les 14 pré-tests et 51 dialogues de test) nous ont permis de définir la fréquence de cooccurrence de chaque acte. Nous avons obtenu 1071 actes dont 1035 actes en réponse (c'est-à-dire actes formulés par l'allocutaire en réponse au locuteur). Nous ne présenterons pas ici la matrice de cooccurrence des bi-grammes {Acte-en-cours, Acte-suivant}. En revanche, cette matrice nous fournira la probabilité de cooccurrence  $P(\text{Acte-courant}, \text{Acte-suivant})$  mais aussi la probabilité d'apparition des actes,  $P(\text{Acte-courant})$ , de l'agent et du client.

Par la formule de Bayes, nous obtenons la probabilité d'apparition de chaque acte en contexte,  $P(\text{Acte-suivant} | \text{Acte-courant})$ . La matrice (**Tableau 36**) présente la fréquence d'apparition de chaque Acte-suivant (en colonne) en fonction de l'acte énoncé (en ligne) :

	$F_A(p)$	$F_A^d(p)$	$F_A^f(p)$	$F_A^{fd}(p)$	$F_A^{fs}(p)$	$F_A^p(p)$	$F_A^s(p)$	$F_C(p)$	$F_C^d(p)$	$F_C^f(p)$	$F_C^{fd}(p)$	$F_C^{fs}(p)$	$F_C^p(p)$	$F_C^s(p)$	Total
<s>						1,25	13,75			11,25		72,50		1,25	100
$F_A(p)$			3,51		7,02	5,26	3,51	17,54		21,05		21,05	19,30	1,75	100
$F_A^d(p)$							50,00						50,00		100
$F_A^f(p)$				2,33	2,33		5,81	44,19		9,30		17,44	2,33	16,28	100
$F_A^{fd}(p)$	2,38							4,76	71,43	7,14		14,29			100
$F_A^{fs}(p)$							0,62			2,48		6,21		90,68	100
$F_A^p(p)$	2,86		5,71					2,86		28,57		17,14	5,71	37,14	100
$F_A^s(p)$			0,91			1,82		10,00		28,18	1,82	30,00	9,09	18,18	100
$F_C(p)$	19,35		16,13	3,23	38,71	3,23	16,13	3,23							100
$F_C^d(p)$	43,33		36,67		6,67	6,67	6,67								100
$F_C^f(p)$	32,47		10,39	7,79	24,68	5,19	16,88			1,30				1,30	100
$F_C^{fd}(p)$		100													100
$F_C^{fs}(p)$	3,55		16,31	8,51	34,75	2,13	33,33					0,71		0,71	100
$F_C^p(p)$	4,35		30,43		26,09	26,09	13,04								100
$F_C^s(p)$	15,46		17,01	10,82	36,08	7,73	11,86					0,52	0,52		100
Total	7,66	0,19	8,59	3,92	15,31	3,45	10,55	5,88	2,80	7,28	0,19	13,26	2,52	18,39	100

**Tableau 36** : Pourcentages d'actes suivants par rapport à chaque acte (PVE)

Dès lors, nous pouvons proposer un modèle simple de prédiction d'acte à partir de l'acte énoncé. En effet, à partir de cette matrice, nous pouvons définir l'acte suivant le plus probable (celui qui maximise la probabilité). Dès lors, l'acte prédit est simplement celui qui a la plus haute probabilité : Acte-prédit =  $\operatorname{argmax}_{\text{acte} \in \{\text{Ens. Actes}\}} P(\text{Acte-suivant} \mid \text{Acte-courant})$ . La prédiction sera  $F_C^s(p)$  si l'on a  $F_A^{fs}(p)$  car cet acte obtient 90,68% de probabilité du succéder à l'acte courant.

## 5.2. Attentes

L'approche fondée sur la théorie des attentes considère l'importance pour la machine de comprendre les attentes de l'utilisateur afin d'y répondre au mieux. Pour cela, nous nous intéressons, comme pour le dialogue humain, à l'acte fourni en réponse (1035 actes en réponse observés dans le corpus). Les bigrammes dont les éléments portent sur le même locuteur ne nous intéressent pas. Nous ne retiendrons donc que les changements de locuteur (les bigrammes  $\{F_A(p), F_B(p)\}$  et  $\{F_B(p), F_A(p)\}$ ). En outre, dans l'historique, seuls nous intéressent les énoncés où l'interlocuteur pose une attente : les cas de demande d'action, d'information ou d'engagement. Les règles de gestion de ces attentes suivent la gestion des buts. Nous

ne retiendrons donc que les séquences d'actes dans lesquelles figure un acte posant un but. Dans le cas des bigrammes, seules sont conservées les séquences {Acte-courant Acte-suivant} avec Acte-courant posant un but. Enfin, nous ne considérons pour l'expérience que l'attente la plus probable. Les résultats sont présentés sur le **Tableau 37**.

	$F_A(p)$	$F_A^d(p)$	$F_A^f(p)$	$F_A^{fd}(p)$	$F_A^{fs}(p)$	$F_A^p(p)$	$F_A^s(p)$	$F_C(p)$	$F_C^d(p)$	$F_C^f(p)$	$F_C^{fd}(p)$	$F_C^{fs}(p)$	$F_C^p(p)$	$F_C^s(p)$	Total
<s>						1,25	13,75			11,25		72,50		1,25	100
$F_A(p)$								21,74		26,09		26,09	23,91	2,17	100
$F_A^d(p)$													100		100
$F_A^f(p)$								49,35		10,39		19,48	2,60	18,18	100
$F_A^{fd}(p)$								4,88	73,17	7,32		14,63			100
$F_A^{fs}(p)$										2,50		6,25		91,25	100
$F_A^p(p)$								3,13		31,25		18,75	6,25	40,63	100
$F_A^s(p)$								10,28		28,97	1,87	30,84	9,35	18,69	100
$F_C(p)$	20,00		16,67	3,33	40,00	3,33	16,67								100
$F_C^d(p)$	43,33		36,67		6,67	6,67	6,67								100
$F_C^f(p)$	33,33		10,67	8,00	25,33	5,33	17,33								100
$F_C^{fd}(p)$		100													100
$F_C^{fs}(p)$	3,60		16,55	8,63	35,25	2,16	33,81								100
$F_C^p(p)$	4,35		30,43		26,09	26,09	13,04								100
$F_C^s(p)$	15,63		17,19	10,94	36,46	7,81	11,98								100
Total	7,73	0,19	8,41	3,86	15,27	3,09	10,05	5,99	2,90	7,44	0,19	13,53	2,51	18,84	100

**Tableau 37** : Pourcentages d'attente par rapport à chaque acte (PVE)

Au moment de prédire un acte de A, nous allons donc vérifier que son historique contient un acte posé par B et portant une attente et comparer, si tel est le cas, l'acte prédit et celui posé. Associé au mécanisme de gestion des buts, ce système devrait offrir de meilleurs résultats que la prédiction seule.

Pour tenter de prédire ce que doit répondre la machine, les statistiques ont été plus poussées pour les bi-grammes  $\{F_A, F_A\}$  et  $\{F_C, F_A\}$ . Le **Tableau 38** montre les statistiques de bi-grammes {Acte, Acte-suivant} estimées selon la méthodologie précédente en considérant un acte par énoncé. Les colonnes représentent les statistiques d'attentes suivant l'acte courant (en ligne). Certains actes du client posent des buts. Selon nous, ces même actes posent des attentes. Ils correspondent, dans le tableau, aux lignes grisées. Ce sont les requêtes d'action, d'information ou de confirmation.

Acte\Attentes	$F_A$	$F_A^d$	$F_A^f$	$F_A^{fd}$	$F_A^{fs}$	$F_A^p$	$F_A^s$	# total
$F_A$			14,3		28,6	42,9	14,3	14
$F_A^d$							100	1
$F_A^f$				22,2	22,2		55,6	9
$F_A^{fd}$	100							1
$F_A^{fs}$							100	1
$F_A^p$	25		50			25		4
$F_A^s$			33,3			66,7		3
$F_C$	20		16,7	3,3	40	3,3	16,7	30
$F_C^d$	43,3		36,7		6,7	6,7	6,7	30
$F_C^f$	32,9		10,5	7,9	25	6,6	17,1	76
$F_C^{fd}$		100						2
$F_C^{fs}$	3,5		16,8	8,4	34,3	3,5	33,6	143
$F_C^p$	4,2		29,2		25	29,2	12,5	24
$F_C^s$	15,4		16,9	10,8	35,9	9,2	11,8	195
début						82,8	17,2	64
Total	13,7	0,3	15,6	7	27,5	16,8	19,1	597

**Tableau 38** : Statistiques de bigrammes (%) et attentes en fonction de chaque acte (PVE)

Avec cette méthode, le taux de bonne prédiction passe à 65.6% ce qui signifie pour nous que la considération de la notion d'attentes permet un meilleur contrôle du dialogue. Nous allons voir dans le prochain chapitre une comparaison plus fine avec la dernière passe d'annotation et une taille d'historique variant de 1 à 8.

### 5.3. Comparaison, taux de bonne prédiction

Comme pour le cadre du dialogue homme-homme, nous avons fait l'hypothèse que le deuxième système devrait offrir de meilleures performances que le premier, considérant que la compréhension est plus fine. Des corpus recueillis, nous avons retiré le corpus pilote homme-homme car nous voulions tester du dialogue homme-machine. Nous avons regroupé le corpus de 14 dialogues de pré-tests et 51 dialogues du corpus de 72 dialogues de tests (la collecte en magicien d'Oz proprement dite). Ce regroupement a servi de corpus d'apprentissage pour entraîner les probabilités de succession d'actes et d'attentes. Ces 65 dialogues sont composés d'environ 1500 actes de dialogue et concernent 80 tâches (hors tâche d'ouverture introduite artificiellement). Les 21 dialogues restant ont servi de corpus d'évaluation. Ils sont composés d'environ 400 actes et concernent 25 tâches (toujours hors tâche d'ouverture introduite artificiellement). Cette répartition permet d'équilibrer les proportions de chaque tâche dans chacun des corpus, comme nous pouvons le constater sur le **Tableau 39** ci-dessous :

Tâches	Apprentissage	Evaluation	Total
Redirection, Standard	16	5	21
Rendez-vous	9	3	12
Réservation de salle	12	3	15
Communication d'information	11	3	14
Gestion d'un agenda partagé	15	5	20
Envoi de document	17	6	23
Ouverture	65	21	86

**Tableau 39** : Description du corpus

Une première comparaison des deux approches dans le cadre du domaine homme-machine et de l'application cible (le projet PVE) nous fournit des résultats prometteurs (**Tableau 40**) :

Prédiction	Au total	Assistant	Utilisateur
Actes	138/293 (47,10%)	47/141 (33,33%)	95/151 (62,91%)
Attentes	78/133 (58,65%)	24/65 (36,92%)	54/68 (79,41%)

**Tableau 40** : Résultats de prédiction d'acte versus attentes (PVE)

Les actes et attentes de l'utilisateur (le client) sont mieux prédits que ceux de l'assistant (l'agent). Comme dans le cas du dialogue homme-homme, les rôles ne sont pas équivalents. L'assistant a tendance à diriger le dialogue tandis que l'utilisateur a tendance à le suivre. Ainsi la stratégie de l'utilisateur sera-t-elle souvent réactive, celui-ci répondant simplement aux questions du système. *A contrario*, l'assistant pose plus souvent des questions à propos des paramètres qu'il sait manquants. La prédiction de l'acte et de l'attente les plus probables est donc plus ardue, ces probabilités étant moins réparties.

La prédiction des attentes demeure, comme en dialogue humain, supérieure à celle des actes. Nous allons donc valider ces résultats après la dernière passe de réannotation. L'évaluation porte alors sur cette dernière annotation et sur des tailles d'historiques variant de un (les unigrammes, c'est-à-dire la répartition statistique des actes dans le corpus) à huit (les octogrammes, la prédiction du huitième acte connaissant les 7 précédents). Les résultats sont donnés dans le **Tableau 41** ci-après :

Historique	Actes				Attentes			
	Assistant Virtuel		Utilisateur		Assistant Virtuel		Utilisateur	
	Taux	Nb	Taux	Nb	Taux	Nb	Taux	Nb
n=1	39,3	178	49,2	244	39,3	178	49,2	244
n=2	55,6	178	52,1	244	69,3	49	91,8	97
n=3	48,3	178	50,4	244	53,2	62	61,6	130
n=4	51,1	178	49,2	244	50,7	73	56,8	162
n=5	51,7	178	46,3	244	53,6	84	52,2	186
n=6	51,7	178	45,9	244	53,5	86	49,5	200
n=7	51,1	178	46,3	244	52,8	91	48,4	215
n=8	50,0	178	46,3	244	52,2	92	48,2	220

**Tableau 41** : Taux (%) de prédictions d’acte et de prédiction d’attentes et nombre à prédire.

Les résultats sont plus faibles qu’en dialogue humain. Néanmoins, si la différence est moins importante, elle n’est pas nulle. Les attentes sont mieux prédites que les actes et en particulier celles qui portent sur l’utilisateur. L’approche par les attentes offre donc un gain non négligeable dans la prédiction de l’acte en particulier celle du client qui est ici plutôt coopératif.

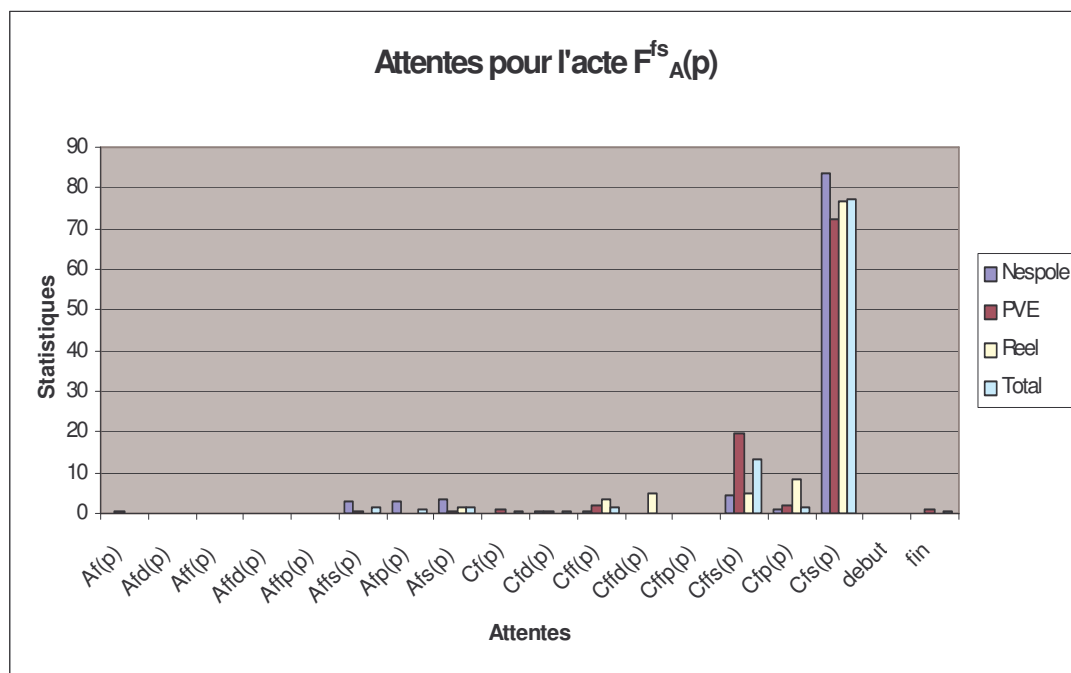
### **Conclusion en dialogue home-machine**

Dans cette partie, nous nous sommes intéressés aux dialogues coopératifs de secrétariat virtuel avec des rôles non équivalents. Nous avons montré que, dans ce cadre, l’acte suivant correspond le plus souvent à l’attente la plus probable. Dans le cadre du dialogue homme-machine, c’est, si possible, cet acte qui serait choisi, engendrant parfois quelques erreurs. Associé à un mécanisme de gestion des attentes, nous pourrions, en revanche, disposer d’une plus grande souplesse dans la compréhension de l’énoncé. Dès lors, les actes moins probables seraient envisageables. Le premier acte n’étant pas toujours possible (nous ne comprenons pas tout), le deuxième pourrait être sélectionné. Nous pourrions donc proposer toutes les attentes avec leur probabilité d’apparition voire affiner rétroactivement le contexte (intentionnel, cognitif et social) avec l’acte effectivement choisi.

## Conclusion d'une modélisation stochastique

Suivant alors le protocole décrit dans [Fouquet 01], nous avons pu mettre en relief avec davantage de données la notion de rôle. Par exemple, nous avons pu souligner, que lorsque l'agent pose un faire-faire le client répond par un faire ou un faire-savoir. Mais lorsque c'est le client qui pose le faire-faire, l'agent peut aussi répondre par un faire-faire-savoir (8%).

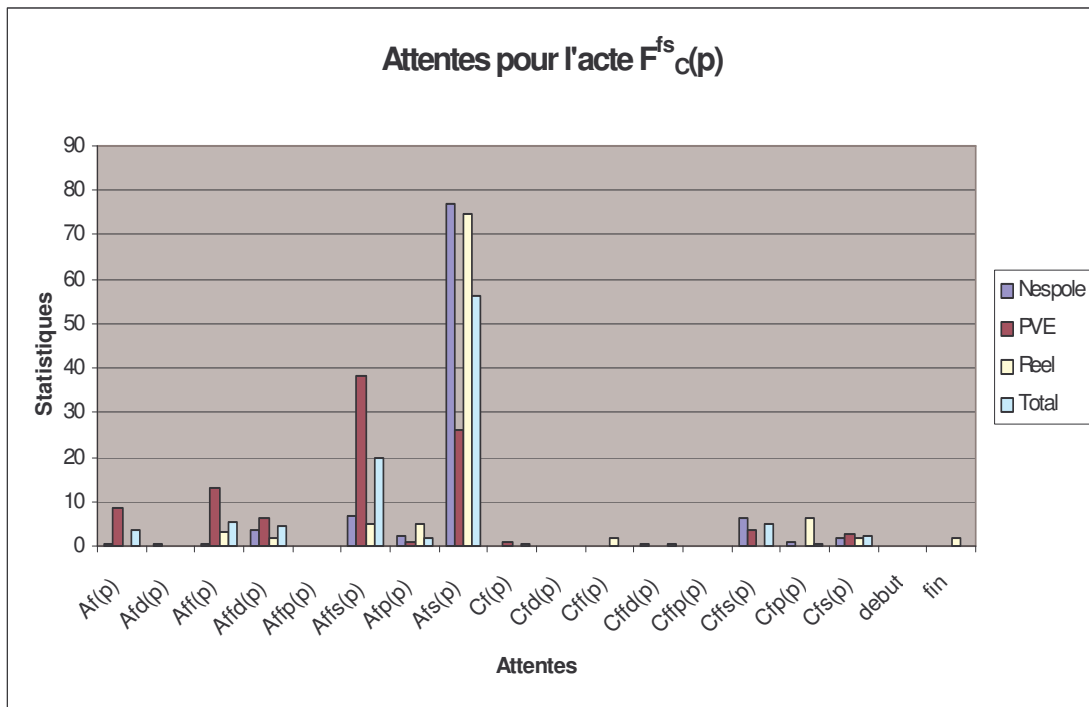
Les résultats acte par acte sont présentés en détail en annexe. Ils comprennent les statistiques effectuées sur le corpus de dialogues homme-homme (pour Nespole! et pour PVE), sur le corpus de dialogues homme-machine dans le cadre des prétests et sur le corpus de dialogues homme-machine dans le cadre des tests. En outre, des combinaisons de ces corpus ont été effectuées. Nous renvoyons le lecteur à l'annexe pour une comparaison des attentes, acte par acte. Notons cependant à titre d'exemple les **Figure 25** et **Figure 26** ci-dessous :



**Figure 25** : Comparaison de divers *corpus* pour les attentes après une demande d'information de l'agent

Cette figure confirme la coopérativité du client (en partie due à l'effet expérimentation) qui adopte une stratégie plutôt réactive dans les trois types de dialogue.





**Figure 26** : Comparaison de divers *corpus* pour les attentes après une demande d'information du client

Cette figure est intéressante car on peut y observer une différence de comportement entre le dialogue homme-homme et homme-machine. En effet, lorsque le client demande une information, les deux résultats de dialogue homme-homme (Nespole! et Pve\_Réel) sont similaires. L'agent qui est alors humain est plutôt réactif et répond à cette demande. En revanche, les résultats de PVE en dialogue homme-machine montrent que l'agent (le compère simulant la machine) est davantage directif et demande davantage de précisions. En fait, le nombre de cas trouvés est relativement proche pour les deux réponses préférées (réponse et demande de précision), mais cette situation nous montre bien une différence du dialogue homme-machine surtout orienté par la tâche. En effet, L'agent est dirigé par cette tâche qu'il connaît et demande directement des précisions pour la résoudre de manière efficace.



## **Conclusions et perspectives**

Nous avons montré l'intérêt de notre approche ainsi que son originalité à travers la prise en compte des attentes du locuteur (et non plus seulement de celles de la machine). La notion d'attente peut apporter des informations non seulement au niveau de la **compréhension automatique** mais aussi au niveau du **contrôle du dialogue**. Elle est donc intéressante pour le **dialogue homme-machine** dans lequel la machine doit pouvoir détecter les attentes de l'humain pour mieux y répondre. Les attentes seront donc considérées comme des indices pertinents pour la machine. Elles sont également fondées en **dialogue homme-homme** médiatisé par la machine, dans lequel elles apportent des indices au compère humain, puis à la machine.

Cette approche est issue de différentes disciplines. La sémiotique a offert la notation par combinaison des modaux de Greimas. La philosophie du langage a

apporté une représentation des attentes sous une forme dérivée des actes de langage. Opérer à un niveau pragmatique permet en outre de rester indépendant de la tâche effectuée. L'ethnométhodologie a apporté la notion de rôles que nous avons représentée de manière statistique. Enfin, la linguistique nous a permis de voir les attentes comme une liste d'écarts possibles incidents à l'axe régissant de Luzzati.

Nous pensons donc que certains actes indirects posent des attentes tandis que d'autres y répondent. Pour les premiers, nous avons établi des règles de la forme :

$$\begin{array}{l}
 \frac{F^{\text{fs}}_A(p)}{\text{ConnaissancesSupposées}_A(p) \wedge \text{Croyances}_A(\text{Coopère}_B)} \Rightarrow \frac{F^{\text{s}}_B(p')}{\text{ConnaissancesSupposées}_B(p') \wedge \text{Stratégie}_B(\text{réactif}) \wedge \text{Coopère}_B} \\
 \Rightarrow \frac{F^{\text{fs}}_B(p')}{\text{ConnaissancesSupposées}_B(p') \wedge \text{Stratégie}_B(\text{Directif} \vee \text{Coopératif}) \wedge \text{Coopère}_B} \\
 \Rightarrow \frac{\neg F^{\text{s}}_B(p')}{\text{ConnaissancesSupposées}_B(p') \wedge (\text{ConnaissancesMisesAJour}_B(p') \vee \neg \text{Coopère}_B)}
 \end{array}$$

Un acte dans un certain contexte cognitif produit un certain nombre d'attentes en fonction du contexte cognitif supposé de l'allocutaire, la stratégie qu'il va employer, etc.

L'introduction d'une partie statistique nous a permis de classer les attentes en fonction des rôles et stratégies et ainsi d'être plus efficace dans leur gestion. Nous avons alors présenté une approche à base de corpus dans le dialogue humain et dans le dialogue homme-machine. Les corpus obtenus ont été décrits avant de présenter l'expérimentation et la plate-forme de simulation en magicien d'Oz qui ont permis leur collecte. Pour les deux magiciens développés, nous avons montré que la théorie des attentes semble être une perspective intéressante pour améliorer l'identification des actes de dialogue et le contrôle du dialogue par la prédiction de l'acte le plus attendu et des autres possibles. A partir des dialogues collectés, nous avons montré que les attentes dépendent des rôles et de la stratégie adoptée.

L'évaluation de la partie statistique du modèle a démontré l'intérêt de notre approche. Elle a mis en avant la coopérativité du client qui a adopté une stratégie plutôt réactive. Elle a souligné une différence entre dialogue humain et dialogue homme-machine : notamment, suite à une demande d'information, les réponses de la machine seraient davantage directives, c'est à dire des demandes de précision. Nous pensons que ce résultat est lié au souci d'efficacité davantage présent en dialogue homme machine où l'écart est peu permis et difficilement récupérable.

Ainsi, nous pouvons adopter une approche générique, avec une séparation de la tâche rendue possible. Notre approche par les attentes du locuteur est une première brique pour un modèle générique de dialogue. Un tel modèle pourra s'insérer dans de nombreuses applications dans lesquelles un dialogue peut améliorer la qualité d'interaction entre l'utilisateur et la machine. En outre, une telle modélisation des attentes permettra d'effectuer un suivi de dialogue dans un contexte homme-homme. Ce suivi montre tout son intérêt dans des cadres comme celui du projet ERIM dans lesquels un individu (ici l'interprète) souhaite s'adapter à une conversation en cours.

### **Limites**

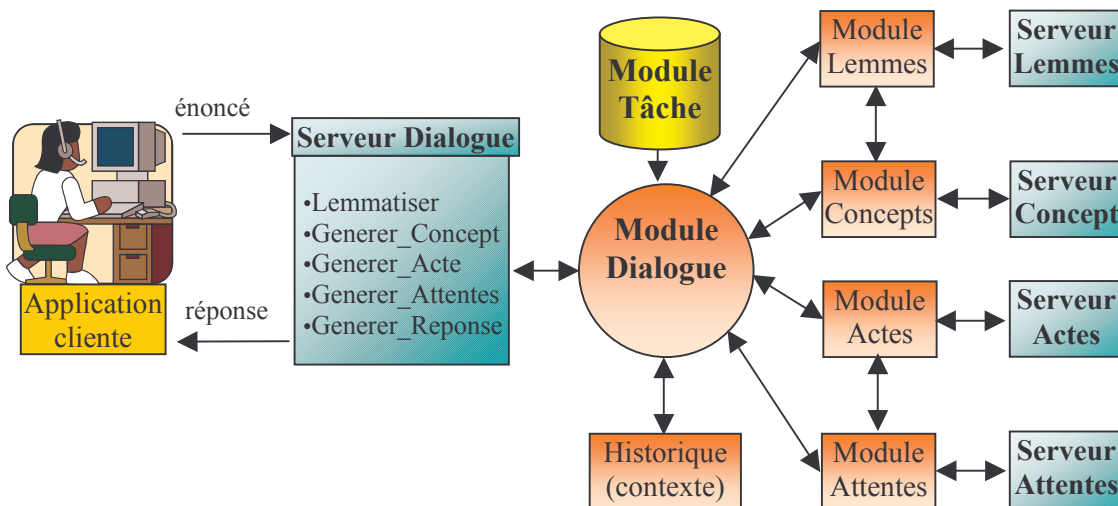
Un tel modèle avoue cependant quelques limites. Travailler à un niveau pragmatique permet de s'affranchir au maximum de la tâche effectuée. Mais cela nécessite un fastidieux travail d'annotation qui reste arbitraire, notamment au niveau du choix de l'acte indirect. Nous avons par exemple commencé une annotation *in vivo* en associant l'énoncé reçu à un unique acte indirect le représentant le mieux et correspondant le mieux à la suite supposée du dialogue. Nous avons alors observé une stratégie directive (question) là où il y avait souvent une stratégie coopérative (réponse puis question). Nous avons alors recommencé le travail d'annotation en associant chaque portion remarquable de l'énoncé à l'acte indirect lui correspondant. Les statistiques obtenues sont donc dépendantes de l'annotation en amont.

Le fait même d'utiliser une méthode statistique apporte des faiblesses quant à la quantité de données nécessaires à la définition des probabilités. Ainsi, changer de domaine nous oblige à recommencer à suivre la méthodologie de collecte de corpus réel, annotation, analyse, construction d'énoncés, élaboration d'un magicien d'Oz, expérimentation, collecte de données homme-machine, annotation et analyse des dialogues obtenus. Nous pensons qu'on devrait pouvoir définir des classes d'applications pour lesquelles les répartitions d'actes et les probabilités des attentes seraient similaires. Nous pensons notamment au modèle maître-esclave dans lequel le **maître** aura une stratégie davantage directive tandis que l'**esclave** sera davantage réactif et au modèle maître-maître dans lequel la stratégie de négociation est nécessaire.

Enfin, les statistiques dépendent aussi des types d'énoncés préétablis pour la machine.

## Perspectives (mise en œuvre)

Un dernier obstacle à notre modèle concerne la réalisation de l'analyse en actes. En effet, nous avons vu l'intérêt d'une notation pragmatique, mais sa réalisation est loin d'être triviale. Nous avons choisi une approche par serveurs opportunistes. En effet, nous pensons que c'est une bonne solution qui nous garantit un outil modulaire et réutilisable. De plus, cette architecture nous permet de nous affranchir de la tâche et de nous y adapter simplement puisque, opportuniste, notre modèle ne renvoie que le résultat de son calcul à un instant donné, tenant malgré tout compte des contextes intentionnel, cognitif et social comme montré dans notre ébauche de modèle. Nous pouvons observer cette architecture à travers le schéma suivant qui nous montre bien des serveurs opportunistes (**Figure 27**) :



**Figure 27** : Architecture du système - serveurs opportunistes

Dans cette architecture, l'application cliente converse uniquement avec le serveur de dialogue qui traite le contrôle du dialogue à partir de l'historique mais surtout des attentes. Devant la nécessité d'utiliser une notation pragmatique, nous devons donc analyser les énoncés sous cet angle. C'est pourquoi nous envisageons un serveur d'actes qui, à partir d'un énoncé, devrait nous fournir l'acte associé. Pour celui-ci, actuellement aléatoire pour les tests, nous pouvons utiliser l'approche de Colineau [Colineau 98] fondée sur les réseaux neuronaux pour évaluer l'acte énoncé. Nous pouvons aussi utiliser une approche plus stochastique à travers une analyse sémantique latente (LSA). Ces approches, quoique demandant beaucoup de données afin d'établir les liaisons, nous semblent suffisamment génériques car liées à un apprentissage qui peut être indépendant de la tâche. En revanche, nous avons besoin

de lemmatiser l'énoncé (i.e. de le transformer de sa forme originelle vers une forme générique non fléchie, donc de le normaliser) puis d'en établir les concepts qui devraient nous permettre d'élaborer les actes.

L'avantage d'une telle architecture est son extrême adaptabilité à l'application cliente. Nous le montrerons à travers l'insertion du suivi de dialogue dans le projet ERIM dans lequel nous pourrions proposer l'historique en termes d'énoncés mais surtout en termes d'actes et d'attentes des participants. Dès lors, l'interprète pourra s'adapter plus vite à la conversation en cours lorsqu'il connaîtra les buts et attentes des participants qui ne se comprennent plus.

En parallèle, nous étudions pour le projet PVE les possibilités offertes par un nouveau standard vocal VoiceXML. En effet, celui-ci offre une architecture dans laquelle nous pouvons créer notre dialogue. Pour le moment, il nous semble trop proche des grammaires pour nous permettre d'offrir un véritable dialogue qui ne soit pas uniquement directif. En revanche, y intégrer nos attentes nous permettrait d'apporter une grande souplesse à un tel système.

Il nous faut finir la plate-forme modulaire par serveurs opportunistes. Nous nous intéresserons ensuite à la génération des concepts, à celle des actes puis à la gestion de l'historique.

Nous étudions également leur intégration, en dialogue homme-machine, avec le modèle du contrôle de dialogue de Nguyen [Nguyen 03ab] et Caelen [Caelen & Nguyen 04] pour la partie **contrôle du dialogue** ainsi qu'avec le modèle d'interprétation de [Xuereb 04] et Caelen [Xuereb & Caelen 04] pour la partie **interprétation pragmatique**.

## **Perspectives (modèle)**

Outre la mise en place de la gestion d'attentes, nos perspectives sont l'affinage de notre modèle à travers une prise en compte plus fine du contexte (échange, stratégie, but, etc.). Pour le moment, nous utilisons uniquement l'attente la plus probable déduite de l'acte porteur d'attentes (c'est-à-dire posant un but) issu de l'historique. Une analyse plus fine incluant les stratégies révèle la première attente comme correspondant à une stratégie réactive et la deuxième comme correspondant à une stratégie coopérative. Une deuxième étape de l'expérimentation a consisté à mettre en valeur les phrases en fonction de la stratégie que le magicien décidera d'employer. Lorsque l'utilisateur pose un faire-faire-savoir, nous devons alors mettre en valeur un faire-savoir (réactif) mais aussi un faire-faire-savoir (coopératif). Il s'agira donc d'utiliser les attentes déduites de l'acte en cours et de la stratégie employée. Une autre perspective sera ensuite d'utiliser l'acte et la stratégie pour estimer la séquence d'actes qui devrait se dérouler avant d'atteindre le but posé.

Nous l'avons vu dans nos limites, notre modèle nécessite une approche à base de corpus. La charge de travail à fournir lorsque l'on veut développer un système incluant ces attentes est donc loin d'être faible. Or nous pensons que les dialogues peuvent se regrouper en classes afin d'évaluer les probabilités pour les classes de dialogue et non pour chaque tâche. Nous avons par exemple relevé des dialogues de type maître-esclave avec de la coopération et des dialogues de type maître-maître avec de la négociation. Il serait d'ailleurs intéressant de limiter l'usage du magicien d'Oz par de l'apprentissage automatique de la tâche.

Un autre intérêt de notre approche est de mettre en avant les situations de succession ne survenant jamais. Ce sont les zéros des matrices de cooccurrence. Tenir compte de ceux-ci permet d'éviter de développer du temps d'analyse d'actes inutiles.

Enfin, nous pensons que notre théorie peut s'appliquer à la partie **génération** du système de dialogue homme-machine ou du système de dialogue homme-homme médiatisé par la machine. En effet, nous utilisons actuellement des phrases à trous pour générer les énoncés. Nous pensons que les attentes peuvent assouplir ce fonctionnement. En effet, si les attentes de l'utilisateur sont un nom dans une liste de noms, il n'est pas nécessairement utile de lui fournir une phrase d'introduction longue mais il est sans doute plus efficace de lui fournir le nom uniquement. C'est cette efficacité que nous allons ensuite tester avec l'introduction des attentes en **évaluation**.



## Perspectives (communication non verbale)

Une autre perspective nous semble très intéressante : le cas de la **communication non verbale**. C'est en ce sens que nous nous tournons vers le projet ACE (Agent Conversationnel Expressif) de dialogue avec avatar. L'objectif scientifique de ce projet est d'intégrer un ensemble d'éléments de communication non verbale à un dialogue multimodal dans le but d'améliorer l'efficacité de l'interaction entre un agent conversationnel et ses utilisateurs. Son originalité réside dans le développement d'un modèle de communication gestuelle générique, *piloté* par le dialogue et *synchronisé* avec les étapes de celui-ci, qu'il s'agisse de dialogues dirigés par l'agent conversationnel ou dirigés par l'utilisateur. Dans le premier cas, l'agent montre une continuité expressive au fur et à mesure de la progression du scénario de dialogue, dans le second cas, il exprime des émotions en réaction aux répliques de l'utilisateur.

Une simulation avec une comédienne a été réalisée en septembre 2003 pour capturer les paramètres comportementaux de l'agent en situation. Pour l'évaluation ergonomique du nouvel agent communicationnel expressif – évaluation en cours dont les résultats devraient être obtenus fin septembre 2004 -, nous avons développé une plate-forme de type magicien d'Oz intégrant la communication verbale et non verbale. Dans cette plate-forme, un compère manipule l'agent conversationnel, faisant passer ce dernier pour un avatar aux yeux des sujets. L'aide fournie par les attentes devrait être assez intéressante pour le compère mais aussi pour le système.

Des tests unitaires ont été élaborés, explorant l'apport de la notion d'attente dans le domaine de **l'évaluation de systèmes de dialogue** avec une question qu'il convient de résoudre : « le système répond-t-il aux attentes de l'utilisateur ».

Ces tests ont la forme d'un questionnaire qui se déroule comme suit :

- ✓ Une séquence contenant du dialogue et un avatar expressif gesticulant est présentée.
- ✓ Des questions sont ensuite posées sur :
  - la compréhension de la séquence (test DQR)
  - les attendus ; ce que le sujet pense qu'il va se passer ensuite (test DAR)
  - les attentes ; ce que le sujet aimerait qu'il se passe ensuite (tests UARD).

## Compréhension

Les tests de compréhension de la séquence (DQR) peuvent porter sur l'acte ou sur l'état 'psychologique' de l'agent :

```
<dqr id="suggest">
  <question reponse="oui">L'agent a-t'il fait une suggestion ? </question>
  <question reponse="non">L'agent a-t'il fait une proposition ? </question>
  <question reponse="non">L'agent a-t'il émis un avertissement ? </question>
</dqr>
<dqr id="joy">
  <question reponse="oui">L'agent est-il joyeux ? </question>
  <question reponse="non">L'agent est-il triste ? </question>
  <question reponse="non">L'agent est-il stressé ? </question>
</dqr>
```

On présente alors un dialogue D à un utilisateur, puis une question Q dont on connaît la réponse « idéale » R. On la compare avec celle de l'utilisateur R<sub>U</sub>.

D = <Séquence d'ouverture avec un agent joyeux>	ou	D = <Séquence d'ouverture avec un agent joyeux>
Q = est-ce un dialogue d'ouverture ?		Q = l'agent est-il enjoué ?
R = oui		R = oui
	mais	D = <Séquence d'ouverture avec un agent joyeux>
		Q = l'agent est-il triste ?
		R = non

## Attendus

Le test sur les attendus fonctionne comme le test DQR, on demande seulement à l'utilisateur ce qui va se passer après une séquence dialogique et on compare sa réponse avec l'action effectuée par l'agent. Il reflète donc ce que pense l'utilisateur que la machine peut faire.

```
<dar id="1" dialogue="corpus_test_unitaires.xml">
  <attendu reponse="non">Est-ce un préambule à une aide ? </attendu>
</dar>
```

Nous pourrions ainsi proposer des questions de la forme :

D = <Séquence d'ouverture avec un agent joyeux>  
A = est-ce un préambule à une aide ?  
R<sub>U</sub> = oui (R=non, car l'agent est programmé pour demander le profil utilisateur)

## Attentes

Enfin les tests sur les attentes reflètent ce que l'utilisateur voudrait que la machine fasse. L'utilisateur a l'initiative, il produit un énoncé, on lui pose une question sur ses attentes et on compare sa réponse avec l'action de l'agent. Ces questions peuvent être posées sur l'acte attendu, sur l'état 'psychologique' (ou expression) attendu ou sur le geste attendu :

```

<uard id="1" dialogue="corpus_test_unitaires.xml">
  <utilisateur>Bonjour je suis Yannick Fouquet </utilisateur>
  <attente reponse="non">
    Vous attendez vous à une présentation de l'agent ?
  </attente>
</uard>
<uard id="1" dialogue="corpus_test_unitaires.xml">
  <utilisateur>Bonjour je suis Yannick Fouquet </utilisateur>
  <attente reponse="oui">
    Vous attendez vous à une ouverture sérieuse de l'agent ?
  </attente>
</uard>
<uard id="1" dialogue="corpus_test_unitaires.xml">
  <utilisateur>Bonjour je suis Yannick Fouquet </utilisateur>
  <attente reponse="non">
    Vous attendez vous à un mouvement des bras de l'agent ?
  </attente>
</uard>

```

Nous pourrions ainsi proposer des questions de la forme :

U = « bonjour, je suis Yannick Fouquet »

A = vous attendez vous à une présentation de l'agent ?

R<sub>U</sub> = oui

D = <Séquence d'ouverture avec un agent joyeux> (=> R=oui)

Ou

U = « bonjour, je m'appelle Jean Caelen »

A = vous attendez-vous à une ouverture sérieuse de l'agent ?

R<sub>U</sub> = oui

D = <Séquence d'ouverture avec un agent joyeux> (=> R=non)

Ou

U = « bonjour, je m'appelle Brigitte Meillon »

A = vous attendez-vous à un mouvement des bras de l'agent ?

R<sub>U</sub> = oui

D = <Séquence d'ouverture avec un agent joyeux> (=> R=non)

Si les statistiques dépendent des énoncés préétablis pour la machine, ces derniers tests permettent leur évaluation afin d'améliorer leur pertinence et leur efficacité. En effet, les séquences D proposées sont en format XML, incluant une annotation en actes et état affectif. Les questions posées sont alors adaptées automatiquement aux informations recueillies dans les séquences afin de déterminer les faiblesses du système et de les résoudre.

## **Bilan général**

Tout au long de cette thèse, nous nous sommes placé dans une double finalité d'améliorer la qualité des interactions verbales entre les humains et la machine, et d'améliorer celle de la communication langagière entre humains par le biais de la machine. Les attentes du locuteur sont intéressantes pour l'interprétation pragmatique et pour le contrôle. Elles s'appliquent au dialogue homme-homme ainsi qu'au dialogue homme-machine. Notre modèle pluridisciplinaire tente de reproduire au maximum le comportement humain en situation de dialogue finalisé. Il nécessite en revanche une analyse fine de corpus en situation, qu'il convient de généraliser pour des classes de tâches. Ce modèle permet alors d'opérer à un niveau pragmatique et de nous abstraire au maximum de la tâche en cours pour une généralité maximale.

Nous adoptons une méthodologie qui implique la collecte et l'analyse de dialogues réels, puis la collecte et l'analyse de dialogues obtenus par le biais d'une expérimentation en magicien d'Oz. Notre méthodologie de génération des attentes à partir du corpus est la suivante. En premier lieu, l'enregistrement des dialogues permet de récupérer le corpus. L'annotation en actes nous permet dans un troisième temps de récupérer les successions d'actes afin d'effectuer un calcul des occurrences de chaque succession et de les stocker dans un fichier. En procédant ainsi, nous pouvons élaborer automatiquement les règles d'association acte courant / attentes, mais aussi, par leur occurrence dans le corpus, les probabilités liées à ces règles.

Nous avons alors commencé à étudier l'apport des attentes du locuteur dans le cadre de l'évaluation de systèmes de dialogue homme-machine avec la question sous-jacente : « Le système répond-il aux attentes de l'utilisateur ». Nous posons trois types de questions : sur la compréhension (ce que comprend l'utilisateur de la situation ; compétences perçues du système), les attendus (ce à quoi s'attend l'utilisateur ; compétences supposées du système) et les attentes (ce qu'aimerait l'utilisateur ; compétences voulues pour le système).

Enfin, le système utilise actuellement une génération à l'aide d'un certain nombre de phrases à trous issues de la première analyse de dialogues en situation homme-homme. Nous allons essayer d'appliquer notre théorie des attentes dans le cadre de la génération des énoncés.

Ainsi nous pourrions précisément proposer un modèle générique qui prenne en compte le locuteur et la tâche en cours, ce qui rend les perspectives d'applications infinies.

## Références bibliographiques

- [Allen & Litman 87] Allen J.F. & Litman D.J., *A Plan Recognition Model For Subdialogues In Conversations*. Cognitive Science 11, 2, 1987.
- [Allen & Perrault 80] Allen J.F. & Perrault C.R., *Analyzing Intention In Utterances*. Artificial Intelligence, N°15, 1980.
- [Asher 93] Asher N., *Reference to Abstract Objects in Discourse*. Kluwer Academic Publisher, 1993.
- [Asher 00] Asher N., *Discourse Structure and the Logic of Conversation*. 2000.
- [Asher & Lascarides 98] Asher N. & Lascarides A., *Questions in Dialogue*. Linguistics and Philosophy, 1998, 21:237-309.
- [Asher & Lascarides 03] Asher N. & Lascarides A., *Logics of Conversation*. Cambridge University Press, 2003.
- [Austin 62] Austin J.L., *How To Do Things With Words*. Oxford U.P., 1962.
- [Austin 70] Austin J.L., *Quand dire c'est faire*. Edition Du Seuil, Paris 1970.

- [Bagein & al.00] Bagein M., Dutoit T., Malfrere F., Pagel V., Ruelle A., Tounsi N., Wynsberghe D., *EULER, An Open, Generic, Multi-lingual and Multi-platform Text-to-Speech System*. Proceedings from LREC 2000, Athens, pp.563-566, 2000.
- [Bange 92] Bange P., *Analyse conversationnelle et théorie de l'action*. Hatier, Paris, 1992.
- [Besacier & al. 01] Besacier L., Blanchon H., Fouquet Y., Guilbaud J.P., Helme S., Mazenot S., Moraru D., Vaufreydaz D., *Speech translation for French in the Nespole! European project*. Eurospeech, 2001.
- [Bilange 92] Bilange E., *Dialogue personne-machine : modélisation et réalisation informatique*. Paris, Hermès, 1992.
- [Boitet & al. 98] Boitet C., Caelen J., Courtin J., Coutaz J., *Traduction de parole*. In projet C-Star, 1998.
- [Brassac & Trognon 92] Brassac C., Trognon A., *L'enchaînement conversationnel*. Cahiers de linguistique française n°13, 1992, p. 76-107
- [Burger 97] Burger S., Morel C., Coletti P., Besacier L., Metze F., *Nespole! Data Collection*. Status Report, January 2001.
- [Caelen 92] Caelen J., *Compte-rendu du « workshop » Interfaces homme-machine multimodales*. GdR-PRC CHM sur le dialogue, Dourdan, 1992.
- [Caelen 95] Caelen J., *Vers une logique dialogique*. Séminaire international de pragmatique, Jérusalem, 1995.
- [Caelen 96a] Caelen J., *Reconnaître et comprendre la parole*. In Nouvelles interfaces homme-machine, Observatoire français des techniques avancées, Ofta, Paris, décembre 1996.
- [Caelen 96b] Caelen J., *Reconnaître la parole c'est la comprendre*. In Numéro Spécial de la revue *La recherche* N°285, "L'ordinateur au doigt et à l'œil", p. 62-65, mars 1996.
- [Caelen 97] Caelen J., *Interaction verbale*, Editions CEPADUES, 1997.
- [Caelen & Nguyen 04] Caelen J. & Nguyen H., *Gestion de buts de dialogue*. Actes de TALN'2004, Fès, avril 2004.
- [Caelen & Villaseñor 97] Caelen J. & Villaseñor L., *Dialogue homme-machine et apprentissage. Apprentissage par l'interaction*. (ed.) Khaldoun Zreik, p 83-117, Europa Productions, Paris, 1997.
- [Carberry 90] Carberry S., *Plan Recognition In Natural Language*. MIT Press, Cambridge, Mass., 1990.

- [Chevallier 92] Chevallier R., *Mise en œuvre d'un modèle dynamique de dialogue dans un tuteur intelligent*. Thèse de l'université du Mans, 1992.
- [Chicoisne & Pesty 99] Chicoisne G. & Pesty S., *Modèle de conversation & agents rationnels socialement corrects*. In Actes de l'atelier la langue dans l'interaction personnes-machines, TALN'99, p. 91-104, Cargèse, 12-17 juillet 1999.
- [Clarkson & Rosenfeld 97] Clarkson P. & Rosenfeld R., *Statistical Language Modeling using the CMU-Cambridge Toolkit*. Eurospeech, 1997.
- [Cohen & Perrault 79] Cohen P.R. & Perrault C.R., *Elements Of A Plan-Based Theory Of Speech Acts*. Cognitive Science, 3(3) : p 177-212, 1979.
- [Cohen & Levesque 90a] Cohen P.R. & Levesque H.J., *Persistence, Intention And Commitment*. In Intentions in Communication, P.R. Cohen, J. Morgan & M.E. Pollack,(Eds.) The MIT Press, 1990.
- [Cohen & Levesque 90b] Cohen P.R. & Levesque H.J., *Rational Interaction As The Basis For Communication*. In Intentions in Communication, P.R. Cohen, J. Morgan & M.E. Pollack,(Eds.) The MIT Press, 1990.
- [Colineau 97] Colineau N., *Etude des marqueurs discursifs dans le dialogue finalisé*. Thèse de doctorat de sciences cognitives, Université Grenoble I, 1997.
- [Dupont 96] Dupont M., *Le modèle des attentes du lecteur dans le calcul de la référence*. Actes de RéciTAL'96, Courcelle, 25-27 septembre 1996, p. 155-160.
- [El-Bèze 93] El-Bèze M., *Les modèles de langage probabilistes: quelques domaines d'application*. Habilitation à diriger des recherches, LIPN (PARIS XIII), janvier 1993, 85p.
- [Falzon 89] Falzon P., *Ergonomie cognitive du dialogue*. Presses Universitaires de Grenoble, 1989.
- [Foltz 96] Foltz P., *Latent Semantic Analysis for text-based research*. Behavior Research Methods, Instruments & Computers, 28(2): p. 197-202, 1996.
- [Fouquet 01] Fouquet Y., *De l'étude de dialogues oraux dans une langue non maternelle*. Actes de Rencontre Jeunes chercheurs en Parole, Mons, septembre 2001.
- [Fouquet 02] Fouquet Y., *Un modèle de dialogue par les attentes du locuteur*, Actes de TALN'2002, Nancy, juin 2002.

- [Fouquet 03] Fouquet Y., *Le magicien d'Oz pour du dialogue oral : expérience avec un assistant virtuel en entreprise*. Actes de Rencontre Jeunes chercheurs en Parole, Grenoble, septembre 2003.
- [Fouquet 04] Fouquet Y., *Prédiction d'actes et attentes en dialogue : expérience avec un assistant virtuel simulé*. Actes de TALN'2004, p. 193-198, Fès, avril 2004.
- [Frechet 92] Frechet A.-L., *Analyse linguistique d'un corpus de dialogue oral homme-machine*. Thèse de doctorat linguistique, Université Paris 3 – La Sorbonne Nouvelle, 1992.
- [Gasquet & Herzig 96] Gasquet O. & Herzig A., *From classical to normal modal logics*. In Heinrich Wansing editor, *Proof Theory of Modal Logics*, number 2 in Applied Logic Series, p. 293-311. Kluwer Academic Publishers, 1996.
- [Goffman 73] Goffman E., *La mise en scène de la vie quotidienne*. Éditions de Minuit, Paris, 1973.
- [Goffman 74] Goffman E., *Les rites d'interaction*. Éditions de Minuit, Paris, 1974.
- [Grau et al. 94] Grau B., Sabah G., Vilnat A., *Pragmatique et dialogue homme-machine*. *Technique et science informatique*, Vol. 13, n°1, p. 9-30, 1994.
- [Grice 68] Grice H.P., *Utterer's Meaning, Sentence-Meaning, and Word-Meaning*. *Foundations of Language*, (4):1-18, 1968.
- [Grice 75] Grice H.P., *Logic And Conversation*. in Cole P. and Morgan J.L., *Syntax and Semantic*, vol. 3, *Speech Acts*. Academic Press, p. 41-58, 1975.
- [Grice 79] Grice H.P., *Logique et conversation*. *Communications*, N° Spécial 30, Paris, Seuil, juin 1979, 57-72.
- [Grimaldi 93] Grimaldi L., *Dialogue pédagogique en langue naturelle dans un système intelligent de communication de connaissances : Application à la résolution d'exercices au jeu d'échecs*. Thèse de doctorat d'informatique, Université Paris 6, 1993.
- [Grosz & Sidner 90] Grosz B.J. & Sidner C.L., *Plans Of Discourse*. In *Intentions in Communication*, P.R. Cohen, J. Morgan & M.E. Pollack.(Eds.) The MIT Press, 1990, Pp 417-444.
- [Halpern & Moses 92] Halpern J.Y. & Moses Y., *A guide to completeness and complexity for modal logics og knowledge and belief*. *Artificial Intelligence*, 54, 319-379, Elsevier Science Publishers, 1992.



- [Herzig & Longin 02 ] Herzig A., Longin D., *Intention et principes de coopération pour le traitement des requêtes et des questions fermées au travers des assertifs*. In Proc. 13<sup>ème</sup> Congrès Francophone AFRIF-AIFA de Reconnaissance des Formes et Intelligence Artificielle (RFIA'02), Angers, pages 221-230. AFRIF-AIFA, 8-10 janvier 2002.
- [Kerbrat-Orecchioni 80] Kerbrat-Orecchioni C., *L'énonciation*. Armand Colin éd., Paris, 1980, p19 1ère édition.
- [Kerbrat-Orecchioni 90] Kerbrat-Orecchioni C., *Les interactions verbales*. Tome 1, Armand Colin éd., Paris, 1990.
- [Landauer & Dumais 97] Landauer T. & Dumais S., *A Solution to Plato's Problem : The Latent Semantic Analysis Theory of Acquisition, Induction and Representation of Knowledge*. Psychological Review, 1997, 104(2) p. 211-240.
- [Lehuen 97] Lehuen J., *Un modèle de dialogue dynamique et générique intégrant l'acquisition de sa compétence. Le système Coala*. Thèse de doctorat, Université de Caen, juin 1997.
- [Lemeunier 98] Lemeunier T., *AMI : un système de DHM expérimental qui apprend à s'adapter aux situations interactionnelles itératives*. Actes de RéciTAL'98, Le Mans, septembre 1998.
- [Lemeunier 99] Lemeunier T., *La modélisation des attentes dans le système AMI*. Actes du colloque Récital'99, Cargèse (Corse), pp. 439-443, 1999.
- [Lemeunier 00] Lemeunier T., *L'intentionnalité communicative dans le dialogue homme-machine en langue naturelle*. Thèse de doctorat informatique, Université du Maine, Le Mans, 1<sup>er</sup> décembre 2000.
- [Lemeunier & Lehuen 99] Lemeunier T. & Lehuen J., *Un modèle de génération des intentions de communication pour le dialogue homme/machine*. Actes de TALN'99, Atelier "La langue naturelle dans l'interaction personne-machine", Cargèse, Corse, p. 12-17, juillet 1999.
- [Litman 85] Litman D.J., *Plan Recognition And Discourse Analysis : An Integrated Approach For Understanding Dialogues*. Thèse de l'université de Rochester, 1985.
- [Litman & Allen 87] Litman D. & Allen J.F., *A Plan Recognition Model For Subdialogues In Conversations*. Cognitive Science. n°11, p. 163-200, 1987.

- [Litman & Allen 90] Litman D. & Allen J.F., *Discourse Processing And Commonsense Plans*. In *Intentions in Communication*, P.R. Cohen, J. Morgan & M.E. Pollack, (Eds.) The MIT Press, 1990, P. 365-388.
- [Luzzati 89] Luzzati D., *Recherches sur le dialogue homme-machine : Modèles linguistiques et traitements automatiques*. Thèse de doctorat, Université de la Sorbonne, 1989.
- [Luzzati 95] Luzzati D., *Le dialogue verbal homme-machine, étude de cas*. Collection sciences cognitives, Masson, 1995.
- [Champagne et al. 04] Champagne M., Herzig A., Longin D., Nespoulous J-L., Virbel J., *Indirect speech acts: A multidisciplinary approach*. In Jean Caelen, Denis Vernant and Daniel Vanderveken editors, *Dialogue and Logic*, Kluwer Academic Press, 2004. to appear.
- [Maybury 93] Maybury M.T., *Intelligent Multimedia Interfaces*. AAAI/MIT Press, USA, 1993.
- [Minker 96] Minker W., *Compréhension automatique de la parole spontanée*, 2-7384-7745-3, 1996.
- [Moeschler 89] Moeschler J., *Marques linguistiques, interprétations pragmatiques et conversation*. In *Cahiers de linguistique française n°10*, p.43-76, 1989.
- [Moeschler 91] Moeschler J., *L'analyse pragmatique des conversations*. In *Cahiers de linguistique française n°12*, p.7-31, 1991.
- [Nerzic 93] Nerzic P., *Erreurs et échecs dans le dialogue oral homme-machine : Détection et réparation*. Thèse de doctorat, Université de Rennes, 1993.
- [Nguyen 03a] Nguyen H., *Vers une architecture générique de système de dialogue oral homme-machine*. Actes de Récital 2003, Batz-sur-Mer, 11-14 juin 2003.
- [Nguyen 03b] Nguyen H., *Generic manager for spoken dialog systems*. Proceedings of Diabruck 2003, Saarbrücken, 4-6 sept 2003.
- [Nicolle & Luzzati 99] Nicolle A., Luzzati D., *La langue dans l'interaction personnes/machines*. Atelier thématique de TALN'99, Cargèse, Corse, 1999.
- [Ozkan 94] Ozkan N., *Analyse communicationnelle de dialogues finalisés*. Thèse de 3ème cycle à l'Institut Polytechnique de Grenoble, 1994.
- [Piaget 64] Piaget J., *Development and Learning, Piaget Rediscovered*. Ripple R.E. and Roccastle V.N. ed., School of Education, Cornell University, Ithaca, New York, 1964.

- [Pierrel 87] Pierrel J.M., *Dialogue oral homme-machine*. Hermès, Paris, 1987.
- [Pierrel & Romary 97] Pierrel J.M. & Romary L., *Quelles références dans les dialogues homme-machine ?*. In Sabah G., Vivier J., Vilnat A., Pierrel J.M., Romary L., Nicolle A., Machine, langage et dialogue, figures de l'interaction, L'Harmattan, 1997.
- [Prendinger & Schurz 96] Prendinger H. & Schurz G., *Reasoning about Action and Change, A Dynamic Logic Approach*. Journal of Logic, Language, and Information, 5:209-245, 1996.
- [Rosenfeld 94] Rosenfeld R., *Adaptative Statistical Language Modeling : A Maximum Entropy Approach*. Thesis for the degree of Doctor of Philosophy, Carnegie Mellon University, 1994.
- [Rosenfeld 96] Rosenfeld R., *A Maximum Entropy Approach to Adaptive Statistical Language Modeling*. Computer, Speech and Language, 1996.
- [Rouillard & Caelen 98] Rouillard J. & Caelen J., *Etude du dialogue homme-machine en langue naturelle sur le web pour une recherche documentaire*. Deuxième colloque international sur l'apprentissage personne-système, Caps'98, Caen, juillet 98.
- [Rouillard 99] Rouillard J., *Navigation versus dialogue sur le web, une étude des préférences*. IHM'99, Montpellier, 1999.
- [Rouillard 00] Rouillard J., *Hyperdialogue sur internet ; le système Halpin*. Thèse de doctorat informatique, Université de Grenoble I, 2000.
- [Roulet & al. 85] Roulet E., Auchlin A., Moeshler J., Rubattel C., Schelling M., *L'articulation du discours en français contemporain*. Peter Lang, Berne, 1985.
- [Sabah 89] Sabah G., *L'intelligence artificielle et le langage, processus de compréhension*. vol. 2, Hermès, Paris, 1989.
- [Sabah & al. 97] Sabah G., Vivier J., Vilnat A., Pierrel J.M., Romary L., Nicolle A., *Machine, langage et dialogue, figures de l'interaction*. L'Harmattan, 1997.
- [Sabah 97a] Sabah G., *La langue et la communication homme-machine, état et l'avenir*. In Sabah G., Vivier J., Vilnat A., Pierrel J.M., Romary L., Nicolle A., Machine, langage et dialogue, figures de l'interaction, L'Harmattan, 1997.
- [Sabah 97b] Sabah G., *Apprentissage et traitement automatique des langues*. dans Apprentissage par l'interaction, K. Zreik, Europa Productions, Paris, 1997.

- [Sadek 96] Sadek D., *Le dialogue homme-machine : de l'ergonomie des interfaces à l'agent intelligent dialoguant*. In Nouvelles interfaces homme-machine, Série ARAGO, Observatoire français des technologies avancées, Paris, 1996.
- [Searle 72] Searle J.R., *Les actes de langage : essai de philosophie du langage* (trad. française par H. Pauchard). Paris, Hermann, 1972.
- [Shanks & Abelson 77] Shanks R. & Abelson R., *Scripts plans goals and understanding*, 1997.
- [Siroux & al. 89] Siroux J., Gilloux M., Guyomard M., Sorin C., *Le dialogue homme-machine en langue naturelle : un défi ?*. Annales télécommunication, 44, n° 1-2, 1989.
- [Sperber & Wilson 89] Sperber D. & Wilson D., *La pertinence, communication et cognition*. Les Éditions de Minuit, Paris, 1989.
- [Stein & Maier 95] Stein A. & Maier E., *Structuring collaborative information-seeking dialogues, knowledge-based systems*. Special issue on human-computer collaboration, vol.8, Elsevier Science B.V., March 1995.
- [Suchman 87] Suchman L.S., *Plans and situated actions*. Cambridge university press, 1987.
- [Trognon 95] Trognon A., *Structures interlocutoires*. Les cahiers de linguistique française, 17, p. 79-98, 1995.
- [Van der Hoek & al. 94a] Van der Hoek W., Van Linder B. & Meyer Ch. J.-J., *A logic of capabilities*. Proceedings of the third international symposium of logic foundations of computer science. LFCS, 94. (eds.) A. Nerode & Yu. V. Matiyasevich., St. Petersburg, july 11-14, Springer-Verlag, 1994.
- [Van der Hoek & al. 94b] Van der Hoek W., Van Linder B. & Meyer Ch. J.-J., *Unraveling nondeterminism : on having the ability to choose*. Proceedings of the sixth international conference on artificial intelligence : methodology, systems, applications, AIMS'94. (eds.) P. Jorrand & V. Sgurev., Sofia, septembre 21-24, World Scientific, 1994.
- [Van der Hoek & al. 94c] Van Linder B., Van der Hoek W. & Meyer Ch. J.-J., *Communicating rational agents*. Proceedings of the 18<sup>th</sup> german annual conference on artificial intelligence, KI-94 : Advances in artificial intelligence (eds.), B. Nebel & L. Dreschler-Fischer, Saarbrücken, septembre 18-23, Springer-Verlag, 1994.
- [Vanderveken 88] Vanderveken D., *Les actes de discours*. Pierre Margada éd. Bruxelles, 1988.

- [Vanderveken 90] Vanderveken D., *La logique illocutoire*. Mandarga éd. Bruxelles, 1990.
- [Vernant 92] Vernant D., *Approche actionnelle et modèle projectif du dialogue informatif*. In *Du dialogue, recherches sur la philosophie du langage*, Vrin, Paris, n° 14, p. 295-314, 1992.
- [Vernant 94] Vernant D., *Dialectique, forme dialogale et dialogique*. Actes du colloque *Le dialogique, sur les formes philosophiques, littéraires, linguistiques et cognitives du dialogue*, Le Mans, 1994.
- [Vernant 97a] Vernant D., *Du discours à l'action : études pragmatiques*. Presses Universitaires de France, Paris, 1997.
- [Vernant 97b] Vernant D., *Notes de cours de Pragmatique du dialogue*, novembre 1997.
- [Villaseñor-Pineda 99] Villaseñor-Pineda L., *Contribution à l'apprentissage dans le dialogue homme-machine*. Thèse de doctorat informatique, Université de Grenoble I, 1999.
- [Vilnat 97] Vilnat A., *Quels processus pour les dialogues homme-machine ?*. In Sabah G., Vivier J., Vilnat A., Pierrel J.M., Romary L., Nicolle A., *Machine, langage et dialogue, figures de l'interaction*, L'Harmattan, Paris, 1997.
- [Vilnat & Nicaud 92] Vilnat A. & Nicaud L., *Un système de dialogue homme-machine : STANDIA*. Actes du séminaire *Dialogue du GdR-PRC Communication homme-machine*, Dourdan, 1992.
- [Vivier & Nicolle 97] Vivier J. & Nicolle A., *Questions de méthode en dialogue homme-machine : l'expérience Compèrobot*. In Sabah G., Vivier J., Vilnat A., Pierrel J.M., Romary L., Nicolle A., *Machine, langage et dialogue, figures de l'interaction*, L'Harmattan, 1997.
- [Weizenbaum 66] Weizenbaum J., *Eliza - A Computer Program For The Study Of The Natural Language Communication Between Man And Machine*. CACM n°9, 1966.
- [Wittgenstein 59] Wittgenstein L., *Les jeux de langage*. 1959.
- [Wittgenstein 58] Wittgenstein L., *Philosophical investigations*. Oxford, Basil Blackwell ed., 1958.
- [Woszczyna & al. 93] Woszczyna M., Coccaro N., Eisele A., Lavie A., McNair A., Polzin T., Rogina I., Rose C., Sloboda T., Tomita M., Tsutsumi J., Aoki-Waibel N., Waibel A., Ward W., *Recent Advances in JANUS : A Speech Translation System*. Eurospeech, 1993, vol. 2, p. 1295-1298.

- [Lamel & al. 91] Lamel L.F., Gauvain J.L., Eskénazi M., *BREF, a Large Vocabulary Spoken Corpus for French*. Eurospeech, Gênes, Italy, vol 2, p. 505-508, 24-26 september 1991.
- [Xuereb & Caelen 04] Xuereb A., Caelen J., *Un modèle d'interprétation pragmatique en dialogue homme-machine basé sur la SDRT*. Workshop SDRT, TALN 2004, Fès, 19-21 avril 2004.
- [Xuereb 04] Xuereb A., *Pragmatique du dialogue homme-machine*. rapport de Master informatique, Université Joseph Fourier, Grenoble, 2004.

## Références non citées

### Analyse de parole :

**Lopez P., Fay-Varnier C., Roussalany A.,** *Sous-langages d'application et LTAG : le système EGAL*. TALN 99.

**Roussel D.,** *Intégration de prédictions linguistiques issues d'applications à partir d'une grammaire d'arbres hors-contexte. Contribution à l'analyse de la parole*. Thèse de sciences cognitives, Grenoble, 1999.

### Compréhension de parole :

**Antoine J.-Y.,** *Coopération syntaxe-sémantique pour la compréhension de la parole spontanée*. Thèse de doctorat informatique, Grenoble, 1994

### Dialogue :

**Colineau N.,** *Vers une compréhension des actes de discours*. Rapport de DEA Sciences Cognitives, Grenoble, juin 1994.

**Faure R., Herzig A., Longin D.,** *Actes indirects et évolution des croyances*. In Actes des 5èmes journées nationales du PRC-GDR Intelligence Artificielle, september 2000. Poster Session.

**Ghiglione R. & Trognon A.,** *Où va la pragmatique? : de la pragmatique à la psychologie sociale*.

Grenoble, PUG , 1993 . - Collection : Vies sociales ; 7

**Gouiaa M., Delisle S., Moulin B.,** *analyse sémantique de dialogues oraux transcrits*. TALN 99.

**Herzig A., Longin D., Virbel J.,** *Towards an analysis of dialogue acts and indirect speech acts in a BDI framework*. In Massimo Poesio and David Traum, editors, Proc. 4th Workshop on the Semantics and Pragmatics of Dialogue (GÖTALOG 2000), june 2000.

**Herzig A. & Longin D.,** *C&L intention revisited*. In Didier Dubois and Chris Welty editors, Proc. KR2004, 2004.

**Hollard S.**, *Analyse de dialogues finalisés dans le contexte de C-STAR*. Journées Thématiques d'Archamps Informatique et Linguistique, Archamps, 1998.

**Hollard S. & Caelen J.**, *Dialogues finalisés et organisation des connaissances en mémoire des locuteurs*. Coling, 1998.

**Imberdis L.**, *Génération d'actes de dialogue*. rapport de DEA informatique, juin 1995.

**Lehuen J., Nicolle A., Luzzati D.**, *Un modèle hypothético-expérimental dynamique pour la gestion des dialogues homme-machine*. RFIA, Rennes, éditeur Hermès Paris, 1996.

**Lehuen J. & Luzzati D.**, *Acquisition coopérative d'une compétence langagière interprétative en dialogue homme machine*. TALN 99, 1999.

**Lemeunier T.**, *L'usage de l'humour en informatique*. Rapport de DEA, Le Mans, 1996.

**Nicolle A. & Vivier J.**, *Dialogue et apprentissage : humain/humain, humain/machine, machine/machine*. CAPS'96, Caen, juillet 1996.

**Rouillard J.**, *Hyperdialogue homme-machine sur le World Wide Web : le système HALPIN*. ERGOIA'98, Biarritz, novembre 1998.

**Rouillard J.**, *Contribution à l'étude du dialogue Homme-Machine à travers le Web : la personnalisation*. RECITAL'98, Le Mans, septembre 1998.

**Vivier J.**, *Psychologie du dialogue homme machine en langage naturel*. ISBN 2-909285-06-5, europia 1996.

**Wlodarczyk A.**, *La validation informative des énoncés et la quantification linguistique*. 1999.

**Wlodarczyk A.**, *Sur les homotopies du topique et du focus*. 2000.

Traduction de parole :

**Boitet C., Caelen J., Fafiotte G., Keller E., Lafourcade M., Wehrli E.**, *Integrating French within C-STAR II*. Second report & demos of the CLIPS++ group., 1998, CLIPS, IMAG.

Site officiel de C-Star : <http://www.c-star.org>

Site officiel de NESPOLE ! : <http://nespole.itc.it/>





## **Annexes**



## Statistique de cooccurrence des bi-grammes {Acte, Acte-suivant}

Réels HH de PVE : 29 dialogues - 4548 actes

	Af(p)	Afd(p)	Aff(p)	Affd(p)	Affp(p)	Affs(p)	Afp(p)	Afs(p)	Cf(p)	Cfd(p)	Cff(p)	Cffd(p)	Cffp(p)	Cffs(p)	Cfp(p)	Cfs(p)	fin	Total
Af(p)	0	0	9	1	0	1	1	9	2	1	2	0	0	8	0	53	11	87
Afd(p)	1	0	0	0	0	1	0	17	0	0	1	3	0	8	4	64	0	72
Aff(p)	3	0	0	7	0	1	1	19	52	0	0	1	0	0	0	14	0	69
Affd(p)	0	0	0	0	0	0	0	3	1	86	0	0	0	3	0	7	0	97
Affp(p)																		0
Affs(p)	0	0	0	0	0	3	3	4	0	1	1	0	0	4	1	84	0	165
Afp(p)	1	0	0	1	0	1	2	7	8	1	1	1	0	5	25	47	1	173
Afs(p)	2	0	1	1	0	3	2	7	1	2	1	2	0	10	1	68	0	1563
Cf(p)	9	0	8	15	0	2	5	32	0	2	0	0	0	0	2	5	23	66
Cfd(p)	4	0	2	5	0	11	2	60	0	0	0	1	0	2	2	11	0	125
Cff(p)	68	0	0	0	0	8	0	16	0	0	0	0	0	0	0	8	0	25
Cffd(p)	0	89	0	0	0	0	2	2	0	0	0	0	0	2	2	0	2	47
Cffp(p)	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	1
Cffs(p)	0	0	0	3	0	7	2	77	0	0	0	1	0	6	1	2	0	325
Cfp(p)	10	2	0	2	0	1	42	19	0	0	0	2	0	12	2	6	2	109
Cfs(p)	1	2	2	3	0	4	3	67	0	0	1	1	0	6	1	9	0	1536
debut	0	0	0	0	0	0	59	2	0	0	0	2	0	3	31	2	0	59
	2	2	2	2	0	4	4	34	1	3	1	1	0	7	2	34	1	4548

Tests de PVE : 86 dialogues - 1604 actes

	Af(p)	Afd(p)	Aff(p)	Affd(p)	Affp(p)	Affs(p)	Afp(p)	Afs(p)	Cf(p)	Cfd(p)	Cff(p)	Cffd(p)	Cffs(p)	Cfp(p)	Cfs(p)	fin	Total
Af(p)	2	0	0	0	0	13	2	5	13	0	6	0	15	13	9	22	104
Afd(p)	0	25	0	0	0	0	0	0	0	0	0	0	25	0	50	0	4
Aff(p)	1	0	0	3	0	4	3	6	45	0	1	0	21	1	13	3	78
Affd(p)	2	0	0	0	0	4	0	0	7	67	4	0	11	2	2	0	46
Affp(p)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	1
Affs(p)	0	0	0	0	0	0	0	0	1	0	2	0	20	2	72	1	250
Afp(p)	1	0	0	0	0	0	1	0	9	0	2	0	69	1	13	3	98
Afs(p)	1	0	1	0	0	2	0	2	8	0	16	1	39	8	20	2	165
Cf(p)	12	0	5	1	0	18	4	10	2	0	0	0	0	0	2	45	82
Cfd(p)	34	0	19	0	3	13	13	13	0	0	0	0	3	0	0	3	32
Cff(p)	33	0	2	7	0	24	0	26	2	0	0	0	0	0	0	5	42
Cffd(p)	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Cffs(p)	9	0	13	6	0	38	1	26	1	0	0	0	3	0	3	0	233
Cfp(p)	8	0	14	3	0	31	8	14	3	0	0	0	0	0	3	17	36
Cfs(p)	15	0	11	9	0	35	6	18	0	0	0	0	2	0	1	2	262
debut	0	0	0	0	0	6	75	18	0	0	0	0	0	1	0	0	85
Total	6	0	5	3	0	16	6	10	5	2	3	0	15	2	16	5	1604

Réels HH + Pré-tests de PVE : 40 dialogues - 1184 actes

	Af(p)	Afd(p)	Aff(p)	Affd(p)	Affs(p)	Afp(p)	Afs(p)	Cf(p)	Cfd(p)	Cff(p)	Cffd(p)	Cffs(p)	Cfp(p)	Cfs(p)	fin	Total	
Af(p)	0	0	0	5	5	5	5	0	0	0	0	5	5	5	63	19	
Afd(p)	0	0	0	0	3	6	0	0	0	0	6	11	17	57	0	35	
Aff(p)	0	0	0	7	0	20	0	67	0	0	0	0	0	7	0	15	
Affd(p)	0	0	0	0	0	0	0	0	91	0	0	0	4	4	0	23	
Affs(p)	0	0	0	0	0	0	2	0	0	3	5	5	8	77	0	60	
Afp(p)	1	0	1	0	1	0	1	13	0	0	4	9	23	46	0	99	
Afs(p)	0	0	1	2	5	2	3	0	3	1	4	5	11	60	1	282	
Cf(p)	0	0	0	0	4	0	16	0	0	0	4	8	4	4	60	25	
Cfd(p)	3	0	3	3	3	7	53	0	0	3	0	3	7	10	3	30	
Cff(p)	33	0	0	11	22	0	11	0	0	0	0	0	22	0	0	9	
Cffd(p)	0	94	0	0	3	0	3	0	0	0	0	0	0	0	0	32	
Cffs(p)	0	0	3	2	5	5	75	0	0	2	0	0	6	2	2	63	
Cfp(p)	14	0	0	2	1	25	21	1	0	0	9	8	0	18	0	87	
Cfs(p)	1	2	2	3	6	14	55	0	0	1	1	7	3	5	0	323	
debut	0	0	2	0	33	31	12	0	0	0	0	0	0	0	2	17	42
Total	2	3	1	2	5	8	24	2	3	1	3	5	7	27	3	1184	

Total PVE : 113 dialogues - 5994 actes

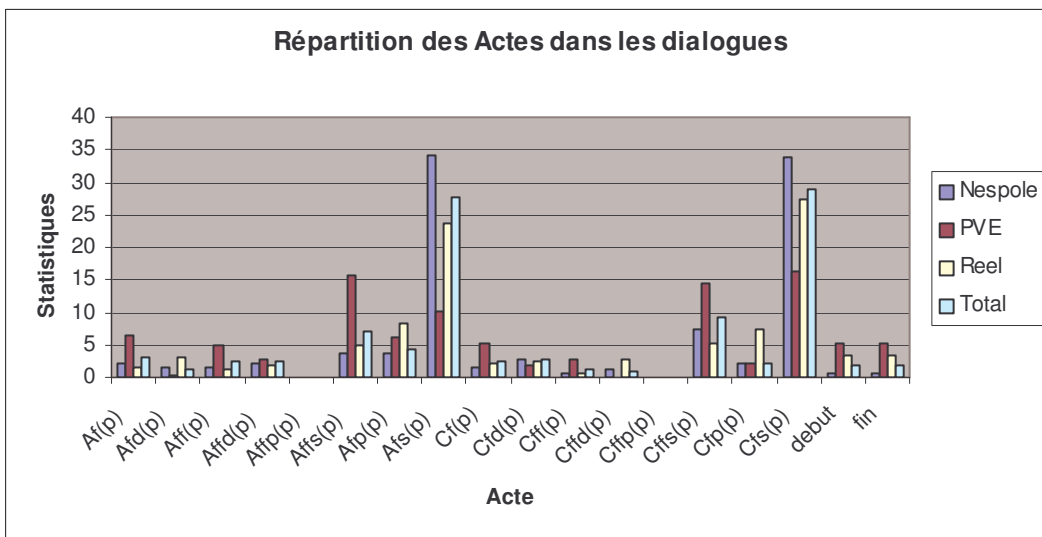
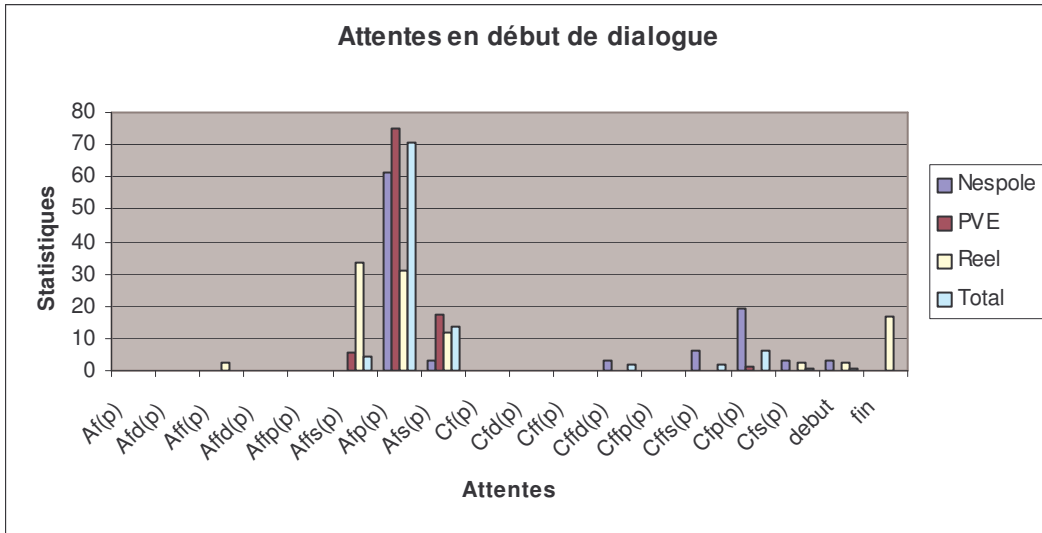
	Af(p)	Afd(p)	Aff(p)	Affd(p)	Affp(p)	Affs(p)	Afp(p)	Afs(p)	Cf(p)	Cfd(p)	Cff(p)	Cffd(p)	Cffs(p)	Cfp(p)	Cfs(p)	fin	Total
Af(p)	1	0	4	1	0	8	2	7	8	1	4	0	12	7	29	17	192
Afd(p)	1	1	0	0	0	1	0	16	0	0	1	3	10	4	62	0	79
Aff(p)	2	0	0	5	0	3	2	12	49	0	1	1	11	1	13	1	149
Affd(p)	1	0	0	0	0	1	0	2	3	80	1	0	6	1	6	0	145
Affp(p)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	1
Affs(p)	0	0	0	0	0	1	1	2	1	0	1	0	13	2	77	1	419
Afp(p)	1	0	0	0	0	1	1	4	9	1	1	1	30	19	31	2	255
Afs(p)	2	0	1	1	0	3	2	6	1	2	2	2	13	2	63	0	1656
Cf(p)	11	0	6	9	0	11	4	19	1	1	0	0	0	1	33	5	150
Cfd(p)	10	0	6	4	1	11	4	50	0	0	0	1	2	1	9	1	159
Cff(p)	46	0	1	4	0	18	0	22	1	0	0	0	0	0	3	3	67
Cffd(p)	0	88	0	0	0	0	4	2	0	0	0	0	2	2	0	2	52
Cffp(p)	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	1
Cffs(p)	4	0	6	5	0	20	2	56	0	0	0	0	5	1	2	0	562
Cfp(p)	10	2	4	2	0	9	28	20	1	0	0	1	10	1	6	6	136
Cfs(p)	3	1	4	4	0	9	4	58	0	0	0	1	6	1	7	0	1741
debut	0	0	0	0	0	4	71	14	0	0	0	2	2	6	1	0	117
Total	3	1	2	2	0	7	4	28	3	3	1	1	9	2	29	2	5994

## Statistiques sur les attentes

PVE réel : 29 dialogues - 4548actes

	Af(p)	Afd(p)	Aff(p)	Affd(p)	Affp(p)	Affs(p)	Afp(p)	Afs(p)	Cf(p)	Cfd(p)	Cff(p)	Cffd(p)	Cffp(p)	Cffs(p)	Cfp(p)	Cfs(p)	Total
Af(p)									2	1	2	0	0	8	0	53	87
Afd(p)									0	0	1	3	0	8	4	64	72
Aff(p)									52	0	0	1	0	0	0	14	69
Affd(p)									1	86	0	0	0	3	0	7	97
Affp(p)																	0
Affs(p)									0	1	1	0	0	4	1	84	165
Afp(p)									8	1	1	1	0	5	25	47	173
Afs(p)									1	2	1	2	0	10	1	68	1563
Cf(p)	9	0	8	15	0	2	5	32									66
Cfd(p)	4	0	2	5	0	11	2	60									125
Cff(p)	68	0	0	0	0	8	0	16									25
Cffd(p)	0	89	0	0	0	0	2	2									47
Cffp(p)																	1
Cffs(p)	0	0	0	3	0	7	2	77									325
Cfp(p)	10	2	0	2	0	1	42	19									109
Cfs(p)	1	2	2	3	0	4	3	67									1536
debut	0	0	0	0	0	0	59	2	0	0	0	2	0	3	31	2	59
fin																	29
	2	2	2	2	0	4	4	34	1	3	1	1	0	7	2	34	4548

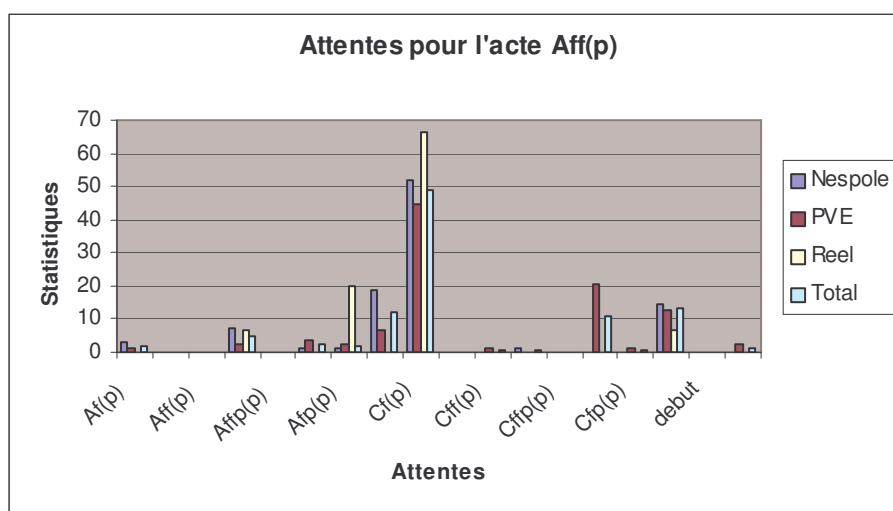
## Analyse des résultats d'attentes acte par acte pour les différents corpus



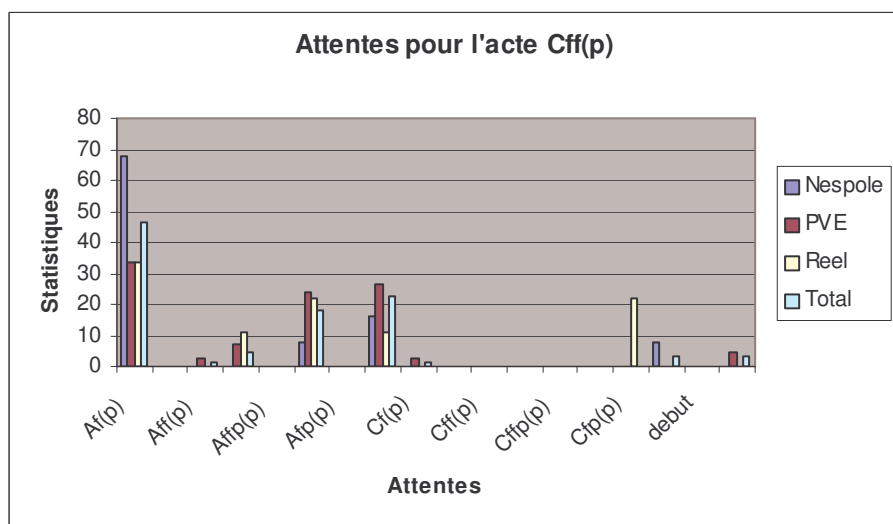
## 1. Actes posant des attentes

### Faire-faire

Côté Agent :

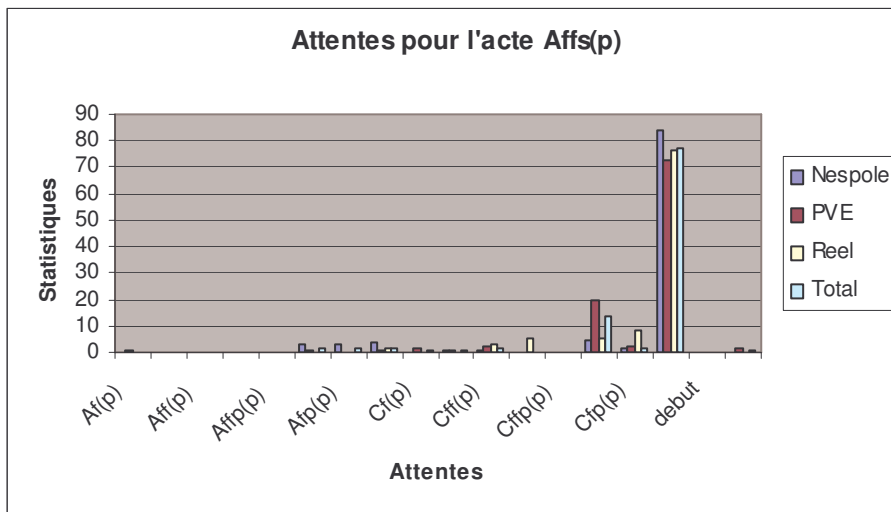


Côté client :

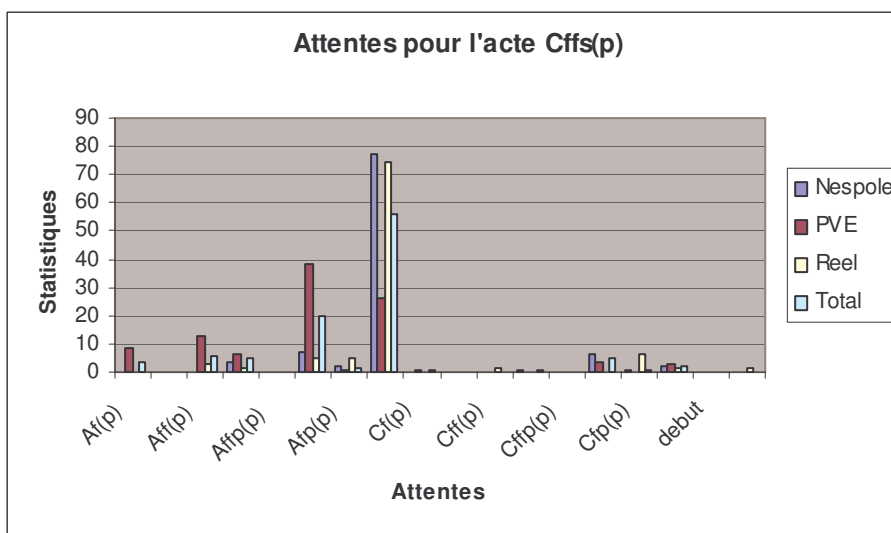


**Faire-faire-savoir**

Côté agent :



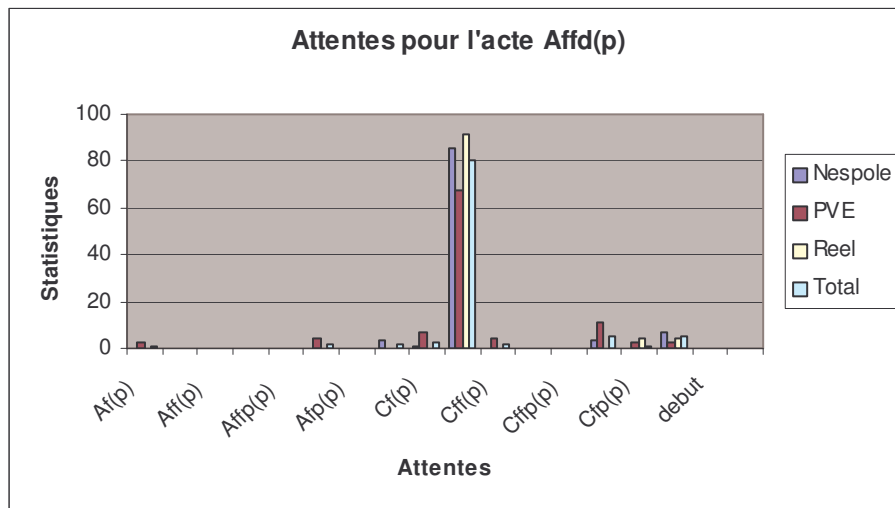
Côté client :



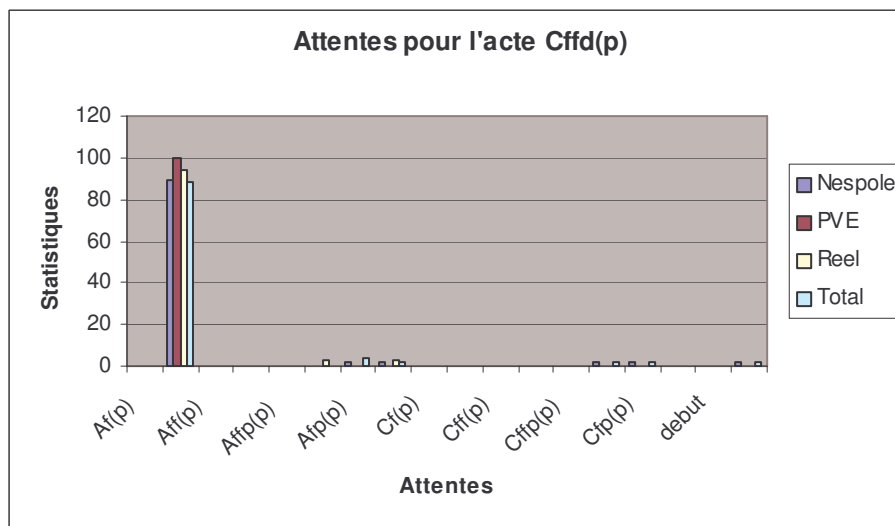


## Faire-faire-devoir

Côté agent :

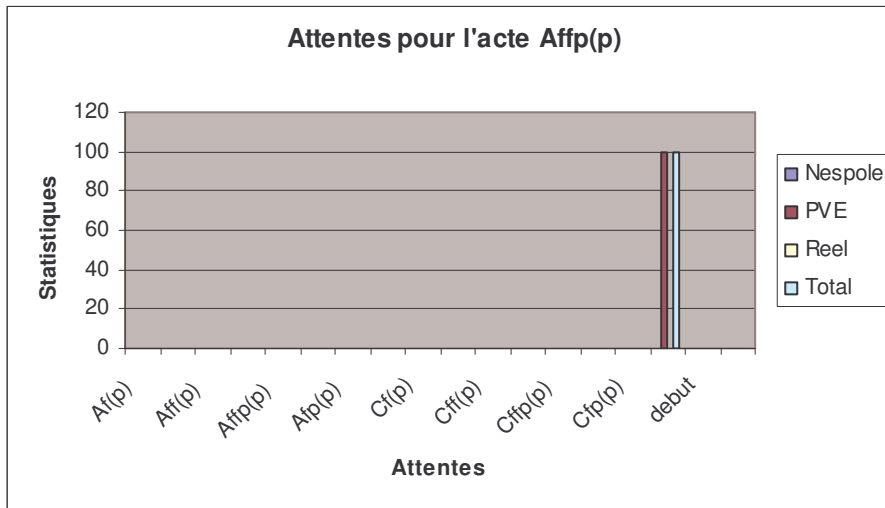


Côté client :

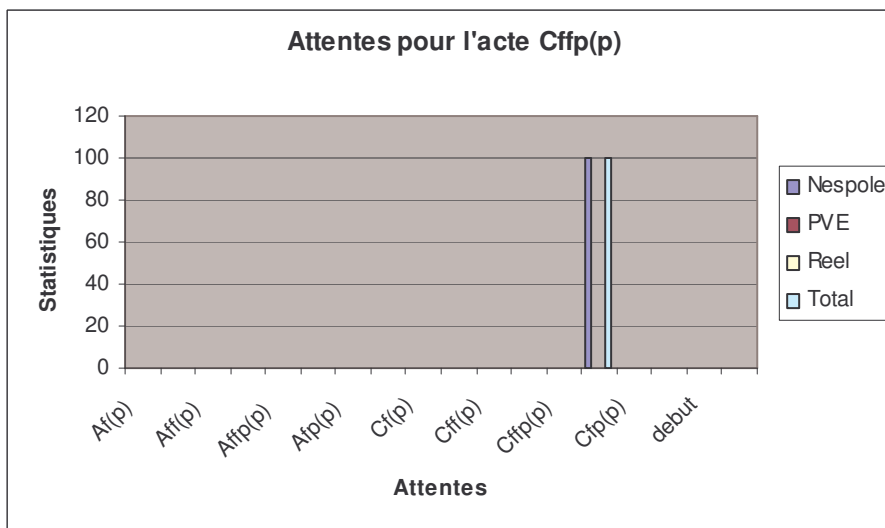


**Faire-faire-pouvoir** (quasiment inexistant dans nos dialogues)

Côté agent :



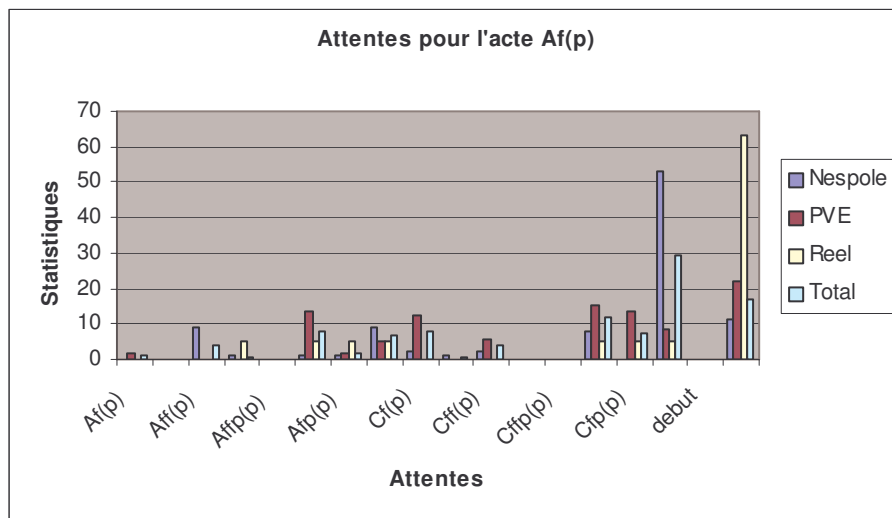
Côté client :



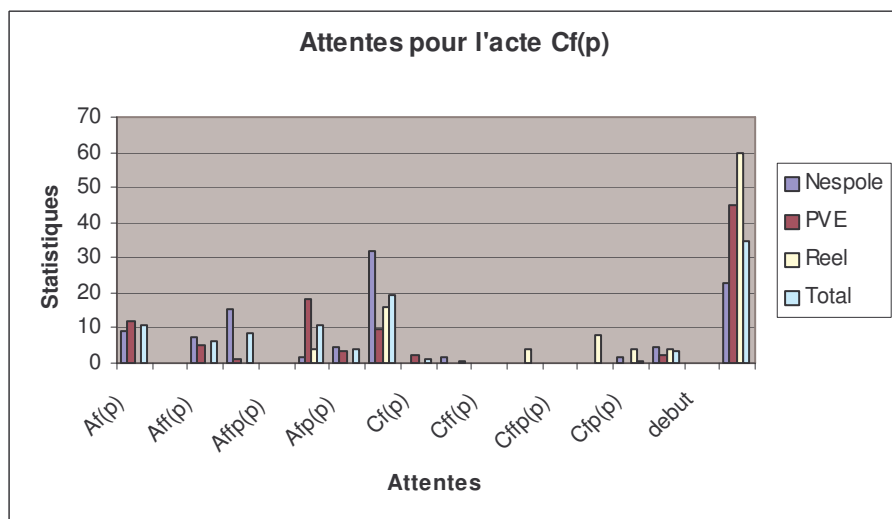
## 2. Actes répondant à des attentes

### Faire-faire

Côté Agent :

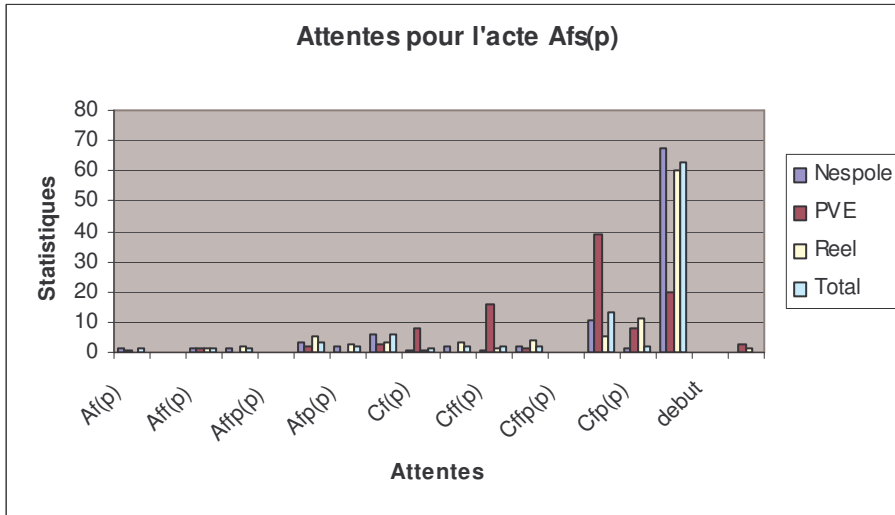


Côté client :

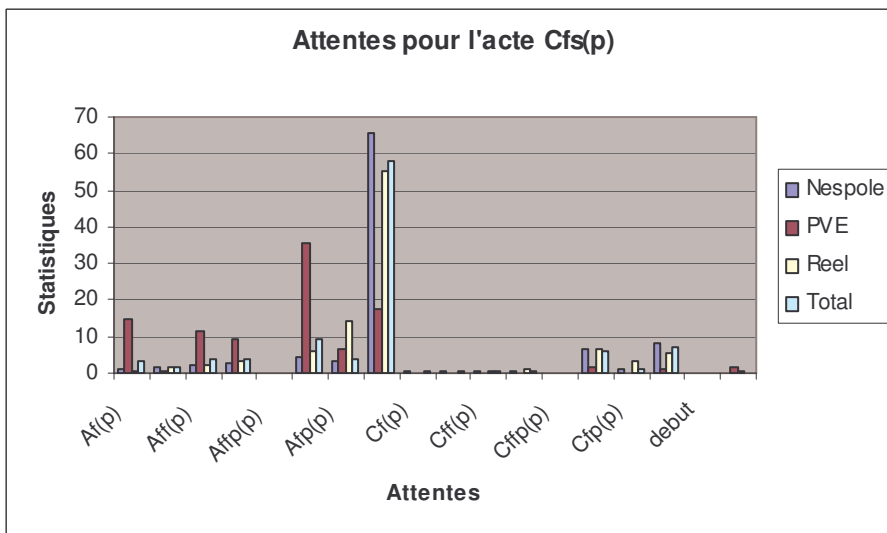


**Faire-savoir**

Côté Agent :

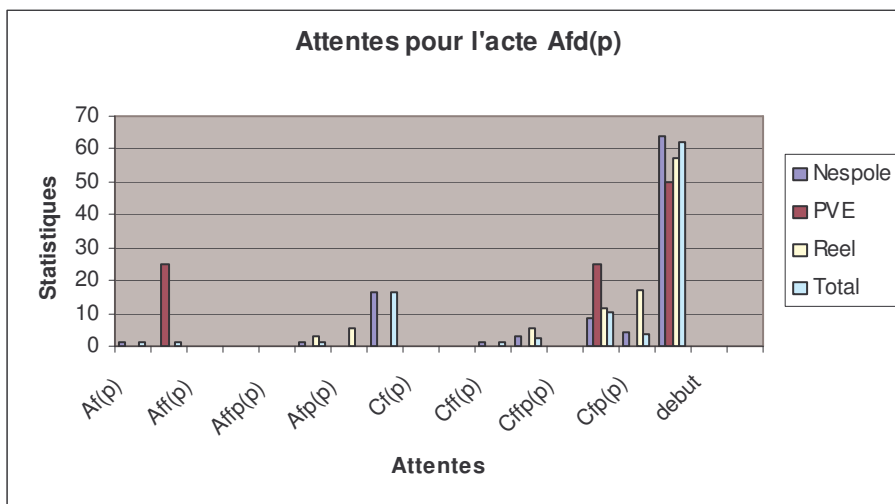


Côté client :

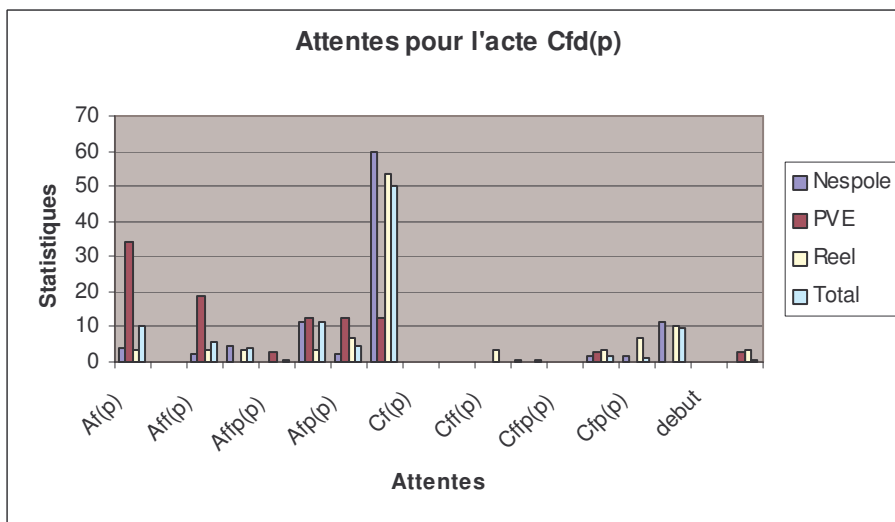


## Faire-devoir

Côté Agent :

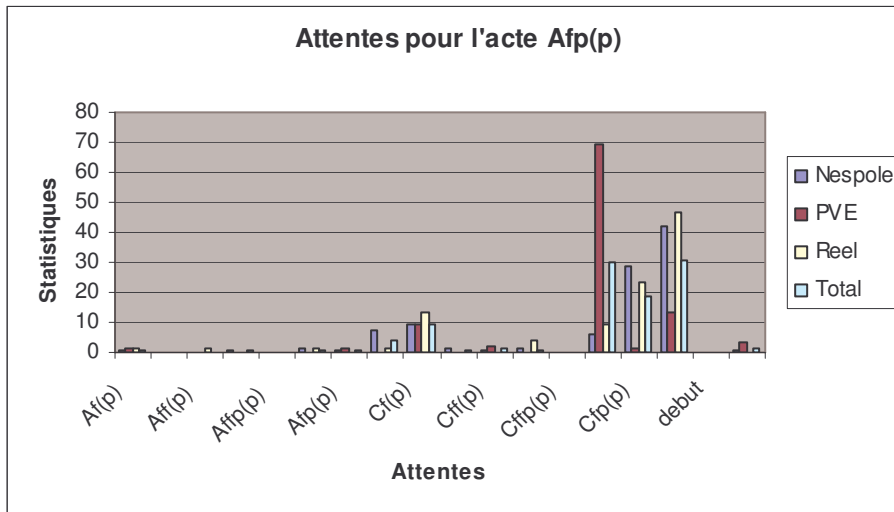


Côté client :

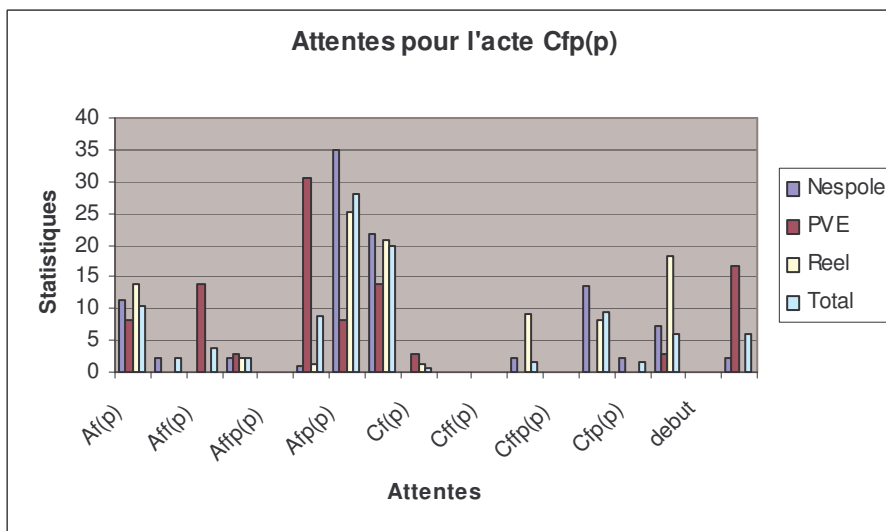


## Faire-pouvoir

Côté Agent :



Côté client :



## Instructions à suivre pour le système PVE

1. Sélectionner le nom de machine correspondant au serveur d'actes
2. Se connecter au serveur d'actes (une fois connecté le bouton reste enfoncé)
3. Decrocher (quand on est prêt)
4. Attendre que le client ait décroché (client passe au vert)
5. Sélectionner la tâche en cours
6. Sélectionner la phase (ouverture, clôture...)
7. Sélectionner le label voulu
8. Sélectionner la phrase correspondante (elle s'affichera dans la zone du bas)
9. Choisir éventuellement les paramètres
10. Envoyer
11. Attendre réponse
12. Raccrocher (pve passera au rouge)

The screenshot displays the 'Magicien d'Oz du système PVE - Vocalisa' interface. At the top, a status bar shows '129.88.41.52 > actes décrocher client pve actes'. Below this is a calendar for January 2003 with the 30th highlighted. To the right of the calendar are time selection fields: '[heure] : 18 heures', '[minutes] : 00', and '[moment] : le matin'. A list of contacts is visible, including 'monsieur, BLANC', 'monsieur, DUPOND', 'monsieur, DUPUIS', 'monsieur, DURAND', 'madame, BLANC', 'madame, DUPONT', 'madame, DUPUIS', 'mademoiselle, BLANC', 'mademoiselle, DUPOND', and 'mademoiselle, DUPUIS'. The main area shows a task selection interface with tabs for '[MULTITACHE]', '[STANDI]', '[RE]', '[RESA]', and '[COM-info]'. The 'Ouverture (D)' phase is selected, and the 'Ouvre' label is highlighted. A 'LEXIQUE' list on the right contains terms like 'ifs : demande d'information', 'fs : information', 'ff : demande d'action', 'f : action', 'fp : invite, ouverture', 'ffd : demande confirmation', and 'fd : confirmation, clôture'. A 'Envoyer' button is at the bottom right.

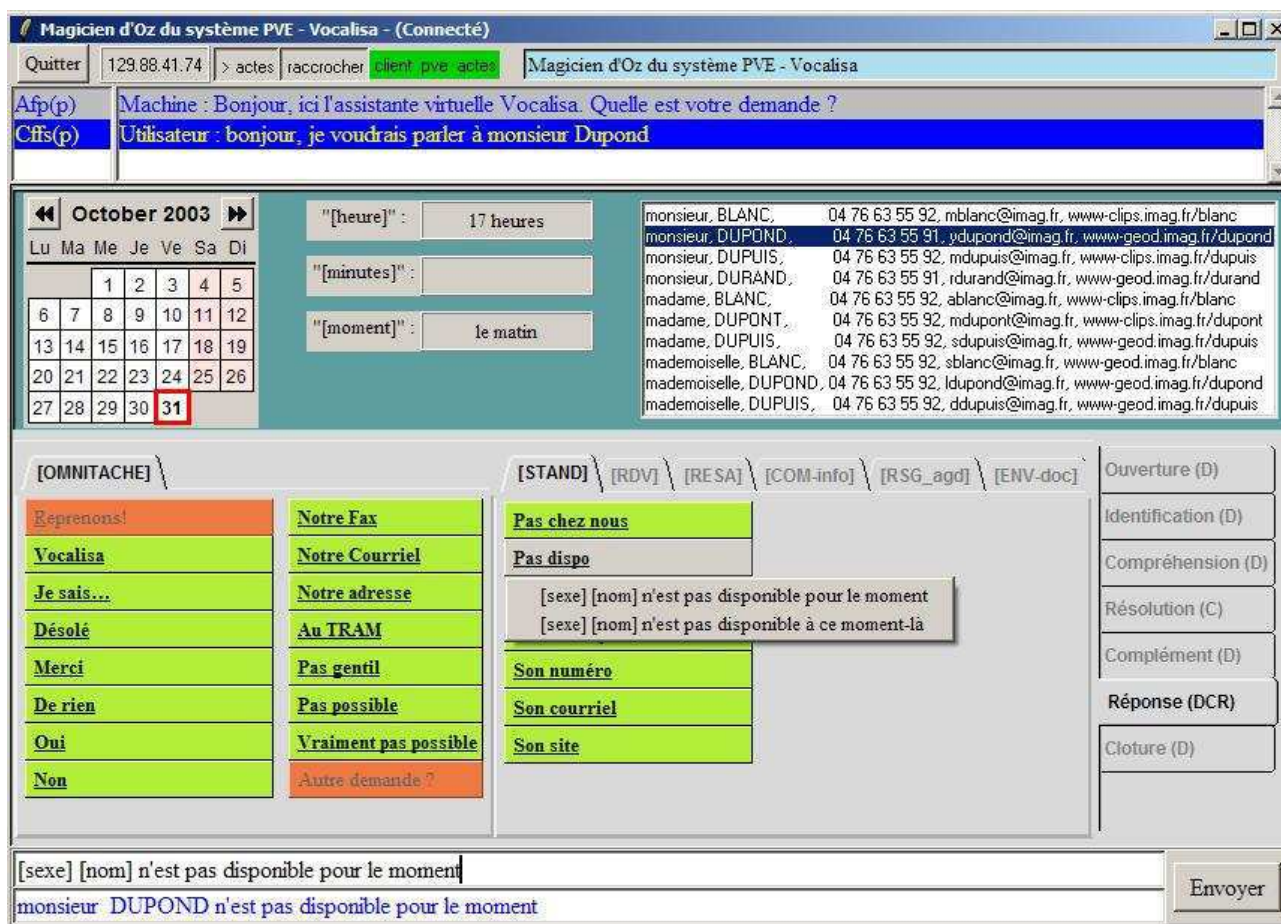
L'expérimentateur qui simule le système exécute le script concernant la partie magicien-secrétaire du système. L'interface ci-dessus lui est alors proposée. Il doit alors, dans une perspective d'annotation *in vivo* en actes de dialogue, se connecter au serveur d'actes. Pour cela, il sélectionne le nom de la machine sur laquelle tourne ce serveur (1), et s'y connecte en appuyant sur le bouton *ad-hoc* (2). Si une telle annotation n'est pas souhaitée (pour les tests par exemple), ces deux premières étapes ne sont pas nécessaires. D'une manière générale, il devra décrocher lorsqu'il sera en mesure de dialoguer (3) et raccrocher lorsqu'il estime que le dialogue est terminé (12). Son avatar (pve) changera alors de statut pour indiquer aux autres participants qu'il est ouvert au dialogue (vert) ou pas (rouge). Lorsqu'il a indiqué qu'il était en mesure de dialoguer (donc pve en vert), l'expérimentateur attend que le client aie décroché, c'est à dire que son avatar (client) sur l'interface passe au vert (en mesure de dialoguer). S'ensuit alors le dialogue entre les deux interlocuteurs.

L'expérimentateur sélectionne dès que nécessaire la tâche en cours (5). Par défaut, il aura les informations voulues au début du dialogue (à l'ouverture) dans la partie omnitâche en attendant que l'utilisateur émette le souhait d'aborder une nouvelle tâche (à n'importe quel moment du dialogue). Par la suite, il n'aura à sélectionner la tâche que si l'utilisateur en a émis le souhait.

Il sélectionne alors la phase dans laquelle se situe cette portion de dialogue (6). Par défaut, au début du dialogue, la phase d'ouverture est sélectionnée. Par la suite, la phase est conservée jusqu'à ce que l'expérimentateur souhaite en changer.

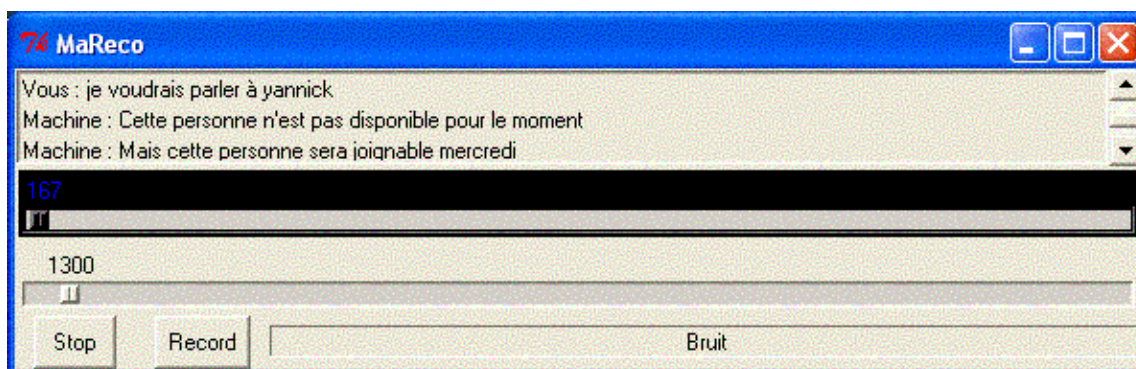
S'affichent alors les étiquettes concernant les énoncés possibles en fonction de la tâche et de la phase sélectionnées. Elles ont une forme condensée afin de diminuer la charge cognitive de l'expérimentateur. Par exemple, il verra « pas dispo » au lieu de « [sexe][nom] n'est pas disponible pour le moment ; [sexe][nom] n'est pas disponible à ce moment là ». Il doit donc sélectionner l'étiquette voulue (7). Un menu contextuel apparaîtra montrant les énoncés correspondants (cf. exemple ci-dessus) lui permettant de choisir précisément l'énoncé qu'il souhaite envoyer à l'utilisateur (8). Cette phrase s'affiche alors dans la zone du bas (par exemple « [sexe][nom] n'est pas disponible pour le moment »). Une autre zone en dessous permet de visualiser la phrase obtenue après remplacement des paramètres par leur valeur (par exemple « monsieur Dupond n'est pas disponible pour le moment »). Les paramètres ont des valeurs par défaut. Il n'est donc pas toujours nécessaire de les changer. Mais dans le cas contraire, il suffit de sélectionner les paramètres et leur valeur (ici choisir monsieur Dupond dans l'annuaire) pour que ces paramètres soient modifiées (9). La zone du bas affichera donc l'énoncé prêt à être envoyé. Il suffit alors, après vérification, d'envoyer l'énoncé à l'utilisateur (10) puis d'attendre la réaction de l'utilisateur (11). L'énoncé de l'utilisateur s'affichera alors dans l'historique et l'expérimentateur reprendra la boucle de dialogue à partir du (5). A la fin du dialogue, c'est à dire lorsque l'utilisateur a raccroché (l'icône client est passée au rouge) considérant que le dialogue est terminé, l'expérimentateur raccroche (12) à son tour (son avatar passe au rouge). Il recommencera au (3) lorsqu'il sera de nouveau disponible.



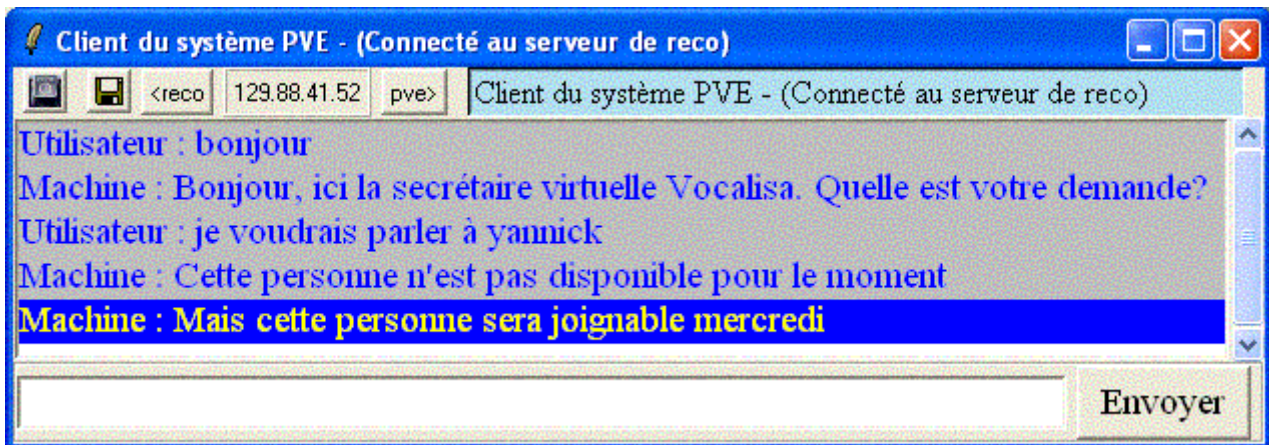


Les interlocuteurs sont en ligne (leur avatar est vert et le bouton raccrocher est enfoncé). L'énoncé de l'utilisateur est dans l'historique. L'expérimentateur a sélectionné Pas dispo parmi les étiquettes proposées (et soulignées) pour faire apparaître le menu lui permettant de choisir la phrase qu'il souhaite répondre à l'utilisateur. Il a ensuite sélectionné la première. Celle-ci s'est affichée en bas tandis que le menu s'est masqué. Il a alors sélectionné monsieur Dupond dans l'annuaire (dans notre exemple monsieur Dupond n'était pas encore sélectionné sinon cette étape est inutile) pour construire sa réponse. Il lui suffit maintenant d'envoyer l'énoncé au client.

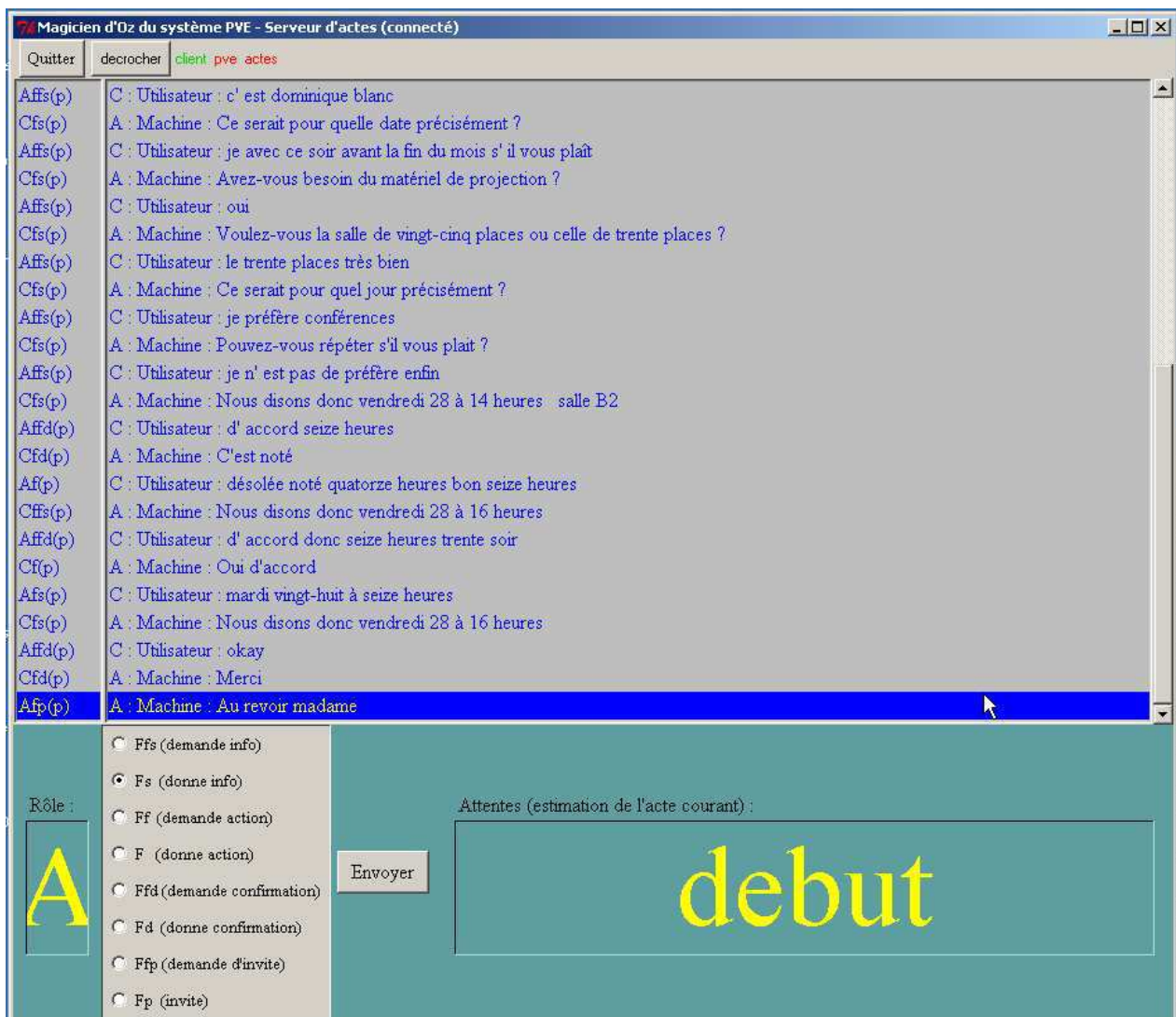
## Système de Reconnaissance + Synthèse



## Interface « Client »



## Dialogue complet vu de l'annotateur



## Enoncés proposés au contrôleur

[OMNITACHE]

0=D=Afp(p)=Ouverture="Bonjour, ici l'assistante virtuelle Vocalisa. (Quelle est votre demande ?/Quel est le motif de votre appel, s'il vous plait ?)"

1=D=Affs(p)=Nom?="(C'est à quel/Quel est votre) nom ?"

1=D=Affs(p)=UnNom?="indiquez moi l'un des noms"

1=D=Affs(p)=Epelez!="Pouvez-vous épeler s'il vous plait ?"

2=D=Affs(p)=Répétez!="(Pouvez-vous répéter s'il vous plait ? /Veuillez m'excuser, je vais vous demander de répéter à nouveau.)"

5=D=Afp(p)=Reprenons!="Alors, reprenons, si vous le voulez bien"

6=D=Afd(p)=Au Revoir="Au revoir ([sexe])"

2=D=Afs(0)=--Je regarde agenda--="je consulte l'agenda"

2=D=Aff(p)=--Attendez!--="(Alors) attendez une minute s'il vous plait"

4=C=Affs(p)=Vos coordonnées="Au cas où j'aurais besoin de vous joindre, quel est votre (numéro de téléphone/adresse électronique) ?."

4=C=Affs(p)=Nos coordonnées?="Avez-vous nos coordonnées ?"

5=R=Afs(p)=Vocalisa="Je m'appelle Vocalisa"

5=R=Afs(p)=Je sais...="Je m'occupe du standard, des prises de rendez-vous et des réservations de salle, etc."

5=R=Afs(p)=Désolé="(Excusez-moi/Désolée)"

5=R=Afs(p)=Merci="Merci"

5=R=Afs(p)=De rien="Je vous en prie"

5=R=Afs(p)=Oui="(Oui/okay) (d'accord)"

5=R=Afs(p)=Non="non"

5=R=Afs(p)=Notre Fax="C'est le 04 76 63 55 52"

5=R=Afs(p)=Notre Courriel="C'est vocalisa arobasse imag point f r"

5=R=Afs(p)=Notre adresse="C'est sur le domaine universitaire de Grenoble (au 220, rue de la chimie au premier étage)."

5=R=Afs(p)=Au TRAM="C'est au terminus de la ligne B du tramway."

5=R=Afs(p)=Pas gentil="(Ce n'est pas gentil!/Vous m'avez coincée! Vous l'avez fait exprès ?)"

5=R=Afs(p)=Pas possible="Je suis désolée, (ça ne va pas être possible/je ne peux pas répondre à votre demande)."

5=R=Afs(p)=Vraiment pas possible="(Désolée, je n'ai pas les éléments pour vous répondre. Voulez-vous que je vous passe [sexe] [nom], la personne compétente./Je suis obligée d'interrompre notre entretien ! Au revoir !)"

5=D=Affs(p)=Autre demande ?="(Avez-vous une autre demande/Desirez-vous autre chose) ?"

[STAND]

3=D=Aff(p)=MSG (début...fin)="(Je vous écoute, veuillez parler après le bip sonore/Votre message est enregistré)"

2=D=Affs(p)=Qui?="qui voulez-vous joindre ?"

2=D=Affs(p)=C'est Perso?="C'est personnel (ou professionnel) ?"

2=D=Affs(p)=De qui?="C'est de la part de qui ?"

3=D=Affs(p)=Rappelez + tard!="Pouvez-vous rappeler plus tard ?"

3=C=Affs(p)=Son poste direct?="Avez-vous essayé son poste direct ?"

3=C=Afs(p)=Mais joignable + tard="Cette personne sera joignable ([joursemaine] [jourmois]/[moment])"

3=C=Afs(p)=Mais autre pers="Je peux vous passer un autre membre de son équipe, par exemple [sexe] [nom]."

3=C=Afs(p)=Mais MSG ?="Si vous voulez, je peux prendre un message."

5=R=Afs(p)=Pas chez nous="(Cette personne n'est pas chez nous/Ah ! Ce n'est pas ici !)"

5=R=Afs(p)=Pas dispo="[sexe] [nom] n'est pas disponible (pour le moment/ à ce moment-là)"

5=R=Afs(p)=Je vous le passe="Un instant, je vous passe votre correspondant(e) ([sexe] [nom])."

5=R=Afs(p)=C'est occupé="(son poste est occupé/ça ne répond pas)."

5=R=Afs(p)=Son numéro="Je vais vous donner son numéro (de portable/ de poste). C'est le [annuaire]"

5=R=Afs(p)=Son courriel="Je vous donne son (mail/ courriel). C'est [nom] arobasse imag point f r"

5=R=Afs(p)=Son site="Je vais vous donner le site internet. C'est [adressesite]"

[RDV]

2=D=Affs(p)=Avec Qui?="Avec quelle personne ?"

2=D=Affs(p)=C'est Perso?="C'est personnel (ou professionnel) ?"

2=D=Affs(p)=De qui?="C'est de la part de qui ?"

2=D=Affs(p)=Date/jour?="(Ce serait pour) quelle (date/jour) ?"

2=D=Affs(p)=Heure?="(Ce serait pour) quelle heure ?"

2=D=Affs(p)=Moment?="à quel moment ?"

2=D=Affs(p)=Combien de tps?="(Ce serait pour) combien de temps ?"

5=D=Affd(p)=Résumé?="Nous disons donc [joursemaine] [jourmois] à [heure] [minutes] (salle B2) (avec [sexe] [nom]) ?"

3=C=Affs(p)=RDV prop°="Je peux vous proposer [joursemaine] ([jourmois] à [heure] [minutes]/ [moment]). (Qu'est-ce qui/Est-ce que cela) vous conviendrait ?"

3=C=Affs(p)=RDV 10 min.="Si ça ne dépasse pas dix minutes, je peux vous mettre [joursemaine] ([jourmois] à [heure] [minutes]/ [moment])"

3=C=Affs(p)=RDV autre pers="Vous pourriez rencontrer un autre membre de son équipe, par exemple [sexe] [nom]."

5=R=Afs(p)=Pas chez nous="(Cette personne n'est pas chez nous/Ah ! Ce n'est pas ici !)"

5=R=Afs(p)=Pas dispo="Cette personne n'est pas disponible (pour le moment/ à ce moment-là)"

5=R=Afs(p)=Les dispos sont="Les disponibilités de [sexe] [nom] sont (lundi et mardi/en fin de semaine)"

5=R=Afs(p)=Les créneaux pris sont="Les créneaux déjà pris sont [joursemaine] ([jourmois] à [heure] [minutes]) (avec [sexe] [nom])"

5=R=Afp(p)=Résumé="Nous disons donc [joursemaine] [jourmois] à [heure] [minutes] (salle B2) (avec [sexe] [nom])"

5=R=Af(p)=C'est noté="C'est noté"

[RESA]

2=D=Affs(p)=Date?="(Ce serait pour) quelle date (précisément) ?"

2=D=Affs(p)=Jour?="(Ce serait pour) quel jour (précisément) ?"

2=D=Affs(p)=Heure?="(Ce serait pour) quelle heure (précisément) ?"

2=D=Affs(p)=Moment?="à quel moment ?"

2=D=Affs(p)=Combien de tps?="(Ce serait pour) combien de temps ?"

2=D=Affs(p)=Matos?="Avez-vous besoin du matériel de projection ?"

5=D=Affd(p)=Confirmez!="Pouvez-vous m'envoyer une confirmation par fax ou par courriel, s'il vous plaît ?"

5=D=Affd(p)=Résumé?="Nous disons donc [joursemaine] [jourmois] à [heure] [minutes] (salle B2) ?"

3=C=Affs(p)=Mais pour 25 (prof)="J'aurais une salle pour vingt-cinq personnes [joursemaine] ([jourmois] à [heure] [minutes] / [moment]), sinon la salle de trente personnes sera libre (lundi/mardi) prochain."

3=C=Affs(p)=Mais sans matos(étud.)="J'aurais une salle à ce moment là mais sans rétro-projecteur, sinon la salle avec rétro-projecteur sera libre [joursemaine] ([jourmois] à [heure] [minutes]/ [moment])"

3=C=Affs(p)=Salle?="Voulez-vous la salle de vingt-cinq places ou celle de trente places ?"

5=R=Afs(p)=SalleB2="La salle B2."

3=R=Afs(p)=Matos="Cette salle possède seulement un rétroprojecteur"

5=R=Afp(p)=Résumé="Nous disons donc [joursemaine] [jourmois] à [heure] [minutes] (salle B2)"

5=R=Af(p)=C'est noté="C'est noté"

[COM-info]

0=D=Afp(p)=PresentVocalisa="Bonjour, ici Vocalisa, la secrétaire virtuelle. J'ai un message à vous transmettre"

1=D=Aff(p)=Identification="Pouvez-vous prononcer (votre nom/identificateur) s'il vous plaît ?"

1=D=Aff(p)=Repetition?="Pouvez-vous répéter votre (nom/identificateur) s'il vous plaît ?"

5=R=Afs(p)=Info="Vous êtes convoqué à une réunion d'équipe pour le projet PVE, samedi prochain, de neuf heures à midi, dans la salle de conférences."

5=R=Afs(p)=Info\_jour="Samedi prochain, le 22"

5=R=Afs(p)=Info\_heure="De neuf heures à midi"

5=R=Afs(p)=Info\_ou="Dans la salle de conférences."

5=R=Afs(p)=Info\_resume="(réunion PVE) samedi prochain, le 22, de neuf heures à midi, dans la salle de conférences."

5=R=Afs(p)=Info\_qui="Toute l'équipe sera là sauf monsieur Dupuis."

4=C=Afs(p)=Complement-info="Il s'agit de revoir le cahier des charges du projet PVE pour l'adapter à un nouveau contrat proposé par un partenaire chinois"

2=D=Affs(p)=Présent?="Serez-vous présent à cette réunion ?"

5=C=Affs(p)=NotéPrésent="Je note donc que vous serez présent samedi prochain. Est-ce correct ?"

5=C=Affs(p)=NotéPasPrésent="Je note donc que vous (ne serez pas présent/serez absent) samedi prochain. Est-ce correct ?"

[RSG\_agd]

1=D=Aff(p)=identification="Pouvez-vous me donner votre (nom/identificateur) s'il vous plaît ?"

1=D=Aff(p)=répétition?="Pouvez-vous répéter votre (nom/identificateur) s'il vous plaît ?"

2=C=Aff(p)=Précis/Identif="Pouvez vous m'en dire un peu plus s'il vous plait ?"

1=C=Affs(p)=confirmation="Pouvez-vous s'il vous plait confirmer cette opération en prononçant le mot : valide ?"

2=C=Afs(p)=bienvenue="Bienvenu [sexe] [nom]. Quelle opération voulez-vous faire ?"

2=C=Afp(p)=opérations="Vous pouvez effectuer des opérations telles que : consulter, ajouter, annuler, déplacer, modifier un horaire d'une réunion ou d'un rendez-vous. Quelle opération voulez-vous faire ?"

5=C=Afs(p)=autre modification?="Voulez-vous mettre à jour autre chose dans votre agenda ?"

5=C=Afs(p)=liste\_disponibilités="Vos (disponibilités/créneaux pris) sont toute la semaine sauf [joursemaine]."

5=C=Affs(p)=Prévenir?="Votre information est enregistrée. Si vous souhaitez que je me charge de la transmettre, veuillez me dire à qui. Je vous écoute"

5=C=Affs(p)=Transmettre?="Voulez-vous que je me charge de transmettre l'information ?"

5=C=Affs(p)=Rendre compte?="Entendu! Je contacterai (cette personne/ces personnes). Dois-je vous rappeler pour vous en rendre compte ?"

5=R=Afs(p)=Les dispo sont="Les disponibilités de [sexe] [nom] sont (lundi et mardi/en fin de semaine)"

5=R=Afs(p)=Les créneaux pris sont="Les créneaux déjà pris de [sexe] [nom] sont [joursemaine] ([jourmois] à [heure] [minutes]) (avec [sexe] [nom])"

0=C=Afd(p)=Cpte rendu OK="Bonjour, ici Vocalisa, la secrétaire virtuelle. Votre information a bien été transmise à vos correspondants."

0=C=Afd(p)=Cpte rendu Pb="Bonjour, ici Vocalisa, la secrétaire virtuelle. Votre information a bien été transmise à vos correspondants, sauf à monsieur Dupont, qui n'est pas joignable avant le mois prochain."

3=C=Afs(p)=mais autre créneau="Je peux vous proposer [joursemaine] [jourmois] à [heure] par exemple."

5=C=Afd(p)=consult\_RDV/réunion="(Ce rendez-vous est prévu/cette réunion est prévue) pour [joursemaine] [jourmois] à [heure] avec [sexe] [nom]."

0=R=Afs(p)=info-modif="Je vous informe que (la réunion prévue/le rendez-vous prévu) [joursemaine] [jourmois] à [heure] avec [sexe] [nom] est (annulé/déplacé)."

1=R=Afs(p)=code bon\_érroné="Ce code est (bon/érroné)."

5=R=Afs(p)=Mr Pas dispo="[sexe] [nom] (est/n'est pas) disponible à ce moment-là"

5=R=Afs(p)=dispo\_déjà pris="Ce créneau est (disponible/déjà pris)."

2=D=Affs(p)=Date/jour?="(Ce serait pour) quelle (date/jour) ?"

2=D=Affs(p)=Heure?="(Ce serait pour) quelle heure ?"

2=D=Affs(p)=Moment?="à quel moment ?"

2=D=Affs(p)=Combien de tps?="(Ce serait pour) combien de temps ?"

2=D=Affs(p)=Avec Qui?="Avec quelle personne ?"

5=D=Affd(p)=Résumé?="Nous disons donc [joursemaine] [jourmois] à [heure] [minutes] (salle B2) (avec [sexe] [nom]) ?"

3=C=Affs(p)=RDV prop°="Je peux vous proposer [joursemaine] ([jourmois] à [heure] [minutes]/ [moment]). (Qu'est-ce qui/Est-ce que cela) vous conviendrait ?"

3=C=Affs(p)=RDV 10 min.="Si ça ne dépasse pas dix minutes, je peux vous mettre [joursemaine] ([jourmois] à [heure] [minutes]/ [moment])"

3=C=Affs(p)=RDV autre pers="Vous pourriez rencontrer un autre membre de son équipe, par exemple [sexe] [nom]."

5=R=Af(p)=C'est noté="C'est noté"

5=R=Af(p)=confirm="(Votre/l') agenda est mis à jour."

5=R=Afd(p)=confirmRemplace="[sexe] [nom] vous remplacera [joursemaine] [jourmois] à [heure] [minutes]"

[ENV-doc]

2=D=Affs(p)=Envoi où?="(Où dois-je envoyer ce document/Pouvez-vous préciser l'adresse) ?"

5=C=Affs(p)=Transmettre?="Voulez-vous que je me charge de transmettre votre document ?"

5=R=Afs(p)=Envoi ok="(Entendu je m'occupe de cet envoi/Oui, je vais le faire)."

5=R=Afs(p)=Attente doc="J'attends votre document."

2=C=Afd(p)=dernière version="La dernière version date d'aujourd'hui (à [heure] [minutes])"

Liste des mots phonétisés et classes

{S} {SIL}  
 {} {SIL}  
 {} {SIL}  
 {} {SIL}  
 {} {SIL}  
 {} {SIL}  
 {<UNK>} {SIL}  
 {FERMETURE} {{b WB} p t k g {f WB}}  
 {FERMETURE(adieu)} {{a WB} d j {2 WB}}  
 {FERMETURE(adios)} {{a WB} d j O {s WB}}  
 {FERMETURE(au\_revoir)} {{o WB} R 2 v w a {R WB}}  
 {FERMETURE(bonne\_journée)} {{b WB} O n Z u R n {e WB}}  
 {FERMETURE(bonne\_soirée)} {{b WB} O n s w a R {e WB}}  
 {FERMETURE(bye)} {{b WB} a {j WB}}  
 {FERMETURE(bye\_bye)} {{b WB} a j b a {j WB}}  
 {FERMETURE(bye\_bye\_2)} {{b WB} a b a {j WB}}  
 {FERMETURE(ciao)} {{t WB} Z a {o WB}}  
 {FERMETURE(ciao\_2)} {{t WB} Z {o WB}}  
 {FERMETURE(bonsoir)} {{b WB} o n s w a A {R WB}}  
 {FERMETURE(salut)} {{s WB} a l {y WB}}  
 {FERMETURE(à\_bientôt)} {{a WB} b i a n t {o WB}}  
 {FERMETURE(à\_ce\_soir)} {{a WB} s 2 s w a {R WB}}  
 {FERMETURE(à\_cet\_après\_midi)} {{a WB} s e t a p R E m i d {i WB}}  
 {FERMETURE(à\_demain)} {{a WB} d 2 m {in WB}}  
 {FERMETURE(à\_la\_semaine\_prochaine)} {{a WB} l a s 2 m E n p R o S E {n WB}}  
 {FERMETURE(à\_plus)} {{a WB} p l y {s WB}}  
 {FERMETURE(à\_plus\_tard)} {{a WB} p l y t a {R WB}}  
 {FERMETURE(à\_tout\_à\_l'heure)} {{a WB} t u t a l 9 {R WB}}  
 {FERMETURE(à\_tout\_de\_suite)} {{a WB} t u d 2 s y i {t WB}}  
 {FERMETURE(à\_toute\_allure)} {{a WB} t u t a l y {R WB}}  
 {GENRE} {{b WB} p t k g {f WB}}  
 {GENRE(madame)} {{m WB} a d a {m WB}}  
 {GENRE(madame\_2)} {{m WB} a A d a A {m WB}}  
 {GENRE(madame\_3)} {{m WB} a A d a A m {& WB}}  
 {GENRE(mademoiselle)} {{m WB} a d m w a z E l {l WB}}  
 {GENRE(mademoiselle)} {{m WB} a d & m w a z E l {l WB}}  
 {GENRE(mesdames)} {{m WB} E d a {m WB}}  
 {GENRE(messieurs)} {{m WB} e s j {2 WB}}  
 {GENRE(monsieur)} {{m WB} & s j {2 WB}}  
 {MOMENT} {{b WB} p t k g {f WB}}  
 {MOMENT(année)} {{a WB} n {e WB}}  
 {MOMENT(année\_2)} {{a WB} n {e WB}}  
 {MOMENT(après-demain)} {{a WB} p R E d 2 m {in WB}}  
 {MOMENT(après-midi)} {{a WB} p R E m i d {i WB}}  
 {MOMENT(aujourd'hui)} {{o WB} Z u R d y {i WB}}  
 {MOMENT(demain)} {{d WB} & m {in WB}}  
 {MOMENT(demain\_2)} {{d WB} m {in WB}}  
 {MOMENT(hier)} {{j WB} E {R WB}}  
 {MOMENT(hier\_2)} {{i WB} j E {R WB}}  
 {MOMENT(hier\_3)} {{i WB} j E {R WB}}  
 {MOMENT(jour)} {{Z WB} u {R WB}}  
 {MOMENT(matin)} {{m WB} a t {in WB}}  
 {MOMENT(matin\_2)} {{m WB} a A t {in WB}}  
 {MOMENT(mois)} {{m WB} n {a WB}}  
 {MOMENT(mois\_2)} {{m WB} w a {z WB}}  
 {MOMENT(mois\_3)} {{m WB} w a A WB}}  
 {MOMENT(nuit)} {{n WB} H {i WB}}  
 {MOMENT(soir)} {{s WB} w a {R WB}}  
 {MOMENT(soir\_2)} {{s WB} w a A {R WB}}  
 {MOMENT(soirée)} {{s WB} w a R {e WB}}  
 {MOMENT(tout\_à\_l'heure)} {{t WB} u t a l 9 {R WB}}  
 {MONAIE} {{b WB} p t k g {f WB}}  
 {MONAIE(dollars)} {{d WB} O l a {R WB}}  
 {MONAIE(euro)} {{9 WB} R {o WB}}  
 {MONAIE(euros)} {{9 WB} R {o WB}}  
 {MONAIE(franc)} {{f WB} R {an WB}}  
 {MONAIE(franc\_2)} {{f WB} R an {k WB}}  
 {MONAIE(francs)} {{f WB} R {an WB}}  
 {MONAIE(francs\_2)} {{f WB} R an {z WB}}  
 {NOM\_JOUR} {{b WB} p t k g {f WB}}  
 {NOM\_JOUR(dimanche)} {{d WB} i m a n {S WB}}  
 {NOM\_JOUR(dimanche\_2)} {{d WB} i m a n S {& WB}}  
 {NOM\_JOUR(jeudi)} {{Z WB} 2 d {i WB}}  
 {NOM\_JOUR(jeudi\_2)} {{Z WB} 29 d {i WB}}  
 {NOM\_JOUR(lundi)} {{l WB} u n d {i WB}}  
 {NOM\_JOUR(mardi)} {{m WB} a R d {i WB}}  
 {NOM\_JOUR(mardi\_2)} {{m WB} a A R d {i WB}}  
 {NOM\_JOUR(mercredi)} {{m WB} E R k R & d {i WB}}  
 {NOM\_JOUR(mercredi\_2)} {{m WB} e R k R & d {i WB}}  
 {NOM\_JOUR(mercredi\_3)} {{m WB} E R k R & d {i WB}}  
 {NOM\_JOUR(samedi)} {{s WB} a m & d {i WB}}  
 {NOM\_JOUR(samedi\_2)} {{s WB} a m d {i WB}}  
 {NOM\_JOUR(samedi\_3)} {{s WB} a A m & d {i WB}}  
 {NOM\_JOUR(vendredi)} {{v WB} a n d R & d {i WB}}  
 {NOM\_JOUR(vendredi\_2)} {{v WB} a n d R & d {i WB}}  
 {NOM\_LIEU} {{b WB} p t k g {f WB}}  
 {NOM\_LIEU(avenue)} {{a WB} v & n {y WB}}  
 {NOM\_LIEU(avenue\_2)} {{a WB} v & n {y WB}}  
 {NOM\_LIEU(bastille)} {{b WB} a s t i {j WB}}  
 {NOM\_LIEU(centre\_ville)} {{s WB} a n t R & v i {l WB}}  
 {NOM\_LIEU(cinéma)} {{s WB} i n e m {a WB}}  
 {NOM\_LIEU(cinéma\_2)} {{s WB} i n e E m {a WB}}  
 {NOM\_LIEU(cours)} {{k WB} u {R WB}}  
 {NOM\_LIEU(gare)} {{g WB} a {R WB}}  
 {NOM\_LIEU(gare\_2)} {{g WB} a A {R WB}}  
 {NOM\_LIEU(gare\_3)} {{g WB} a A R {& WB}}  
 {NOM\_LIEU(jardin)} {{Z WB} a R d {in WB}}  
 {NOM\_LIEU(jardin\_2)} {{Z WB} a A R d {in WB}}  
 {NOM\_LIEU(l'aéroport)} {{l WB} a A e R o O p O {R WB}}  
 {NOM\_LIEU(mer)} {{m WB} E {R WB}}  
 {NOM\_LIEU(montagne)} {{m WB} o n t a {NJ WB}}  
 {NOM\_LIEU(montagne\_2)} {{m WB} o n t a A {NJ WB}}  
 {NOM\_LIEU(montagne\_3)} {{m WB} o n t a A NJ {& WB}}  
 {NOM\_LIEU(musée)} {{m WB} y z {e WB}}  
 {NOM\_LIEU(rue)} {{R WB} {y WB}}  
 {NOM\_LIEU(stade)} {{s WB} t a {d WB}}  
 {NOM\_LIEU(stade\_2)} {{s WB} t a A {d WB}}  
 {NOM\_LIEU(stade\_3)} {{s WB} t a d {& WB}}  
 {NOM\_LIEU(terrasse)} {{t WB} e E R a A {s WB}}  
 {NOM\_LIEU(terrasse\_2)} {{t WB} e E R a A s {& WB}}  
 {NOM\_LIEU(théâtre)} {{t WB} e a t {R WB}}  
 {NOM\_LIEU(corenc)} {{k WB} o R {an WB}}  
 {NOM\_LIEU(biviers)} {{b WB} i v j {e WB}}  
 {NOM\_LIEU(clips)} {{k WB} l i p {s WB}}  
 {NOM\_LIEU(fourrier)} {{f WB} u R j {e WB}}  
 {NOM\_LIEU(imag)} {{i WB} m a {g WB}}  
 {NOM\_MACHINE} {{b WB} p t k g {f WB}}  
 {NOM\_MACHINE(vocaliza)} {{v WB} o k a l i z {a WB}}  
 {NOM\_MACHINE(pve)} {{p WB} e v e {& WB}}  
 {NOM\_MOIS} {{b WB} p t k g {f WB}}  
 {NOM\_MOIS(août)} {{u WB} {t WB}}  
 {NOM\_MOIS(août\_2)} {u}  
 {NOM\_MOIS(août\_3)} {{u WB} t}  
 {NOM\_MOIS(avril)} {{a WB} v R i {l WB}}  
 {NOM\_MOIS(décembre)} {{d WB} e s a n b {R WB}}  
 {NOM\_MOIS(décembre\_2)} {{d WB} e s a n b R {& WB}}  
 {NOM\_MOIS(décembre\_3)} {{d WB} e E s a n b {R WB}}  
 {NOM\_MOIS(décembre\_4)} {{d WB} e E s a n b R {& WB}}  
 {NOM\_MOIS(février)} {{f WB} e v R i {e WB}}  
 {NOM\_MOIS(février\_2)} {{f WB} e E v R i {e WB}}  
 {NOM\_MOIS(janvier)} {{Z WB} a n v j {e WB}}  
 {NOM\_MOIS(juillet)} {{Z WB} H i j {e WB}}  
 {NOM\_MOIS(juillet\_2)} {{Z WB} H i j {e E WB}}  
 {NOM\_MOIS(juin)} {{Z WB} H {in WB}}  
 {NOM\_MOIS(juin\_2)} {{Z WB} o {in WB}}  
 {NOM\_MOIS(juin\_3)} {{Z WB} w {in WB}}  
 {NOM\_MOIS(mai)} {{m WB} {e E WB}}  
 {NOM\_MOIS(mars)} {{m WB} a A R {s WB}}

{NOM\_MOIS(novembre)} {{n WB} O v a n b {R WB}}  
 {NOM\_MOIS(novembre\_2)} {{n WB} O v a n b R {& WB}}  
 {NOM\_MOIS(novembre\_3)} {{n WB} oO v a n b {R WB}}  
 {NOM\_MOIS(novembre\_4)} {{n WB} oO v a n b R {& WB}}  
 {NOM\_MOIS(octobre)} {{O WB} k t O b {R WB}}  
 {NOM\_MOIS(octobre\_2)} {{O WB} k t O b R {& WB}}  
 {NOM\_MOIS(octobre\_3)} {{oO WB} k t O b {R WB}}  
 {NOM\_MOIS(octobre\_4)} {{oO WB} k t O b R {& WB}}  
 {NOM\_MOIS(septembre)} {{s WB} E p t a n b {R WB}}  
 {NOM\_MOIS(septembre\_2)} {{s WB} E p t a n b R {& WB}}  
 {NOM\_PAYS} {{b WB} p t k g {f WB}}  
 {NOM\_PAYS(france)} {{f WB} R a n {s WB}}  
 {NOM\_PAYS(italie)} {{i WB} t a l {i WB}}  
 {NOM\_PAYS(japon)} {{Z WB} a p {on WB}}  
 {NOM\_PAYS(suisse)} {{s WB} y i {s WB}}  
 {NOM\_PAYS(suisse\_2)} {{s WB} H i s {& WB}}  
 {NOM\_PAYS(usa)} {{y WB} E s {a WB}}  
 {NOM\_PERSONNE} {{b WB} p t k g {f WB}}  
 {NOM\_PERSONNE(adenot)} {{a WB} d 2 n {o WB}}  
 {NOM\_PERSONNE(akbar)} {{a WB} k b a {R WB}}  
 {NOM\_PERSONNE(andr e)} {{an WB} d R {e WB}}  
 {NOM\_PERSONNE(andry)} {{an WB} d R {i WB}}  
 {NOM\_PERSONNE(argeles)} {{a WB} R Z 2 l E {s WB}}  
 {NOM\_PERSONNE(argentier)} {{a WB} R Z a n t j {e WB}}  
 {NOM\_PERSONNE(becquard)} {{b WB} e k {E WB}}  
 {NOM\_PERSONNE(besacier)} {{b WB} & z a s i {e WB}}  
 {NOM\_PERSONNE(besacier\_2)} {{b WB} 2 z a s i {e WB}}  
 {NOM\_PERSONNE(blanc)} {{b WB} l {an WB}}  
 {NOM\_PERSONNE(boisdon)} {{b WB} w a d {on WB}}  
 {NOM\_PERSONNE(boitet)} {{b WB} w a t {e WB}}  
 {NOM\_PERSONNE(bougnoux)} {{b WB} u N J {u WB}}  
 {NOM\_PERSONNE(caelen)} {{k WB} a E l E {n WB}}  
 {NOM\_PERSONNE(carton)} {{k WB} a R t {on WB}}  
 {NOM\_PERSONNE(chabuel)} {{S WB} a b y E {l WB}}  
 {NOM\_PERSONNE(coquelle)} {{k WB} O k E {l WB}}  
 {NOM\_PERSONNE(cornut)} {{k WB} O R n {y WB}}  
 {NOM\_PERSONNE(cott e)} {{k WB} O t {e WB}}  
 {NOM\_PERSONNE(courteau)} {{k WB} u R t {o WB}}  
 {NOM\_PERSONNE(cuccaro)} {{k WB} k a R {o WB}}  
 {NOM\_PERSONNE(demongeot)} {{d WB} 2 m o n Z {o WB}}  
 {NOM\_PERSONNE(degery)} {{d WB} e Z 2 R {i WB}}  
 {NOM\_PERSONNE(descalle)} {{d WB} E s k a {l WB}}  
 {NOM\_PERSONNE(descalle\_2)} {{d WB} E z k a {l WB}}  
 {NOM\_PERSONNE(dupond)} {{d WB} y p {on WB}}  
 {NOM\_PERSONNE(dupuis)} {{d WB} y p y {i WB}}  
 {NOM\_PERSONNE(durand)} {{d WB} y R {an WB}}  
 {NOM\_PERSONNE(fagot)} {{f WB} a g {o WB}}  
 {NOM\_PERSONNE(ferrand)} {{f WB} e R {an WB}}  
 {NOM\_PERSONNE(fouquet)} {{f WB} u k {E WB}}  
 {NOM\_PERSONNE(garrel)} {{g WB} a R E {l WB}}  
 {NOM\_PERSONNE(gatia)} {{g WB} a s j {a WB}}  
 {NOM\_PERSONNE(gehard)} {{Z WB} e a {R WB}}  
 {NOM\_PERSONNE(giroud)} {{Z WB} i R {u WB}}  
 {NOM\_PERSONNE(gorius)} {{g WB} O r i y {s WB}}  
 {NOM\_PERSONNE(grandvaux)} {{g WB} R a n v {o WB}}  
 {NOM\_PERSONNE(gravil)} {{g WB} R i v E {l WB}}  
 {NOM\_PERSONNE(hinaux)} {{i WB} n {o WB}}  
 {NOM\_PERSONNE(hollard)} {{o WB} l a {R WB}}  
 {NOM\_PERSONNE(ittel)} {{i WB} t E {l WB}}  
 {NOM\_PERSONNE(joseph)} {{Z WB} O z E {f WB}}  
 {NOM\_PERSONNE(joubert)} {{Z WB} u b E {R WB}}  
 {NOM\_PERSONNE(kirshner)} {{k WB} i R S n E {R WB}}  
 {NOM\_PERSONNE(millien)} {{m WB} i l j {in WB}}  
 {NOM\_PERSONNE(mollard)} {{m WB} O l a {R WB}}  
 {NOM\_PERSONNE(nombret)} {{n WB} o n b R {E WB}}  
 {NOM\_PERSONNE(noraz)} {{n WB} O R a {z WB}}  
 {NOM\_PERSONNE(p ene)} {{p WB} E n {& WB}}  
 {NOM\_PERSONNE(piolat)} {{p WB} j o l {a WB}}  
 {NOM\_PERSONNE(piolat\_2)} {{p WB} i o l {a WB}}  
 {NOM\_PERSONNE(qualizza)} {{k WB} a l i z {a WB}}  
 {NOM\_PERSONNE(qualizza\_2)} {{k WB} w a l i z {a WB}}  
 {NOM\_PERSONNE(reverdi)} {{R WB} 2 v E R d {i WB}}  
 {NOM\_PERSONNE(schlecht)} {{a WB} k b a {R WB}}  
 {NOM\_PERSONNE(serignat)} {{s WB} E r i N J {a WB}}  
 {NOM\_PERSONNE(vaufreydaz)} {{v WB} o f R e d a {z WB}}  
 {NOM\_VILLE} {{b WB} p t k g {f WB}}  
 {NOM\_VILLE(chamb ery)} {{S WB} a n b e R {i WB}}  
 {NOM\_VILLE(gen eve)} {{Z WB} 2 n E {v WB}}  
 {NOM\_VILLE(grenoble)} {{g WB} R & n O b {l WB}}  
 {NOM\_VILLE(lyon)} {{l WB} j {on WB}}  
 {NOM\_VILLE(nantes)} {{n WB} a n t {& WB}}  
 {NOM\_VILLE(new-york)} {{n WB} u j o R {k WB}}  
 {NOM\_VILLE(new-york\_2)} {{n WB} y u j o R {k WB}}  
 {NOM\_VILLE(new-york\_3)} {{n WB} y u j O R {k WB}}  
 {NOM\_VILLE(Paris)} {{p WB} a R {i WB}}  
 {OUVERTURE} {{b WB} p t k g {f WB}}  
 {OUVERTURE(all o)} {{a WB} l {o WB}}  
 {OUVERTURE(bonjour)} {{b WB} o n Z u {R WB}}  
 {OUVERTURE(bonsoir)} {{b WB} o n s w a A {R WB}}  
 {OUVERTURE(salut)} {{s WB} a l {y WB}}  
 {PRENOM\_PERSONNE} {{b WB} p t k g {f WB}}  
 {PRENOM\_PERSONNE(alain)} {{a WB} l {in WB}}  
 {PRENOM\_PERSONNE(alexis)} {{a WB} l E k s {i WB}}  
 {PRENOM\_PERSONNE(alexis\_2)} {{i WB} e t {i WB}}  
 {PRENOM\_PERSONNE(anne)} {{a WB} n {& WB}}  
 {PRENOM\_PERSONNE(anne-claire)} {{a WB} n k l E {R WB}}  
 {PRENOM\_PERSONNE(anne-claire\_2)} {{a WB} n & k l E {R WB}}  
 {PRENOM\_PERSONNE(anne-lise)} {{a WB} n l i {z WB}}  
 {PRENOM\_PERSONNE(anne-marie)} {{a WB} n m a R {i WB}}  
 {PRENOM\_PERSONNE(anne-marie\_2)} {{a WB} n & m a R {i WB}}  
 {PRENOM\_PERSONNE(annie)} {{a WB} n {i WB}}  
 {PRENOM\_PERSONNE(aristide)} {{a WB} R i s t i d {& WB}}  
 {PRENOM\_PERSONNE(aristide\_2)} {{a WB} R i s t i {d WB}}  
 {PRENOM\_PERSONNE(bertrand)} {{b WB} E R t R {an WB}}  
 {PRENOM\_PERSONNE(brieuc)} {{b WB} R i {2 WB}}  
 {PRENOM\_PERSONNE(brigitte)} {{k WB} R i Z i {t WB}}  
 {PRENOM\_PERSONNE(carol)} {{k WB} a R O {l WB}}  
 {PRENOM\_PERSONNE(carole)} {{k WB} a R O {l WB}}  
 {PRENOM\_PERSONNE(carole\_2)} {{k WB} a R O l {& WB}}  
 {PRENOM\_PERSONNE(catherine)} {{k WB} a t 2 R i {n WB}}  
 {PRENOM\_PERSONNE(catherine\_2)} {{k WB} a t 2 R i n {& WB}}  
 {PRENOM\_PERSONNE(c eline)} {{s WB} e l i {n WB}}  
 {PRENOM\_PERSONNE(c eline\_2)} {{s WB} e l i n {& WB}}  
 {PRENOM\_PERSONNE(chlo e)} {{k WB} l o {e WB}}  
 {PRENOM\_PERSONNE(christian)} {{k WB} R i s t i {an WB}}  
 {PRENOM\_PERSONNE(christine)} {{k WB} R i s t i {n WB}}  
 {PRENOM\_PERSONNE(christine\_2)} {{k WB} R i s t i n {& WB}}  
 {PRENOM\_PERSONNE(christophe)} {{k WB} R i s t o {f WB}}  
 {PRENOM\_PERSONNE(christophe\_2)} {{k WB} R i s t o f {& WB}}  
 {PRENOM\_PERSONNE(chrystel)} {{k WB} R i s t E {l WB}}  
 {PRENOM\_PERSONNE(claire)} {{k WB} l E {R WB}}  
 {PRENOM\_PERSONNE(clotilde)} {{k WB} l o t i l {d WB}}  
 {PRENOM\_PERSONNE(clotilde\_2)} {{k WB} l o t i l d {& WB}}  
 {PRENOM\_PERSONNE(danielle)} {{d WB} a n i E {l WB}}  
 {PRENOM\_PERSONNE(danielle\_2)} {{d WB} a n i E l {& WB}}  
 {PRENOM\_PERSONNE(davy)} {{d WB} e v {i WB}}  
 {PRENOM\_PERSONNE(delfine)} {{d WB} E l f i {n WB}}  
 {PRENOM\_PERSONNE(delfine\_2)} {{d WB} E l f i n {& WB}}  
 {PRENOM\_PERSONNE(dominique)} {{d WB} O m i n i {k WB}}  
 {PRENOM\_PERSONNE(dominique\_2)} {{d WB} o m i n i {k WB}}  
 {PRENOM\_PERSONNE(elisabeth)} {{e WB} l i z a b E {t WB}}  
 {PRENOM\_PERSONNE(eric)} {{e WB} R i {k WB}}  
 {PRENOM\_PERSONNE(elys e)} {{e WB} v & l i n {& WB}}  
 {PRENOM\_PERSONNE(franck)} {{f WB} R a n {k WB}}  
 {PRENOM\_PERSONNE(fran ois)} {{f WB} R a n s u {a WB}}  
 {PRENOM\_PERSONNE(fran oise)} {{f WB} R a n s u a {z WB}}  
 {PRENOM\_PERSONNE(fran oise\_2)} {{f WB} R a n s u a z {& WB}}  
 {PRENOM\_PERSONNE(fr ed erique)} {{f WB} R e d e R i {k WB}}  
 {PRENOM\_PERSONNE(fr ed erique\_2)} {{f WB} R e d e R i k {& WB}}  
 {PRENOM\_PERSONNE(fr ed erique\_3)} {{f WB} R e d e R i {d WB}}  
 {PRENOM\_PERSONNE(gu enaelle)} {{g WB} e n a E {l WB}}  
 {PRENOM\_PERSONNE(gu enaelle\_2)} {{g WB} e n a E l {& WB}}  
 {PRENOM\_PERSONNE(h el ene)} {{e WB} l E {n WB}}  
 {PRENOM\_PERSONNE(h el ene\_2)} {{e WB} l E n {& WB}}  
 {PRENOM\_PERSONNE(henriette)} {{an WB} R i E t {& WB}}  
 {PRENOM\_PERSONNE(herv e)} {{E WB} R v {e WB}}

{PRENOM\_PERSONNE(isabelle)} {{i WB} z a b E {l WB}}  
 {PRENOM\_PERSONNE(isabelle\_2)} {{i WB} z a b E l {& WB}}  
 {PRENOM\_PERSONNE(jean)} {{Z WB} {an WB}}  
 {PRENOM\_PERSONNE(jean\_2)} {{d WB} Z i {n WB}}  
 {PRENOM\_PERSONNE(jean-christophe)} {{Z WB} an k R i s t O {f WB}}  
 {PRENOM\_PERSONNE(jean-christophe\_2)} {{Z WB} an k R i s t O f {& WB}}  
 {PRENOM\_PERSONNE(jean-christophe\_3)} {{Z WB} i s {e WB}}  
 {PRENOM\_PERSONNE(jean-francois)} {{Z WB} an f R an s u {a WB}}  
 {PRENOM\_PERSONNE(jean-francois\_2)} {{d WB} Z E {f WB}}  
 {PRENOM\_PERSONNE(jean-luc)} {{Z WB} an l y {k WB}}  
 {PRENOM\_PERSONNE(jean-rené)} {{Z WB} an R & n {e WB}}  
 {PRENOM\_PERSONNE(jérémie)} {{Z WB} e R e m {i WB}}  
 {PRENOM\_PERSONNE(jérôme)} {{Z WB} e R O {m WB}}  
 {PRENOM\_PERSONNE(jérôme\_2)} {{Z WB} e R O m {& WB}}  
 {PRENOM\_PERSONNE(jocelyn)} {{Z WB} O s & l {in WB}}  
 {PRENOM\_PERSONNE(jocelyn\_2)} {{Z WB} O s l {in WB}}  
 {PRENOM\_PERSONNE(joel)} {{Z WB} o E {l WB}}  
 {PRENOM\_PERSONNE(julien)} {{Z WB} y l i {in WB}}  
 {PRENOM\_PERSONNE(juliette)} {{Z WB} y l i E {t WB}}  
 {PRENOM\_PERSONNE(juliette\_2)} {{Z WB} y l i E t {& WB}}  
 {PRENOM\_PERSONNE(karine)} {{k WB} a R i {n WB}}  
 {PRENOM\_PERSONNE(karine\_2)} {{k WB} a R i n {& WB}}  
 {PRENOM\_PERSONNE(laetitia)} {{l WB} e t i s j {a WB}}  
 {PRENOM\_PERSONNE(laure)} {{l WB} o R {& WB}}  
 {PRENOM\_PERSONNE(laurent)} {{l WB} o R {an WB}}  
 {PRENOM\_PERSONNE(mado)} {{m WB} a d {o WB}}  
 {PRENOM\_PERSONNE(maelle)} {{m WB} a E {l WB}}  
 {PRENOM\_PERSONNE(maelle\_2)} {{m WB} a E l {& WB}}  
 {PRENOM\_PERSONNE(marcello)} {{m WB} a R t Z E l {o WB}}  
 {PRENOM\_PERSONNE(marie)} {{m WB} a R {i WB}}  
 {PRENOM\_PERSONNE(marie-agnès)} {{m WB} a R i a N J E {s WB}}  
 {PRENOM\_PERSONNE(marie-madelaide)} {{m WB} a R i m a d & l E {n WB}}  
 {PRENOM\_PERSONNE(marie-odile)} {{m WB} a R i O d i l {& WB}}  
 {PRENOM\_PERSONNE(marie-pierre)} {{m WB} a R i p j E {R WB}}  
 {PRENOM\_PERSONNE(marie-pierre\_2)} {{m WB} a R i p j E R {& WB}}  
 {PRENOM\_PERSONNE(marie-pierre\_3)} {{m WB} a R i {p WB}}  
 {PRENOM\_PERSONNE(marine)} {{m WB} a R i {n WB}}  
 {PRENOM\_PERSONNE(marine\_2)} {{m WB} a R i n {& WB}}  
 {PRENOM\_PERSONNE(marjolaine)} {{m WB} a R Z o l e n {& WB}}  
 {PRENOM\_PERSONNE(marjolaine\_2)} {{m WB} a R Z o l E {n WB}}  
 {PRENOM\_PERSONNE(marjolaine\_3)} {{m WB} a R Z {o WB}}  
 {PRENOM\_PERSONNE(myriam)} {{m WB} i R j a {m WB}}  
 {PRENOM\_PERSONNE(nadège)} {{n WB} a d E {Z WB}}  
 {PRENOM\_PERSONNE(nadège\_2)} {{n WB} a d E Z {& WB}}  
 {PRENOM\_PERSONNE(nathalie)} {{n WB} a t a l {i WB}}  
 {PRENOM\_PERSONNE(nicolas)} {{n WB} i k o l {a WB}}  
 {PRENOM\_PERSONNE(noémie)} {{n WB} o e m {i WB}}  
 {PRENOM\_PERSONNE(odile)} {{o WB} d i {l WB}}  
 {PRENOM\_PERSONNE(odile\_2)} {{o WB} d i l {& WB}}  
 {PRENOM\_PERSONNE(olivier)} {{o WB} l i v j {e WB}}  
 {PRENOM\_PERSONNE(olivier\_2)} {{b WB} u f {i WB}}  
 {PRENOM\_PERSONNE(pascaline)} {{p WB} a s k a l i {n WB}}  
 {PRENOM\_PERSONNE(pascaline\_2)} {{p WB} a s k l i n {& WB}}  
 {PRENOM\_PERSONNE(patrick)} {{p WB} a t r i {k WB}}  
 {PRENOM\_PERSONNE(patrick\_2)} {{p WB} a t O {S WB}}  
 {PRENOM\_PERSONNE(pierre)} {{p WB} j E {R WB}}  
 {PRENOM\_PERSONNE(pierre\_2)} {{p WB} j E R {& WB}}  
 {PRENOM\_PERSONNE(revol)} {{R WB} e v O {l WB}}  
 {PRENOM\_PERSONNE(roland)} {{R WB} o l {an WB}}  
 {PRENOM\_PERSONNE(sabine)} {{s WB} a b i {n WB}}  
 {PRENOM\_PERSONNE(sabine\_2)} {{s WB} a b i n {& WB}}  
 {PRENOM\_PERSONNE(sami)} {{s WB} a m {i WB}}  
 {PRENOM\_PERSONNE(samuel)} {{s WB} a m u E {l WB}}  
 {PRENOM\_PERSONNE(samuel\_2)} {{s WB} a {m WB}}  
 {PRENOM\_PERSONNE(sandrine)} {{s WB} an d R i {n WB}}  
 {PRENOM\_PERSONNE(sandrine\_2)} {{s WB} an d R i n {& WB}}  
 {PRENOM\_PERSONNE(sébastien)} {{s WB} e b a s t j {in WB}}  
 {PRENOM\_PERSONNE(sébastien\_2)} {{s WB} E {b WB}}  
 {PRENOM\_PERSONNE(serge)} {{s WB} E R Z {& WB}}  
 {PRENOM\_PERSONNE(serge\_2)} {{s WB} E R Z {& WB}}  
 {PRENOM\_PERSONNE(serge\_3)} {{s WB} E R Z j {o WB}}  
 {PRENOM\_PERSONNE(serge\_4)} {{s WB} E R t Z j {o WB}}  
 {PRENOM\_PERSONNE(séverine)} {{s WB} e v & R i {n WB}}  
 {PRENOM\_PERSONNE(séverine\_2)} {{s WB} e v & R i n {& WB}}  
 {PRENOM\_PERSONNE(simone)} {{s WB} i m O {n WB}}  
 {PRENOM\_PERSONNE(simone\_2)} {{s WB} i m O n {& WB}}  
 {PRENOM\_PERSONNE(solange)} {{s WB} o l a n {Z WB}}  
 {PRENOM\_PERSONNE(solemn)} {{s WB} o l E {n WB}}  
 {PRENOM\_PERSONNE(sophie)} {{s WB} o f {i WB}}  
 {PRENOM\_PERSONNE(sophie\_2)} {{s WB} O f {i WB}}  
 {PRENOM\_PERSONNE(stephanie)} {{s WB} t e f a n {i WB}}  
 {PRENOM\_PERSONNE(sylvain)} {{s WB} i l v {in WB}}  
 {PRENOM\_PERSONNE(sylvie)} {{s WB} i l v {i WB}}  
 {PRENOM\_PERSONNE(valérie)} {{v WB} a l e R {i WB}}  
 {PRENOM\_PERSONNE(véronique)} {{v WB} e R o n i {k WB}}  
 {PRENOM\_PERSONNE(véronique\_2)} {{v WB} e R {o WB}}  
 {PRENOM\_PERSONNE(yannick)} {{j WB} a n i {k WB}}  
 {PRENOM\_PERSONNE(yannick\_2)} {{f WB} u k {s WB}}  
 {PRENOM\_PERSONNE(yves)} {{i WB} {v WB}}  
 {PRENOM\_PERSONNE(yves\_2)} {{i WB} v {& WB}}  
 {affolant} {{a WB} f O l {an WB}}  
 {affolant(2)} {{a WB} f o O l {an WB}}  
 {affolant(3)} {{a WB} f O l a n {t WB}}  
 {calculé} {{k WB} a l k y l {e WB}}  
 {annulées} {{a WB} n y l {e WB}}  
 {annulées(2)} {{a WB} n y l e {z WB}}  
 {annulés} {{a WB} n y l {e WB}}  
 {annulés(2)} {{a WB} n y l e {z WB}}  
 {appelais} {{a WB} p & l {E WB}}  
 {appelais(3)} {{a WB} p & l E {z WB}}  
 {appelais(2)} {{a WB} p l {E WB}}  
 {appellerait} {{a WB} p E l & R {E WB}}  
 {appellerait(2)} {{a WB} p E l & R E {t WB}}  
 {arobase} {{a WB} R o b a z {z WB}}  
 {aïe} {{a WB} {i WB}}  
 {aïe(2)} {{a WB} i {& WB}}  
 {aïe(3)} {{a WB} j {WB}}  
 {aïe(4)} {{a WB} j {& WB}}  
 {bise} {{b WB} i z {& WB}}  
 {bises} {{b WB} i z {& WB}}  
 {bises(2)} {{b WB} i z e {z WB}}  
 {boulette} {{b WB} u l E {t WB}}  
 {boulette(2)} {{b WB} u l E t {& WB}}  
 {boulettes} {{b WB} u l E t {z WB}}  
 {boulettes(2)} {{b WB} u l E {t WB}}  
 {bousculé} {{b WB} u s k y l {e WB}}  
 {bécane} {{b WB} e k a n {& WB}}  
 {bécanes} {{b WB} e k a n {& WB}}  
 {bécanes(2)} {{b WB} e k a n & {z WB}}  
 {ciao} {{t WB} Z a {o WB}}  
 {copieur} {{k WB} O p j 2 {R WB}}  
 {copieurs} {{k WB} O p j 2 {R WB}}  
 {d'ac} {{d WB} a {k WB}}  
 {d'accord} {{d WB} a A k O {R WB}}  
 {d'accord(2)} {{d WB} a k O {R WB}}  
 {dicter} {{d WB} i k t {e WB}}  
 {dicter(2)} {{d WB} i k t e {R WB}}  
 {débrouiller} {{d WB} E b R u j {e WB}}  
 {démon} {{d WB} E m {o WB}}  
 {démon(2)} {{d WB} E m o {z WB}}  
 {dépanner} {{d WB} e p a n {e WB}}  
 {dépanner(2)} {{d WB} e p a n e {R WB}}  
 {déranger} {{d WB} E R a n Z e {R WB}}  
 {déranger(2)} {{d WB} E R a n Z {e WB}}  
 {email} {{i WB} m E {l WB}}  
 {email\_2} {{i WB} m E j {l WB}}  
 {email\_3} {{k WB} u R j E {l WB}}  
 {email\_4} {{m WB} E {l WB}}  
 {email\_5} {{m WB} E j {l WB}}  
 {embrouillé} {{an WB} b R u j {E WB}}  
 {embête} {{an WB} b E t {& WB}}  
 {embête(2)} {{an WB} b E {t WB}}  
 {embêté} {{an WB} b E t {e WB}}



{embêtée} {{an WB} b E t {e WB}}  
 {ennuyée} {{an WB} n H i j {e WB}}  
 {ennuyé} {{an WB} n H i j {e WB}}  
 {excusé} {{E WB} k s k y z {e WB}}  
 {excusée} {{E WB} k s k y z {e WB}}  
 {externalisé} {{E WB} k s t E R n a l i z {e WB}}  
 {feras} {{f WB} & R {a WB}}  
 {feras(2)} {{f WB} & R a {z WB}}  
 {fff} {{f WB} f {f WB}}  
 {férié} {{f WB} e E R j {e WB}}  
 {férié(2)} {{f WB} E R j {e WB}}  
 {gérait} {{Z WB} E R {E WB}}  
 {gérait(3)} {{Z WB} E R E {t WB}}  
 {gérait(2)} {{Z WB} e R {E WB}}  
 {d'hab} {{d WB} a {b WB}}  
 {indiquerez} {{in WB} d i k & R {e WB}}  
 {indiquerez(2)} {{in WB} d i k & R e {z WB}}  
 {jeanne} {{Z WB} a {n WB}}  
 {jeanne(2)} {{Z WB} a n {& WB}}  
 {labo} {{l WB} a b {O WB}}  
 {login} {{l WB} O g i {n WB}}  
 {maintiens} {{m WB} i n t j {in WB}}  
 {maintiens(2)} {{m WB} i n t j i n {z WB}}  
 {mettrai} {{m WB} E t R {E WB}}  
 {mettrais} {{m WB} E t R {E WB}}  
 {mettrais(2)} {{m WB} E t R E {s WB}}  
 {ok} {{O WB} k {E WB}}  
 {ok(2)} {{O WB} k {e WB}}  
 {okay} {{O WB} k {E WB}}  
 {okay(2)} {{O WB} k {e WB}}  
 {okai} {{O WB} k {E WB}}  
 {okai(2)} {{O WB} k {e WB}}  
 {ouh} {u}  
 {parles} {{p WB} a R l {& WB}}  
 {parles(2)} {{p WB} a R l & {s WB}}  
 {pff} {{p WB} f {f WB}}  
 {plantée} {{p WB} l a n t {e WB}}  
 {poisse} {{p WB} w a s {& WB}}  
 {poisse(2)} {{p WB} w a {s WB}}  
 {pressez} {{p WB} R E s e {z WB}}  
 {pressez(2)} {{p WB} R E s {e WB}}  
 {promettent} {{p WB} R O m E t {& WB}}  
 {promettent(2)} {{p WB} R O m E {t WB}}  
 {promettent(3)} {{p WB} R O m E t & {t WB}}  
 {prévienne} {{p WB} R E v j E {n WB}}  
 {prévienne(2)} {{p WB} R E v j E n {& WB}}  
 {raccroche} {{R WB} a k R O S {& WB}}  
 {raccroche(2)} {{R WB} a k R O {S WB}}  
 {raccroché} {{R WB} a k R O S {e WB}}  
 {rassurées} {{R WB} a s y R e {z WB}}  
 {rassurées(2)} {{R WB} a s y R {e WB}}  
 {rassurée} {{R WB} a s y R {e WB}}  
 {rassuré} {{R WB} a s y R {e WB}}  
 {redémarrer} {{R WB} 2 d E m a R {e WB}}  
 {redémarrer(3)} {{R WB} 2 d E m a R e {R WB}}  
 {redémarrer(2)} {{R WB} 2 d e m a R {e WB}}  
 {regardes} {{R WB} & g a R {d WB}}  
 {regardes(2)} {{R WB} & g a R d {& WB}}  
 {renégocier} {{R WB} 2 n E g O s j {e WB}}  
 {renégocier(3)} {{R WB} 2 n E g O s j e {R WB}}  
 {renégocier(2)} {{R WB} 2 n e g O s j {e WB}}  
 {reprenne} {{R WB} & p R E {n WB}}  
 {reprenne(2)} {{R WB} & p R E n {& WB}}  
 {réceptionniste} {{R WB} E s E p s j o n i s {t WB}}  
 {réceptionniste(2)} {{R WB} E s E p s j o n i s t {& WB}}  
 {sauriez} {{s WB} O R j {e WB}}  
 {sauriez(2)} {{s WB} O R j e {z WB}}  
 {sss} {{s WB} s {s WB}}  
 {tit} {{t WB} i {t WB}}  
 {tit(2)} {{t WB} {i WB}}  
 {tondeuse} {{t WB} o n d 2 {z WB}}  
 {tondeuse(2)} {{t WB} o n d 2 z {& WB}}  
 {transmette} {{t WB} R a n s m E t {& WB}}  
 {transmette(2)} {{t WB} R a n s m E {t WB}}  
 {trompée} {{t WB} R o n p {e WB}}  
 {ttendez} {{t WB} a n d e {z WB}}  
 {ttendez(2)} {{t WB} a n d {e WB}}  
 {verras} {{v WB} E R {a WB}}  
 {verras(2)} {{v WB} e R {a WB}}  
 {visioconférences} {{v WB} i z j o k o n f E R a n {s WB}}  
 {visioconférences(3)} {{v WB} i z j o k o n f E R a n s {z WB}}  
 {visioconférences(2)} {{v WB} i z j o k o n f e R a n {s WB}}  
 {voilou} {{v WB} w a l {u WB}}  
 {yep} {{j WB} E {p WB}}  
 {yop} {{j WB} O {p WB}}



---

## **Modélisation des attentes en dialogue oral**

### Résumé :

Cette thèse porte sur la modélisation des attentes du locuteur en dialogue humain et en dialogue homme-machine. Les attentes du locuteur sont vues comme les résultats espérés de ses actes, selon l'hypothèse qu'un individu en situation de dialogue finalisé (oral ou non, humain ou homme-machine) aura certaines attentes vis-à-vis de son allocutaire suite à son énoncé.

La théorie développée puise ses sources dans la philosophie analytique et sa logique dialogique dérivée des actes de langage, dans l'éthnométhodologie et la théorie des faces, dans les cognosciences et la théorie de la pertinence, et enfin dans la linguistique. Une notation en actes de dialogue a été établie pour représenter les connaissances, les rôles et les actes de langage des interactants. La modélisation considère des règles d'association acte-attentes assorties de probabilités, les stratégies employées et les buts dialogiques.

Pour évaluer les règles et les probabilités, deux corpus ont été collectés : l'un en dialogue humain (dans un cadre de traduction de parole), l'autre en dialogue homme-machine (pour développer un assistant d'entreprise virtuel). Pour la collecte du deuxième, une plate-forme à base de magicien d'Oz a été réalisée afin de simuler le système final. Les corpus recueillis ont été transcrits, annotés et analysés en vue de montrer l'intérêt de considérer les attentes plutôt que de faire de la simple prédiction d'acte.

Les résultats sont intéressants pour le dialogue humain et pour le dialogue homme-machine : la prise en compte des attentes améliore l'identification des actes de dialogue des interlocuteurs, et le contrôle du dialogue.

Mots-clés : attentes, dialogue homme-homme, dialogue homme-machine, actes de langage, magicien d'Oz

---

## **Modelization of Expectations in Oral Dialogue**

### Summary:

This thesis aims at modelling speaker's expectations in human dialogue and in man-machine dialogue. Speaker's expectations are the expected results of its acts, assuming that someone in finalized dialogue (oral or not, human or man-machine) has certain expectations from its addressee following his statement.

The theory developed draws its sources from analytic philosophy and its speech acts, from ethnomethodology through the faces theory, from cognosciences and the theory of relevance, and from linguistics. A notation of dialog acts has been made to represent knowledge, roles and speech acts of participants. The model considers rules - associating acts and expectations - supplied with probabilities, strategies and dialogic goals.

For rules and probabilities evaluation, two corpora were collected: first with human dialogue (within a framework of speech translation), second with man-machine dialogue (to develop a virtual assistant). For the collection of the second one, a wizard of Oz based platform has been made in order to simulate the final system. The corpora collected were transcribed, annotated and analysed in order to show the interest of considering expectations rather than making simple speech acts prediction.

The results are interesting for both human and man-machine dialogue: considering the expectations improves identification of interlocutor dialogue acts, and dialogue management.

Keywords: expectations, human dialogue, man-machine dialogue, speech act, wizard of Oz