Analyse de dialogues finalisés et simulés

J. Caelen
CLIPS-IMAG, Grenoble
Jean.Caelen@imag.fr

1. Introduction

Les études sur le dialogue homme-machine nécessitent une analyse fine des processus de dialogue et du langage utilisé par les usagers placés en situation de communication avec la machine. Trois grands problèmes se posent pour recueillir des corpus significatifs et les analyser objectivement :
1) la mise en situation et la capture d'énoncés "naturels",
2) la transcription et le codage des éléments pertinents du dialogue,
3) la méthode d'analyse.

Nous proposons dans cet article une contribution à ces problèmes à partir de nos propres expériences.

2. Capture d'énoncés

Les conditions de mise en situation sont bien connues par la psychologie expérimentale pour isoler un certain nombre de variables au moyen d'expériences finement contrôlées. Dans le dialogue les variables sont nombreuses et il n'est pas envisageable d'espérer obtenir des données significatives à l'aide d'un seul type d'expérience. Généralement on peut distinguer des situations dans lesquelles :

- l'observateur est dans la boucle d'interaction et participe au dialogue — c'est le cas des méthodes de verbalisation ou d'élicitation de connaissances (on distingue en outre la verbalisation en cours de tâche de la verbalisation en dehors de toute activité),
- l'observateur est hors de la boucle d'interaction et ne participe pas au dialogue proprement dit.

Cette deuxième méthode est apparemment la moins biaisée mais ne permet pas d'orienter la conversation, ce qui peut avoir deux effets opposés chez les sujets : la prolixité pour les uns, un certain mutisme pour les autres.

Pour étudier et modéliser le dialogue homme-machine, seule une méthode itérative est possible puisqu'on ne dispose pas de machine *a priori* : on est obligé de simuler totalement ou en partie des situations de dialogue homme-machine de plus en plus réalistes en partant de dialogues humains. Les différentes étapes pour l'approche d'un dialogue homme-machine réel sont :

- le dialogue humain finalisé (appelé aussi opératif),
- le dialogue homme-machine simulé (par une technique dite de Magicien d'Oz),
- le dialogue homme-machine proprement dit.

Rien ne prouve d'ailleurs que l'étude du dialogue humain ou du dialogue simulé soit pertinente pour le dialogue homme-machine et justifie cette démarche. De leur côté, les éthnométhodologues ne conseillent pas de recourir à de telles méthodes qui biaisent la communication. La psychologie expérimentale préconise au contraire des protocoles précis pour analyser les variables du problème.

Les variables que nous avons retenues sont :

- la complexité de la tâche,
- la richesse sémantique du thème du dialogue,

• les rôles des partenaires dans l'interaction.

La première variable *complexité de la tâche* permet de mesurer l'influence de la tâche sur le dialogue (le raisonnement sous-tendu par la planification de la tâche induit-il à des structures de dialogue particuliers ? Ou inversement, en quoi le dialogue peut-il offrir un cadre structurant au raisonnement ? Retrouve-t-on des marqueurs, des points d'articulation, des stratégies utilisées dans la tâche pour le dialogue ?). La deuxième variable *richesse sémantique* permet de mesurer les effets des connaissances d'arrière-plan dans la compréhension d'un dialogue, et la troisième variable *rôle des partenaires* permet de relativiser les observations obtenues sur des dialogues à des groupes ou des classes d'individus ou d'usagers. Les consignes qui leur sont données en début de session expérimentale jouent en effet pour une bonne part dans les résultats obtenus. Un entretien après la session est donc nécessaire avant l'interprétation des données enregistrées.

Les conditions d'enregistrement influencent le comportement des sujets de l'expérience. Mais pour un traitement ultérieur il est nécessaire de posséder des enregistrements de bonne qualité — par exemple l'enregistrement en chambre sourde nécessaire au traitement automatique de la parole peut enlever une certaine crédibilité à l'expérience pour un sujet. Souvent un enregistrement vidéo permet de compléter les données contextuelles, de noter les expressions faciales ou corporelles et les gestes.

1ère expérience

Pour cette expérience nous avons choisi une tâche de manipulation de signaux à l'aide de plusieurs logiciels d'édition de signal (PTS et Snorri). Ces logiciels ont une interface graphique avec une présentation par menus déroulants et palette d'outils spécialisés. Cette tâche nécessite un niveau d'expertise tant en ce qui concerne le domaine du traitement du signal que la manipulation de l'interface graphique. L'utilisateur est successivement mis en face des deux interfaces qui bien que différentes lui permettent de faire (qualitativement) les mêmes tâches. On lui demande (a) de visualiser un signal enregistré, (b) d'obtenir un sonagramme de ce signal, (c) de visualiser au mieux (en réglant les couleurs, le contraste, etc.) ce sonagramme sur l'écran et (d) d'y faire quelques mesures (d'amplitude, de fréquence). Il peut faire des incidences sur d'autres tâches dans la mesure où il exécute celles-ci dans cet ordre. Les utilisateurs sont des experts du domaine et/ou de l'interface, des occasionnels et des novices. L'observateur n'intervient pas dans les aides ni sur la tâche proprement dite car il n'est pas lui-même compétent dans la réalisation de la tâche ; son rôle se limite à celui d'incarner un apprenant potentiel.

Le corpus a été recueilli sur magnétophone visible par l'utilisateur. Son mode d'expression n'était pas contraint. Il avait en charge la réalisation des tâches dont il a été question ci-dessus sans obligation stricte de les verbaliser toutes. L'échantillon d'utilisateurs allait du novice (trois) au plus expert (sept) en passant par un niveau intermédiaire d'utilisateurs occasionnels ne connaissant, à l'exclusion l'un de l'autre, que l'informatique ou le domaine de la parole (quatre). Le but était de tester la variable *complexité de la tâche* en liaison avec le langage et la structure du dialogue.

2ème expérience

Dans une 2ème expérience au contraire, la complexité de la tâche a été réduite à la description de figures spatiales simples. Le choix de cette focalisation sur la description de figures est motivé par trois facteurs : (a) tout d'abord, parce que la description d'une figure suit un plan décomposable en actes de désignation des objets de cette figure, (b) en deuxième lieu parce que la désignation est le point de convergence entre les modes verbal, gestuel et visuel, modes qui nous intéressent particulièrement pour

la communication homme-machine multimodale, (c) en troisième lieu et plus fondamentalement, parce qu'il nous apparait que le domaine de la désignation spatiale dans un cadre actionnel est particulièrement intéressant pour l'étude du langage, car il permet une confrontation entre le langage et ce qu'il tente d'exprimer. L'expérience met en scène un instructeur et un manipulateur sur deux postes de travail distants. Tous les deux sont des sujets de l'expérience, l'observateur n'est pas visible.

L'expérience est divisée en quatre phases : une première phase préparatoire, où les sujets prennent connaissance des consignes et des contraintes pour chacune des tâches à effectuer et essaient un exemple de chaque tâche, et une phase correspondant à chacune des trois tâches de description. Les tâches consistent essentiellement en des descriptions de figures : l'instructeur doit énoncer des instructions afin que le manipulateur reproduise, au fur et à mesure de ces instructions, chaque figure à l'ordinateur. Les sujets disposent d'un éditeur de dessin fonctionnant en collecticiel.

• La première tâche consiste à décrire six figures abstraites (fig. 1a), deux de chaque type défini par Levelt, dans le but de les faire reproduire à l'écran par le manipulateur.
• La seconde tâche consiste à décrire six figures qui sont structurellement identiques à celles de la première tâche, mais qui représentent des plans de locaux (toujours dans le but de les faire reproduire à l'écran par le manipulateur). Ceci est explicite dans les consignes : les carrés figurent des pièces, les segments sont remplacés par des icônes figurant des portes et des icônes figurant des fenêtres sont rajoutés (fig. 1b).
• La troisième tâche consiste à décrire la position des meubles se trouvant dans les locaux représentés par les plans de la deuxième tâche (fig. 1c). Les icônes "meublant" les pièces figurent des tables ou des chaises.
Nous avons donc affaire d'une tâche à l'autre à un *enrichissement sémantique* du monde de l'application ; c'est la variable-test.

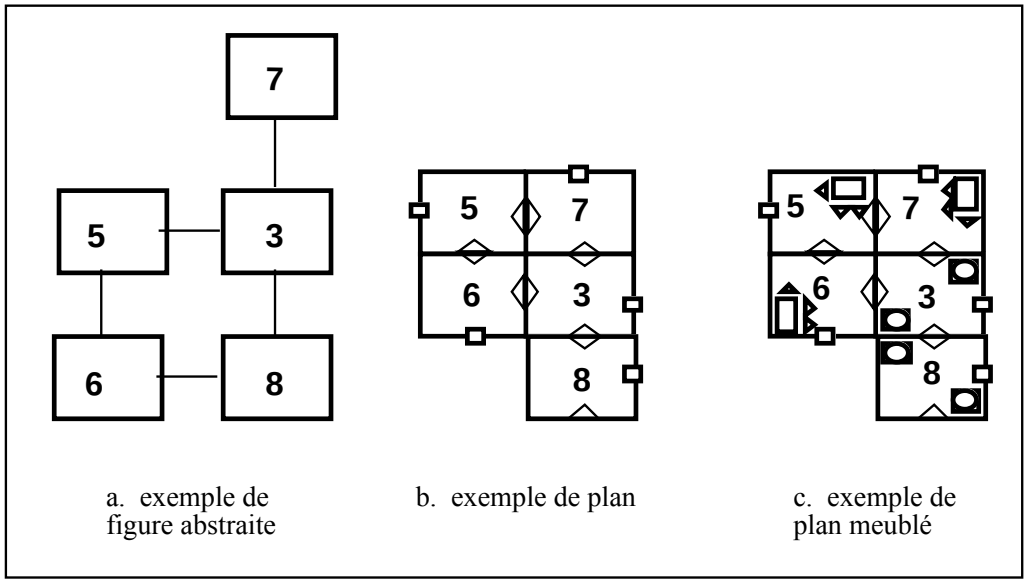

Fig. 1 : Exemples de figures présentées aux sujets avec enrichissement sémantique pour une même structure de forme.

3ème expérience

La 3ème expérience se déroule dans les mêmes conditions que la 2ème. Les figures à reproduire

sont ici des scènes (personnages stylisés, maisons, paysages) toutes construites avec les mêmes éléments géométriques de base (cercles, rectangles, triangles, segments de droite). Des éléments intrus sont parfois ajoutés aux figures de manière à provoquer des ruptures de dialogue entre l'instructeur qui voit les figures mais ne sait par si le manipulateur a les outils nécessaires pour les reproduire et ce dernier qui ne connait pas les figures (les couples de sujet sont renouvelés à chaque session). Ici on cherche à tester la composition de variables *complexité sémantique - rôles dialogiques.*

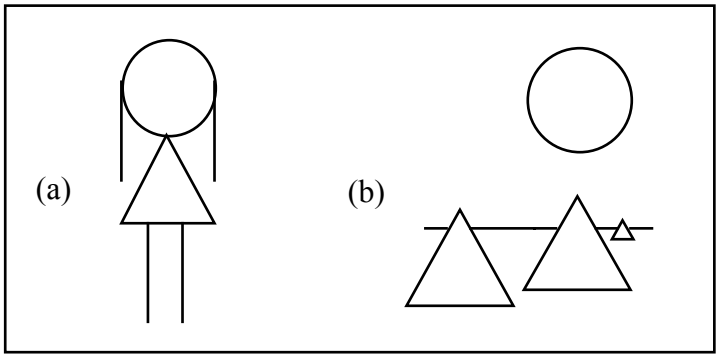

Fig. 2 : Exemples de figures que doivent reproduire les sujets. Ces figures sont toutes constituées des mêmes éléments de base, traits, cercles, triangles, rectangles.

3. Transcription et codage

• Parole

Nous avons transcrit les enregistrements pour obtenir des textes correctement orthographiés, propres à l'analyse automatique et représentatifs de la langue orale liée à l'application. L'objectif de la transcription était d'obtenir une version écrite aisément lisible, aussi fidèle que possible à l'enregistrement sur cassette et exploitable par un logiciel d'analyse de la langue naturelle. Par conséquent, le code phonétique (API) n'a pas été retenu, au profit des mots correctement orthographiés. Pourtant, il a été nécessaire d'introduire un code de signes supplémentaires pour rendre compte de faits prosodiques, de la nature de certaines phrases, de l'altération de segments, de bruits divers, de silences et d'événements extérieurs à l'enquête proprement dite. Ces signes (diacritiques, para-discursifs, extra-discursifs, méta-discursifs) sont les suivants :

< >	métadiscours (l'utilisateur se parle à lui-même)
[ ]	métadiscours (l'utilisateur parle à l'enquêteur ou quelqu'un d'autre)
? !	marque que ce qui précède a été perçu comme interrogatif ou exclamatif par le transcripteur. L'absence de marque indique un énoncé assertif
( )	les caractères notés entre parenthèses n'ont pas été prononcés et sont destinés à faciliter la lecture
e	généralement transcrit par "euh" dans la graphie traditionnelle
:	allongement d'un son ; plus il y a de points, plus l'allongement est important (:/::/:::/)
°	prononciation inattendue d'un son habituellement non prononcé, ex. rouge°, plus°. Exception : le"e" muet en finale d'un mot n'est généralement pas prononcé, cependant pour plus de commodité, nous conservons sa graphie. Lorsqu'il est prononcé, il est suivi du signe °
/	interruption brusque du son précédent, interruption en coup de glotte
**	sons prononcés rapidement, à peine audibles
(+)	claquement de langue
(=)	rire
(m)	son prononcé lèvres closes
(h)	inspiration ou expiration buccale audible

[h]	inspiration ou expiration nasale audible
(f) (pf)	inspiration ou expiration buccale bilabiale
(&)	toux
(°)	sifflement
,	pause
,,	longue pause, silence
/,,/	écoute du signal
/*/	sonnerie
§	intervention d'une autre voix, pouvant provoquer un recouvrement
§§	fin du recouvrement des voix
gras	prononciation appuyée : accent d'intensité, accent d'insistance.
" "	termes propres aux menus du logiciel et aux noms de fenêtre.

Voici un extrait du corpus pour la 1ère expérience :
« (h) j'appelle le programme "PTS", à l'aide de: du clavier (h) et j'appuie sur "return", encore faut-il savoir que: (h) que ** que l(e) programme c'est "PTS" (=) (h) donc j(e) veux une "console graphique" "oui" et j(e) lance° par "return" j(e) veux une po la "position standard des fenêtres" "oui" (h) j'appuie sur "return°" bon j(e) laisse le: le fichier se: se mettre en place pour pouvoir choisir l(e) domaine où je veux aller (h),, (h) donc là: à partir de là (h) j(e) peux choisir e: (h) j(e) dois choisir le domaine d'étude (h) mais e: à partir de la souris puisque j(e) peux pas l'appeler par° commande vocale, choisis l(e) signal temporel puisque c'est c(e) qu'on m(e) fait étudier, alors il faut qu(e) j'appelle il faut qu(e) j'appuie sur le milieu de la souris mais e: ça il faut l(e) savoir aussi donc ça serait quand même plus facile de de parler, (+) et j(e) vais aller dans: "lecture" puisque j(e) vais étudier un: un son, dans l(e) "corpus public" »

- **Gestes**

La deuxième étape du codage consiste à segmenter la parole en actes de langage et à coder les gestes effectués en parallèle (pour cela une exploitation de l'enregistrement vidéo est nécessaire). Un langage de codage des actes de désignation est utilisé dont quelques éléments apparaissent dans l'exemple ci-dessous —dK est enfoncement du boutement de la souris, fK le relâchement, depl () est le déplacement d'un objet et l'opérateur '+' indique la séquentialité des commandes. Nous obtenons alors un tableau tel que celui-ci pour la 3ème expérience :

Voix	Geste - souris	Commentaires
I1: •alors là c'est sensé être des pyramides dans le désert • e ... une à ce niveau...	I2 dK + depl(tri_gd) + fK	I place les pryramides
I3 •et une là •e il faut mettre une barre entre les deux •je vais la mettre mais je crois qu'il faudra déplacer la pyramide	I4 •dK + depl(tri_gd) + fK •dK + depl(hor_gd) + fK	la deuxième pyramide est largement à gauche pour laisser la place : I anticipe le dessin de l'horizon •place une horizontale collée à la pyramide de droite
I5 voilà il faut déplacer la pyramide parce qu'elle doit coller à la barre mais à la même hauteur	M1 dK + depl(tri_gd)	M ajuste la pyramide de gauche, en la déplacant de manière continue
M2 à la même hauteur que celle de droite ?	M3 depl(tri_gd)	

I6	non plus bas plus bas plus bas mais e ...	M4 depl(tri_gd)	
I7	voilà comme ça	M5 fK	la pyramide de gauche est placée
I8	•voilà c'est la même distance •ensuite une pyramide petite •il faut qu'il y ait des distances égales à la petite barre entre la petite pyramide et la grande	I9 dK + depl(tri_pt)	I place petite pyramide sans lâcher la pyramide
I10	e je vais mettre la petite barre	I11 fK	I lâche la pyramide
M6	mettre la petite barre peut être ouais	I12 dK + depl(hor_pt) + fK	I ajuste l'horizon à droite des grandes pyramides
I13	•voilà •je te laisse déplacer la petite pyramide		

Tableau I : séquence des actes langagiers et gestuels dans un dialogue multimodal.

• Actes de langages

La dernière étape de codage fournit pour chaque acte (langagier ou non) ses composantes locutoire, illocutoire et perlocutoire [Austin, 62], [Searle, 69], avec les définitions suivantes* :

Acte : α, locutoire
| actants : {}
| signification {a, o}
| contraintes temporelles
| liens spatiaux

Acte : α, illocutoire
| Type : {f, ff, fs, fp, fd}
| sens : E -> {E, D,m}
| intention en action : vType(b)
| pile_buts : {B}
| effet -> ε
| conséquent : {A}

Acte : α, perlocutoire
| sens : E -> {D,m}
| effet-déontique -> {d(), p()}
| stratégie-attendue : δ
| conséquent : {K}
| conséquent : {c(E)}

*en utilisant la notion de schéma [Minsky, 75]

Composante locutoire

α : nom de l'acte (identificateur),
actants
| objet (quoi)
| destinataire (à qui)
| agent (qui)
| manière (comment)
| instrument (avec quoi)
| lieu (où), etc.

Les valeurs des actants sont déterminées par analyse pragmatique de l'acte
contraintes temporelles : choisies parmi les 13 relations de Allen (avant, après, etc.). Ces contraintes doivent être marquées linguistiquement dans l'acte pour être codées
liens spatiaux : marqués par les prépositions ou les adverbes.

Composante illocutoire

la composante illocutoire caractérise l'action que l'acte dénote
type : f (faire), ff (faire-faire), fs (faire-savoir), fc (faire-croire), fp (faire-pouvoir), fd (faire-devoir).
sens : de l'énonciateur E au destinataire D, à lui-même E ou au monde m
intention en action : précise l'objectif intrinsèque de l'acte noté v()(b) (veut)
pile des buts : pile des buts restant à atteindre
effet : résultat attendu sur la variable d'écartement ε, ε–, ε+, ε––, ε++, ε=0
conséquent : liste d'actions impliquées pour une exécution effective de l'acte

Composante perlocutoire

la composante perlocutoire caractérise les effets attendus par l'énonciateur E sur son destinataire D. Ces effets sont de deux ordres : les modifications des connaissances (K) de D, les croyances (c) de D sur E notamment sur son rôle vis-à-vis de D.
sens : de l'énonciateur E au destinataire D ou sans effet Ø
effet-déontique : sont les devoirs ou pouvoirs imposés ou donnés à D par E = df (par ex. agir), dfs (par ex. répondre), ds(par ex. remise à jour des K), pf (par ex. peut faire), etc.
effet : c'est la manière dont D devrait se comporter dans l'action suivante en réponse à cet acte, c'est donc la stratégie δ, δd, δr, δc, δn, δi, δk (directive, réactive, coopérative, de négociation, dirigée par les intentions, constructive) avec laquelle E entend que le dialogue pourra se dérouler
conséquent : modifications dans le modèle de l'utilisateur (croyances) et modifications dans les connaissances (K)
Les connaissances (K) portent sur la tâche (tâche), les objets de la tâche (monde-t), les mondes d'arrière-plan (mondes), le plan d'activité (plan), les états mentaux ou les rôles sociaux ou le meta-discours (com).

Exemple de codage d'un dialogue :

I1 : "alors il faut faire deux maisons de chaque côté d'une route"

Acte : I1, locutoire	Acte : I1, illocutoire	Acte : I1, perlocutoire
qui : ?	Type : {ff}	sens : I -> M
quoi : scène	sens : I -> {M, I}	effet-déontique -> ds
	intent. en act : B1(monde_t)	stratégie-attendue : {δi, δc, δk}
	pile_buts : {B1}	cons. : K(tâche), K(monde_t)
faire(scène)	effet -> ε1	conséquent : c(I=instructeur)
entre(route, 2 maisons)	conséquent : dessiner	

M1 : "e..."

Acte : M1, locutoire	Acte : M1, illocutoire	Acte : M1 perlocutoire
	Type : fs	sens : M -> I
	sens : M -> I	effet-déontique -> dfs
	intention-en-action : vfs(?)	stratégie-attendue : δc
	pile_buts : {B1}	
	effet -> ε1++	conséquent : c(M=naïf)

I2 : "Alors c'est deux carrés... "

Acte : I2, locutoire	Acte : I2, illocutoire	Acte : I2, perlocutoire

| qui : scène | Type : fs | sens : I -> M
| quoi : carrés | sens : I -> M | effet-déontique -> ds
| combien : 2 | int.-en-action : vfs(monde_t) | stratégie-attendue : δn, δk, δr
| | pile_buts : {B1} | conséquent : K(monde_tâche)
| forme(carré, maison) | effet -> ε1– |

I3 : "Faut faire la route avant, faut faire la route avant"

Acte : I3, locutoire	Acte : I3, illocutoire	Acte : I3, perlocutoire
qui : ?	Type : fs	sens : I -> M
quoi : route	sens : I -> M	effet-déontique -> ds
	intent. : vfs(plan), vfs(monde-t)	stratégie-attendue : δr,δn, δc
	pile_buts : {B1, B2}	conséquent : K(tâche), K(plan)
faire(route)	effet -> ε2	
T : avant	conséquent : dessiner	

I4 : click + dépl + déclick (vert_gd)

Acte : I4, locutoire	Acte : I4, illocutoire	Acte : I4, perlocutoire
qui : I	Type : f	sens : I -> M
quoi : vert_gd	sens : I -> m	effet-déontique -> {ds, dsf}
où : (x0, y0)	intention-en-action : vf(B2)	stratégie-attendue : δc
combien : 1	pile_buts : {B1, B2}	conséq. : K(plan), K(monde-t)
	effet -> ε2–	

I5 : "voilà "

Acte : I5, locutoire	Acte : I5, illocutoire	Acte : I5, perlocutoire
	Type : fs	sens : I -> M
	sens : I -> M	effet-déontique -> ds
	intent. : vfs(monde-t) ou vfs(B2)	stratégie-attendue : {δk, δi}
métadiscours	pile_buts : {B1, B2}	conséquent : K(plan)
	effet -> ε2 ou ε2––	

4. Analyse

L'ensemble des matériaux recueillis, transcrits et codés constitue une base de données pour les différentes analyses que l'on peut envisager sur ces corpus : analyses linguistiques, analyse de l'activité, analyse de l'usage des modes, déroulement du dialogue, etc. Nous décrivons ci-après quelques analyses effectuées sur ces corpus (ce travail s'est déroulé sur plusieurs années).

Analyse de fréquences lexicales

Pour l'étude du lexique, quantitative (calcul des fréquences de morphèmes) et qualitative (répartition des entités lexicales en classes), nous avons utilisé le logiciel PILAF (Procédures Interactives Linguistiques appliquées au français), logiciel développé par l'équipe TRILAN [Courtin, 87, 91]. PILAF associe à chaque forme lexicale une "classe grammaticale" qui permet le traitement des formes lexicales sans confusion homographique.

L'analyse lexicale se réalise en trois étapes qui sont les suivantes :

– pour la première étape, les textes en version ASCII ont été soumis à l'analyse morphologique qui consiste à attribuer une classe grammaticale et les paramètres correspondants à chaque entité lexicale. Les mots du texte sont traités séquentiellement. Le module PILMOR propose toutes les différentes possibilités d'analyse morphologique d'une chaîne de caractères. Les mots non identifiables portent la classe "pas de résultat". Il s'agit en grande majorité de mots incomplets ou d'erreurs d'élocution,

– la seconde étape est le filtrage manuel des différentes possibilités proposées pour l'analyse d'un mot, par exemple "porte" peut être un substantif ou un verbe, pour ne retenir que la bonne solution. Ce filtrage manuel fournit le corpus d'apprentissage qui permet la mise en oeuvre d'un filtre automatique fondé sur les modèles de Markov [Menézo, 92].

– PILGEN, le module de génération, associe alors à chaque occurrence (forme telle qu'elle apparaît dans le texte) sa forme canonique, qui est une occurrence particulière, obtenue selon les règles suivantes : pour un verbe conjugué ou à la forme participe, il s'agit de restituer sa forme infinitive, pour un nom, sa forme au singulier et pour les adjectifs leur forme au masculin singulier. Si les participes fonctionnent comme préposition, nom ou adjectif, leur reconstruction se plie aux règles relatives à leur classe grammaticale.

A l'issue de cette analyse morphologique, chaque occurrence d'un item lexical est accompagné de sa forme canonique, de sa classe grammaticale et des différents paramètres morphologiques. Ces résultats ont donné lieu à divers calculs statistiques [Fréchet, 92] que nous résumons ci-après :

• la *fréquence* absolue et relative. Ce calcul de fréquences nous permet de mesurer l'utilité d'une entité lexicale en tant qu'outil de communication. La fréquence d'un mot est le nombre de réalisations de ce mot dans un corpus. Nous appelons *fréquence absolue* le nombre de réalisations de ce mot dans un ensemble d'entités donné et *fréquence relative* le pourcentage obtenu par le quotient : nombre de réalisations/nombre d'entités d'un ensemble. Les objets de ce calcul sont les rubriques, les formes canoniques, les occurrences, les classes grammaticales, les lemmes et les formes flexionnelles. La méthode de calcul est la suivante : pour chaque liste d'entités lexicales, on calcule les fréquences absolues et relatives puis on range les entités lexicales par ordre décroissant des fréquences, pour faciliter les comparaisons. A partir de ces listes, nous déduisons,

• le *vocabulaire partiellement utilisé* en soustrayant à la liste du vocabulaire total la liste du vocabulaire commun,

• la *banalité* [Falzon, 89] du vocabulaire qui est la mesure du nombre de textes dans lesquels une entité lexicale est attestée.

Les listes d'entités lexicales d'oral finalisé peuvent revêtir plusieurs aspects en fonction de la nature des occurrences combinées à leur classe et du calcul opéré (union ou intersection du vocabulaire des textes). Pourtant, malgré le caractère finalisé du vocabulaire de nos listes, les mots les plus fréquents sont, pour la plupart, attestés dans d'autres listes similaires dont l'historique a été fait par Gougenheim (1956) et repris par Malécot (1976). Le tableau II associe à chaque lemme de notre liste le rang correspondant chez Gougenheim et chez Malécot.

rang	Fréchet (1991)	(nombre de réalisations)	rang chez Gougenheim (1956)	rang chez Malécot (1976)
1	pper-je (j')	1423	4	5
2	prep-de	448	3	2

3	det -le	451	11	6
4	apdi-e	442		
5	det -la	413	7	9
6	xal -aller	386		
7	pnp -on	378	12	13
8	apdi-donc	342		
9	coco-et	300	10	1
10	pper-il	265	5	7
11	xav -avoir	220	2	
12	ce -ça	220	15	15
13	pas -pas	216	8	10
14	apdi-alors	208	22	27
15	apdi-bon	202		
16	det -les	200	16	17
17	det -un	191	14	12
18	cocs-que	186	17	8
19	det -une	177	23	21
20	apdi-là	176		9
21	prep-sur	175		
22	pres-c'est	175		
23	det -l'	165		
24	subc-fenêtre	164		
25	subc-signal	162		
26	prep-dans	157		23
27	conp-voilà	155		
28	verb-pouvoir	147		
29	verb-avoir	147	2	
30	infi-faire	139	19	

Tableau II : rangs dans les listes de mots

La classe "appui du discours"

Elle se compose de quinze éléments présentés par ordre décroissant de fréquence. Le premier nombre de la parenthèse indique la fréquence absolue et le second la répartition (selon Engwall 1984) ou la banalité [Falzon, 89] c'est-à-dire le nombre de textes dans lesquels ce terme est attesté. Nous pouvons établir trois groupes d'appuis du discours : le groupe A contient ceux dont l'occurrence est la plus élevée : e (442/14) ; donc (342/13) ; alors (208/14) ; bon (202/13) ; là (176/13). Le groupe B contient des éléments dont le nombre de réalisations est supérieur à 20 : ben (39/10) ; d'accord (25/7) ; enfin (23/8). Dans le groupe C, nous avons rassemblé des éléments très peu fréquents : ok (6/6) ; quoi (5/4) ; bref (4/3) ; déjà (3/2) ; d'abord (4/3) ; du coup (3/2) ; bien (2/1). "e" et "ben", les premiers termes des séries A et B sont plutôt à envisager comme des pauses sonores ayant une fonction phatique. Le rôle de la fonction phatique est habituellement de maintenir le contact entre le locuteur et le destinataire. Ils marquent ici une hésitation montrant la volonté du locuteur de garder la parole tout en réfléchissant à ce qu'il va dire. "je peux choisir e: (h) je dois choisir le domaine d'étude". Nous récapitulons les différents groupes dans le tableau III.

APPUIS DU DISCOURS	
groupe A	e (442/14) ; donc (342/13) ; alors (208/14) ; bon (202/13) ; là (176/13)
groupe B	ben (39/10) ; d'accord (25/7) ; enfin (23/8)
groupe C	ok (6/6) ; quoi (5/4) ; bref (4/3) ; déjà (3/2) ; d'abord (4/3) ; du coup (3/2) ; bien (2/1)

Tableau III : représentants de la classe des "appuis du discours"

Les connecteurs pragmatiques

Les connecteurs pragmatiques [Morel, 88, 89], au nombre de 18, sont des expressions qui ponctuent l'action. Le connecteur "voilà" avec 155 réalisations se détache des autres du point de vue de la fréquence. Loin derrière, vient en deuxième position le connecteur "c'est ça" avec 14 occurrences. Quinze connecteurs pragmatiques ont la particularité d'être des formes verbales introduites par ça ou c'. Deux connecteurs sont introduits par les pronoms "je" et "on" désignant toujours le locuteur. Les quatre connecteurs les plus fréquents (voilà, c'est ça, ça y est, c'est bon) sont phonétiquement des dissyllabes. Nous classons les connecteurs pragmatiques en trois groupes selon leur lien avec l'action. Le premier groupe rassemble les connecteurs relatifs à la réalisation de l'action du point de vue de l'utilisateur, le second contient ceux montrant la réalisation de l'action du point de vue de la machine et le troisième contient le terme "voilà" dont la valeur est mixte.

- *marqueurs de satisfaction* :
c'est ça (14/5) : acquiescement ou prise de conscience de l'utilisateur
ça y est (9/4) : réalisation de l'action de la machine, action escomptée par l'utilisateur
c'est bon (9/7)
- *marqueurs de déception et de désapprobation*
c'est pas ça (7/4) : inadéquation de l'action ou l'objet choisi
ça va pas (3/2) : interrogation à l'enquêteur
- *marqueurs de résignation*
ça fait rien (4/3) : Pour deux occurrences, le logiciel ne répond pas à l'attente de l'utilisateur et le "ça" signifie la machine. Dans deux autres cas, il s'agit d'une locution verbale signifiant l'abandon de l'action.
je laisse tomber (3/3) : abandon d'un média (la souris), de l'action en cours ou abandon de l'intention de réaliser une action.
- *marqueurs de début d'action*
c'est parti (3/3) : relevé en début de session
on y va (3/3) : intention de réaliser une action (deux réalisations), en début de session (une réalisation)
- *marqueurs de fin d'action*
c'est fini (2/2) : fin de la session
c'est terminé (1/1)
- *marqueurs d'action réussie*
ça marche (9/5) : locution verbale (quatre réalisations), proposition subordonnée du type "je ne sais pas si ça marche" (quatre réalisations) et dans un cas, la question "ça marche ?" est posée.
ça a marché (3/1)
- *marqueurs d'action échouée*
ça marche pas (5/5)
ça ne marche pas (1/1)
- *marqueurs de limitation*
ça suffit (2/2) : changement d'activité (un cas) ou limitation de paramètre (un cas).
exemple : "un seul écran ça suffit"
- *marqueurs de continuation*
c'est pas fini (1/1)

Le connecteur voilà

Le connecteur voilà (155/14) a une valeur mixte car il est utilisé pour désigner à la fois une action de l'utilisateur ou de la machine. Il marque le fait que l'utilisateur a réussi son action : "je le cherche voilà". Mais l'action réussie de l'utilisateur implique la plupart du temps l'achèvement d'une action de la machine attendue ou subie par l'utilisateur : "je vais sélectionner "listen selection" /„/ voilà donc là j'ai écouté toute la sélection". L'utilisateur fait une action qui implique une réponse immédiate de l'application.

Les marqueurs de négation

Dans l'état actuel de notre travail, nous avons recensé 2 614 phrases (délimitées de manière experte) dans le corpus et nous n'avons relevé que 226 phrases négatives (soit 8,55 %). Nous remarquons que pour plus de la moitié d'entre elles (129 phrases, soit 57 %), la première partie de la négation "ne" est élidée. La deuxième partie de la négation est, par ordre de fréquence : pas (216), pas du tout (8) et jamais (2).

Constitution du plan d'activité de désignation et stratégies cognitives

L'analyse de la séquence des actes langagiers et non-langagiers permet de constituer le plan d'activité du sujet. La constitution du plan d'activité facilite l'analyse des stratégies cognitives de description. Dans ce domaine, rappelons certains des résultats qui nous intéressent particulièrement : Levelt [Levelt, 82b] a distingué deux grands groupes de stratégies de description (a) la description structurelle de la figure[1] et (b) la linéarisation de la figure par la description successive de chaque noeud selon les liens. Dans de nombreux exemples cependant, le plan d'activité indique que la stratégie du sujet n'appartient ni à l'une ni à l'autre de ces catégories, mais qu'elle est mixte. Il s'agit en fait d'une décomposition de la figure en trois parties, décrites successivement par linéarisation.

La constitution du plan d'activité facilite également le repérage de l'évolution des stratégies de linéarisation en cours de tâche, et donc de l'apprentissage du sujet. Par exemple, nous pouvons comparer le plan d'activité pour la dernière figure décrite à la tâche 1 de l'exéprience 2 au plan d'activité du sujet pour la première figure de cette même tâche . La stratégie du sujet en début de tâche est la linéarisation, qui peut se qualifier de stratégie opportuniste du fait qu'elle ne requiert pas de planification. En fin de tâche, la description est basée sur les régularités de la figure observées par le sujet. De fait, la grande majorité des sujets passent d'une stratégie opportuniste à une stratégie de planification de la description en fonction des caractéristiques structurelles de chaque figure.

Levelt a observé qu'un même sujet adopte toujours la même stratégie. Puisque le mode de linéarisation est invariant pour un sujet, Levelt l'assimile donc à un "style cognitif". Des travaux plus récents sur le sujet [Montarnal, 91] vont dans le sens des observations que nous avons effectuées sur notre corpus et raffinent ces conclusions: la stratégie de linéarisation semble être facteur non seulement du sujet, mais de la figure et des figures linéarisées auparavant. Ce thème de recherche est toujours à l'étude.

Evolution des connaissances partagées

Chez la majorité des sujets, les descriptions à la fin de la session sont beaucoup plus économiques que les description en début de session. Certaines connaissances explicites deviennent partagées entre l'instructeur et le manipulateur au cours de leur interaction, pour s'assimiler au contexte commun

---

[1] Par exemple : "C'est une figure en forme de T avec trois carrés liés formant la ligne du T et trois carrés liés formant la colonne du T" etc.

d'interprétation des actes locutoires. Nous examinons ici quelles sont ces connaissances.

Chez certains sujets, la progression des connaissances partagées sur les relations entre objets au cours de la session est très marquée. Comparons par exemple les deux extraits suivants des productions verbales de la tâche 1 (expérience 2) pour un même sujet. Le premier extrait se situe en début de tâche et le second en fin de tâche. Tous deux expriment la position d'un carré.

> « ... tu choisis un troisième carré que tu viens placer dans l'alignement donc à la suite de ce segment horizontal et donc tu viens placer le côté gauche de telle façon que son milieu coïncide avec le segment de droite ... »

> « tu viens placer à la suite de ce segment de droite un nouveau carré donc le prolongement »

Ces connaissances implicites en fin de tâche concernent la complétude de la relation définie entre les éléments de la figure. En terme de prédicats spatiaux, nous avons:

$\Phi(arg_1, arg_2, ... arg_n)$

{*relation ensembliste: intérieur,extérieur, etc.*
distance en x et en y entre les centres des arguments
direction: haut,bas,gauche,droite, etc.
référentiel: ego,pattern }

où les attributs mis en italique sont implicites. Seule la direction est jugée suffisante et nécessaire par le sujet pour définir la relation. Si nous nous penchons sur les attributs implicites, nous observons que :
- la relation ensembliste peut être inférée à partir de la définition de classe des arguments (contenant ou contenu),
- les distances en x et en y entre les arguments sont toujours les mêmes d'une figure à l'autre,
- le référentiel est toujours le même au cours d'une tâche.

Ainsi, nous remarquons que les connaissances répétitives ainsi que les connaissances pouvant être inférées à partir d'éléments par ailleurs connus deviennent partagées.

Ruptures et changement de monde de référence

• *Inter-tâches*

Un acte locutoire peut renvoyer à plusieurs mondes de référence. Il s'agit là d'un phénomène communicationnel courant qui ne provoque pas de rupture d'interprétation si les mondes de référence sont compatibles. Dans le cas de notre expérience, les tâches se succédant dans le temps constituent des mondes distincts et exclusifs. Nous observons toutefois des actes locutoires tels que

> « ... on place une première porte au niveau du côté inférieur de ce *carré* au milieu voilà ... »

à la tâche 2, expérience 2. Le mot en italique n'appartient pas au monde de la tâche en cours, car, strictement parlant, l'objet "Carré" n'existe pas à la tâche 2. Ceci est attribuable, ici comme dans la majorité des cas observés, au phénomène d'amorçage, où un objet appartenant à un monde référencé précédemment (dans une tâche précédente) reste présent dans l'esprit du sujet.

Pour éviter une rupture dans le dialogue, l'interprétation de ces actes locutoires se fait en tenant compte des liens sémantiques entre les objets appartenant à des mondes différents. L'identité de l'objet du

monde de la tâche en cours sera inférée en parcourant les liens sémantiques de l'objet référencé dans l'acte locutoire.

Le dialogue peut alors se poursuivre avec ou sans confirmation de l'interprétation. Dans le premier cas, il y a rupture : le manipulateur initiera une phase de négociation dans le dialogue ("par carré, veux-tu dire pièce ?"). Dans le second cas, la rupture peut être évitée si les interprétations des interlocuteurs coïncident: le manipulateur exécutera alors l'instruction du sujet sans confirmer son interprétation. Le sujet initiera la phase de négociation que si l'état résultant de l'interface personne-machine ne correspond pas à ses attentes ("non, par carré je veux dire pièce").

• *intra-tâches*

L'apparition, dans les productions verbales, d'un objet nouveau, qui n'appartient à aucun des mondes des tâches, constitue un autre type d'enchevêtrement de mondes de référence. Par exemple :

« ... et pose-le au-dessus du carré numéro 3 avec un décalage d'un *carreau* ... »

Le *carreau* en question est un élément du quadrillage de fond de la zone de travail. Ces objets nouveaux peuvent appartenir au monde de la tâche, au monde de l'interface, etc.

Rôles des partenaires et stratégies de dialogue

La 3ème expérience nous a permis de mettre en évidence les stratégies de dialogue en fonction des rôles des partenaires et de définir une certaine *rhétorique de l'inter-action* définie par rapport au but : maintenir un but, le déplacer, faire converger le dialogue vers un but commun, proposer un nouveau but, le mettre en attente, etc.

Notons B1 le but de U1 et B2 le but de U2.

• *Stratégies non-inférentielles*

Ce type de stratégie ne met pas en jeu la recherche du but de l'interlocuteur.

— mode directif
l'initiative reste toujours du côté du partenaire U1, il y a en général réduction progressive du focus et,
U1 impose son but B1 en ignorant B2
U1 impose à U2 une réponse réactive ou négociée
U1 est dominant et égocentré

— mode réactif
dans ce mode chaque interlocuteur réagit le plus complétement possible au dernier échange. S'il s'agit d'une commande, celle-ci est toujours interprétée et exécutée (prise de décision par défaut). Il y a maintien du focus et par exemple quand U1 est réactif,
U1 accepte le but de U2 en cours de dialogue (B1<-B2)
U1 est dominé et exocentré

• *Stratégies inférentielles*

Ce type de stratégie passe par la recherche du but de l'interlocuteur ou met en relation les connaissances supposées et partagées.

— mode coopératif

dans ce mode le locuteur se fait obligation de fournir un maximum d'informations pour aider et orienter son interlocuteur. Mais fournir trop d'informations augmente sa charge cognitive ainsi que celle de son interlocuteur. La règle est donc de fournir l'information la plus pertinente eu égard à la situation et aux interlocuteurs eux-mêmes (principe de la pertinence de Sperber & Wilson, maximes de coopération de Grice).

Proposer une (ou des) solution qui convient au mieux à l'interlocuteur revient à : évaluer la situation, présenter une explication, des exemples, des aides ou des arguments pertinents et offrir un choix fermé (parce que facile cognitivement). Le mode coopératif procède par recherche d'un optimum dans un espace de possibles logiques (cohérence, pertinence, etc.). Il y a en général élargissement du focus et,

U1 fait sien le but de U2, (B1=B2)
U1 ouvre toutes les stratégies pour U2 ainsi que les incidences
U1 est dominé et exocentré

— mode négocié

par opposition au mode précédent, ce mode consiste à minimiser l'espace de concession de son partenaire. Le locuteur maintien son but et ne cède qu'après argumentation ou réfutation de son interlocuteur. Il y a en général maintien ou déplacement du focus et,

U1 impose son but ou accepte un compromis (B1≠B2)
U1 et U2 sont égaux et égocentrés

— mode dirigé par les intentions

c'est un mode coopératif fondé sur le but en priorité. Cela consiste à comprendre et interpréter les intentions de l'interlocuteur pour déduire ses objectifs.

U1 comprend (ou s'informe sur) le but de U2 pour continuer le dialogue (B1<-B2)
U1 ouvre toutes les stratégies à U2 sauf les incidences
U1 et U2 sont égaux et exocentrés

— mode constructif

le principe de ce mode est d'apporter des informations nouvelles hors du focus du discours dans le but de provoquer une rupture dialogique si possible enrichissante vis-à-vis des connaissances partagées. Cela peut procéder par présentation d'exemples. Il y a déplacement systématique du focus et,

U1 déplace temporairement le but de U2, (B1 <- B2+)
U1 ouvre toutes les stratégies à U2
U1 est dominant et égocentré

5. Conclusion et perspectives

La technique d'observation que nous avons choisie (tâche artificielle, situation de laboratoire), très proche des méthodes expérimentales classiques, peut sembler être en contradiction avec un cadre théorique où le contexte d'action est primordial. Il est clair que toute recherche-terrain exige un compromis entre le caractère naturel de la situation observée d'une part et la précision et la pertinence du matériel recueilli d'autre part. Nous avons favorisé ce dernier facteur, tout en tentant de préserver une situation d'observation qui soit la plus naturelle possible. Les expérimentations sont graduellement plus complexes partant de simples conversations humaines pour aboutir à des dialogues homme-machine simulés et parfaitement contrôlés. Les enregistrements audio et vidéo sont soumis ensuite à un traitement de codage et d'interprétation manuel.

Les études que l'on peut faire sur des corpus sont très variées et souvent très longues. Dans les expériences de communication orale décrites ici, nous avons étudié des aspects lexicaux, syntaxiques, pragmatiques mais aussi multimodaux, dialogiques et communicationnels. L'automatisation d'un certain nombre de processus semble maintenant indispensable pour une analyse systématique de corpus. Des plateformes Magicien d'Oz doivent être développées pour un contrôle systématique des expériences et une interprétation assistée des protocoles.

Références

[Austin 62] J.L. Austin (1962), How to do Things with Words. Oxford University Press. Version française : Quand dire c'est faire (1970). Seuil, Paris.

[Caelen 92] J. Caelen, J. Coutaz (1992), Interaction homme-machine multimodale : quelques problèmes. Bull. de la Com. Parlée n°2, p. 125-140.

[Courtin, 77] J. Courtin (1977). Algorithmes pour le traitement interactif des langues naturelles. Thèse d'Etat, Grenoble I.

[Courtin, 87] J. Courtin , D. Dujardin, D. Genthial, I. Kowarski, B. Cohard, V. Strube de Lima (1987). Le Système PILAF. Journées du PRC-CHM, Paris.

[Courtin, 91] J. Courtin, D. Dujardin (1991).Paramètres linguistiques du français dans le système PILAF. Rapport RT 67 LGI-IMAG, Grenoble.

[Eluerd, 85] R. Eluerd, (1985). *La pragmatique linguistique*. Paris, Nathan, coll. linguistique générale.

[Engwall, 84] G. Engwall, (1984). *Vocabulaire du roman français (1962-1968) Dictionnaire des fréquences.,* Stockholm, Almqvist & Wiksell International, 426 p.

[Falzon, 89] P.Falzon (1989). *Ergonomie Cognitive du Dialogue,* Grenoble, PUG, 175 p.

[Fréchet 92] A.L. Fréchet, M.A. Morel, D. Dujardin, J. Caelen (1992), Analyse lexicale d'un langage opératif. Bull. de la Com. Parlée n°2, p. 167-182.

[Gougenheim, 56] G. Gougenheim, R. Michéa, P. Rivenc, A. Sauvageot (1956). *L'élaboration du français élémentaire*. Paris, Didier.

[Grévisse, 64] M. Grévisse (1964). *Le bon usage*. Paris, Hatier.

[Levelt, 82a] W.J.M. Levelt (1982). Linearization in describing spacial networks. Dans S. Peters et E. Saarinens (Eds). *Processes, beliefs and questions* . D. Reidel Publishing Company, pages 199 à 220.

[Levelt, 82b] W.J.M. Levelt (1982). Cognitive Styles in the Use of Spacial Direction Terms. Dans R.J. Jarvella et W. Klein (Eds). *Speech, Place and Action* . John Wiley and Sons, pages 251 à 268.

[Malécot, 76] A. Malécot (1976). Fréquence d'occurrence des mots dans la conversation. *Revue d'Acoustique*, 38, pp. 200-205.

[Menézo, 92] J. Menézo, J. Courtin, J. Caelen (1992). *Désambiguïsation lexicale par filtrages*.. Séminaire Lexique, GRD-PRC, 21 et 22 janvier 1992, IRIT-UPS Toulouse.

[Minsky, 75] M. Minsky (1975). *A Framework for Representing Knowledge.* in The Psychology of Computer Vision, P.H. Winston, ed., McGraw-Hill, New York, N.Y.

[Minsky, 83] M. Minsky (1983). *The society of Mind*. MIT Press, Mass.

[Montarnal, 91] C. Montarnal (1991). *Etude expérimentale de l'activité de linéarisation dans les textes descriptifs de configurations spatiales*. Notes et Rapports de Recherche du CRISS no. 23, Université Mendès France, Grenoble.

[Morel, 88] M.A. Morel (1988). *Dialogue Homme-Machine, Premier Corpus : Centre de renseignements SNCF à Paris*. Paris : Publications de la Sorbonne Nouvelle, 292 p.

[Morel, 89] M.A. Morel (1989). *Analyse linguistique d'un corpus, Deuxième corpus : Centre d'Information et d'orientation de l'université de Paris V*. Paris : Publications de la Sorbonne Nouvelle, 331 p.

[Mounin, 70] G. Mounin (1970). *Introduction à la sémiologie*. Ed. de Minuit, Paris.

[Ozkan 92] N. Ozkan, J. Caelen (1992), Vers un modèle de dialogue adaptatif. Actes IHM'92, p. 72-78.

[Searle, 69] J.R. Searle (1969). *Speech Acts*. Cambridge University Press, UK.

[Searle, 83] J.R. Searle (1983). *Intentionality*. Cambridge University Press, UK.

[Vandeloise, 86] C. Vandeloise (1986). *L'espace en Français*. Le Seuil, Paris.

[Vernant, 86] D. Vernant (1986). *Introduction à la philosophie de la logique*. Philosophie et Langage, Pierre Mardaga éditeur, Bruxelles.

[Winograd, 86] T. Winograd et F. Flores (1986). *Understanding Computers and Cognition; A New Foundation for Design*.. Ablex Publihing Corporation, New-Jersey.