

Vers une meilleure intégration de la parole dans des systèmes de communication homme-machine

*Towards the integration of speech and natural
language into human-computer interaction
Systems*



J.-M. PIERREL

CRIN/INRIA-Lorraine BP 239
F-54506 Vandœuvre-Lès-Nancy Cedex

Jean-Marie PIERREL, 38 ans, est professeur d'informatique à l'université de Nancy I. Depuis plus de 15 ans, il travaille en reconnaissance et compréhension de la parole continue au sein du CRIN (Centre de Recherche en Informatique de Nancy), URA 262 du CNRS. Après avoir successivement défini et mis en œuvre les systèmes MYRTILLE I et II, travail qui fut couronné par le prix scientifique IBM-France en informatique, il orienta ses recherches plus particulièrement vers le dialogue oral homme-machine puis aujourd'hui vers le dialogue multi-mode. Il est aujourd'hui directeur adjoint du CRIN et responsable, au sein de l'équipe *Reconnaissance des Formes et Intelligence Artificielle*, d'un projet de recherche « Parole et dialogue » commun au CRIN et à l'INRIA.



N. CARBONELL

CRIN/INRIA-Lorraine BP 239
F-54506 Vandœuvre-Lès-Nancy Cedex

Noëlle CARBONELL est Maître de Conférences à l'Université de Nancy II. Elle est membre du CRIN-INRIA Lorraine et ses activités de recherche portent sur la reconnaissance et la compréhension de dialogues oraux homme-machine finalisés en langue naturelle. Elle s'intéresse également aux problèmes d'ergonomie des logiciels et plus particulièrement aux interfaces de dialogue homme-machine multi-média.



J.-P. HATON

CRIN/INRIA-Lorraine BP 239
F-54506 Vandœuvre-Lès-Nancy Cedex

Jean-Paul HATON, agrégé de l'université, est professeur à l'Université de Nancy I où il enseigne divers aspects de l'informatique et de l'intelligence artificielle. Il est responsable, au sein du Centre de Recherche en Informatique de Nancy, de l'équipe *Reconnaissance des Formes et Intelligence Artificielle*. Son principal domaine d'intérêt concerne, depuis plus de quinze ans, la communication homme-machine et l'intelligence artificielle : reconnaissance de la parole, reconnaissance de caractères et de graphismes, traitement et interprétation d'images, systèmes experts et systèmes à bases de connaissances.

Jean-Paul Haton est directeur du GRECO-PRC « Communication Homme-Machine ». Il est actuellement détaché à l'INRIA (Institut National de la Recherche en Informatique et en Automatique) comme directeur des recherches à l'INRIA-Lorraine où il est responsable du projet SYCO (Systèmes de Compréhension et Bases de Connaissances). Il est l'auteur de plus de 250 ouvrages et articles dans le domaine de l'intelligence artificielle et de la communication homme-machine.



K. SMAILI

CRIN/INRIA-Lorraine BP 239
F-54506 Vandœuvre-Lès-Nancy Cedex

Kamel SMAÏLI titulaire d'un diplôme d'ingénieur d'état en informatique de l'université des sciences Houari Boumedienne (USTHB) d'Alger (1986) et d'un DEA de l'université de Nancy I (1987). Il a rejoint l'équipe RFIA au CRIN/INRIA Lorraine où il travaille dans le cadre d'une thèse sur les aspects de la reconnaissance de la parole continue et plus particulièrement sur la réalisation d'une machine à dicter destinée aux grands vocabulaires.

RÉSUMÉ

Cet article se propose de dresser une synthèse et une classification des diverses applications du dialogue oral homme-machine. Il présente, dans une première partie, les avantages et les limites de la parole comme moyen de communication entre un utilisateur et un système automatisé. Le problème essentiel pour le développement d'interfaces utilisateur à composante orale, à côté du choix d'applications appropriées, est la reconnaissance de la parole, spécialement de la parole continue. Comme les approches diffèrent suivant le type d'applications, nous présentons tout d'abord les problèmes et les techniques spécifiques à l'entrée orale de données et, à titre d'exemple, nous décrivons rapidement l'approche que nous avons adoptée pour la machine à dicter que nous développons dans notre laboratoire. Ensuite nous abordons la compréhension et la gestion de dialogues oraux. Pour illustrer cette présentation, nous présentons l'architecture et

les fonctionnalités de divers prototypes que nous avons mis en œuvre : système de messagerie électronique, dialogue pour la commande d'une console sonar et dialogue entre un chirurgien et un système d'aide au diagnostic.

Enfin, nous détaillons le gestionnaire de dialogues DIAL, en cours de développement, dont l'objectif est d'aider et de guider un utilisateur dans des activités cognitives complexes telle la recherche de renseignements administratifs.

MOTS-CLÉS

Reconnaissance de la parole continue/compréhension du langage naturel/dialogue oral homme-machine/machine à dicter/interfaces orales homme-machine.

SUMMARY

This paper first surveys and classifies applications of voice human-computer dialogue. The advantages and limits of speech as a means of communication between users and software are then considered. A major problem in the development of user interfaces with a voice component, besides the choice of appropriate applications, is speech recognition, especially continuous speech recognition. As approaches differ according to the type of application, we first summarize problems and techniques specific to voice data input; as an example, we briefly describe the speech recognition approach that we have adopted for the dictation machine that we are developing in our Laboratory.

Then, the case of voice dialogue understanding and management is considered. To illustrate the discussion, the architecture and functionalities of some prototypes that we have implemented are presented: for instance, an E-mail system and a Sonar control software. Finally, we present a dialogue manager DIAL capable of helping/guiding the user in complex cognitive activities that we are currently developing and implementing.

KEYWORDS

Continuous speech recognition/Natural language understanding/Voice human-computer dialogue/dictation machines/Voice human-computer interfaces.

1. Introduction

Actuellement, sous l'impulsion, entre autres, des récentes avancées de la recherche en Intelligence Artificielle, on assiste à l'essor d'applications informatiques interactives concernant des tâches cognitives complexes. Parallèlement, le rôle dévolu au logiciel dans l'interaction homme/ordinateur évolue. Les systèmes informatiques ne sont plus aujourd'hui confinés à l'exécution, en réponse à des commandes humaines, de tâches ponctuelles sans autre lien entre elles que l'ordre chronologique; leur autonomie et leurs capacités d'initiative ne cessent de croître, de sorte que ces systèmes deviennent capables d'assister l'utilisateur dans des activités de raisonnement variées et complexes :

— soit en le guidant et en lui fournissant les connaissances qui lui font défaut, c'est-à-dire, en jouant auprès de lui un rôle d'expert ;

— soit en prenant en charge la résolution de sous-problèmes (tâches) précis(es); c'est le cas, par exemple, de certaines formes de logiciels d'aide au diagnostic ou à la décision; Millot [1] appelle ce type d'assistance : coopération horizontale ;

— soit encore en proposant des outils, spécifiques ou non, que l'on peut utiliser et combiner entre eux à son gré, dans le cadre de ses propres stratégies cognitives; c'est ce type de coopération, qualifiée par Millot de verticale, qu'offrent, entre autres, les boîtes à outils des environnements de programmation, les logiciels de constitution et de consultation de bases de données ou de représentation de connaissances etc.

Cette évolution des applications informatiques, analysée en particulier par Rasmussen [2], s'accompagne d'une modification profonde des modalités d'interaction homme/ordinateur. Aux modes de communication désormais classiques, à savoir : suites de menus ou de formulai-

res/grilles, séquences de questions/réponses ou de commandes, se substitue un véritable **dialogue coopératif finalisé**.

Même dans le cadre d'un univers pragmatique (ou application) restreint, réaliser un dialogue convivial et efficace soulève de nombreuses difficultés en raison de l'insuffisance des connaissances actuelles sur le dialogue, sur son fonctionnement en particulier ; il reste encore beaucoup à découvrir sur :

— les diverses stratégies possibles de conduite du dialogue ainsi que les mécanismes permettant de prendre le contrôle du dialogue, d'informer l'interlocuteur de son intention de prendre ou non l'initiative des échanges ;

— les stratégies susceptibles de fournir, à partir de l'interaction et du dialogue, des informations sur les capacités cognitives et les connaissances de l'interlocuteur, ses buts et sa démarche, ses attentes ;

— la pertinence du modèle de la communication homme/homme pour le dialogue homme/machine [3], en raison, principalement, du nombre restreint et de la faible taille des corpus de dialogues homme/machine actuellement disponibles.

Cette insuffisance des connaissances actuelles sur le dialogue constitue l'une des principales pierres d'achoppement des recherches visant à la conception et à la mise en œuvre d'**interfaces orales de dialogue en langue naturelle**.

En effet, la langue naturelle, en particulier sous forme orale, représente le mode de communication humain le plus naturel. Par conséquent, si l'on propose à l'utilisateur d'utiliser un mode d'expression spontané pour communiquer avec un système informatique, on ne peut lui imposer un dialogue fortement contraint, voire artificiel.

Il n'est donc pas surprenant que les chercheurs qui s'intéressent à la **compréhension automatique de la parole continue dans un environnement multilocuteur** en vue de l'élaboration d'interfaces orales, s'interrogent depuis plusieurs années déjà sur ce qu'est un **dialogue finalisé naturel** et s'efforcent de déterminer en quoi consiste une **gestion naturelle** du dialogue homme/machine.

Hélas, les performances actuelles des systèmes de reconnaissance et de compréhension de la parole sont un frein au développement d'interfaces orales homme-machine efficaces et confortables. Les taux de reconnaissance de mots isolés ou connectés dépassent actuellement 90 %, qu'il s'agisse de systèmes monolocuteurs (avec phase d'apprentissage pour passer d'un locuteur à l'autre) ou véritablement multilocuteurs. Mais ces performances peuvent chuter si l'environnement est bruyant ou le locuteur sous l'empire d'une émotion (anxiété, par exemple) ou dans une situation de stress. En ce qui concerne la parole continue, les systèmes de compréhension multilocuteurs obtiennent des taux de reconnaissance de mots bien inférieurs, entre 70 % et 80 %, tout en imposant au locuteur des contraintes d'expression sévères : pas ou peu d'hésitations ni de reprises, élocution régulière et relativement lente, articulation soignée, restrictions syntaxiques importantes, vocabulaire limité (de l'ordre de mille mots). Leur utilisation requiert donc de l'utilisateur un certain

apprentissage ou tout au moins une phase d'adaptation. Les capacités de ces systèmes sont très diverses selon le type de langage qu'ils admettent en entrée : des langages artificiels à syntaxe pauvre, rigide et à vocabulaire très limité — c'est le cas des langages de commande par exemple — à des sous-ensembles plus ou moins restreints de la langue naturelle allant même jusqu'à des langages quasi naturels qui permettent une expression sans contrainte dans un univers pragmatique limité, circonscrit par l'ensemble des tâches à effectuer.

On peut donc s'interroger sur l'opportunité, à moyen terme, de développer des interfaces homme-machine orales ou multimédia intégrant la parole.

Nous répondons à cette question dans une première partie qui s'appuie sur les résultats d'études ergonomiques portant, d'une part, sur la communication orale en situation de travail et d'autre part, sur les applications potentielles de la parole en tant que support de communication homme-machine. Cela nous permettra de dégager les profils des différentes classes d'utilisateurs potentiels auxquels peuvent s'adresser les systèmes de compréhension et de gestion de dialogues oraux finalisés. Ensuite, nous mettrons en évidence les contraintes précises qu'imposent à l'expression de l'utilisateur les limites de ces systèmes en compréhension.

Enfin, nous présenterons différents systèmes de dialogues oraux, actuellement en cours de développement dans notre laboratoire, qui fournissent une bonne idée de l'état de l'art en ce domaine, d'une part ceux, à l'état de prototypes, qui admettent en entrée des langages artificiels, d'autre part, un système plus ambitieux, pour l'instant sous forme d'un ensemble de maquettes autonomes, susceptible de gérer des dialogues oraux finalisés en langue quasi naturelle, dans un univers pragmatique restreint.

Nous insisterons, au cours de la présentation de ces systèmes, sur le rôle joué par la composante de gestion de dialogue, qui peut permettre à la fois de :

— faciliter et améliorer la reconnaissance et la compréhension des énoncés du locuteur,

— prendre en compte l'utilisateur, ses buts et ses attentes, pour obtenir une meilleure efficacité dans la réalisation de la tâche.

2. Applications potentielles du dialogue oral homme-machine

2.1. LA COMMUNICATION ORALE EN SITUATION DE TRAVAIL

Des études menées jusqu'à présent sur la communication orale en situation de travail, il ressort que :

— l'échange oral entre professionnels met en œuvre un sous-ensemble restreint du langage naturel, tant au point de vue lexical que syntaxique [4] ;

— dans le cas où le volume des échanges est important et joue un rôle prépondérant dans la réalisation de la tâche, le souci d'efficacité et une pratique fréquente conduisent de manière naturelle au développement de véritables **sous-langages opératifs** [5] qui optimisent la communication en minimisant :

- d'une part, le volume des énoncés, grâce à l'adoption d'une syntaxe et d'un lexique qui favorisent la concision ;
- d'autre part, leur complexité, de façon à éviter les ambiguïtés et donc les demandes de répétition ou de reformulation ; en effet, des structures stéréotypées et standardisées, un vocabulaire restreint et spécifique [6], en rendant la forme et le contenu des énoncés de l'interlocuteur partiellement prévisibles, limitent les incompréhensions et les erreurs de reconnaissance.

Par ailleurs, des études sur le dialogue oral finalisé entre un opérateur expert et un usager naïf, par exemple dans le cadre du renseignement téléphonique, montrent que l'opérateur, en adoptant une **stratégie experte de conduite du dialogue**, peut réduire le volume et la complexité des interventions de l'utilisateur ainsi que le nombre des échanges [4].

Il semble donc que, dans les situations de travail qui font intervenir la communication orale, que ce soit entre professionnels ou entre un opérateur expert et un utilisateur naïf, les capacités actuelles des systèmes de compréhension de la parole continue autorisent la mise en œuvre, sous forme orale, des échanges d'informations entre l'utilisateur et le système. Mais ce n'est pas pour autant qu'il faille adopter ce mode d'interaction homme-machine ni qu'il soit préférable à un dialogue classique par clavier/écran : la communication homme-homme n'est pas nécessairement un modèle pertinent pour la communication homme-machine [7].

Voyons quelles sont les caractéristiques des applications dans lesquelles le dialogue entre l'utilisateur et le système pourrait s'effectuer oralement avec profit, au moins partiellement.

2.2. LA COMMUNICATION ORALE HOMME-MACHINE : INTÉRÊT ET LIMITES

Les travaux qui portent sur la communication orale homme-machine mettent en évidence, au moins en ce qui concerne la commande de machines ou de processus, le faible débit d'informations (dû à une redondance très élevée) et le manque de fiabilité de ce mode d'interaction par rapport à d'autres, le clavier ou un organe de commande motrice [8], [9].

Plusieurs études ont comparé l'entrée vocale à d'autres dispositifs d'entrée : clavier, touches fonctions, mécanismes de sélection ; en l'absence de système opérationnel de traitement de la parole continue, l'entrée vocale est généralement limitée, dans ces expériences, à des mots isolés ou enchaînés. Ainsi, Murray [10] a montré que le déplacement du curseur par commande vocale était deux fois plus lent que par touches ; en outre, les sujets préfèrent le mode de contrôle par touches. Des sujets non

familiers avec le clavier préfèrent, au bout de quatre séances de pratique, l'utilisation de l'éditeur « *ed* » d'Unix à la commande vocale, pour corriger « en ligne » des documents, le groupe témoin constitué de sujets habitués au clavier préférant, dès la première session, l'éditeur à la commande vocale [11]. Néanmoins, des programmeurs professionnels ont préféré un éditeur vocal à un éditeur classique clavier/écran, bien que l'utilisation de l'éditeur vocal ait augmenté le temps d'exécution des tâches de test [12]. L'expérimentation dans ce domaine mérite donc d'être poursuivie.

Bien que la voix constitue un mode d'interaction homme-machine lent (en particulier pour le traitement des erreurs) et spécifique de la communication humaine, elle peut s'avérer utile comme support principal ou auxiliaire des échanges d'informations entre l'utilisateur et le système, dans un certain nombre d'applications, outre celles destinées aux handicapés :

a) En tant que support auxiliaire de communication

C'est l'exemple des situations de travail où les autres modes de communication sont soit déjà utilisés, soit indisponibles, par exemple lorsque l'opérateur a les mains ou les yeux occupés ailleurs, qu'il doit se déplacer ou encore que l'environnement interdit l'emploi d'un terminal classique (manque de place, travail à l'extérieur dans des conditions difficiles, ...).

L'interaction dans ce type de contexte étant professionnelle, fréquente plutôt qu'occasionnelle, les échanges oraux s'effectuent généralement dans un langage simple et stéréotypé (cf. paragraphe 2.1) qui peut être représenté par un langage artificiel, comportant un vocabulaire restreint, quelques centaines de mots, une syntaxe rigide formalisable à l'aide de règles non contextuelles, une sémantique non ambiguë. Le langage support des échanges peut même se réduire à un ensemble de suites de mots clefs.

b) En tant que moyen de communication unique

Les terminaux classiques sont en général satisfaisants pour les utilisateurs occasionnels sans connaissances informatiques ni expérience de la tâche, à condition qu'ils offrent la possibilité d'un dialogue sous forme de menus ou la manipulation directe des éléments d'une représentation métaphorique de l'univers de la tâche.

Mais lorsque le nombre ou la taille des menus est élevé et que l'on ne dispose d'aucune analogie susceptible de représenter adéquatement les objets et les actions associés à la tâche, le dialogue oral en langue naturelle est préférable. C'est le cas des applications de consultation/renseignement où le système tiendrait le rôle de l'expert que l'on interroge sur un domaine spécifique de connaissances. En ce qui concerne ces applications, seule une interface orale admettant la langue naturelle en entrée permettrait d'éviter une phase coûteuse et fastidieuse d'apprentissage aux usagers occasionnels qui constituent la majorité des utilisateurs potentiels de tels systèmes destinés essentiellement au grand public.

Le caractère finalisé des dialogues, les limites de la tâche elle-même, une stratégie de conduite du dialogue appro-

prisée permettraient certainement de maintenir l'expression de l'utilisateur à l'intérieur d'un sous-ensemble de la langue naturelle dont la complexité n'excéderait pas les capacités de décodage des systèmes actuels (cf. paragraphe 3.1) ; on peut espérer, en particulier, que le vocabulaire de l'ensemble des usagers ne dépasserait pas quelques milliers de mots.

L'utilisation de la parole continue pour des applications grand public semble, en outre, devoir recueillir l'adhésion des usagers potentiels. L'analyse des opinions exprimées par les sujets d'une expérimentation soutenue par le GRECO-PRC « Communication homme-machine » sur le renseignement SNCF le suggère. Les conclusions de cette étude sont reproduites en annexe 1.

Une autre classe d'applications prometteuses de la parole continue en tant que mode de communication homme-machine, sinon unique du moins prépondérant, concerne l'entrée orale de données souvent appelée *machine à dicter* lorsqu'il s'agit de données textuelles.

2.3. VERS UNE TYPOLOGIE DES APPLICATIONS UTILISANT LA PAROLE

a) De la saisie de données aux demandes de renseignements

L'objectif essentiel des études sur la reconnaissance et la compréhension automatique de la parole est de permettre, à terme, une interaction plus naturelle entre l'homme et la machine, dans le cadre d'une application spécifique. En ce sens, ces recherches sont nettement finalisées, même si elles nécessitent d'aborder des problèmes fondamentaux liés à la langue orale, support de la communication. Bien entendu, personne n'a jamais prétendu étudier des systèmes capables de comprendre la langue orale dans son intégralité. Les applications potentielles de la reconnaissance et compréhension automatique de la parole font apparaître divers types de communication orale possibles entre l'homme et la machine, qui se différencient par l'utilisation faite de la parole, le niveau de langue traité et, par conséquent, les traitements mis en œuvre. On peut regrouper l'ensemble des études menées dans ce domaine sous deux rubriques principales : la saisie de données et le dialogue homme-machine. Chaque domaine peut, de nouveau, être redécomposé suivant le niveau de langage traité pour finalement aboutir à quatre domaines spécifiques d'application :

— La saisie de données à format ou structure pré-défini,

Que ce soit pour la saisie de données numériques ou le remplissage de grilles de saisie prédéfinies, le vocabulaire est en général assez restreint (quelques centaines de mots) et la structure ou syntaxe des données très figée. Des systèmes de reconnaissance de mots isolés ou enchaînés, tels ceux actuellement sur le marché, permettent de prendre en charge de telles applications.

— La saisie de textes ou machine à dicter,

Avec comme objectif de permettre la saisie orale de textes en machine, la machine à dicter est sans doute l'un des domaines les plus prometteurs du traitement de la parole. Plusieurs laboratoires à travers le monde ont abordé ce

problème dès le début des années 80. On peut citer par exemple IBM Yorktown aux États-Unis [13], le CERFIA à Toulouse [14], IBM Paris [15], le LIMSI à Orsay [16], le CRIN à Nancy [17] et bien d'autres. A vrai dire, les Japonais ont été les premiers à aller dans ce sens avec l'annonce par leurs principaux constructeurs électroniques (Hitachi, Toshiba, Matsushita) de l'arrivée imminente de tels systèmes. Il faut relativiser ce résultat en raison des particularités de la langue japonaise : elle n'est en effet composée que d'une centaine de syllabes principales. Le problème est plus difficile pour les langues indo-européennes où le nombre de syllabes est de plusieurs milliers.

Dans la littérature, on confond souvent système de reconnaissance de parole et machine à dicter ; nous verrons au paragraphe 3 que si la machine à dicter est fortement dépendante du système de reconnaissance de la parole elle n'en possède pas moins ses spécificités propres. Quoi qu'il en soit, cette application nécessite l'utilisation de très gros vocabulaires (plusieurs milliers de mots) et ne permet pas d'imposer de contraintes au locuteur autres que d'élocution.

— Le dialogue et la communication à l'aide de langages artificiels

Particulièrement bien adaptée, pour des spécialistes, à des applications de type commande de processus ou d'interrogation de bases de données, cette communication correspond le plus souvent à l'activation d'automates de dialogues relativement simples et peut se contenter d'un niveau de reconnaissance de type mots enchaînés ou, en parole continue, à des systèmes guidés par la syntaxe [18].

— Le dialogue en langue naturelle limité à un domaine d'expertise

Nécessaire pour aborder des applications grand public, tels les centres de renseignements téléphoniques, ce type de communication doit utiliser comme support la langue naturelle avec un minimum de restrictions, afin de limiter le plus possible la phase d'apprentissage du langage. Il conduit alors à la mise en place de processus de compréhension et de gestion de dialogues beaucoup plus complexes que dans le cas précédent [19].

b) Des applications à l'usage de professionnels aux applications grand-public

Pour conclure cette première partie, soulignons que la communication orale homme-machine fait apparaître, lorsqu'on étudie les utilisateurs potentiels de ce type de technique, deux classes principales d'applications :

— des applications spécialisées à l'usage des professionnels,

— des applications destinées au grand-public.

Ces deux familles d'applications déterminent deux populations distinctes d'utilisateurs du dialogue oral homme-machine, que l'on peut caractériser comme suit :

— les utilisateurs de la première classe d'applications étant en contact fréquent avec le système dans l'exercice de leur profession, accepteront plus facilement de contraindre leur expression linguistique ; ils supporteront sans difficulté les désagréments d'une phase plus ou moins

longue d'adaptation qui leur sera facilitée par la pratique régulière du système ; ils se plieront même à un apprentissage systématique, à condition que l'utilisation de la parole leur permette une plus grande efficacité. Des langages artificiels mais suffisants pour exprimer l'ensemble des actions et des objets liés à la tâche ne constituent pas pour eux un obstacle à l'utilisation du système ;

— en revanche, la majorité des usagers de la seconde famille d'applications utiliseront le système de manière très occasionnelle et en dehors de leur activité professionnelle ; leur souci majeur ne sera donc certainement pas l'efficacité mais plutôt le confort d'utilisation. Ils risquent d'être rebutés par des contraintes linguistiques avec lesquelles ils n'auront pas l'opportunité de se familiariser ; si un apprentissage est nécessaire, l'intervalle de temps entre deux utilisations successives sera trop long pour permettre une capitalisation des acquis d'une session à l'autre. Il apparaît donc nécessaire de leur proposer un langage d'interaction avec le système, aussi proche que possible de la langue naturelle, dont les limites soient cernables intuitivement et correspondent à des contraintes d'expression que le contexte pragmatique et la stratégie de conduite du dialogue suffisent à imposer naturellement à l'utilisateur.

Comme on le verra au cours de la présentation des systèmes développés dans notre laboratoire, les architectures de ces deux classes d'interfaces orales, la nature et le volume des connaissances qu'elles mettent en œuvre sont très différents, de même que les algorithmes de reconnaissance/compréhension de la parole. Une évolution progressive d'un type de système à l'autre n'est pas envisageable, du moins actuellement ; on pourrait simplement réunir les deux types d'interfaces au sein d'un même système de dialogue oral. Mais ce type de solution ne présente pas un grand intérêt du point de vue ergonomique.

Il est encore utile de préciser les problèmes et les difficultés spécifiques que l'on rencontre lorsqu'on tente de développer une interface orale homme-machine. C'est l'objet de la partie 3 qui sert à mieux faire comprendre nos choix concernant l'architecture, les sources de connaissances et les stratégies de contrôle mises en œuvre dans ces systèmes de dialogue.

3. Problèmes liés au développement d'interfaces orales

3.1. ASPECTS SPÉCIFIQUES DE L'ENTRÉE DE DONNÉES OU MACHINE A DICTER

Comme nous l'avons déjà signalé, on confond souvent système de reconnaissance de parole et entrée orale de données. S'il est vrai qu'une machine à dicter (MAD) est fortement dépendante du système de reconnaissance de la parole (SRP) qu'elle utilise, elle a besoin d'un certain nombre d'outils pour pouvoir fonctionner en tant que telle.

Lorsque l'on dicte une lettre administrative, par exemple, il arrive très souvent qu'on ait envie de remplacer un mot par un autre ou tout simplement de supprimer ce mot ; un simple SRP ne serait pas en mesure d'assumer cette tâche. Par conséquent, nous pouvons dire qu'une MAD doit comporter au moins deux composants : un SRP très puissant et performant et un éditeur permettant d'effectuer facilement les modifications ou corrections sur le texte reconnu.

Les systèmes de reconnaissance et leurs problèmes

Il nous faut, tout d'abord, distinguer deux types de système de reconnaissance suivant le mode d'élocution choisi :

- continue ou spontanée : le locuteur parle normalement devant la machine comme il le ferait avec une personne ;
- par mots isolés : le locuteur doit alors séparer deux mots consécutifs par une courte pause.

Le but des recherches actuelles en parole est évidemment d'arriver à mettre en œuvre des systèmes utilisant le mode d'élocution spontanée. Malheureusement, beaucoup de problèmes liés à la reconnaissance de la parole continue restent posés et rendent la reconnaissance de la parole très difficile. Parmi ces difficultés on peut citer :

- La continuité du signal temporel qui rend difficile la détermination des frontières de mots.
- Les phénomènes de coarticulation ; chaque phonème est fortement influencé par son contexte d'élocution.
- La variabilité inter- et intra-locuteur.
- La redondance du signal acoustique qui véhicule plusieurs informations comme la structure syntactico-sémantique, l'identité, le sexe du locuteur, etc.

Un bon système de reconnaissance doit tenir compte de ces points très importants. Outre le mode d'élocution, la taille du vocabulaire contribue à différencier les SRP. Ce sont des SRP différents qui traitent les grands vocabulaires (parfois plus de 10 000 mots) et les vocabulaires de quelques dizaines de mots. Dans ce dernier cas, plusieurs images pour un même mot (acoustiques ou phonétiques) peuvent être stockées en mémoire ce qui facilite la reconnaissance. En revanche les SRP utilisant de grands vocabulaires ne peuvent se permettre de faire la même chose pour des raisons évidentes de place mémoire et de temps d'apprentissage.

Par ailleurs, les performances d'un système de reconnaissance seront meilleures si l'utilisateur accepte de se soumettre à une phase initiale, dite d'apprentissage, qui, suivant les cas, peut durer de quelques minutes à plusieurs heures. A l'heure actuelle, une reconnaissance monolocuteur avec apprentissage ou adaptation au locuteur est beaucoup plus fiable qu'un décodage véritablement multilocuteur.

Proposer à l'utilisateur une telle phase d'apprentissage est envisageable dans le cas d'une utilisation fréquente, mais apparaît comme peu réaliste et inefficace vis-à-vis d'utilisateurs occasionnels.

Dans le cas d'un vocabulaire de taille importante, il est courant qu'on fasse prononcer à chaque nouveau locuteur un certain nombre de phrases représentatives sur les plans

phonémique ou syllabique, selon que l'unité de base est respectivement le phonème ou la syllabe. C'est le cas du système Parsifal, développé par IBM-France, où chaque nouveau locuteur doit d'abord passer par une phase d'apprentissage qui nécessite la lecture de 400 phrases courtes [41]. Dans le cas de la réalisation d'une machine à dicter, il est tout à fait concevable de passer par une telle phase d'apprentissage, étant donné qu'il s'agit d'un outil spécifique intégré à une station de travail multimédia personnelle.

La recherche de mots dans un grand vocabulaire pose, elle aussi, des problèmes délicats, il est évident qu'on ne peut faire une recherche lexicale séquentielle du mot à reconnaître. Pour remédier à ce problème, Zue et Aull [24] ont étudié l'utilisation des classes phonétiques majeures pour permettre de réduire le domaine de recherche. Mais les traits robustes en parole continue sont peu discriminants. En effet, dans une phase de présélection de sous-vocabulaires utilisant une telle technique, beaucoup de mots ne sont pas détectés avec les classes phonétiques majeures, de plus, en parole continue, beaucoup d'erreurs (fausse détection) apparaissent à la jonction entre plusieurs mots [20]. En ce qui nous concerne, dans le système MACDIC [21] nous avons écarté l'idée d'utiliser les classes phonétiques majeures, étant donné que notre objectif est la reconnaissance de la parole continue.

Le choix de l'unité de décision dans un système de reconnaissance est aussi particulièrement délicat ; ce problème est clairement posé dans [40] en termes d'avantages et d'inconvénients. Pour Parsifal, l'unité de base choisie est la syllabe ; pour notre part, nous avons préféré choisir le phonème en tant qu'unité de décision, à cause du nombre réduit des phonèmes en français et de la qualité des résultats d'APHODEX [22], décodeur acoustico-phonétique (DAP) conçu par notre équipe autour d'un système expert. Malheureusement, et ceci quelle que soit la qualité des DAP existants, le treillis phonétique engendré est toujours entaché d'erreurs, ce qui rend très difficile la tâche du module linguistique qui est de reconnaître les phrases.

Encore une fois, les techniques utilisées diffèrent selon la taille du vocabulaire et le domaine d'application. Dans le cas de langages artificiels, il est possible d'écrire une grammaire [19] qui modélise le langage spécifique de l'application. En revanche, lorsque l'on s'intéresse au langage naturel, le problème est beaucoup plus difficile. Des études fondamentales sur des modèles syntaxiques et sémantiques de représentation du langage sont alors nécessaires. S'il est possible d'obtenir des résultats probants pour la représentation de langages finalisés [23], dans le cas de la machine à dicter l'approche la plus utilisée est l'approche probabiliste. En effet, le but de la syntaxe en reconnaissance de la parole est de participer au choix des candidats d'une phrase à transcrire. Comme la séquence de mots d'une phrase obéit à des contraintes syntaxico-sémantiques, il est tout à fait naturel d'utiliser un modèle probabiliste ; la probabilité d'un mot dans une phrase prononcée dépend conditionnellement de toute la première partie de cette phrase. En pratique, il est très difficile d'estimer cette probabilité, c'est pourquoi Jelinek

[13] restreint le champ de calcul de la probabilité de la chaîne de mots au calcul de la probabilité des suites de trois mots. Le problème qui se pose dans ce cas est que, lorsque le vocabulaire est de taille importante, il est hors de question de calculer des probabilités sur des séquences de 3 mots car cette opération nécessiterait un corpus d'apprentissage de taille démesurée. Pour remédier à ce problème, l'idée utilisée est de revenir aux matrices de précédentes fréquentielles [25] et de se fonder non plus sur des suites de mots pour calculer des probabilités de séquences de mots mais sur des séquences de classes syntaxiques. Pour notre part, nous avons construit un modèle probabiliste avec 200 classes syntaxiques. Ce modèle est de plus augmenté à l'aide de modèles syntaxiques locaux complémentaires qui permettent de filtrer les phrases reconnues par le système [21].

L'éditeur associé à la machine à dicter

Comme nous l'avons précisé plus haut, il est indispensable dans le cas d'une saisie orale de données d'associer au SRP un éditeur pour effectuer les modifications ou corrections nécessaires sur le texte reconnu. Deux approches sont alors possibles suivant que l'on choisit un éditeur vocal ou non.

La première solution est difficile à mettre en œuvre car il faut pouvoir distinguer du flot des données dictées les mots du langage de commande de l'éditeur.

Quant à la deuxième approche, elle est beaucoup plus abordable surtout dans une station de travail multi-média. Il suffit de permettre au locuteur d'interrompre le système de reconnaissance avec, par exemple, la souris et de donner la main à un éditeur classique qui permettra de cliquer sur le mot à modifier. Un système de reconnaissance de mots isolés pourra être déclenché pour remplacer le mot sélectionné. Une solution plus facile consisterait à modifier le mot désigné par le clavier ou à sélectionner, par la souris, le mot correct parmi les autres hypothèses fournies par le système [26].

3.2. ASPECTS SPÉCIFIQUES DE LA COMPRÉHENSION ET GESTION DE DIALOGUES ORAUX

Nous venons de voir que la reconnaissance de la parole pour la saisie de données pose de nombreux problèmes. La compréhension et la gestion de dialogues oraux, qu'il s'agisse de langages artificiels ou de langues quasi naturelles, soulèvent des difficultés supplémentaires spécifiques :

— comprendre et gérer un dialogue ne consiste pas simplement à reconnaître des énoncés ; or, la structuration des échanges oraux, l'enchaînement des tours de parole, les stratégies de contrôle du dialogue oral ont été jusqu'à présent relativement peu étudiés ;

— la communication orale homme-machine ne fonctionne certainement pas comme le dialogue oral homme-homme. Certaines études portant sur des dialogues homme-machine simulés indiquent que les attentes et le comportement de l'utilisateur ne sont pas identiques face à une machine et face à un interlocuteur humain (cf. le tableau donné en annexe 1) ; mais surtout, la fiabilité du canal de communication (ici la parole) est bien inférieure dans le cas du

dialogue homme-machine. Dans un système informatique de compréhension et de gestion de dialogues oraux, la fonction de **gestion du canal de communication** prend donc une importance particulière.

Nous présentons dans la suite plus en détails nos études sur le dialogue qui montrent l'importance d'une bonne intégration de la reconnaissance et de la gestion de la tâche. Pour faciliter la lecture des paragraphes consacrés à la présentation des systèmes que nous développons, il est nécessaire de récapituler les différentes fonctions que doit assurer un module de gestion de dialogues oraux finalisés et d'indiquer la nature des connaissances nécessaires pour réaliser ces fonctions de manière intelligente et efficace.

Gestion du canal de communication

Deux fonctions distinctes doivent être prises en charge :

- *la validation des résultats de la reconnaissance phonétique* : le système devra, en fonction des cas, engendrer différents types de messages ; nous donnons sous forme de tableau un exemple de stratégie possible.

Qualité de la reconnaissance	Type de messages à générer
Ambiguïtés lexicales, syntaxiques et sémantiques	Pas de message, choix de la meilleure hypothèse jusqu'à contestation de la part de l'utilisateur
Score de reconnaissance faible	Demande de confirmation
Mauvaise reconnaissance	Demande de répétition ou de reformulation

- *la gestion du canal utilisateur-machine* : deux types de messages sont associés à cette fonction, messages de relance pour encourager l'utilisateur à poursuivre et messages de maintien de la communication pour faire patienter l'utilisateur dans le cas d'un traitement interne long.

Gestion de l'application

Cette fonction présente deux aspects essentiels :

- à partir des énoncés de l'utilisateur, détection et, éventuellement, négociation de ses buts et des actions à entreprendre sur l'application, (liaison utilisateur-application),
- exécution des actions demandées par l'utilisateur sur l'application, à condition qu'elles relèvent du domaine de la tâche et de la compétence du système (liaison système-application) ;

Compréhension et gestion du dialogue

- aspect **machine-canal-tâche** : sont inclus sous ce titre, le choix et la mise en œuvre, par le système, d'une stratégie de dialogue ; ce qui implique, en particulier, la classification des énoncés de l'utilisateur en fonction de leur contenu sémantique (relatif à la tâche ou au canal de communication) ainsi que la gestion d'un historique du dialogue.
- aspect **utilisateur-canal-tâche** : c'est-à-dire, entre autres, la résolution par le système des ellipses et des anaphores présentes dans le discours de l'utilisateur, la détection et la prise en compte des changements de thème.

La nature des fonctions nécessaires pour gérer efficacement un dialogue oral détermine celle des connaissances à mettre en jeu. Les domaines sur lesquels doit porter le savoir du système sont donc :

- le support de la communication orale, c'est-à-dire le langage utilisé, ses structures, sa sémantique, son lexique,
- la tâche et l'application,
- le type de dialogue,
- l'utilisateur.

Dans la suite de cet article, nous présenterons les solutions que nous avons retenues, pour doter de stratégies naturelles de dialogue les différents prototypes d'interfaces orales admettant la langue naturelle en entrée, développés par notre équipe au cours des trois dernières années. Nous ne décrirons pas les techniques de reconnaissance de parole mises en œuvre dans de tels systèmes dont [18] et [19] fourniront une description assez complète mais nous illustrerons difficultés et solutions rencontrées lors de la définition des dialogues, en prenant nos exemples dans deux familles très différentes d'applications, donc deux types d'interfaces orales nettement différenciées :

— d'une part, des applications classiques où l'ordinateur intervient essentiellement comme exécutant et qui mettent en jeu un univers pragmatique restreint : nous avons retenu à titre d'exemple un système oral de messagerie électronique, PARTNER ;

— d'autre part, des applications concernant des activités cognitives plus complexes dans un univers pragmatique plus riche : le renseignement téléphonique grand public ; nous avons choisi de discuter l'architecture et les stratégies d'un système de gestion de dialogues oraux finalisés, DIAL, destiné au renseignement administratif.

4. Dialogue naturel oral et applications interactives classiques

Dans le cas d'une interaction portant sur des tâches relativement simples comme l'édition et le traitement de textes, l'utilisateur n'a pas besoin de l'aide du système pour sélectionner puis ordonner les actions élémentaires appropriées à la réalisation de ses buts ; le logiciel peut donc fonctionner comme un simple exécutant et se dispenser de toute initiative.

Il paraît donc inutile, dans ce type d'application, de donner au système de gestion de l'interface utilisateur la capacité de détecter les objectifs de l'utilisateur comme de déterminer ses aptitudes, ses connaissances et son fonctionnement cognitif. L'adaptation à des usagers de compétence et de comportement différents peut s'effectuer de manière statique au moyen d'un modèle a priori des différentes classes d'utilisateurs potentiels [7]. Il semble, en outre, que la communication entre l'utilisateur et le système se réduise naturellement, par le seul effet des contraintes pragmatiques, à une suite d'échanges élémentaires indépendants les uns des autres, et que le contrôle de la communication puisse être, sans inconvénient, confié entièrement à l'un des deux interlocuteurs.

En bref, il apparaît superflu de mettre en place les structures nécessaires à la gestion d'un véritable dialogue coopératif finalisé ; un modèle fondé sur la notion d'échange élémentaire semble suffisant.

Mais l'usage de la langue naturelle, même dans le cadre d'un univers pragmatique restreint, impose à chaque interlocuteur :

- d'une part, de mémoriser au moins le dernier échange,
- d'autre part, de s'efforcer de déterminer le(s) but(s) courant(s) de l'utilisateur et de conserver en mémoire de travail une trace du (des) dernier(s) but(s) satisfait(s) (ou actifs),

pour être en mesure de résoudre les références pronominales, plus généralement les **anaphores** ainsi que les **ellipses** (des exemples sont donnés au début du paragraphe 4.1.2). Ces phénomènes linguistiques apparaissent fréquemment dans les énoncés constitués de plusieurs propositions [27] et donc, a fortiori, dans les dialogues en langue naturelle, à l'intérieur d'un échange ou même d'un sous-dialogue. Et il a été montré que, dans les situations simples où l'utilisateur se borne à lancer des commandes en séquence, la quasi-totalité des ambiguïtés créées par la présence d'anaphores et d'ellipses pouvait être facilement levée par la simple prise en compte du contexte immédiat et des connaissances pragmatiques relatives au domaine d'application [28].

En outre, dans le cas du dialogue oral homme/ordinateur, la gestion de la communication proprement dite revêt une importance particulière, en raison de la présence d'énoncés bruités partiellement reconnus ou compris, ce qui rend les sous-dialogues de clarification nettement plus fréquents que dans le dialogue homme/homme ; une hiérarchie des échanges apparaît donc, qui oblige à construire et à maintenir un historique plus ou moins complet du dialogue et de l'activité en cours.

Signalons que le mécanisme de traitement des ellipses peut être appliqué avec succès aux énoncés incomplets du fait des limites de la reconnaissance/compréhension, afin de combler des lacunes de la reconnaissance.

Ce sont donc essentiellement des phénomènes d'ordre linguistique qui justifient la présence, dans les interfaces orales destinées à des applications classiques peu complexes, d'un mécanisme de détection des buts de l'utilisateur d'une part, d'un historique du dialogue et de l'interaction d'autre part. Nous indiquons dans la suite du paragraphe comment nous avons traité le problème des anaphores et des ellipses dans notre système oral de messagerie électronique, PARTNER ; nous précisons en particulier les structures de données que nous avons dû créer, au niveau du dialogue et de la tâche, pour y parvenir.

4.1. UN EXEMPLE SIGNIFICATIF : LE SYSTÈME PARTNER [29]

4.1.1. Architecture

L'architecture d'ensemble de PARTNER, qui a été développée dans le cadre d'un projet COST de la C.E.E. avec l'Université de Munich, est fortement modulaire et permet

de paramétrer le système à la fois par l'application et le langage. Ainsi, l'application de messagerie orale électronique fonctionne-t-elle actuellement en français et en allemand. En outre, bien que la messagerie ait été la première application-test mise en œuvre, il a suffi de quelques jours pour développer, à l'aide de PARTNER, une interface de commande de robot.

PARTNER peut donc être intégré en tant qu'interface orale à des systèmes classiques en vue d'assurer un dialogue en langue naturelle entre l'utilisateur et des applications relativement simples.

Le système se compose de trois modules hiérarchisés de traitement :

- reconnaissance des énoncés de l'utilisateur (CASSIS),
- gestion du dialogue (DIALOG),
- gestion de l'activité/tâche (VOC).

A noter que le module DIALOG met en œuvre des mécanismes généraux et agit en étroite interaction avec le module de gestion de la tâche, ce qui le rend complètement indépendant de l'application et facilite le passage d'une application à une autre.

La figure 1 fournit un schéma simplifié de l'architecture du système.

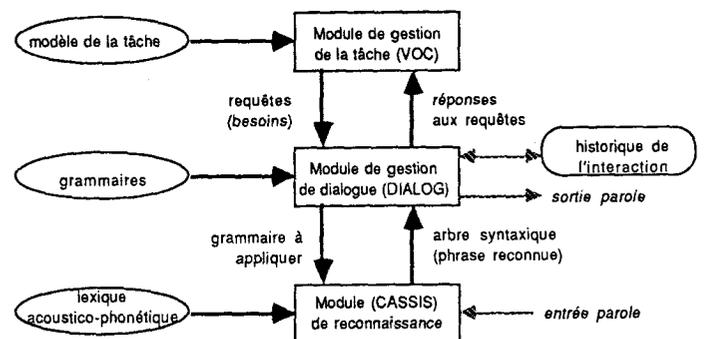


Figure 1. — Architecture de PARTNER.

Représentation des connaissances et fonctionnement global

Dans le cadre d'une approche d'analyse descendante, CASSIS construit une représentation syntaxico-sémantique de l'énoncé courant, en utilisant une sous-grammaire sémantique appropriée [29], sélectionnée par le gestionnaire de dialogue en fonction du contexte pragmatique (avancement de la tâche en cours) et du dialogue antérieur (nature des derniers échanges).

Exemple : à la phrase : « Je voudrais envoyer un message à DURAND, » CASSIS associe l'arbre syntaxique : PH1 = <ph> (<politesse> (Je voudrais) <verbe-envoi-infinif> (envoyer message) <utilisateur> (DURAND)).

Dès qu'un but de l'utilisateur, indiqué le plus souvent sous forme prédicative par un verbe, a été identifié, le gestionnaire de la tâche entreprend sa réalisation en sélectionnant et en activant, dans le modèle de la tâche où chaque commande est représentée par un automate, celui qui

correspond au but de l'utilisateur ; il envoie ensuite au gestionnaire de dialogue des **requêtes** ou **besoins** pour obtenir les informations nécessaires à l'exécution de la commande activée. Si l'utilisateur a omis de préciser certaines de ces informations, DIALOG, pour répondre aux requêtes besoins correspondantes émises par VOC, décide s'il doit interroger l'utilisateur ou si une consultation, soit de l'historique, soit de l'ensemble des valeurs par défaut, suffit. VOC détermine le but courant de l'utilisateur en envoyant au gestionnaire de dialogue une requête vide qui conduira ce dernier à générer, par exemple, le message « *Que voulez-vous faire ?* » à l'intention de l'utilisateur qui précisera alors un nouveau but.

VOC exprime ses requêtes au moyen de deux primitives : EXTEND qui permet d'obtenir la classe ou valeur sémantique d'une unité d'analyse produite par CASSIS et VALUE qui rend la fraction d'énoncé correspondant à une telle unité ; ainsi, EXTEND (<verbe>, PH1) renvoie « envoi » qui désigne l'action d'envoyer des messages, tandis que VALUE (<utilisateur>, PH1) sélectionne la chaîne de caractères « DURAND ».

VOC contrôle donc l'ensemble des traitements effectués par PARTNER, sauf en cas de divergence manifeste entre les actions lancées par le système et les exigences exprimées par l'utilisateur ; DIALOG prend alors l'initiative. Ce module peut en effet déclencher une resynchronisation de VOC (changement de l'état courant de l'automate actif ou activation d'un autre automate), chaque fois que l'utilisateur conteste une initiative du système, interrompt ou annule une action en cours ou encore désire reprendre une activité suspendue.

4.1.2. Le traitement des ellipses et des anaphores dans PARTNER

PARTNER permet de traiter quelques cas simples d'ellipses ou d'anaphores. C'est au gestionnaire de dialogue que revient naturellement cette tâche, puisqu'elle exige, pour être menée à bien, la connaissance des derniers énoncés échangés et du contexte pragmatique (action du système, buts de l'utilisateur) dans lequel ils s'inscrivent. DIALOG s'appuie, pour résoudre anaphores, ellipses et énoncés incomplets (dus aux limites actuelles de la reconnaissance automatique) sur un historique de l'interaction comportant deux types d'informations chronologiquement ordonnées :
 — les représentations syntaxico-sémantiques des n derniers énoncés,
 — la liste des réponses fournies par DIALOG aux p dernières requêtes/besoins exprimées par le gestionnaire de la tâche ; nous appelons cette liste « trace des besoins satisfaits ».

Pour l'instant, n et p sont fixés de manière arbitraire, mais nous envisageons, dans une étape ultérieure, une ré-initialisation « intelligente » de l'historique, par exemple en association avec chaque changement de thème. On notera que, en raison de la simplicité de la tâche, l'exécution d'une commande/fonction se confond, pour le système, avec la satisfaction d'un but de l'utilisateur ; la trace des besoins satisfaits peut donc être assimilée à celle des **buts successifs de l'utilisateur**.

Les représentations syntaxico-sémantiques des énoncés antérieurs permettent d'interpréter les anaphores, tandis que la trace des besoins satisfaits est utilisée pour hypothétiser les éléments sous-jacents aux énoncés elliptiques.

Cette approche est illustrée par l'exemple suivant extrait de l'application messagerie électronique qui utilise un vocabulaire d'une centaine de mots avec un facteur de branchement maximal de 30 (un exemple de dialogue complet entre l'utilisateur et PARTNER est reproduit dans l'annexe 2) :

Sujet1 : *Est-ce que j'ai reçu un message de DURAND ?* → Partner1 : Oui
 Sujet2 : *Et de SCHMITT ?* (ellipse) → Partner2 : Non, ctmaintenant ?
 Sujet3 : *Je voudrais écouter le message de DURAND* → Partner3 : Voici ce message
 réunion RFIA
 lundi prochain.

Sujet4 : *Détruis-le.* (anaphore)

L'anaphore :

La représentation dans l'historique des deux derniers énoncés de l'utilisateur se présente comme suit :

PH3 = <ph> (<politesse> (*Je voudrais*) <verbe-écoute-infinitif> (*écouter*) <msg> (*message* (<utilisateur> (*DURAND*))))
 PH4 = <ph> (<verbe-suppression-impératif>) (*Détruis*).

Lors de l'exécution de la commande de destruction, la nature de l'action à effectuer sera détectée par VOC grâce à la requête EXTEND (<verbe>, PH4) à l'intention du gestionnaire de dialogue. Pour répondre correctement à la requête VALUE (<msg>, PH4) et préciser à VOC quel objet il doit détruire, DIALOG consultera la représentation syntaxico-sémantique PH3 dans l'historique, puisqu'aucune désignation de message ne figure ni dans PH4 ni dans l'ensemble des valeurs par défaut associées à l'application.

L'ellipse :

La trace, dans l'historique, des besoins satisfaits contient, après traitement des deux premiers énoncés, les informations suivantes :

action	objet	utilisateur	temps
exist	msg	DURAND	any
$n1$	$n2$	$n3$	$n4$

où n_i ($1 < i < 4$) désigne, dans l'automate associé à la commande de recherche d'un message, l'état atteint après obtention de l'information correspondante. En ce qui concerne l'énoncé suivant : « *Et de SCHMITT ?* », qui donne lieu à une arborescence syntaxico-sémantique incomplète :

PH2 = <ph> (<utilisateur>) (*SCHMITT*)

le gestionnaire de dialogue supplée aux lacunes de la représentation en consultant dans l'historique la trace des derniers besoins satisfaits, ce qui lui permet de répondre aux requêtes de VOC relatives à PH2 et d'enrichir la trace des besoins satisfaits par la séquence :

action	objet	utilisateur	temps
exist	msg	SCHMITT	any
$n1$	$n2$	$n3$	$n4$

Cette stratégie permet de traiter efficacement non seulement les ellipses mais aussi certains énoncés incomplets (dû à un échec partiel de la reconnaissance) et les changements de thème.

4.2. BILAN DE CE TYPE DE SYSTÈME

PARTNER, actuellement sous forme de prototype, met donc en jeu la plupart des structures d'informations et des mécanismes indispensables à une gestion évoluée du dialogue, c'est-à-dire :

— la détection et le suivi des buts de l'utilisateur,
— la construction puis la mise à jour, au fur et à mesure de l'interaction :

- d'un historique des échanges d'informations entre l'utilisateur et le système,
- d'une trace de l'activité de l'ensemble homme/système (ou interaction), qui peut être confondue avec la trace des buts successifs de l'utilisateur.

Toutefois, les tâches susceptibles d'être exécutées sont actuellement trop simples et l'environnement pragmatique encore trop pauvre pour que toutes les possibilités offertes par ces outils soient exploitées :

— le guidage de l'utilisateur dans le choix et l'enchaînement des actions à exécuter par le système s'avère inutile ;
— il est également superflu d'implanter des stratégies flexibles de contrôle du dialogue ; un modèle du type question/réponse, qui assimile le dialogue à une suite d'échanges élémentaires indépendants, est suffisant, pour peu que, comme nous l'avons fait dans PARTNER [29], on y intègre une procédure de re-synchronisation en cas de contestations de l'utilisateur, et conduit à une gestion du dialogue, qui peut être considérée comme naturelle dans le contexte considéré.

Historique du dialogue et trace de l'interaction (ou des buts successifs de l'utilisateur) servent donc simplement, dans ce type d'interface, à interpréter les anaphores, les ellipses et les lacunes éventuelles dues aux erreurs de reconnaissance.

Il serait aisé de doter les interfaces de communication homme/ordinateur classiques, qui dialoguent avec l'utilisateur par l'intermédiaire de menus, d'un langage de commande artificiel ou d'une combinaison des deux, de mécanismes capables de traiter anaphores et ellipses. En effet, dans le cas d'applications mettant en jeu un univers pragmatique restreint et des tâches simples sur le plan cognitif, la solution adoptée dans PARTNER conviendrait parfaitement.

C'est ce que nous avons mis en œuvre dans deux prototypes, SATIC et DIAPASON, développés conjointement avec des industriels et dont les caractéristiques communes sont :

- vocabulaire limité d'environ 100-150 mots,
- facteur de branchement variant de 15 à 30,
- tâche bien structurée, fonctionnant sans parole avec une interface classique de type menus hiérarchisés.

Ces deux systèmes diffèrent fondamentalement par l'approche choisie pour la reconnaissance qui est de type globale dans SATIC et de type analytique dans DIAPASON.

Le système SATIC avec COGNITECH [30]

Utilisant en entrée une carte de reconnaissance de mots enchaînés, ce système permet à un chirurgien de dialoguer avec un système expert de conseil et de génération automatique de comptes rendus médicaux. Le domaine choisi est celui de la chirurgie caelioscopique et plus particulièrement du traitement des grossesses extra-utérines et des adhérences. L'objectif est de permettre au chirurgien de construire automatiquement en cours d'opération un compte rendu d'intervention et d'obtenir des aides sous forme de conseils diagnostiques et thérapeutiques. Développé en lien avec le Professeur Bruhat de la Polyclinique de Clermont-Ferrand, SATIC a été exposé avec succès au congrès mondial de chirurgie obstétrique en juin 1989.

Voici, à titre d'exemple, un dialogue obtenu avec SATIC (C : indique une phrase prononcée par le chirurgien ; S : une phrase prononcée en synthèse vocale par SATIC) :

C : Début SATIC

C : Diagnostic GEU ampulaire gauche

S : Confirmation diagnostic Grossesse Extra-Utérine localisation ampulaire gauche

C : OK

C : Traitement

S : Confirmation traitement

C : D'accord

S : Changement de chapitre non autorisé

S : Taille ?

C : trente

S : Taille en centimètre

C : Trois centimètres, rompue

S : Confirmation taille trois centimètres, rompue

C : OK

C : Suite ?

S : Complication ?

C : Hémopéritoine de 40 cc

S : Confirmation Hémopéritoine de 400 cc

C : Étiologie antécédent de GEU

S : Autre diagnostic annexe adhérence ?

C : Annexe controlatérale gauche normale

S : Incohérence entre annexe gauche normale et GEU gauche

C : Non, pas annexe gauche normale

C : Annexe droite normale

Le système DIAPASON avec THOMSON-SINTRA-DASM [31]

Ce système avait pour objectif de montrer l'apport d'un dialogue oral avec une console de visualisation incluant à la fois un système de reconnaissance analytique, multilocuteur et très robuste et un dialogue souple. L'étude a été menée dans le cas réel de la console d'un sonar opérationnel dont les commandes sont complexes. La comparaison entre le dialogue de commande du Sonar réel et la maquette implantée sur PC a eu lieu sur la plate-forme d'intégration des sonars chez Thomson. Les résultats obtenus ont montré l'apport indéniable de la parole en complément des autres modes de communication : graphisme-écran, souris, clavier.

Cette expérimentation a montré que l'intégration aux interfaces classiques de mécanismes de traitement des anaphores et des ellipses améliorerait sensiblement leur efficacité et leur convivialité, notamment en permettant d'accroître la concision des dialogues par l'élimination de la majorité des redondances et en contribuant au confort de l'utilisateur ainsi qu'au naturel du dialogue.

Par ailleurs, de nombreuses expériences ergonomiques de simulation de dialogues homme/ordinateur en langage naturel (écrit ou oral) ont montré que, si les sujets ont tendance à simplifier leur expression (syntaxe et lexique essentiellement) lorsqu'ils pensent dialoguer avec une machine [32, 33, 34] allant même parfois, au moins à l'écrit, jusqu'à se créer un pseudo-langage de commande [35], ils ne renoncent jamais ni à l'usage des anaphores ni à celui des ellipses [33, 34], en dépit de la réelle complexité de ces procédés. De tels résultats expérimentaux suggèrent que anaphores et ellipses constituent des mécanismes puissants d'élimination des redondances dont l'expression humaine sait difficilement se passer. Ils peuvent, de ce fait, être considérés comme des modes d'expression naturelle, même en dehors du cadre de la langue naturelle.

5. Vers un système de gestion de dialogues pour le renseignement téléphonique : DIAL

Le renseignement téléphonique, une des applications à très long terme les plus prometteuses des systèmes de compréhension de la parole continue, dans la mesure où il est destiné au grand public, requiert des opérateurs une double expertise [4]. En effet, ceux-ci doivent posséder une maîtrise suffisante du domaine technique ou scientifique concerné, mais aussi être capables d'assister l'utilisateur dans sa recherche d'informations, grâce à une gestion du dialogue efficace, visant à l'économie des échanges.

On est donc en présence d'une famille d'applications où l'utilisateur, pour atteindre ses objectifs, doit mettre en œuvre des stratégies cognitives qu'il maîtrise mal, où la réalisation de ses buts implique la combinaison d'actions élémentaires et où, contrairement aux applications évoquées en 4, il est utile d'assister l'utilisateur dans sa démarche et de l'aider à choisir et à ordonner les actions nécessaires à la réalisation de ses buts. Par conséquent, les systèmes de renseignement ne doivent pas se borner à satisfaire les requêtes des utilisateurs mais doivent également les aider à préciser leurs objectifs, donc être capables de gérer avec efficacité de véritables dialogues coopératifs finalisés. En outre, comme la population visée est le grand public, il faut envisager un dialogue non seulement naturel mais aussi en langue naturelle, afin d'épargner à l'utilisateur tout apprentissage préalable de la mise en œuvre de l'interface homme/ordinateur.

Nous décrivons ci-après les structures d'informations et les procédures qui assurent dans le système DIAL un dialogue oral, naturel, coopératif et finalisé, c'est-à-dire un dialogue qui :

— assiste l'utilisateur dans sa recherche d'informations,

— assure un contrôle souple du dialogue, capable d'évoluer au cours de l'interaction en fonction du comportement de l'utilisateur et de l'avancement de la tâche en cours, afin de minimiser le nombre et le volume des échanges.

Nous passerons sous silence l'architecture d'ensemble de DIAL déjà décrite par ailleurs [37, 38], pour détailler davantage la composante de gestion de dialogue. De même, nous évoquerons brièvement le traitement des anaphores, des ellipses et des énoncés partiellement reconnus, pour insister davantage sur la mise en œuvre de stratégies de contrôle des échanges en vue d'un dialogue à la fois naturel et efficace.

Précisons toutefois que les différentes composantes de DIAL ont été testées séparément avec des données simulées complexes. Il resterait, maintenant, à intégrer les différents modules, ce que nous ne ferons vraisemblablement pas pour l'instant, en raison des temps d'exécution prohibitifs sur le matériel dont nous disposons. En revanche, nous projetons de tester chaque module avec des données réelles incluant les informations en provenance des autres modules.

5.1.1. Architecture du gestionnaire de dialogue du système DIAL

DIAL admet en paramètre un domaine de renseignement plus ou moins vaste. Nous avons choisi comme application-test le renseignement administratif (cf. les pages roses de l'annuaire), en raison de la complexité et de la richesse de l'univers pragmatique correspondant. Les horaires SNCF, par exemple, offrent des situations beaucoup plus pauvres sur le plan de l'activité cognitive, qui donnent lieu à des dialogues peu différents, en termes de complexité de ceux gérés par PARTNER.

Le gestionnaire de dialogue de DIAL est un système à bases de connaissances multiples : nous présenterons en parallèle les fonctionnalités des différents modules qui le composent et les sources de connaissances utilisées par chacun d'eux (cf. le schéma de la figure 2). Nous ne détaillerons pas l'étape de génération des réponses du système à partir d'un contenu informationnel défini préalablement, car elle relève uniquement du traitement des langues naturelles.

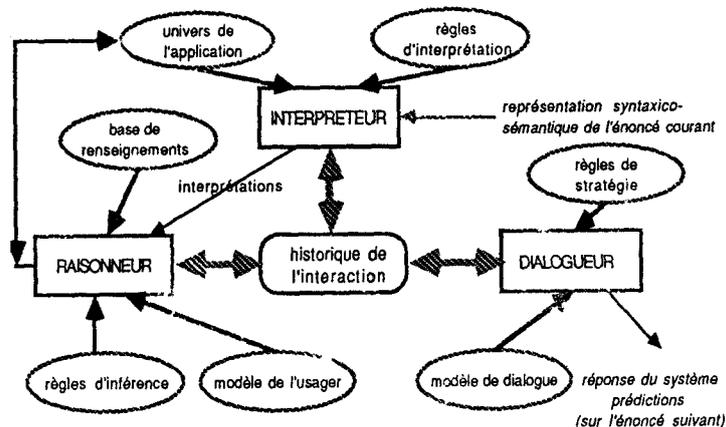


Figure 2. — Architecture de DIAL.

Interprétation des énoncés de l'utilisateur

Les modules qui assurent dans ce système la reconnaissance et la compréhension des interventions orales en langue naturelle de l'utilisateur fournissent à l'interpréteur, pour chaque énoncé de l'utilisateur, les différentes **représentations syntaxico-sémantiques** possibles de cet énoncé, sous forme de prédicats générés par une grammaire de cas [36]. À chaque représentation est associé un score de plausibilité qui rend compte de la qualité des résultats de la reconnaissance/compréhension.

L'**interpréteur** résout anaphores, ellipses et représentations incomplètes à l'aide d'une **base de règles de production contextuelles d'interprétation** qui font appel essentiellement, d'une part, à des connaissances pragmatiques et linguistiques, d'autre part, à l'historique du dialogue et de l'activité (ceci sera décrit plus loin). Il construit, pour chaque énoncé de l'utilisateur, une **interprétation** représentée par un prédicat logique du premier ordre.

En raison de la richesse de l'univers pragmatique associé à l'application et de la complexité des activités cognitives permises, l'interprétation des anaphores et des ellipses est beaucoup plus difficile que dans le cadre des applications prises en charge par PARTNER, ce qui justifie l'expression, sous forme déclarative, des connaissances nécessaires pour traiter ces mécanismes linguistiques et des traitements eux-mêmes.

Raisonnement sur les informations fournies par l'utilisateur

Le **raisonneur**, constitué essentiellement par un moteur d'inférence spécifique, déduit de l'énoncé courant de l'utilisateur ou, plus précisément, d'une interprétation de cet énoncé sélectionnée en fonction de sa plausibilité, les informations utiles au système pour satisfaire les buts de l'utilisateur. Il vérifie également que les informations obtenues sont cohérentes avec celles contenues dans l'historique de l'interaction et avec les connaissances pragmatiques relatives au domaine d'application.

Les **règles d'inférence** s'appuient sur les connaissances statiques décrivant l'univers pragmatique, sur les informations dynamiques conservées dans l'historique du dialogue et de l'activité ainsi que sur le modèle de l'utilisateur.

Les **connaissances pragmatiques** sont regroupées en deux ensembles :

— une base de données (base de renseignements) qui contient les informations relatives à l'application (exemple : la durée de validité d'un visa, le prix d'un timbre fiscal, une adresse) ;

— un ensemble d'objets structurés décrivant les concepts liés à l'application (exemple : obtenir un document administratif, être majeur, étranger, etc.).

La présence d'un tel module de raisonnement semble liée à la spécificité même de la famille d'applications considérée ici le renseignement.

Pour l'instant, la structure et le rôle du **modèle de l'utilisateur** sont comparables à ceux de PARTNER, c'est-à-dire que la structure d'information correspondante se réduit à un ensemble de valeurs par défaut, propres à l'application. Le raisonneur peut utiliser ces valeurs pour suppléer aux

lacunes que comportent les informations fournies au cours des échanges par l'utilisateur. Néanmoins, les données qu'il contient sont plus riches et plus diversifiées que dans PARTNER, du fait de la complexité supérieure du contexte pragmatique. Nous donnons, à titre indicatif, un exemple :

tout usager qui ne précise pas sa nationalité est considéré comme français jusqu'à preuve du contraire (incohérence détectée par le raisonneur entre cette valeur par défaut et les informations fournies par l'utilisateur).

5.2. GESTION DU DIALOGUE

À partir des informations contenues dans l'historique, le **dialogueur** détermine le contenu de la prochaine intervention orale du système en direction de l'utilisateur. Les sources de connaissances qu'il utilise se composent d'un **modèle de dialogue** et d'un ensemble de **règles de stratégies** qui permettent la mise en œuvre des informations contenues dans l'historique, en vue d'assurer le contrôle du dialogue et la réalisation de la tâche.

Nous évoquerons d'abord les informations dynamiques, c'est-à-dire l'historique, puis les sources de connaissances (informations statiques), à savoir le modèle de dialogue et l'ensemble des règles qui définissent les stratégies du dialogueur.

L'historique

Cette mémoire à moyen terme contient trois types d'informations :

— d'une part, une base de faits comprenant, à l'instar de l'historique de PARTNER :

- les informations ou **faits**, obtenus au cours des échanges antérieurs et validés par le module de raisonnement. En raison de la complexité de l'univers pragmatique de l'application-test choisie, un historique dont la portée serait réduite aux derniers échanges du dialogue comme dans PARTNER ne suffit pas. DIAL conserve donc les interprétations (sous forme de prédicats) correspondant à la totalité des énoncés produits par l'utilisateur au cours du dialogue, ainsi que les prédicats déduits de ces interprétations par le raisonneur ;

- une trace des **buts courants** de l'utilisateur sous forme d'une arborescence de prédicats. La détection de ces buts est plus complexe ici que dans les applications prises en charge par PARTNER et nécessite la mise en œuvre des connaissances pragmatiques ou des informations conservées dans le modèle de l'utilisateur, car celui-ci peut exprimer ses requêtes de manière indirecte (exemple : « J'ai perdu ma carte d'identité. ») ou formuler des demandes incomplètes (exemple : « Je dois aller à Boston. Est-ce que j'ai besoin d'un visa ? »).

— d'autre part, une description sous forme arborescente également, de la **structure d'ensemble du dialogue** en cours avec, pour chaque énoncé, l'indication de son type : requête, apport d'informations (réponse, par exemple), contestation, etc.

Le modèle de dialogue

La présence de cette dernière structure d'informations, absente de l'historique de PARTNER, est indispensable pour gérer des dialogues coopératifs dans le cadre d'une interaction homme/ordinateur mettant en jeu des activités cognitives complexes. En effet, l'analyse de corpus de dialogues oraux homme/homme finalisés [23] a montré que les situations de renseignement où le dialogue porte sur un univers pragmatique complexe et où l'expert assiste l'utilisateur dans sa recherche d'informations donnent lieu à des dialogues fortement hiérarchisés en raison, principalement, de la multiplication des sous-dialogues initialisés par l'expert, soit pour élucider les buts de l'utilisateur, soit pour obtenir de lui les informations nécessaires à la satisfaction de sa requête. Il ne s'agit donc plus simplement, comme dans PARTNER, de gérer les quelques sous-dialogues nécessaires pour corriger les erreurs de la reconnaissance/compréhension ou gérer le canal de communication.

L'analyse du dialogue en cours s'effectue à l'aide d'un modèle de dialogue exprimé au moyen d'une grammaire hors contexte. Ce modèle très simple indique par exemple que, si l'utilisateur pose une question, le système doit y répondre de façon à terminer le sous-dialogue initialisé par la question de l'utilisateur, et que cette réponse peut être suivie ou non (en fonction des stratégies du système) par l'ouverture d'un nouveau sous-dialogue à l'initiative du système, par exemple pour satisfaire un des buts actifs de l'utilisateur.

Les règles de stratégie

Les stratégies de contrôle du dialogue sont exprimées sous forme déclarative de façon à faciliter leur modification. Leur définition exige une attention particulière car ces règles conditionnent dans une large mesure l'efficacité et la convivialité du dialogue, donc de l'interaction. Pour obtenir une stratégie souple permettant une assistance et un guidage efficaces de l'utilisateur tout en préservant un dialogue naturel, nous avons retenu les principes suivants qui résultent d'une analyse fine de dialogues oraux homme/homme de renseignement [4] :

— lorsque l'utilisateur explicite incomplètement l'objet d'une requête au système ou lorsque les informations disponibles sont insuffisantes pour satisfaire une de ses requêtes, ne pas essayer d'obtenir les informations manquantes à l'aide de questions ouvertes mais consulter le modèle de l'utilisateur pour combler ces lacunes et proposer, dès que possible, une solution que ce dernier pourra contester au cas où les hypothèses du système seraient erronées ;

— tant que le but courant de l'utilisateur n'est pas satisfait, réduire les interventions de ce dernier, par exemple en évitant les questions ouvertes ou encore en ouvrant, après chaque énoncé marquant la fin d'un sous-dialogue, un nouveau sous-dialogue en liaison directe avec la satisfaction du but courant ; il pourra s'agir d'un sous-dialogue d'assistance, de focalisation ou de recentrage, selon la nature du but courant et l'état d'avancement de sa réalisation (le thème du sous-dialogue sera déterminé par consultation, dans l'historique, de la trace des buts actifs et des faits déjà acquis).

L'idée sur laquelle repose cette stratégie est qu'une assistance et un guidage efficaces de l'utilisateur par le système impliquent que ce dernier possède la maîtrise du dialogue. À noter que le modèle de dialogue adopté fournit un moyen simple, pour le système, de prendre l'initiative des échanges, en autorisant un même interlocuteur à fermer un sous-dialogue et à en ouvrir un autre immédiatement après, au sein d'une même réplique.

En outre, si l'utilisateur a le contrôle du dialogue, ses interventions seront plus longues et plus complexes, donc plus difficiles à analyser et à comprendre, d'autant que le système ne pourra émettre aucune hypothèse précise ni sur leur structure ni sur leur contenu sémantique. Or de telles **prédictions** simplifient considérablement la compréhension d'un énoncé et diminuent les risques d'erreurs de reconnaissance et de compréhension en permettant, entre autres, une analyse (partiellement au moins) descendante de l'énoncé.

Les avantages de cette **stratégie d'anticipation** qui donne au système la conduite du dialogue sont discutés de manière détaillée dans [4] ; nous citons uniquement les principaux d'entre eux, à savoir : d'une part, la réduction du nombre et du volume des échanges donc de la complexité du dialogue et de l'interprétation des énoncés de l'utilisateur, d'autre part, une assistance efficace de l'utilisateur dans la réalisation de ses buts.

Nous donnons un bref exemple de dialogue produit par DIAL, afin d'illustrer la mise en œuvre de cette stratégie (on trouvera dans l'annexe 3 un exemple de dialogue commenté avec, en parallèle, une description des traitements effectués par DIAL) :

Utilisateur : *Faut-il un visa pour aller en Angleterre ?*

Système : Si vous êtes citoyen de l'un des pays membres de la CEE, le visa est inutile.

Utilisateur : *Mais je suis Marocain !*

Dans ce cas, ...

5.3. BILAN DE CE TYPE D'ÉTUDE

Pour la gestion du dialogue, IDAL se distingue de PARTNER essentiellement par :

— la complexité et la richesse supérieures du modèle de l'utilisateur,

— la portée plus importante de l'historique,

— et surtout, la présence d'un modèle de dialogue et la mise en œuvre de stratégies souples de contrôle des échanges, stratégies dont la complexité nécessite une formulation déclarative, alors que, dans PARTNER, l'initiative étant entièrement à l'utilisateur, il est inutile d'expliciter la stratégie du système.

Le développement de DIAL a montré que, dans le cadre d'un univers pragmatique complexe, si l'on veut fournir à l'utilisateur une assistance adéquate, il est nécessaire de pouvoir :

— gérer des dialogues fortement hiérarchisés, donc disposer d'un modèle de dialogue,

— implanter des stratégies de contrôle du dialogue souples, c'est-à-dire mettre en œuvre des dispositifs qui permettent au système de conduire le dialogue de manière naturelle, sans imposer à l'utilisateur des contraintes explicites comme dans le cas des dialogues à base de menus ; en particulier, disposer d'un modèle de dialogue et d'un mécanisme qui donne au système les moyens de prendre naturellement l'initiative. Ce modèle devrait être suffisamment riche, pour que l'on puisse appliquer une stratégie d'anticipation susceptible de réduire le volume et le nombre des échanges en autorisant l'utilisateur à expliciter incomplètement ses buts et à fournir une partie seulement des informations nécessaires à leur réalisation.

Le modèle de dialogue développé dans le cadre du projet DIAL peut être mis en œuvre avec profit dans toutes les applications qui visent à assister l'utilisateur interactivement dans la réalisation de ses objectifs, quels que soient le support et le mode de communication choisis, dans la mesure où cette source de connaissances permet de gérer de manière naturelle et efficace des dialogues fortement hiérarchisés.

En outre, la stratégie de contrôle du dialogue que nous proposons, stratégie qui met en jeu des mécanismes d'anticipation à partir d'un modèle de l'utilisateur suffisamment informatif, est applicable également à la gestion de dialogues en langage artificiel. Son utilisation pour gérer les échanges entre une application et des utilisateurs naïfs ou occasionnels conduirait à des dialogues à la fois efficaces et naturels, donc conviviaux pour ce type de population. Une telle stratégie conviendrait particulièrement bien pour les interfaces de consultation de bases de données.

6. Conclusion

Il semble donc, au terme de cette brève analyse des études que nous menons dans le domaine de la communication orale homme/ordinateur, plus particulièrement, dans celui de la conception d'interfaces orales de dialogue en langue naturelle, que certains des dispositifs développés dans le cadre de la recherche pourraient être implantés avec profit dans les interfaces classiques de dialogue homme/ordinateur, en vue d'obtenir un dialogue plus naturel, donc plus confortable, pour l'utilisateur. Il s'agit principalement des structures d'informations et des mécanismes permettant :

- le traitement des anaphores et des ellipses,
- la gestion de dialogues fortement hiérarchisés,
- l'implantation de stratégies souples de contrôle du dialogue, stratégie d'anticipation en particulier.

Mais de nombreux obstacles restent à surmonter avant que l'on soit en mesure de concevoir et de développer des interfaces homme/ordinateur véritablement « intelligentes ». La reconnaissance doit encore faire des progrès significatifs pour que de tels systèmes soient possibles. Les autres améliorations que nous proposons visent à la fois l'efficacité de l'interaction et son naturel. Elles portent essentiellement sur le dialogue, donc le style de l'interaction et ne concernent qu'indirectement son contenu, c'est-

à-dire la fonction d'assistance à l'utilisateur que doit assurer tout système interactif moderne. Les principaux obstacles sur lesquels achoppent encore actuellement les recherches, sont les suivants :

— la détection des buts de l'utilisateur peut s'avérer très complexe ; car, dans certaines applications, l'utilisateur n'est pas tenu d'explicitement ses buts ; le système doit donc les déterminer à partir de la seule observation/analyse de son comportement ;

— un modèle statique de l'utilisateur est insuffisant dans un bon nombre d'applications dont la complexité au niveau cognitif impose de construire un modèle du fonctionnement cognitif de l'utilisateur capable de rendre compte de sa démarche et de ses stratégies effectives, c'est-à-dire un modèle qui puisse être appliqué dynamiquement pour analyser son comportement au cours de l'interaction et prévoir ses réactions ultérieures.

Néanmoins, comme nous l'avons montré, il est possible, dès aujourd'hui, d'intégrer un dialogue oral à une interface homme/machine dans le cas de tâches simples (le choix de l'application est alors particulièrement important) et nous pensons qu'à terme, grâce à des dialogues plus naturels, l'efficacité de la communication homme/machine pourra être nettement améliorée.

Manuscrit reçu le 13 octobre 1989.

Annexe 1 : ÉTUDE DE SATISFACTION POUR L'APPLICATION SNCF [39]

CRITERE	PHASE II	PHASE III
(a) gêne du locuteur	20% oui 78% non 2% ?	19% oui 79% non 2% ?
(b) confiance	86% oui 2% non 12% ?	63% oui 5% non 33% ?
(c) comportement différent	48% oui 49% non 3% ?	62% oui 32% non 6% ?
(d) préférence	56% homme 4% machine 40% ?	67% homme 9% machine 24% ?
(e) technique adaptée	94% oui 2% non 4% ?	77% oui 8% non 15% ?

Phase II : simulation d'un dialogue homme-machine sans limitation de compréhension (utilisation de voix synthétique).

Phase III : simulation d'un dialogue homme-machine avec limitation de compréhension (utilisation de voix synthétique).

Résultats obtenus auprès d'utilisateurs ayant accepté de participer à une expérience de simulation pour la SNCF.

**Annexe 2 : EXEMPLE DE DIALOGUE
AVEC LE SYSTÈME PARTNER**

- > Bonjour, ici le système de messagerie du CRIN.
Quel est votre LOGIN ?
 - *Mon Login est Martin.*
- > Bienvenue Martin, que voulez-vous ?
 - *Est-ce que j'ai reçu un message de Durand ?*
- > Oui, le message No 2 du 14-09-86, et maintenant ?
 - *Et de Schmitt ?*
- > Non, et maintenant ?
 - *Je voudrais écouter le message No 2.*
- > Voici le message de Durand : 'Il y aura une réunion lundi prochain.' Et maintenant ?
 - *Sauve-le*
- > Dans quel fichier ?
 - *Dans mailbox*
- > Ok, c'est fait, et maintenant ?
 - *Je veux envoyer une lettre.*
- > Quel est le texte ?
 - *'Ok pour la réunion'*
- > J'envoie ce message à Durand ?
 - *Oui.*
- > Ok, c'est fait, et maintenant ?
 - *De quoi parle le message No 1 ?*
- > Ce message de Pierre parle de : Date du colloque RF-IA, et maintenant ?
 - *Détruis le message No 2.*
- > Je détruis le message No 2 de Durand ?
 - *Euh non, le message de Pierre.*
- > Je détruis le message No 1 de Pierre ?
 - *Oui.*
- > Destruction effectuée, et maintenant ?
 - *Détruis maintenant le message No 3.*
- > Désolé, cette spécification est incorrecte, et maintenant ?
...

Annexe 3 : EXEMPLE DE DIALOGUE AVEC LE SYSTÈME DIAL

Dialogue	Commentaires
S : Renseignements administratifs bonjour.	<-- salutation
U : Bonjour, je voudrais une carte d'identité, à qui dois-je m'adresser SVP ?	--> but (locuteur, obtenir (locuteur, carte_d_iden.) s_adresser (locuteur, but_courant ?) (.sélection dans la base de renseignements de la structure correspondant au but courant et remplacement dans cette structure de " personne" par "locuteur" ; sélection dans cette structure de la formule correspondant à la requête de l'utilisateur, en utilisant comme valeur par défaut (modèle de l'usager) : nationalité (locuteur, France))
S : Si vous êtes français, adressez-vous à la mairie ou au commissariat.	<-- ou (mairie, commissariat)
U : Et quelles sont les formalités ?	--> formalité (but_courant, ?) (. sélection, dans cette structure, de l'expression conditionnelle ; . pas de valeur pour la variable mineur/majeur (ni dans l'historique, ni dans le modèle de l'usager) ; donc on pose la question.)
S : Etes-vous mineur ?	<-- ? mineur (locuteur)
U : Non, non	--> non (mineur (locuteur)) (. <u>non</u> (mineur (locuteur)) est ajouté à l'historique et les formules correspondant à <u>non</u> (mineur (locuteur)) sont sélectionnées dans la structure)
S : Vous devez vous présenter avec votre livret de famille ou une fiche d'état civil et apporter un timbre fiscal.	<-- apporter (locuteur, <u>et</u> (timbre_fiscal, <u>ou</u> (livret_famille, fiche_état_civil)))
U : Bien, merci.	--> satisfaction (. suppression du but courant . demande d'un autre but (puisque pas de but actif))
S : Désirez-vous autre chose ?	<-- but (locuteur, ?)
U : Non merci, au revoir.	--> but (locuteur, 1) <u>et</u> prise_de_congé
S : Au revoir	<-- prise_de_congé

N.B. : Tout ce qui apparaît à l'écran est en caractères gras, les commentaires sont en italique ; pour faciliter la lecture, nous avons donné la forme de surface des énoncés plutôt que leur représentation sémantique.

BIBLIOGRAPHIE

- [1] P. MILLOT, Coopération homme-machine dans les tâches de supervision des procédés automatisés, *Thèse de Doctorat de l'Université de Valenciennes*, octobre 1987.
- [2] J. RASMUSSEN, Cognitive engineering, *INTERACT'87*, Second IFIP International Conference on Human-Computer Interaction, H. G. Bullinger, K. Kornwachs and B. Shackel Eds., North Holland, Amsterdam, Stuttgart, XXV-XXX, septembre 1987.
- [3] R. AMALBERTI, N. CARBONELL, P. FALZON, Communication orale homme-homme et communication orale homme-machine : un même modèle ?, *Colloque de l'ARC*, Toulouse, 1988.
- [4] P. FALZON, R. AMALBERTI, N. CARBONELL, Dialogue control strategies in oral communication, in *Foundation for human-computer communication*, K. T. Hopper, I. A. Newman Eds., North Holland, 1966.
- [5] M.-J. KELLY and A. CHAPANIS, Limited Vocabulary Natural Language Dialogue, *International Journal of Man-Machine Studies*, Vol. 9, pp. 479-501, 1983.
- [6] L. GUIBERT, *La spécificité du terme scientifique et technique*, *La langue Française* (17) 2, 1973.
- [7] B. SHNEIDERMAN, *Designing the user interface : strategies for effective human-computer interaction*, Addison-Wesley, 1987.
- [8] A. CHAPANIS, Words, words, words., *Human Factors* (7), 1-17, 1965.
- [9] A. CHAPANIS, R. B. OCHSMAN, R. N. PARRISH, G. D. WEEKS, The effects of four communication modes on the behaviour of teams during cooperative problem-solving, *Human Factors* (14), pp. 487-509, 1972.
- [10] J. T. MURRAY, J. VAN PRAAG, D. GILFOIL, Voice versus Keyboard control of cursor motion, *Proc. of Human Factors Society*, 27th Annual Meeting, 103, 1983.
- [11] D. L. MORISSON, T. R. GREEN, A. C. SHAW, S. J. PAYNE, Speech-controlled text-editing : effects of input modality and command structure, *International Journal of Man-Machine Studies*, 21, pp. 49-63, 1984.
- [12] J. LEGETT and G. WILLIAMS, An empirical Investigation of Voice as an Input Modality for Computer Programming, *International Journal of Man-Machine Studies* 21, pp. 493-520, 1984.
- [13] F. JELINEK, R. L. MERCER, L. R. BAHL, Continuous Speech Recognition : Statistical Methods, *CSR group, IBM TJ Watson research center*, Yorktown Heights NY 10598, 1982.
- [14] G. PERENNOU, A. FATHOLAHZADEH, M. DE CALMES, Le filtrage syntaxique dans l'entrée vocale de texte, in *Syntaxe et sémantique en compréhension de la parole*, J. P. Haton, J. M. Pierrel et P. Quinton Eds., GALF-AFCET, pp. 75-91, 1980.
- [15] A. M. DEROUAULT, B. Merialdo, Natural language modeling for phoneme to text transcription, *IEEE Transactions on pattern analysis and machine intelligence*, Vol. 8, n° 6, November 1986.
- [16] J. J. MARIANI, Hamlet : un prototype de machine à écrire vocale, *16^e journées d'étude sur la parole*, Société Française d'Acoustique, Hammamet 5-9 octobre 1987.
- [17] F. CHARPILLET, Un système de reconnaissance de parole continue pour la saisie de textes lus, *Thèse de Docteur d'université en informatique*, Nancy I, 1985.
- [18] CALLIOPE, *La parole et son traitement automatique*, Collection technique et scientifique des télécommunications, Masson, 718 p., 1989.
- [19] J.-M. PIERREL, *Dialogue oral homme-machine*, Hermès, 240 p., 1987.
- [20] G. ADDA, Reconnaissance de grands vocabulaires : une étude syntaxique et lexicale, *Thèse de Docteur-ingénieur en informatique*, Paris XI, 1987.
- [21] J. P. HATON, K. SMAILL, A dictation machine based upon 35 000 french words, *Rapport interne CRIN*, September 1989.
- [22] N. CARBONELL, D. FOHR, J. P. HATON, Aphodex, an acoustic-phonetic decoding expert system, *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 1, No. 2, 1987.
- [23] G. DEVILLE, Modelization of Task-Oriented Utterances in a Man-Machine Dialogue System, Phd Thesis, University of Antwerpen, 1989.
- [24] V. M. ZUE, A. M. AULL, Lexical stress determination and its application to large vocabulary speech recognition, *Proceedings of the IEEE International Conference on acoustics, speech and signal Processing*, April 1985.
- [25] F. DEBILI, Traitements syntaxiques utilisant des matrices de précedence fréquentielles construites automatiquement par apprentissage, *Thèse de docteur ingénieur, spécialité électronique, option traitement de l'information*, Université de Paris VI, Septembre 1977.
- [26] J. M. BAKER, Dragon Dictate : Natural Language Speech Recognition With 30 000 Words, *Eurospeech 89*, CEP Consultants Ltd, pp. 167-164, 1989.
- [27] E. RICH, *Artificial Intelligence*, McGraw-Hill, 1983.
- [28] M. A. MOREL *et al.*, Analyse linguistique d'un corpus d'oral finalisé, *Rapport final d'étude*, CNRS-GRECO « Communication parlée », Nancy, 1985.
- [29] P. MORIN, J.-M. PIERREL, PARTNER : un système de dialogue oral homme-machine, *COGNITIVA 87 : Machines, Réseaux intelligents*, Paris, 1987.
- [30] L. BRUNESSAUX, *SATIC : un système de dialogue d'aide au diagnostic et de génération de compte rendu en obstétrique*, COGNITECH, Paris, 80 p., septembre 1989.
- [31] P. ALINAT, E. GALLAIS, J.-M. PIERREL, G. SOUVAY, Intégration d'un système de reconnaissance analytique de la parole dans une console Sonar : vers un dialogue naturel, *Rapport CRIN*, Nancy, septembre 1989.
- [32] M. A. RICHARDS, K. UNDERWOOD, Talking to machines : how are people naturally inclined to speak ? in *Contemporary Ergonomics*, E. D. Megaw Ed., Taylor and Francis, Londres, 1984.
- [33] A. KENNEDY, A. WILKES, L. ELDER, W. MURRAY, Dialogue with machines, *Cognition*, 30 (1), October 1988.
- [34] A. HAUPTMANN, A. RUDNICKY, Talking to computers : an empirical investigation, *Technical Report*, CMU-CS-87-186, Dpt of Computer Science, Carnegie-Mellon University, Pittsburgh, 1987.
- [35] N. BORENSTEIN, The design and evaluation of online help systems, *Ph. D. Dissertation*, Carnegie-Mellon University, Pittsburgh, 1985.
- [36] C. FILLMORE, The case for case, in *Universals in Linguistic Theory*, E. Bach, R. T. Harn Eds., Rinehart and Winston, New York, 1968.
- [37] N. CARBONELL, J.-M. PIERREL, Task-oriented dialogue processing in human-computer voice communication, in *Recent Advances in Speech Understanding and Dialog Systems*, H. Niemann, M. Jang, G. Sagerer Eds., NATO ASI Series, Springer-Verlag, pp. 491-496, 1988.
- [38] N. CARBONELL, J.-M. PIERREL, Prise en compte de l'utilisateur dans la conception de systèmes de communication orale homme-machine, *ERGO-IA'88, Colloque Européen Ergonomie et Intelligence Artificielle*, Biarritz, pp. 258-274, 1988.
- [39] J. C. SPERANDIO, C. LETANG-FIGEAC, *Simulation expérimentale de dialogues oraux en communication homme-machine*, Rapport de fin d'études, GRECO-CNRS Communication parlée, Nancy, 50 p., 1986.
- [40] W. A. LEA, J. E. SHOUP, Contributions of the ARPA-SUR project, in *Trends in speech recognition*, Prentice-Hall, 1980.
- [41] H. CERF-DANON, A. M. DEROUAULT, M. ELBEZE, B. Merialdo, Speech Recognition in French With a Very Large Dictionary, *EUROSPEECH-89*, CEP Consultants Ltd, pp. 150-153, 1989.