

Rapport INRIA 1994 — Programme 3

Dialogue homme-machine à forte composante  
langagière

Projet DIALOGUE

3 mai 1995



Projet DIALOGUE

---

# Dialogue homme-machine à forte composante langagière

---

**Localisation :** *Nancy*

**Mots-clés :** compréhension de texte (1), dialogue homme-machine (1), intelligence artificielle (1), langage naturel (1), reconnaissance de la parole (1), reconnaissance des formes (1).

## 1 Composition de l'équipe

### **Responsable scientifique**

Jean-Marie Pierrel, Professeur, Univ. Henri Poincaré

### **Responsable permanent**

Noëlle Carbonell, Professeur, Univ. Henri Poincaré

### **Secrétariat**

Chantal Cridlig, Assistante-Ingénieur, Univ. Henri Poincaré

### **Personnel INRIA**

Yannick Toussaint, Chargé de Recherche

Eric Vannesson, Ingénieur expert

### **Personnel CNRS**

Bertrand Gaiffe, Chargé de Recherche CNRS

Yves Laprie, Chargé de Recherche CNRS

Anne Reboul, Chargée de Recherche CNRS 34ième section

Laurent Romary, Chargé de Recherche CNRS  
Gilles Souvay, Ingénieur sur contrat CNRS

### **Personnel Université**

Daniel Coulon, Professeur, Ecole des Mines de Nancy (INPL<sup>1</sup>)  
Christine Fay-Varnier, Maître de Conférences, E.N.S.G. (INPL)  
Denis Finck, PRAG, Univ. Henri Poincaré  
Samy Kouniali, ATER, INPL, jusqu'au 31/8/94  
Joseph di Martino, Maître de Conférences, Univ. H. Poincaré  
Azim Roussanaly, Maître de Conférences, Univ. Nancy II

### **Collaborateurs extérieurs**

Nabil Hathout, Chargé de Recherche CNRS, INaLF  
Jean Claude Junqua, chercheur à STL, Santa Barbara USA  
Philippe Morin, chercheur à STL, Santa Barbara USA  
Xavier Pouteau, Ingénieur Thomson-LCR, convention CIFRE

### **Chercheurs doctorants**

Jean Manuel Alvarenga, Boursier INRIA - Région Lorraine,  
jusqu'au 14/11/94  
Yolande Anglade, ATER Univ. Henri Poincaré, jusqu'au  
31/08/94  
Nadia Bellalem, Boursière MESR  
Emmanuel Bouyer, Boursier DGA/CNRS  
Laurent Buniet, Boursier MESR  
Laurent Chapelier, Boursier MESR  
Jean-Christophe Dubois, Boursier MESR  
Fabrice Duermael, Boursier DGA/CNRS  
Dominique François, ATER, Univ. Henri Poincaré, jusqu'au  
1/11/94  
Jean-Luc Husson, Boursier MESR  
Ramez Hajislam, Boursier Syrien, jusqu'au 1/04/94  
Bruno Mathieu, Boursier MESR, à compter du 1/10/1994  
François-Arnould Mathieu, Boursier MESR  
Christophe Mignot, Boursier MESR  
Vincent Pagel, Boursier MESR  
Daniel Schang, Boursier MESR

---

<sup>1</sup>INPL : Institut National Polytechnique de Nancy

Remarques : Projet commun à l'INRIA-Lorraine et au Centre de Recherche en Informatique de Nancy (CRIN), laboratoire des universités de Nancy I, Nancy II et de l'INPL, unité associée au CNRS (UA 262).

## 2 Présentation du projet

Concevoir des interfaces agréables et surtout adaptées à leurs conditions d'emploi pour des tâches particulières demeure une gageure dans la majorité des cas. La langue sous forme orale ou écrite a incontestablement un rôle majeur à jouer dans un tel cadre. Les recherches conduites au sein du projet DIALOGUE structurent un ensemble d'études visant à intégrer la langue dans des systèmes de dialogues finalisés par une tâche donnée. Dès lors, on y retrouvera des actions de recherche autour des niveaux classiques de l'étude de la langue : reconnaissance de la parole jusqu'à un premier niveau linguistique (phonème ou mot selon les cas), structuration de surface, sémantique depuis le lexique jusqu'à l'énoncé et enfin pragmatique avec en particulier des études sur la référence aux objets et aux actions.

Si toutes ces études sont nécessaires, elles ne sont en revanche pas suffisantes. Dans bien des cas, parce que la langue est peu adaptée à certaines commandes, on y adjoint la possibilité de désigner par le geste ; nous nous intéressons donc à ce qu'il est convenu d'appeler le dialogue multimodal. Enfin, et l'intitulé même du projet l'illustre, on ne peut envisager la langue dans une interface que dans la perspective du dialogue : c'est pourquoi nous abordons la gestion de sous-dialogues de négociation ou de correction et un certain nombre de phénomènes, tels l'anaphore, qui relèvent d'une structuration faisant appel à un niveau supérieur à celui de l'énoncé.

Lorsqu'on place ces études dans une perspective d'implantation, on se trouve confronté à une importante nécessité de robustesse aux différents niveaux. Cette nécessité a longtemps été considérée comme annexe et peu importante au niveau théorique ; les choses sont cependant en train de changer. L'étude fine des différences entre les conditions de laboratoire et les environnements réels d'utilisation fera (et commence à faire) la différence entre des systèmes utilisés et des gadgets.

La validation des objectifs que nous venons de décrire s'opère donc sur deux plans complémentaires :

- d'une part, via des interactions avec d'autres disciplines telles que la phonétique, la linguistique, la psychologie, l'ergonomie et la neurobiologie, nous concevons des modèles informatiques (en utilisant des méthodes de la reconnaissance des formes ou de l'intelligence artificielle) et nous les validons vis-à-vis de ces disciplines ;
- d'autre part nous implantons ces modèles dans des systèmes de dialogue vérifiant ainsi à la fois leur calculabilité et leur adéquation à des environnements réels.

### **3 Actions de recherche**

#### **3.1 Reconnaissance de la parole**

Les activités du projet Dialogue sont scindées en deux types d'approches :

- la prise en compte de manière plus réaliste des mécanismes de production de la parole dans le cadre de méthodes globales. Il s'agit de modéliser les aspects temporels dans un système neuromimétique, de déclencher un perceptron multicouches à l'endroit où son utilisation sera la plus discriminante possible, et enfin d'utiliser des automates sans topologie fixée a priori qui permettent d'exploiter au mieux le corpus d'apprentissage.
- l'utilisation directe de connaissances phonétiques ou articulatoires. Ces connaissances peuvent être utilisées dans le cadre d'un système de décodage acoustico-phonétique et nécessitent la mise au point de détecteurs automatiques d'indices acoustiques (notamment le suivi de formants, la fréquence fondamentale et la segmentation du signal de parole en unités phonétiques). Par ailleurs, il est clair que les connaissances articulatoires sont amenées à jouer un rôle clé dans le mécanisme de décodage acoustico-phonétique. Cela nous a donc incités à travailler sur l'extraction et le suivi des positions successives des articulateurs dans une suite d'images radiocinématographiques en vue, notamment, de tester différents modèles articulatoires et, à plus long terme, d'améliorer la fidélité des modèles articulatoires. La prosodie, enfin, puisqu'elle peut contribuer à la segmentation du signal de parole et qu'elle représente le lien entre le décodage acoustico-phonétique et les processus de plus

haut niveau, fait l'objet d'une approche triple : étude de la contribution des variations prosodiques à la structuration perceptive des énoncés, définition et repérage de marques prosodiques à l'aide de modèles de Markov.

### 3.1.1 Reconnaissance de mots isolés et enchaînés

*Participants* : Yolande Anglade, Laurent Buniet, Jean-Claude Junqua, Joseph di Martino, Bruno Mathieu, Jean-Marie Pierrel

#### RECONNAISSANCE DE VOCABULAIRES DIFFICILES

Dans le cadre d'un projet en lien avec la Sollac (cf 4.5), nous nous sommes intéressés à la reconnaissance, dans un environnement de travail réel, de mots que l'on confond facilement, tels que les lettres de l'alphabet (épellation). Dans un second temps, nos efforts se sont portés sur l'adaptation de cette méthode à la reconnaissance de mots enchaînés, en travaillant toujours sur les vocabulaires des chiffres et des lettres de l'alphabet. L'idée consiste à effectuer la reconnaissance en plusieurs étapes, en restreignant progressivement le nombre des candidats, lettres ou chiffres, à partir des voyelles identifiées dans la chaîne de mots à reconnaître.

#### L'AMÉLIORATION DE LA ROBUSTESSE

Dans le but d'améliorer la robustesse de la reconnaissance automatique de la parole, nous avons mené deux types d'expériences : le premier consiste à entraîner le système avec de la parole bruitée à un certain rapport signal/bruit (S/B), puis à effectuer des tests de reconnaissance de parole bruitée à d'autres niveaux du rapport S/B. Dans le deuxième type d'expérience, nous avons mélangé plusieurs niveaux de rapport S/B pour faire l'apprentissage. Les résultats obtenus montrent qu'une technique d'adaptation au bruit améliorerait davantage la robustesse d'un tel système qu'une approche de débruitage avant la reconnaissance.

En relation avec Jean-Claude Junqua, nous avons également poursuivi notre travail sur l'analyse acoustique de l'effet Lombard. Nous avons déjà pratiqué des tests sur la langue anglo-américaine. Ils avaient mis en évidence des changements acoustiques entre de la parole Lombard et de la parole propre, intervenant sur de nombreux paramètres acoustiques

utilisés en reconnaissance automatique de la parole. Dans le cadre du projet Esprit ROARS, nous avons étendu cette étude à la langue française. Après avoir constitué une base de données française de parole Lombard composée de six locuteurs (trois hommes, trois femmes), nous avons repris la même procédure de test. Les résultats obtenus concordent avec ceux obtenus sur l'anglais-américain[37, 14]. L'évolution des paramètres acoustiques étudiés confirme la nécessité d'en tenir compte pour des systèmes de reconnaissance fonctionnant dans un environnement difficile.

UTILISATION DE RÉSEAUX NEUROMIMÉTIQUES POUR LA RECONNAISSANCE DE PAROLE BRUITÉE <sup>2</sup>

Une autre voie étudiée pour aborder le problème de la reconnaissance de petits vocabulaires difficiles en milieu bruité est l'utilisation de réseaux connexionnistes. L'utilisation de perceptrons multicouches a en effet montré que la reconnaissance était encore possible avec un rapport signal sur bruit de 6 dB. Nous avons donc imaginé un système structuré en trois niveaux [24] destiné à la reconnaissance de mots enchaînés en milieu bruité : un premier réseau segmente le signal en extrayant des plages de type voisé (les voyelles et les autres sons pour lesquels les cordes vocales vibrent) ; un deuxième réseau reconnaît la voyelle prononcée (cette voyelle se trouve toujours dans un segment de type voisé si le premier réseau est suffisamment fiable) ; un troisième réseau permet de faire la distinction entre deux mots si la voyelle reconnue peut figurer dans deux mots différents du vocabulaire. Cette méthode a cependant des limitations : elle perd très vite sa validité. Des rapports signal sur bruit de 0 ou -6 dB dégradent très fortement la qualité finale des réponses. Après étude des sorties des réseaux de type perceptron multicouches, il s'est avéré que la dégradation était due à l'incapacité, pour de telles architectures, d'avoir des notions temporelles. De tels réseaux n'ont en effet aucune connaissance de leur état au pas de temps précédent. On peut alors avoir des durées de voisement (première étape) qui sont totalement irréalistes. Par conséquent, il a fallu introduire dans ces réseaux une notion de temporalité et donc s'orienter vers les réseaux connexionnistes récurrents. De tels réseaux, qui gardent une trace du passé grâce à la mémorisation de leur(s) état(s) précédents(s), devraient

<sup>2</sup>Action commune avec le Projet SYCO



permettre d'améliorer la première étape du décodage et, par conséquent, conduire à des systèmes exploitables en milieu industriel.

#### RECONNAISSANCE DE MOTS ISOLÉS À L'AIDE D'AUTOMATES NON PROBABILISTES

De nombreuses méthodes de reconnaissance de mots isolés ont vu le jour depuis une dizaine d'années. Parmi les plus populaires et les plus performantes à l'heure actuelle, on peut noter l'approche à base de modèles de Markov et l'approche connexionniste. Ces deux méthodes fournissent d'excellents résultats. Toutefois, elles ne permettent pas un apprentissage incrémental. C'est-à-dire que lorsqu'un mot est mal reconnu, ou lorsqu'un mot nouveau apparaît, il n'est pas possible de l'ajouter simplement dans l'espace d'apprentissage.

Nous avons donc développé une méthode à base d'automates [29] où chaque mot du vocabulaire est représenté par un automate, chaque chemin de l'automate décrivant une variation d'élocution de ce mot. Contrairement aux deux méthodes précédentes, notre approche permet facilement un apprentissage incrémental. En effet, lorsqu'un mot est mal reconnu, il suffit d'ajouter un chemin représentant ce mot dans l'automate. La reconnaissance d'un mot inconnu est réalisée par programmation dynamique.

Au cours de cette année, nous avons réalisé des expériences montrant qu'il est possible d'effectuer une reconnaissance en temps réel dans le cadre de la reconnaissance de mots isolés multi-locuteurs. Pour cela, un algorithme de quantification de formes, voisin de l'algorithme classique de quantification vectorielle, a été implanté. Un taux de 99.6% obtenu sur la base TiDigits de Texas Instruments et un temps de réponse de 1,1 s ont validé avec succès notre approche.

#### 3.1.2 Décodage acoustico-phonétique

*Participants* : Noëlle Carbonell, Dominique François, Ramez Hajislam, Jean-Luc Husson, Yves Laprie, Vincent Pagel, Jean-Marie Pierrel

#### ETUDE DES APPORTS MUTUELS DES SYSTÈMES À BASE DE CONNAISSANCES ET DES SYSTÈMES AUTO-ORGANISATEURS

Poursuivant nos études sur le décodage acoustico-phonétique, nous avons été amenés cette année à terminer des travaux tendant à unifier dans un même système une démarche à base de connaissances caractéristique du système APHODEX défini conjointement avec le Projet SYCO et une démarche reposant sur des classifieurs locaux à base de réseaux neuro-mimétiques [4]. Au cours de cette année, nos efforts ont porté sur l'évaluation du système ainsi obtenu, qui a montré une amélioration notable des résultats obtenus jusqu'alors par APHODEX. Ces études, complémentaires du système SYNDAP décrit dans le paragraphe suivant, nous ont permis d'analyser les points forts et les points faibles respectifs de l'approche neuronale et de l'approche à base de connaissances, et de proposer des solutions pour faire coopérer ces deux approches au sein d'une structure hybride intégrant les aspects positifs de chaque approche [4, 6].

Le système SYNDAP (SYstème Neuronal de Décodage Acoustico-Phonétique) fonctionne en deux étapes [6] :

- une étape de segmentation multi-niveaux qui consiste à construire un treillis de segments acoustiquement homogènes,
- une étape d'identification phonétique de ces segments réalisée à l'aide d'un ensemble de 17 réseaux de neurones organisés hiérarchiquement. Chaque réseau a une tâche précise qui consiste à mesurer le degré d'appartenance d'un segment à une classe phonétique donnée.

#### UTILISATION D'INDICES ACOUSTIQUES ET DU RAISONNEMENT HYPOTHÉTIQUE <sup>3</sup>

Ce projet, conduit en commun avec le projet SYCO [13], porte également sur le décodage acoustico-phonétique. Le but que nous poursuivons est de construire un système de DAP dont le principe est d'expliquer de manière cohérente les observations acoustiques en fonction de nos connaissances des phénomènes de production de la parole ; il s'agit donc d'une approche abductive du décodage acoustico-phonétique.

Pour tester l'intérêt de cette approche, nous avons abordé le problème de la reconnaissance automatique des occlusives sourdes [9, 30] qui couvre la plupart des difficultés rencontrées en DAP.

<sup>3</sup>Action menée en collaboration avec le Projet SYCO

Les observations acoustiques sont des indices (caractéristiques de la barre d'explosion et trajectoires formantiques) extraits automatiquement du signal. Ces indices peuvent avoir un pouvoir de discrimination plus ou moins fort et apparaître plus ou moins clairement dans le signal. Cela conduit à deux types d'indices définis en fonction de leur puissance de discrimination des réalisations acoustiques. Cette approche, testée avec succès dans le cadre du système DAPHNÉ, (cf. SYCO) sera étendue à d'autres sons, en particulier à la détection de la nasalité qui reste un problème très délicat pour un système de décodage acoustico-phonétique.

#### SUIVI AUTOMATIQUE DE FORMANTS

Notre algorithme de suivi automatique de formants est conçu pour être utilisé dans le cadre de l'approche abductive du décodage acoustico-phonétique (cf. le paragraphe précédent). Son principe général est d'interpréter au mieux et de manière cohérente les lignes de pics du spectrogramme en termes de formants de la cavité orale [38]. Les formants sont ensuite régularisés grâce à un algorithme de Snake développé dans le projet MOVI. Cet algorithme permet en outre d'évaluer la qualité finale du suivi.

Cette année, l'algorithme a été complété de manière à pouvoir utiliser le lissage spectral avec les Snakes. Ceci nécessite en effet que le spectrogramme puisse être recalculé localement pour séparer deux formants très proches. Nous poursuivons par ailleurs l'amélioration de la stratégie d'interprétation.

#### SEGMENTATION DU SIGNAL DE PAROLE ET EXTRACTION DE LA FRÉQUENCE FONDAMENTALE

De nombreuses méthodes de segmentation reposent sur une mesure unique de discontinuité dans le signal mais se heurtent à la difficulté de calculer dynamiquement (localement) le seuil qui minimise globalement les effets de sur et sous-segmentation. Nous avons choisi d'utiliser la technique du dendrogramme qui effectue une segmentation multi-niveaux, en regroupant dans une structure unique les différentes décompositions d'un même segment de parole. Notre idée est que l'exploitation combinée de connaissances de haut niveau (indices acoustiques, connaissances phonétiques, modèles articulatoires) peut permettre simultanément de

sélectionner le meilleur chemin -segmentation- dans le dendrogramme et d'identifier les sons produits.

La première source d'informations pour la segmentation en unités phonétiques est la fréquence fondamentale puisqu'elle permet de séparer les zones de parole voisées des zones non voisées. Or il n'existe aucun algorithme d'extraction de F0 qui fonctionne de manière optimale dans toutes les conditions de production. Ainsi, la méthode du peigne fréquentiel de Ph. Martin intégrée dans Snorri n'est pas toujours fiable, car elle risque de ne pas détecter F0 dans les régions peu énergétiques et où une seule harmonique est présente (cas des fricatives et occlusives voisées).

Aussi, pour pallier les problèmes induits par l'utilisation d'une technique unique, nous avons proposé et mis en œuvre dans Snorri une méthode générale permettant la coopération de plusieurs techniques concurrentes pour l'extraction de F0. Pour tester cette coopération, nous avons implanté l'algorithme d'extraction de F0 temporel AMDF dont le comportement est meilleur dans les conditions qui défavorisent l'algorithme de Ph. Martin.

La coopération entre les deux techniques est effectuée en post-traitement grâce à un algorithme de programmation dynamique, et à une homogénéisation des résultats fournis par les deux algorithmes. L'évaluation de la méthode coopérative sur différents corpus a mis en évidence une réduction significative du nombre de dévoisements abusifs des occlusives et fricatives voisées, et un accroissement de la précision temporelle et fréquentielle en milieu bruité.

#### SUIVI DE LA LANGUE DANS UNE SÉQUENCE D'IMAGES RADIOCINÉMATOGRAPHIQUES

Les informations articulatoires constituent une autre source de connaissances importante en décodage acoustico-phonétique, dans la mesure où elles permettent d'améliorer le traitement des phénomènes de coarticulation. Pour étudier la manière dont est produite la parole, un grand nombre de locuteurs ont été filmés par radiographie dans le but de suivre les mouvements des différents articulateurs (lèvres, langue, palais, voile du palais, pharynx, larynx) et d'étudier des modèles de production de la parole. Malheureusement ces films (datant des années 70-80) n'ont été que très partiellement utilisés car il faut "dessiner à la main" les arti-

culateurs puis ensuite digitaliser ces croquis, ce qui est très fastidieux d'une part et peu fiable d'autre part.

Nous développons donc actuellement, en collaboration avec le Projet MOVI, une méthode permettant de suivre le plus automatiquement possible les principaux articulateurs (lèvres, langue, voile du palais et larynx), en vue d'obtenir des connaissances susceptibles d'améliorer les modèles articulatoires existants et donc une meilleure prise en compte des effets acoustiques de la coarticulation en décodage acoustico-phonétique (parole continue et grands vocabulaires). Nous avons commencé par la langue car il s'agit de l'articulateur le plus important, le plus mobile et par conséquent le plus difficile à suivre. Le lecteur trouvera de plus amples détails dans le chapitre concernant le projet MOVI. Pour l'instant, l'algorithme n'effectue aucun contrôle pour s'assurer que le contour de la langue est cohérent d'un point de vue articulatoire. Notre objectif est donc d'utiliser un modèle articulatoire approché pour interpréter le contour de la langue obtenu et vérifier qu'il est bien correct.

#### MACROPROSODIE ET RECONNAISSANCE DE LA PAROLE CONTINUE

Les variations temporelles des paramètres prosodiques (rythme, intonation, intensité) au cours de la production d'un énoncé contribuent, dans une large mesure, à l'identification de la structure syntaxique et discursive de cet énoncé par l'auditeur. Notre objectif est :

- dans un premier temps, de construire des algorithmes robustes de détection et d'interprétation des informations linguistiques et énonciatives véhiculées par la prosodie en parole continue ;
- d'intégrer ensuite ces algorithmes dans un système (à base de connaissances) de reconnaissance de la parole continue, en vue d'améliorer la segmentation du continuum sonore en unités lexicales, syntaxiques et discursives.

Malgré des travaux comme ceux de l'I.P.O. (Eindhoven), on manque, aujourd'hui encore, d'indications précises sur la perception et l'interprétation humaines de la mélodie d'un énoncé. Notre effort a porté, dans un premier temps, sur la mise en évidence expérimentale des contributions respectives des différents indices prosodiques à la structuration perceptive d'un signal de parole (en unités lexicales et supra-lexicales).

Une première expérience, dont le protocole et les résultats sont décrits dans [41], a porté sur la mélodie et l'intensité. Les conclusions de cette expérience ont permis notamment la conception et l'implantation d'une méthode de lissage "perceptif" des courbes de F0. Cette méthode, testée sur l'ensemble des corpus de parole continue à notre disposition, atténue significativement les variations de F0 liées à la nature des sons prononcés ; or ces variations, qui n'apportent aucune information sur la structure d'un énoncé, se confondent facilement avec les marques intonatives de structuration linguistique.

Pour poursuivre, dans le cadre d'une collaboration avec J. Vaissière, (Professeur à Paris III), l'étude de la perception et de l'interprétation des phénomènes prosodiques, nous avons testé différentes méthodes de synthèse de la parole. Nous avons retenu, pour constituer les stimuli des expérimentations futures, la méthode mise au point par T. Dutoit, et implanté cette méthode dans Snorri [39]. Actuellement, nous travaillons à la définition du protocole d'une expérience de perception qui se déroulera prochainement à l'Institut de Phonétique de Paris III. Parallèlement, nous étudions la modélisation (par des HMM) des variations intonatives et rythmiques porteuses d'informations structurelles, en vue d'obtenir une segmentation automatique robuste des énoncés en intonèmes (unités intonatives supra-segmentales).

## 3.2 Traitement de la langue

### 3.2.1 Référence

*Participants* : Nadia Bellalem, Emmanuel Bouyer, Jean-Christophe Dubois, Fabrice Duermael, Bertrand Gaiffe, François-Arnould Mathieu, Jean-Marie Pierrel, Xavier Pouteau, Anne Reboul, Laurent Romary, Daniel Schang

L'objectif de ces travaux est de mettre en évidence les mécanismes intervenant dans l'interprétation, en situation de dialogue, d'énoncés en langage naturel. Notre approche repose d'une part sur une analyse fine des indices linguistiques disponibles et d'autre part sur la prise en compte du caractère fortement localisé (dans le temps et l'espace en particulier) des interprétations auxquelles ces indices conduisent. Il n'est donc pas question d'aborder des points de théorie linguistique qui ne conduisent pas à des phénomènes de sens pertinents en situation de dialogue.

**Un modèle du calcul de la référence** Les travaux menés jusqu'ici au sein du projet ont conduit à une proposition de modèle pour l'interprétation des expressions référentielles dans un dialogue homme-machine qui repose sur l'hypothèse que ces expressions ne visent pas directement des référents dans un historique de dialogue ou un environnement perceptuel partagé. Au contraire, elles doivent s'interpréter comme des instructions de sélection à l'intérieur de contextes locaux. Une telle approche semble être en mesure d'expliquer certaines différences d'acceptabilité telles que celles rencontrées dans les exemples 1 et 2, suivant que le discours a introduit ou non de tels contextes.

1. *Dessine une voiture et un camion. Mets la voiture/\*la en rouge.*<sup>4</sup>
2. *Dessine une voiture. Mets la/\*la voiture en rouge.*

Dans ce cadre, on observe par exemple qu'une expression en "le N" réalise une discrimination inter-catégorielle à l'intérieur d'un ensemble, alors que pour une forme telle que "ce N" il s'agit d'une discrimination intra-catégorielle. Nos travaux actuels portent sur différents aspects susceptibles de préciser la nature de ces contextes locaux :

- description plus fine de certaines expressions référentielles (descriptions définies en *le N*, expressions déictiques telles que *ici* pouvant intervenir dans des dialogues multimodaux [47], etc.) ;
- cas des référents en évolution dans une tâche de commande - articulation avec une modélisation du temps ;
- cas des contextes spatiaux (cf. ci-dessous : Espace et référence) ;
- apport d'une approche pragmatique ; la présence au sein du projet d'une linguiste spécialiste de pragmatique nous a permis de confronter constamment nos hypothèses à des cadres théoriques plus généraux (cf. ci-dessous : Pragmatique de la référence) ;
- conséquences d'une telle approche sur le lexique (cf. ci-dessous : Lexique et référence).

**Espace et référence** Notre réflexion actuelle vise à intégrer, sur la base des travaux précédents concernant la référence, l'ensemble des mécanismes qui permettent d'accéder à un espace visuel (graphique) à l'aide d'expressions multimodales à forte composante langagière. Dans

---

<sup>4</sup>l'étoile indique l'inacceptabilité linguistique de certains énoncés ou au moins un écart certain par rapport à une normativité reconnue

ce cadre, le travail de recherche que nous menons porte sur trois aspects complémentaires :

- modélisation du geste de désignation dans le cadre d'énoncés où celui-ci intervient comme complément d'expression langagières à valeur déictique (démonstratifs, déictiques purs). Après avoir mis en évidence, aux cours des années précédentes, un certain nombre d'indices et de procédures permettant de segmenter et de modéliser les trajectoires gestuelles, nos travaux ont porté cette année sur la définition des modes de représentation des objets de la tâche de dialogue qui permettraient d'interpréter au mieux les gestes correspondants. En particulier, et en lien avec l'analyse des références spatiales, nous avons abordé le rôle de l'appropriation perceptive de l'espace de présentation de la tâche [23, 22, 46]. Enfin, l'approche que nous développons est particulièrement originale en ce quelle s'ancre délibérément sur les résultats acquis en matière de calcul de la référence langagière. Ainsi, nous avons mis en évidence la nécessité d'une analyse contrastive des trajectoires gestuelles, de manière à ne conserver que l'information permettant de discriminer les référents entre eux, sur la base des contraintes fournies, par exemple, par le groupe nominal démonstratif associé (pour un syntagme nominal en *ce N*, il s'agira de discriminer entre les référents pouvant être catégorisés comme "N") ;
- interprétation des références spatiales (par exemple : déplace la fenêtre de gauche) susceptibles d'être rencontrées dans des dialogues de commande tels que le positionnement d'objets. Dans ce cadre, nous avons introduit la notion de contexte [49, 50, 51] de référence qui permet de filtrer les informations contextuelles utiles à l'interprétation d'une expression donnée (limitation de l'extension spatiale ainsi que des dimensions pertinentes) ;
- référence à des bases de données images où il s'agit de mettre en évidence le rôle des expressions référentielles dans la discrimination des données retenues par un utilisateur. Au cours de cette année, notre travail a porté sur l'étude des modèles formels de représentation des relations de type partie/tout (méronymie) et classe/instance (hypo/hyperonymie).

**Pragmatique de la référence** L'hypothèse principale sur laquelle porte cet aspect de notre travail est la suivante : on ne peut pas analyser



les expressions référentielles en faisant l'impasse sur la propriété qui leur est commune, à savoir leur capacité à désigner un objet dans le (un?) monde, et donc hors du discours où elles apparaissent. Cette hypothèse découle d'un certain nombre de thèmes communs :

- importance d'un traitement des expressions référentielles qui ne soit pas limité à la linguistique strictement entendue (syntaxe et sémantique)
- nécessité de l'intégration de données non linguistiques, qu'elles relèvent de la connaissance sur le monde ou de la situation de communication, etc.

Dès lors, une approche plus ou moins pragmatique du problème de la référence s'impose.

Pour autant, n'importe quelle pragmatique ne peut faire l'affaire et un premier effort a porté sur la description du cadre pragmatique général dans lequel une pragmatique de la référence pourrait s'insérer [2]. D'autre part, certaines approches actuelles de la référence, qui se veulent "pragmatiques" ou "discursives", sont criticables : il s'agit notamment des approches centrées autour de la notion de cohérence, dont nous avons montré qu'elle souffre d'imprécision et qu'elle échoue de toute façon à rendre compte de l'usage des expressions référentielles [44]. Enfin, les théories qui font de l'accès lexical à l'antécédent d'une anaphore une donnée lexicalement "marquée" dans tel ou tel type d'expression référentielle ne sont pas plus satisfaisantes [45].

Dans la perspective ainsi dégagée, il est possible d'énumérer un certain nombre de projets individuels ou collectifs:

- analyse des procédures (linguistiques) de traitement du pronom [1];
- spécificité du pronom de première personne [43];
- sémantique et pragmatique comparée des constructions copulatives avec le pronom personnel et le pronom démonstratif en français et en anglais;
- des travaux sur les référents évolutifs, concrétisés par des publications de trois membres de l'équipe dans un livre chez Benjamins, qui y est consacré: *Evolving reference and anaphora: time and objects*.

**Référence aux actions** Classiquement, dans le domaine du dialogue de commande homme-machine, les énoncés sont analysés en termes de structures prédicatives, à charge pour la description lexicale d'établir le lien entre les prédicats et les fonctions de l'application à exécuter. Ce type de démarche rend donc compte du sens de l'énoncé par l'instanciation d'un profil de fonction. Des difficultés apparaissent donc dès lors qu'un énoncé réfère à une séquence de fonctions et non plus à une fonction isolée. Ce type de situation apparaît de façon naturelle en particulier lorsqu'il s'agit de créer des objets. Ainsi, la plupart des systèmes actuels seraient conduits à accepter des énoncés tels que "déplace la fenêtre verte" et "crée une fenêtre" mais à rejeter "crée une fenêtre verte" pour autant que l'application sous-jacente ne fournisse pas de fonction de création de fenêtre d'une couleur donnée. Le concepteur de ce genre de systèmes est alors confronté à deux difficultés : soit il souhaite effectivement rejeter l'énoncé "crée une fenêtre verte", et la description du langage devient difficile (on ne peut plus définir la structure du groupe nominal indépendamment de l'action à laquelle il est fait référence), soit il accepte ce type d'énoncé et il est obligé d'ajouter via le modèle de l'application les fonctions élémentaires qui font défaut (*créer\_une\_fenêtre\_colorée*). Nous avons montré dans [32] et [5] que le nombre d'opérateurs à ajouter ainsi est une fonction exponentielle du nombre de propriétés utilisables pour la référence ; cette démarche sans méthodologie pour la guider est donc une activité à haut risque. Nous avons donc proposé de modéliser les énoncés comme référant à un état final, à charge pour un planificateur de calculer la séquence d'actions à mettre en œuvre pour l'atteindre. Notre modèle de référence aux actions s'appuie sur les notions de tension et détension proposées par G. Guillaume et permet de décomposer temporellement la prédication linguistique pour en déduire les propriétés des objets dans cet état final. Il rend compte également, dans le cadre spécifique des actes de langage directifs, de la compositionnalité de l'aspect établie par H. Verkuyl, et s'intègre tout à la fois dans le formalisme des zones temporelles proposées par Laurent Romary et aux schémas référentiels de Bertrand Gaiffe.

**Lexique et référence** L'objectif principal est d'arriver à une définition de structures lexicales adaptées au phénomènes de référence dans le dialogue (un N, le N, ce N). Notre cadre de travail n'est pas sans rappeler celui des travaux axés plus directement sur la référence. Ainsi, nous avons besoin d'une représentation du sens lexical qui permette en amont

de l'interprétation d'une expression référentielle de définir un ensemble de contraintes susceptibles de limiter les résultats possibles (contexte et/ou référents). Par exemple, pour interpréter une description définie en *le N*, nous avons mis en évidence qu'il est important de ne sélectionner que les indices pertinents pour l'opposition inter-catégorielle, notamment en ce qui concerne les propriétés communes partagées par les différentes catégories concernées. Cet aspect de nos recherches a déjà donné lieu à un certain nombre de travaux au sein du projet, qui se sont poursuivis cette année dans le cadre d'une thèse.

### 3.3 Terminologie et textes techniques

*Participants* : Nabil Hathout, Jean-Marie Pierrel, Yannick Toussaint

Cette activité autour de la terminologie pour le traitement de textes scientifiques et techniques a été menée au sein du projet DIALOGUE depuis un an, suite à l'arrivée au sein de l'équipe de Yannick Toussaint. Elle répond à un besoin croissant, tant au niveau industriel que scientifique, de maîtrise de volumes très importants de documentation scientifique et technique. Notre objectif est de développer des méthodologies et des outils informatiques et linguistiques pour le traitement de ces textes. La notion de traitement est prise ici, dans un sens relativement vaste puisqu'il peut s'agir d'apporter de nouvelles données pour des problématiques comme :

- la recherche d'information,
- la synthèse de textes scientifiques,
- l'aide à la rédaction,
- l'analyse automatique de textes techniques en vue de contrôler leur qualité.

Notre démarche analytique repose sur le constat suivant : dans les domaines techniques, la langue utilisée possède toutes les caractéristiques des langues de spécialité et elle est de ce fait beaucoup plus contrainte que la langue commune. Les travaux qui ont été menés sur la documentation technique ou sur les textes scientifiques ont montré que l'information pertinente se trouve localisée préférentiellement dans les groupes nominaux. Notre approche vise à exploiter les analyses linguistiques réalisées sur ces syntagmes dans les différents niveaux de traitement de la langue

(morphologique, syntaxique, sémantique) afin d'en proposer une modélisation informatique. C'est dans cette optique que nous travaillons en collaboration étroite avec des linguistes et informaticiens-linguistes.

Nos travaux s'articulent autour de deux grands axes complémentaires. Le premier porte sur la terminologie et le second sur l'étude d'un type particulier de texte technique que constituent les spécifications de logiciel rédigées en langage naturel. Notre démarche se base sur le traitement de corpus de textes volumineux. Nous avons donc collecté un certain nombre de textes sur lesquels nous pourrions valider nos différentes hypothèses.

**La terminologie :** L'étude des termes d'un domaine technique donné –entité linguistique qui peut être associée à une notion du domaine considéré– est essentielle pour dépasser le niveau purement sémantique du langage et accéder à une véritable interprétation du langage dans le domaine considéré. Notre action se situe au sein de deux projets :

- le projet GOLEM, développé en collaboration avec l'Institut National de l'Information Scientifique et Technique (INIST) et l'Institut National de la Langue Française (INaLF) pour synthétiser de façon automatique l'information contenue dans la littérature grise (thèses, rapports, articles scientifiques...).
- Le projet ICE dont l'objectif est de développer des outils pour l'extraction et la représentation d'une terminologie, simultanément à partir de textes techniques et à partir d'un dictionnaire de langue générale. Il s'agit de montrer que l'analyse des définitions du dictionnaire par rapport à un domaine scientifique ou technique peut fournir des données complémentaires à celles des travaux sur les textes techniques. Le *Trésor de la Langue Française* est le dictionnaire autour duquel s'articule ce projet. Ce projet associe le Projet DIALOGUE à des équipes de l'INaLF et à l'Equipe de Recherche en Sémantique et syntaxe(ERSS, URA-CNRS 1033, Toulouse).

Dans le cadre de ces deux projets, les thèmes scientifiques sur lesquels nous nous focalisons sont les suivants :

- **la structuration des termes :** il s'agit de proposer un modèle informatique de structuration des termes d'un domaine donné pour :

- représenter dans un réseau sémantique les termes-candidats proposés par des logiciels d'identification automatique de termes, proposer des outils de raisonnement sur ce réseau qui permettent d'identifier les termes-candidats non valides et regrouper les termes désignant des concepts du domaine identiques.
- mettre en œuvre des raisonnements logiques sur le réseau de termes et définir ainsi les notions de *cohérence* et de *complétude* de ce réseau par rapport à des critères linguistiques et logiques.
- **la sémantique lexicale pour l'identification des relations entre termes**: il s'agit d'étudier quel est le type d'information sémantique lexicale permettant de repérer les liens entre termes qui participent à leur définition. Ces liens seront donc introduits dans le modèle de structuration des termes.

**L'étude des spécifications de logiciel** : Cette étude fait suite aux travaux de thèse<sup>5</sup> de Yannick Toussaint, en association avec l'Université Polytechnique de Catalogne [12, 27]. Nous avons montré [52] que la terminologie constitue un pré-requis essentiel à la phase de spécification et qu'il est nécessaire de la représenter dans un modèle formel pour pouvoir vérifier certains critères de qualité exigés par les normes de génie logiciel. Les résultats de ce travail ont été également présentés dans le cadre d'un cours, organisé par la SEE<sup>6</sup>, sur l'analyse linguistique et la reformulation d'une spécification de logiciel rédigée en langage naturel (cf. 6.2).

Dans ces deux axes de recherches, nous sommes amenés à traiter des volumes de textes importants et donc, à mettre en œuvre des techniques d'informatique linguistique robustes. La chaîne de traitement qui permet le repérage puis la structuration en base de connaissance des termes d'un domaine technique nous amènera, parallèlement à nos travaux de recherche, à tester des outils existants, notamment pour l'extraction automatique de candidats-termes, afin de déterminer les manques importants dans le processus global.

---

<sup>5</sup>*Méthodes Informatiques et linguistiques pour l'aide à la spécification de logiciel*, Thèse de l'Université Paul Sabatier de Toulouse, octobre 1992

<sup>6</sup>SEE : Société des Electriciens et Electroniciens

### 3.3.1 Lexique et syntaxe

*Participants* : Jean-Manuel Alvarenga, Denis Finck, Jean-Marie Pierrel

Nos travaux sur l'analyse syntaxique et sur la désambiguïisation du lexique sont complémentaires des recherches menées dans le projet sur la référence, sur le traitement de textes techniques et la terminologie. En effet, il s'agit de traiter le problème complexe de l'analyse syntaxico-sémantique de textes en s'intéressant plus particulièrement à deux aspects : d'une part, la mise en œuvre des Tree-Adjoining-Grammars (TAG) en utilisant la notion d'agent et, d'autre part, la désambiguïisation des analyses syntaxiques.

#### Analyse conduite par des agents

Pour pouvoir faire coopérer tous les niveaux linguistiques (syntaxique, sémantique, pragmatique) au cours de l'analyse d'un texte ou d'un dialogue en langue naturelle, nous avons envisagé d'utiliser les notions d'objet et d'agent qui permettent d'introduire très tôt des éléments de compréhension pour l'interprétation de textes ou l'interprétation des ordres dans les dialogues homme-machine. Nous avons implanté un système qui élabore une représentation de la structure et du sens d'une phrase grâce à un analyseur construit à partir d'une grammaire TAG décrivant un sous-ensemble significatif du français. Ce processus est activé par les agents qui représentent les items lexicaux présents dans la phrase [33].

#### Désambiguïisation du lexique

Cette activité nous a permis de faire un lien entre les travaux sur l'analyse syntaxique et les modèles probabilistes utilisés en reconnaissance de la parole. Nous avons étudié les possibilités de prédiction de construction des structures prédicatives en utilisant un modèle grammatical probabiliste (Baker 1970, Lori & Young 1990...). Nous nous sommes intéressés à l'algorithme E-M (Dempster). Ce modèle probabiliste facilite la résolution des ambiguïtés syntaxiques et sémantiques.

### 3.4 Dialogue et applications

Offrir à l'utilisateur la possibilité d'interagir oralement avec un logiciel d'application soulève des problèmes spécifiques non résolus actuellement. Nos travaux portent sur les thèmes suivants :

- l'analyse ergonomique de l'usage de la parole et du geste en situation d'interaction personne-machine, et la modélisation du comportement et des stratégies d'expression multimodale des futurs usagers,
- la conception d'architectures logicielles et d'outils génériques destinés à faciliter le développement d'interfaces utilisateur robustes et conviviales, qui intègrent la parole parmi les modalités d'expression offertes à l'utilisateur.

#### 3.4.1 Modélisation de l'utilisateur

*Participants* : Noëlle Carbonell, Christophe Mignot

Nous étudions les stratégies et formes d'utilisation de la parole et du geste non contraints en situation d'interaction personne-machine, en vue de définir des langages de commande multimodaux artificiels qui soient :

- acceptables par l'utilisateur, c'est-à-dire efficaces (en termes de pouvoir d'expression et de concision) et faciles à apprendre, i.e. compatibles avec les formes naturelles d'expression humaine, en particulier avec les structures de la langue naturelle,
- réalistes, c'est-à-dire traitables de manière robuste par les systèmes actuels de reconnaissance/interprétation de la parole continue et les interprètes d'expression gestuelle.

Pour acquérir les connaissances nécessaires à la définition d'un modèle pertinent de l'utilisation de la parole et du geste par les usagers des futures interfaces de dialogue multimodal, nous avons réalisé, en collaboration avec des ergonomes du CERMA, une étude expérimentale portant sur l'utilisation libre de la parole (langue naturelle) et du geste 2D (sur un écran tactile), dans des situations d'interaction personne-machine classiques. Les sujets réalisaient des tâches de conception ou de contrôle de processus en interagissant avec des logiciels courants dont l'interface utilisateur existante, langage de commande ou manipulation

directe<sup>7</sup> avait été remplacée par une interface multimodale que simulaient des agents humains (cf. le paradigme expérimental du magicien d'Oz). Les résultats de cette expérience (8 sujets, 3 sessions d'une heure en moyenne) sont décrits dans [25] et [26]. Ce travail a été réalisé en partie dans le cadre d'un contrat de recherche soutenu par la DRET (étude G1 91/419 sur la définition d'interfaces multimodales), en collaboration avec THOMSON-CSF/SDC et le CERMA [57].

A partir de l'analyse des stratégies et des styles d'expression multimodale des sujets, nous avons défini un langage artificiel de commande combinant la parole et le geste 2D. Ce langage est traitable automatiquement, puisque la syntaxe des expressions orales est décrite par une grammaire *Context Free* et le vocabulaire compte une centaine de mots (formes lexicales) ; quant aux gestes, ils se réduisent à ceux que prennent en compte les écrans tactiles actuels (pointage, tracer). Des éléments de définition d'un tel langage sont donnés dans [55].

Pour évaluer la tolérance de l'utilisateur aux contraintes syntaxiques et sémantiques imposées par le langage artificiel que nous avons défini, et vérifier s'il peut s'acquérir facilement au cours de l'interaction, i.e. sans apprentissage préalable, nous définissons actuellement, avec P. Dauchy chercheur au CERMA, le protocole d'une nouvelle expérience. Les tâches à effectuer seront identiques à celles proposées dans le cadre de la première expérience afin de permettre les comparaisons, mais les sujets devront utiliser le langage de commande que nous avons conçu pour les réaliser, et la reconnaissance de la parole sera assurée par un logiciel commercialisé.

### 3.4.2 Architectures logicielles pour le dialogue

*Participants* : Laurent Chapelier, Christine Fay, Bertrand Gaiffe, Vincent Morlot, Jean-Marie Pierrel, Azim Roussanaly, Gilles Souvay

#### ENVIRONNEMENT DE DÉVELOPPEMENT DE DIALOGUE HOMME-MACHINE : DIAPASON

Au cours des années précédentes nous avons développé le système de dialogue DIAPASON. Le but de ce système est de mettre en œuvre un dialogue finalisé pour des applications de type commande de machine. A partir d'une phrase prononcée par un opérateur coopérant, le système

<sup>7</sup>i.e. utilisation d'un dispositif gestuel pour manipuler des icônes graphiques ou sélectionner des items dans un système de menus



doit interpréter les hypothèses fournies par le système de reconnaissance en fonction du contexte général du dialogue et de l'état de la tâche à diriger, pour déterminer la commande à exécuter.

Trois axes de travail se sont dessinés dans cette étude : la réalisation d'un système de dialogue indépendant de l'application, le développement d'outils génériques décrits plus loin permettant la mise en place de l'application, et la réalisation d'applications. Cet ensemble a donné naissance à l'environnement de développement DIAPASON.

Le cœur du **système de dialogue** est un automate d'états finis qui fait appel à des connaissances statiques (définition de la tâche, valeurs par défaut...) et à des connaissances dynamiques (phrase prononcée par l'opérateur, historique du dialogue, état de la tâche...) pour réunir les différents éléments nécessaires à l'exécution d'un ordre. DIAPASON tient compte des aspects typiques du dialogue que sont la résolution des ellipses, la correction, l'annulation, la confirmation ou la répétition d'un ordre. Il est capable de choisir une hypothèse de phrase parmi  $n$  propositions fournies par le système de reconnaissance qui lui est connecté, en fonction de l'état du dialogue (attente d'une confirmation, réponse à une question...) ou de l'état de la tâche (détection des incohérences...). Le système est capable de gérer les entrées clavier, orale et souris.

Les **outils génériques** permettent la mise en place des connaissances spécifiques à une application dans une structure d'accueil appelée plateforme DIAPASON. Il existe quatre outils. Le gestionnaire d'applications permet d'initialiser la plate-forme avec une nouvelle application. L'éditeur de structures syntaxico-sémantiques permet de saisir les structures syntaxico-sémantiques des commandes de l'application. Le compilateur prend en entrée ces structures et fournit en sortie des fonctions C pour paramétrer le niveau dialogue et une description générale du vocabulaire et de la syntaxe pour le niveau reconnaissance. Les formateurs permettent d'adapter cette description générale à un analyseur syntaxique donné.

Le type d'application visé par DIAPASON est la commande de console. L'interface de ces applications peut être modélisée à l'aide d'objets et de menus hiérarchiques agissant sur ces objets.

L'environnement DIAPASON a servi de base à l'élaboration du système de dialogue défini dans le cadre du projet Esprit ROARS (cf. 4.1) et sert de support à la nouvelle convention Thomson-DRET (cf. 4.2).

#### DÉFINITION D'UN DIALOGUE MULTIMODAL D'AIDE

Notre objectif est la conception d'une interface multimodale intelligente répartie alliant le vocal, la désignation gestuelle et le graphisme, et apportant une assistance à l'utilisateur lors de l'utilisation effective de systèmes multimédias. Nous envisageons la réalisation d'un prototype dans le cadre du projet de recherche régional IRMA dans lequel nous sommes fortement impliqués (cf. 5.5).

Au cours de cette année, nous avons réalisé une expérience qui consiste à cerner les besoins des utilisateurs face à un logiciel possédant, réellement ou de manière simulée, des possibilités d'interaction multimodale.

Notre choix s'est porté sur le logiciel GOCAD développé au sein de l'équipe Infographie du CRIN. GOCAD est un logiciel de modelage de surfaces principalement utilisé par des géologues dans le but de modéliser des surfaces géologiques à partir de données prélevées sur le terrain.

Ce choix est motivé par les raisons suivantes :

- la complexité et la diversité des fonctionnalités offertes par GOCAD,
- l'utilisation réelle du logiciel dans les milieux professionnels et pédagogiques, et dans des domaines aussi variés que l'exploration pétrolière, l'embryologie ou la parodontologie,
- les améliorations souhaitées par les promoteurs du logiciel dans le domaine de l'assistance et de la documentation,
- la disponibilité du logiciel au laboratoire, ainsi que celle des experts qui travaillent à son développement et qui l'utilisent.

L'expérience s'est déroulée en Novembre 1993. Des psychologues du GRC (Groupe de Recherche sur les Communications de l'Université de Nancy II) ont participé à sa réalisation et à l'analyse du corpus recueilli. Cette expérience est décrite dans [28]. Les transcriptions, le protocole et les analyses détaillées du corpus recueilli ont fait l'objet de trois rapports internes [54, 53, 59]. Actuellement, nous définissons les fonctionnalités et la structure des différents modules ou agents d'un système d'aide à l'utilisation de GOCAD qui tienne compte des conclusions expérimentales. Plus généralement, nous avons entrepris une réflexion sur la définition d'une architecture logicielle adaptée à la conception d'interfaces multimodales intelligentes. Cette réflexion, menée en collaboration avec V. Chevrier et R. Foisel du Projet SYCO, nous conduit à proposer une archi-

tecture multi-agents à plusieurs "couches" qui permette un découpage conceptuel et une mise en œuvre incrémentale du système, ainsi que l'intégration d'un modèle de l'utilisateur. L'architecture ainsi définie servira de plate-forme d'intégration des différentes solutions que nous proposons aux problèmes d'assistance, en particulier de celle que développe Laurent Chapelier.

#### OUTILS POUR MODÉLISER LA TÂCHE DANS PARTNER

Le système Partner, dès le début de son développement, avait mis l'accent sur la nécessaire séparation entre la modélisation de la tâche sur laquelle porte le dialogue et le système de dialogue lui-même. Dans les premières versions, cette tâche était modélisée comme un ensemble d'automates, chacun de ces automates correspondant à l'obtention des informations nécessaires à l'exécution d'une fonction élémentaire de la tâche. Nous nous sommes efforcés de spécifier les requêtes qu'un module de dialogue peut émettre vis-à-vis d'un modèle de gestion de la tâche [58]. Une fois cette spécification obtenue, nous avons implanté des outils génériques de gestion d'un modèle de tâche et des outils permettant au concepteur d'un système de dialogue de décrire confortablement une application particulière.

Cette description devient alors un paramètre du module de gestion de tâche en ce qu'elle décrit des classes dont l'état des objets sur lesquels peut porter le dialogue à un instant donné sont des instances. Par ailleurs, chacun de ces objets contient les actions qui peuvent porter sur lui sous la forme de méthodes provenant de la classe dont il relève.

## 4 Actions industrielles

### 4.1 Projet Esprit Roars

*Participants* : Jean-Marie Pierrel, Gilles Souvay

Le projet Esprit Roars est arrivé à son terme en janvier 1994. Ce projet rassemblait Thomson-ASM, Ena Telecomunicaciones (Madrid) et l'université polytechnique de Valencia (Espagne). L'objectif du projet, en ce qui concerne DIALOGUE était le développement d'outils génériques pour concevoir des systèmes de dialogue. Ces outils concernent la description du langage utilisé (ce qui a permis de décrire le vocabulaire et la syntaxe pour des langages finalisés fondés sur le français, l'anglais

et l'espagnol). Par ailleurs, des outils de description de l'application ont également été conçus et permettent d'assurer la généralité vis-à-vis de l'application à commander via le dialogue. Ces outils sont compatibles avec le système de dialogue DIAPASON ainsi paramétrable par la langue et par l'application. Le système DIAPASON et les outils précédemment décrits ont été validés sur une application de contrôle aérien. Cette démonstration concernait bien entendu l'ensemble du système, reconnaissance de parole développée par Thomson-ASM y compris. Si le système ainsi conçu a donné toute satisfaction, il semble maintenant clair que certains des principes sur lesquels repose l'interaction entre reconnaissance et système de dialogue doivent être remis en cause si l'on souhaite garder des temps de réponse raisonnables sur des applications plus complexes. Cette réflexion justifie une poursuite de la collaboration entre Thomson-ASM et le projet DIALOGUE. Une telle collaboration est décrite au paragraphe 4.2.

#### **4.2 Intégration dialogue-reconnaissance de parole (DRET, Thomson ASM)**

*Participants*: Bertrand Gaiffe, Arnould Mathieu, Jean-Marie Pierrel, Gilles Souvay

Ce nouveau contrat entre Thomson-ASM et le projet DIALOGUE vise à intégrer plus étroitement l'analyse du dialogue et la reconnaissance de la parole. Dans les systèmes de dialogue actuels, et en particulier dans la version courante de DIAPASON, l'interaction entre le système de dialogue et la reconnaissance de la parole est essentiellement fondée sur la syntaxe du langage de commande. Les seules contraintes de nature plus prédictive que peut imposer le module de gestion du dialogue consistent à invalider des règles syntaxiques en fonction du contexte dialogique. L'objectif du projet est donc de faire entièrement prendre en charge l'analyse du langage par le système de dialogue, contrairement au projet précédent dans lequel le système de reconnaissance dispose de la grammaire du langage. Concrètement, cela se traduit par une architecture donnant au système de reconnaissance le rôle de serveur auquel le système de gestion du dialogue fait des requêtes de type "word spotting" d'une part et reconnaissance fine (nécessairement de la gauche vers la droite) d'autre part. De la sorte, le système de gestion du dialogue peut organiser l'analyse du langage en s'appuyant non seulement sur la syntaxe mais, de façon plus globale, sur le contexte courant. Ainsi, on

peut envisager de prédire la possibilité du fragment d'énoncé "la piste verte" et invalider la possibilité de "la piste rouge" dans un contexte où la seule piste présente serait verte. L'application retenue pour valider l'approche est de type Sonar, l'objectif cette fois (par rapport à un précédent contrat) n'étant plus la commande du sonar lui-même, mais la structuration par un opérateur des pistes sonar en contact, les premières correspondant à des signaux sonar et les secondes aux causes de ces signaux (navires essentiellement). Cette application devrait pleinement justifier l'architecture choisie, sa prise en compte par le modèle précédent étant en tout cas clairement hors de portée, en temps réel du moins.

### 4.3 Collaboration avec Thomson LCR dans le cadre d'une convention CIFRE

*Participants* : Jean-Marie Pierrel, Xavier Pouteau, Laurent Romary

Dans le cadre d'une convention CIFRE, le projet DIALOGUE a poursuivi, en lien avec le Laboratoire Central de Recherches de THOMSON-CSF, une réflexion faisant suite à l'étude DRET G1 91/419 sur la définition des interfaces multimodales [57]. Ce travail a contribué à positionner une partie de l'orientation de l'équipe vis-à-vis du thème de recherche concernant les interfaces multimodales. Dans une perspective de dialogue réel entre opérateur et système, il est en effet nécessaire de remettre en cause une logique de fonctionnement de l'interface reposant sur la réactivité des objets graphiques pour autoriser un usage adapté du canal vocal [42], sans quoi l'intégration de la parole n'a un intérêt ergonomique qu'en tant que canal de secours.

En poursuivant dans cette direction, nous avons pu définir un cadre à l'intégration du geste dans le processus d'interprétation des énoncés (pour les coréférences déictiques). Ce cadre permet de faire le lien d'un point de vue référentiel entre les segments linguistiques et les désignations effectuées en prenant en compte les éléments de présentation de l'interface graphique de l'application, en accord avec les recherches menées au laboratoire sur les gestes de désignation<sup>8</sup>.

<sup>8</sup>N. Bellalem et L. Romary "Le dialogue homme machine multimodal, vers la compréhension du geste de désignation", *Actes des 2 èmes journées internationales "L'interface des mondes réels et virtuels"* mars 1993, Montpellier

L. Romary, N. Bellalem et B. Gaiffe "Gestion de la référence dans un dialogue homme-machine", *actes du colloque CNRS "Images et langages"* Paris, avril 1993

#### **4.4 Intégration dialogue et reconnaissance vocale pour avions d'armes (DRET, Sextant avionique)**

*Participants* : Noëlle Carbonell, Yves Laprie, Jean-Marie Pierrel, Eric Vannesson

L'objectif général de ce contrat est, d'une part, d'intégrer des fonctionnalités de l'ordre du dialogue et, d'autre part, d'améliorer la reconnaissance de parole en milieu bruité pour un système utilisable dans un avion d'arme. Dans ce cadre, le système SNORRI d'étude de la parole, développé dans le projet par Yves Laprie et porté sous Motif par Nathalie Flammang (ingénieur contractuel CNRS) a été fourni à Sextant Avionique.

#### **4.5 Tobie-Sol**

*Participants* : Yolande Anglade, Jean-Christophe Dubois, Jean-Marie Pierrel

Le projet Tobie-Sol, avec la SOLLAC est arrivé à son terme en août 1993. L'objectif du projet peut être considéré comme atteint dans la mesure où l'ensemble du système de dialogue et de reconnaissance est opérationnel et utilisé quotidiennement par la standardiste aveugle de Florange. Rappelons par ailleurs, que les idées innovantes liées au projet ont donné lieu à trois brevets déposés par la SOLLAC.

Pour autant, bien que le projet soit arrivé à terme, les apports auxquels il a donné lieu continuent à être décrits dans des publications. Ainsi, en août 1994 il a été présenté à Toronto (International Ergonomics Association) [20] et en octobre dernier au Congrès ERGO'IA à Biarritz [31].

### **5 Actions nationales et internationales**

#### **5.1 Projet Lingua**

*Participants* : Bertrand Gaiffe, Laurent Romary

**Concordances parallèles multilingues appliquées à l'enseignement, à l'apprentissage et à la traduction d'une variété de langues européennes**

Ce projet du bureau Lingua de l'Union Européenne réunit principalement des enseignants en langues étrangères ainsi que des informaticiens principalement regroupés au sein du projet Dialogue et à l'Université de Birmingham. Le projet a pour objectifs :

- la mise au point d'un logiciel de concordances parallèles multilingues, c'est-à-dire capable de mettre en regard des citations provenant de textes parallèles traduits dans diverses langues ;
- la constitution d'un corpus de textes parallèles (traduits dans diverses langues européennes, de façon à la fois idiomatique et fidèle, sans détours inutiles) sous forme électronique. Les langues qui ont été choisies sont le français, l'allemand, l'anglais, l'italien, le danois et le grec ;
- l'élaboration de matériel pédagogique fondé sur des concordances parallèles obtenues à partir de ce corpus.

Dans ce cadre, nous avons tout d'abord défini un cadre de représentation des textes du corpus sur la base de la norme SGML et plus particulièrement des consignes de codage de textes littéraires issues de la *Text Encoding Initiative*. D'autre part, nous avons développé une série d'outils de codage et de manipulation de textes en nous appuyant sur la plate-forme Dilib, développée par ailleurs au sein du laboratoire. Enfin, un outil de concordance parallèle, reposant sur un algorithme de type programmation dynamique, a été mis en oeuvre [19]. Le travail ainsi réalisé, et qui devrait se prolonger durant l'année 95, apporte au projet une expérience en matière de gestion de corpus textuels.

## 5.2 Action Régionale scientifique Cognisciences Grand-Est PIR-CNRS Cognisciences

*Participants* : Noëlle Carbonell, Daniel Coulon, Bertrand Gaiffe, Jean-Marie Pierrel, Anne Reboul, Laurent Romary

Nous participons activement, depuis février 1991, à l'activité du réseau régional Grand-Est du PIR-CNRS Cognisciences. En effet, N. Carbonell est co-responsable (avec J.M. Danion, Université Louis Pasteur, Strasbourg) de ce réseau créé pour développer les recherches pluridisciplinaires en Sciences Cognitives dans l'Est de la France. J.M. Pierrel est membre du bureau et du comité scientifique. Enfin, D. Coulon, également membre du comité scientifique, anime l'un des deux séminaires

de recherche régionaux (le séminaire Scicogn) et assure (avec C. Bursztein, Université Louis Pasteur, Strasbourg) l'édition et la diffusion d'un bulletin mensuel d'information.

Le dynamisme scientifique du réseau Grand-Est s'est traduit en 1994, par le soutien de cinq projets, dont deux propositions nouvelles. L'un des nouveaux projets sélectionnés par le Comité scientifique, qui porte sur l'étude perceptive et linguistique de la prosodie, implique des chercheurs de DIALOGUE et de l'Institut de Phonétique de Nancy. Cette année également, le Programme Cognisciences a renouvelé son soutien au projet "L'anaphore et son traitement", après évaluation nationale (en Décembre 1993) des travaux réalisés au cours de l'année 1993. Des chercheurs de DIALOGUE participent activement à ce projet. Les recherches qui y sont développées sont tout à fait complémentaires de celles présentées en 3.2.1.

### 5.3 Projet inter-PRC DALI

*Participants* : Noëlle Carbonell, Christine Fay-Varnier, Bertrand Gaiffe, Jean-Marie Pierrel, Laurent Romary, Azim Roussalany

A la suite du dernier appel d'offre inter-PRC du MRE sur le thème "langue et dialogue", le projet DALI (Dialogue Adaptatif: Langue et Interaction) défini sous l'impulsion conjointe de Gérard Sabah du LIMSI et de Jean-Marie Pierrel a été retenu par le MESR pour la période septembre 93 - septembre 95. Ce projet implique cinq équipes : LIMSI (G. Sabah, F. Neel, M. Denis), IRIT (M. Borillo, G. Perennou), IRISA (J. Siroux, M. Guyomard), ICP (J. Caelen) et le projet DIALOGUE. Ce projet s'articule autour de six grands thèmes fortement interconnectés entre eux :

- gestion du dialogue et planification,
- étude des phénomènes de référence,
- lexique,
- modèles de tâches et d'utilisateur,
- architectures des systèmes,
- robustesse dans le dialogue homme-machine.



Les chercheurs du projet DIALOGUE, compte tenu des objectifs de recherche du projet (cf. 3), sont plus particulièrement impliqués dans les trois premiers thèmes.

#### 5.4 Projet PRC-Sciences Cognitives

*Participants* : Noëlle Carbonell, Yannick Toussaint

Le projet pluridisciplinaire "Dialogue et coopération", soumis en réponse à l'appel d'offres du PRC-Sciences Cognitives, a été sélectionné en Juillet 1993 par le MESR et bénéficie de son soutien jusqu'en Juin 1996. Il fournit l'occasion à des chercheurs de DIALOGUE de coopérer avec des psychologues, des psycholinguistes et des ergonomes, dont Pierre Falzon (Professeur au CNAM), sur l'étude expérimentale et la modélisation des processus de coopération au sein de dialogues finalisés. L'expérimentation, qui porte sur l'assistance à l'apprentissage de Word dans différentes situations de communication, est terminée ; l'analyse des dialogues, ainsi que la modélisation des comportements et stratégies des partenaires (novice et expert du logiciel de traitement de textes) sont en cours.

#### 5.5 Projet régional IRMA

*Participants* : Noëlle Carbonell, Laurent Chapelier, Christine Fay-Varnier, Azim Roussanaly

Dans le cadre de la préparation du prochain plan Etat - Région Lorraine, le projet DIALOGUE a été fortement impliqué dans la définition d'un projet de recherche IRMA (Interaction Répartie Multimodale pour le multimedia) dont l'objectif final est la conception d'une interface multimodale intelligente répartie alliant le vocal, la désignation gestuelle et le graphisme, et apportant une assistance à l'utilisateur lors de l'utilisation effective de systèmes multimédias. Ce projet, qui bénéficie d'un soutien régional, fédère plusieurs équipes de laboratoires membres du Pôle Technologique Régional IAE+M. Il est placé sous la responsabilité conjointe de Christine Fay-Varnier et de Francis Lepage (CRAN). Noëlle Carbonell assure la responsabilité scientifique de l'action "dialogues naturels multimodaux".

## 5.6 Relations avec la communauté nationale

Nous participons aux GDR-PRC “Intelligence Artificielle” et “Communication Homme-Machine” (J.M. Pierrel est membre du bureau du pôle Langage Naturel et co-responsable du projet DALI cité ci-dessus). De plus, nous participons à divers groupes de travail du GDR-PRC CHM sur le lexique, l'ingénierie linguistique, la prosodie et les interfaces utilisateur multimodales.

Notons par ailleurs que nous sommes impliqués dans plusieurs associations scientifiques nationales ou internationales, en particulier : l'AF CET (J.M. Pierrel, membre du comité technique Intelligence Artificielle et Reconnaissance des Formes), l'ARC (N. Carbonell, membre du bureau), la SFA (Société Française d'Acoustique) Groupe Communication Parlée (Yves Laprie, membre du bureau), l'AFIA (Association Française d'Intelligence Artificielle)...

Citons aussi des relations plus particulières mises en place avec S. Maeda, chercheur à l'ENST, sur l'interprétation automatique de données radiocinématographiques.

Enfin, il faut noter de nombreuses interactions avec plusieurs laboratoires français à travers, entre autres, la participation d'un membre du projet à leur comité scientifique : INALF (Nancy et Paris), URIAH (Valenciennes), LAFORIA (Paris), LIPN (Paris Nord), LIMSI (Orsay), IRIT (Toulouse), LIM (Marseille), ENST (Paris), ICP et GETA (Grenoble), et l'implication forte de l'un des nôtres dans des structures d'évaluation de la recherche (Comité national de la recherche scientifique du CNRS) ou de réflexion sur la recherche (Observatoire de la recherche fondamentale en informatique du MESR).

## 5.7 Relations internationales

Parmi les relations internationales spécifiques de notre projet, outre nos participations au projet Esprit II ROARS, et au réseau d'excellence “Speech and Language” (ELSNET), on peut noter nos liens avec :

- les facultés Notre-Dame de la Paix de Namur (J. Berleur, G. Deville) ;
- le Centre de Recherche du Centre Universitaire du Luxembourg (P. Mousel) : recherches communes sur les analyseurs syntaxico-sémantiques ;

- l'Institut d'Informatique Appliquée (IAI) de Sarrebrück : projets transfrontaliers communs, échanges de logiciels ;
- l'université de Valencia (J.M. Benedi et P. Casacuberta) : accord cadre de coopération et projet Esprit ROARS ;
- Speech Technology Laboratory (STL), Santa Barbara, USA (H. Wakita et J.C. Junqua) : poursuite des recherches communes, échanges de logiciels et, depuis le 1er décembre 1991, accueil à STL de P. Morin, chercheur de l'équipe [7] ;
- l'université de Genève (J. Moeschler) : échanges scientifiques sur la pragmatique du dialogue.
- en outre, nous avons des relations plus ponctuelles avec de grands centres étrangers travaillant sur le dialogue oral homme-machine et le traitement automatique du langage naturel en Europe, aux Etats-Unis et au Japon.
- nous sommes également impliqués dans l'ESCA (European Speech Communication Association), SIGCHI (Special Interest Group on Computer-Human Interaction of the ACM).

## 6 Diffusion des résultats

### 6.1 Diffusion de produits

*Participants* : Yves Laprie, Jean-Marie Pierrel, Gilles Souvay

Le logiciel Snorri (éditeur interactif pour l'étude de la parole) a déjà été diffusé auprès de douze laboratoires ou sociétés en France et à l'étranger dans une première version de laboratoire. Une version plus industrielle sous MOTIF a fait l'objet, cette année, première licence commerciale avec Sextant Avionique. Une version sur PC sous Windows, en cours de développement, sera commercialisée prochainement ; son développement est assuré par un ingénieur de recherche et de valorisation mis à la disposition du projet par le CNRS. D'autre part, les outils pour le dialogue définis au sein de DIAPASON (cf paragraphe 3.4.2) ont servi de support à la définition de l'application ATC (Air Traffic Control) développée dans le projet Esprit ROARS et, dans ce cadre, ont été implantés chez Thomson Sintra DASM et chez ENA Telecomunicaciones (Madrid).

## 6.2 Colloques, congrès, revues

Animation de revues scientifiques : *Intellectica* (D. Coulon, rédacteur en chef jusqu'en Janvier 1994), *TSI* (Technique et Science Informatique), (D. Coulon et J.M. Pierrel, membres du comité de rédaction), *Traitement du Signal*, (J. M. Pierrel, membre du comité de lecture).

Participation à des comités de programme : congrès RF-IA 94 de l'AF-CET (J.M. Pierrel), ERGO-IA 94 (J.M. Pierrel), XX èmes journées d'étude sur la parole de la SFA (Y. Laprie), IHM'94 (N. Carbonell), European Conference on Cognitive Science ECCS'95 (N. Carbonell), International Workshop on Intelligence and Multimodality in Multimedia Interfaces IMMI-1'95 (N. Carbonell), et lecture d'articles pour de nombreux autres colloques ou congrès.

Participation en qualité de conférencier invité à :

- “The way we speak and the structure of the world”, (L. Romary), Université de Sarrebrück, novembre 1993.
- “Le suivi de formants”, (Y. Laprie), Université Libre de Bruxelles, février 1994.
- “Le poids des pères, le choc des fils : prédicats de phase, modificateurs et identification”, (A. Reboul), Journée Scicogn, mars 1994.
- “Représentation formelle de la sémantique d'une spécification de logiciel en langue naturelle”, (Y. Toussaint), Journées ‘Textes de type consigne’, Prescott, Toulouse, avril 1994.
- “Modèles formels et référence”, (B. Gaiffe, L. Romary), Séminaire Cognosciences-Est, avril 1994.
- “On the philosophy and linguistics of identity through time”, (A. Reboul), Workshop ‘L'identité à travers le temps’, Genève, avril 1994.
- “Quels indices pour quel calcul référentiel”, (L. Romary), Atelier ‘Interface Syntaxe-Sémantique’, Neuchatel, mai 1994.
- “Référence et perception du monde : un parallèle nécessaire”, (L. Romary), Séminaire du LIMSI, juin 1994.
- “Langue naturelle et geste en communication homme-machine : Résultats expérimentaux”, (N. Carbonell), Journées du Groupe

de Travail du GDR-PRC-CHM sur les interfaces multimodales, Paris, septembre 1994.

- “Spécification en langue naturelle”, (Y. Toussaint), Séminaire de formation ‘Elaboration et écriture des spécifications’, organisé par la S.E.E. et le Club Génie Logiciel, Paris, 19-21 octobre 1994.

Participation à des Workshops :

- Workshop sur “le dialogue homme-robot en langue naturelle”, Université de Caen, 21-22 octobre 1993.
- Journées Génie Linguistique, Paris, Juillet 1994.

Participation à des écoles de formation :

“European Summer School on Langage, Logic and Information”, Copenhague 1994.

Ecole d’Eté d’Analyse Numérique (EEAM), CEA-INRIA-EDF “Réseaux de neurones et applications”, Bréau-Sans-Nappe, 13-24 juin 1994.

### 6.3 Actions d’enseignement

La présence au sein du projet de huit enseignants-chercheurs explique, avec la participation significative des chercheurs permanents, notre forte implication dans les actions d’enseignement à Nancy et à Metz. Nous citerons simplement les enseignements liés aux activités scientifiques du projet et destinés à des étudiants de troisième cycle ou à des élèves ingénieurs de troisième année :

Informatique linguistique, communication homme-machine (DESS Informatique, Univ. Henri Poincaré, DEA Informatique commun aux trois universités de Nancy).

Représentation des connaissances et intelligence artificielle (DESS Industries de la Langue, Université de Metz, DESS Ingénierie de la Connaissance et de la Communication commun aux universités de Nancy II et de Metz).

Intelligence artificielle (DESS Information Scientifique et Technique commun aux universités de Nancy, ENSG (INPL) troisième année, DEA Sciences du Langage, université de Nancy II).

Ingénierie des interfaces homme-machine (ESIAL troisième année, Univ. Henri Poincaré).

Reconnaissance des formes et intelligence artificielle (SUPELEC Metz, troisième année, option IA).

Pragmatique et référence, DEA Sciences du Langage, Université de Nancy II).

Parmi les nombreuses responsabilités assurées par les membres du projet dans le domaine de l'enseignement, nous mentionnerons uniquement celles qui concernent la formation à et pour la recherche :

Direction de l'Ecole Doctorale IAE+M (Informatique, Automatique, Electronique et électrotechnique, Mathématique) de Nancy, et Direction du FIRTECH Informatique-Automatique de Lorraine (J. M. Pierrel).

Enfin, la présence au sein du projet de plusieurs chercheurs d'expérience et de trois professeurs d'université se traduit aussi par une forte implication dans la formation à et par la recherche : direction de recherche, rapports de thèse, participation à des jurys de thèse à Grenoble, Paris, Orsay, Sophia Antipolis, Valenciennes, etc.

## 7 Publications

### Livres et monographies

- [1] J. MOESCHLER, A. REBOUL, J.-M. LUSHER, J. JAYEZ, *Langage et pertinence, Collection Processus Discursifs Langue et Cognition*, Presses Universitaires de Nancy, 1994.
- [2] J. MOESCHLER, A. REBOUL, *Dictionnaire Encyclopédique de Pragmatique*, Editions du Seuil, 1994.

### Thèses

- [3] Y. ANGLADE, *Robustesse de la reconnaissance automatique de la parole : étude et application dans un système d'aide vocal pour une standardiste mal-voyante*, thèse de doctorat, Université Henri Poincaré, Nancy I, Vandœuvre-lès-Nancy, février 1994.
- [4] D. F. COIS, *Détection et identification des occlusives et fricatives au sein du système indépendant du locuteur APHODEX*, thèse de doctorat, Université Henry Poincaré, Nancy I, 1995, à paraître.

- [5] F. DUERMAEL, *Référence aux actions dans des dialogues de commande Homme-Machine*, thèse de doctorat, Institut National Polytechnique de Lorraine, 21 décembre 1994.
- [6] R. HAJISLAM, *Décodage acoustico-phonétique et robustesse en reconnaissance automatique de la parole*, thèse de doctorat, Université Henri Poincaré, Nancy I, Vandœuvre-lès-Nancy, mars 1994.
- [7] P. MORIN, *PARTNER : Un Système de Dialogue Homme-Machine Multimodal pour Applications Finalisées à Forte Composante Orale*, thèse de doctorat, Université Henri Poincaré, Nancy I, Vandœuvre-lès-Nancy, septembre 1994.

### Articles et chapitres de livre

- [8] P. ALINAT, J.-M. PIERREL, *ROARS : Robust Analytical Speech Recognition System*, édition K.C. Varghese, S. Pflieger, J.P. Lefèvre, 1, Advanced Speech Applications, European Research on Speech Technology (Research Reports SPRIT), 1994.
- [9] A. BONNEAU, F. CHARPILLET, S. COSTE-MARQUIS, J. P. HATON, Y. LAPRIE, P. MARQUIS, «Towards a Multilevel Model for Hypothetical Reasoning in Continuous Speech Recognition», in : *Levels in Speech Communication : Relations and Interactions*, C. Sorin, J. Mariani, et H. Méloni (éd.), Elsevier, 1994.
- [10] M. BORILLO, Y. TOUSSAINT, «Linguistics in Software Engineering», in : *Encyclopedia of Software Engineering*, John Wiley & Son, Inc, 1994, p. 609–613, Vol.1.
- [11] E. BOUYER, D. COULON, «Détection des entités et des procès mis en relation dans les anaphores associatives», in : *L'anaphore associative (Aspect linguistiques, psycholinguistiques et automatiques)*, C. Schnedecker, M. Charolles, G. Kleiber, et J. David (éd.), Librairie Klincksieck, Juin 1994, p. 321–343.
- [12] N. CASTELL, O. SLAVKOVA, Y. TOUSSAINT, T. TUELLS, «Control de calidad de las especificaciones de software escritas en lenguaje natural», *Novatica, Inteligencia Artificial-2*, 109, Mai 1994, p. 33–40.
- [13] D. FOHR, J.-P. HATON, Y. LAPRIE, «Knowledge-based Techniques in Acoustic-Phonetic Decoding of Speech: Interest and Limitations», *International Journal of Pattern Recognition and Artificial Intelligence* 8, 1, 1994, p. 133–153.
- [14] R. HAJISLAM, Y. ANGLADE, J. C. JUNQUA, J. M. PIERREL, «Etude acoustique du réflexe Lombard en vue de la reconnaissance de la parole produite en milieu bruité», *Journal de Physique IV, Colloque C5 4*, Mai 1994, p. 485–488.

- [15] J.-M. PIERREL, B. GAIFFE, L. ROMARY, « Quel lexique pour un traitement automatique de la référence : une première approche », *Cahiers de Linguistique Française*, 14, 1994, p. 37–52.
- [16] J.-M. PIERREL, « Représentations conceptuelles et Intelligence Artificielle », *in: Traitements informatisés de corpus textuels*, E. Martin (éd.), *Collection Etudes de sémantique lexicale*, Didier Erudition, 1994, p. 81–106.
- [17] A. REBOUL, « The description of lies in speech acts theory », *in: Pretending to communicate*, H. Parret (éd.), de Gruyter, Berlin, 1994, p. 292–298.
- [18] A. REBOUL, « Le poids des pères, le choc des fils: prédicats de phase, modificateurs et identification », *Cahiers de Linguistique Française*, 14, 1994, p. 229–246.
- [19] L. ROMARY, N. MEHL, D. WOOLLS, « The Lingua Parallel ConCORDancing Project: Managing Multilingual Texts for Educational Purpose », *Text Technology à paraître*, 1994.

### Communications à des congrès, colloques, etc.

- [20] Y. ANGLADE, J.-C. DUBOIS, J.-M. PIERREL, « TOBIE-SOL: a multimodal system for a telephone switchboard operator center », *in: Proceedings of International Ergonomics Association*, p. 148–150, Toronto (Canada), août 1994.
- [21] Y. ANGLADE, J.-C. DUBOIS, J.-M. PIERREL, « Un poste de travail vocal pour une standardiste malvoyante », *in: Actes séminaire INSERM-ANPEA*, p. 61–73, Paris, novembre 1994.
- [22] N. BELLALEM, L. ROMARY, « Langue et geste pour le dialogue homme-machine finalisé », *in: Actes de 01Design'95 - Aspects communicatifs en conception*, Europa & GDR-PRC Communication homme-machine, janvier 1995.
- [23] N. BELLALEM, « Interprétation de la désignation dans un dialogue homme-machine associant parole et geste », *in: Actes Secondes Rencontres Nationales des Jeunes Chercheurs en Intelligence Artificielle*, p. 279, Marseille, 1994.
- [24] L. BUNNET, D. FOHR, J. C. JUNQUA, « Une méthode connexionniste pour la reconnaissance de mots enchaînés en milieu bruité », *in: Actes 7èmes Journées Neurosciences et Sciences de l'Ingénieur*, P. Marchal (éd.), p. 27–30, Chamonix, Mai 1994.
- [25] N. CARBONELL, C. MIGNOT, « “Natural” Multimodal HCI: Experimental Results on the Use of Spontaneous Speech and Hand Gestures », *in:*



- Actes ERCIM Workshop Reports, Multimodal Human-Computer Interaction*, N. Carbonell (éd.), p. 97–112, 1994.
- [26] N. CARBONELL, C. VALOT, C. MIGNOT, P. DAUCHY, «Etude empirique: usage du geste et de la parole en situation de communication homme-machine», *in: Actes Congrès ERGO-IA '94*, R. Patesson (éd.), p. 128–139, Biarritz, Octobre 1994.
- [27] N. CASTELL, O. SLAVKOVA, T. TUELLS, Y. TOUSSAINT, «Quality control of software specifications written in natural language», *in: actes de la conférence IEA-AIÉ'94 et rapport technique de l'Université Polytechnique de Catalogne, LSI-93-50-R*, 1994.
- [28] L. CHAPELIER, C. FAY-VARNIER, A. ROUSSANALY, V. SAINT-DIZIER, «Recueil et analyse d'un corpus d'interactions multimodales Homme-Machine», *in: Actes ERGO-IA '94*, Biarritz, Octobre 1994.
- [29] J. DI MARTINO, J. F. MARI, B. MATHIEU, K. PEROT, K. SMAÏLI, «Which Model for Future Speech Recognition Systems: Hidden Markov Models or Finite-State Automata», *in: Proceedings International Conference on Acoustics, Speech and Signal Processing*, IEEE, Adélaïde (South Australia), April 1994.
- [30] L. DJEZZAR, A. BONNEAU, Y. LAPRIE, «Effet de la connaissance de la voyelle sur la perception du bruit d'explosion des occlusives», *in: Actes XXèmes Journées d'Etude sur la Parole*, p. 234–238, Trégastel, Juin 1994.
- [31] J.-C. DUBOIS, Y. ANGLADE, J.-M. PIERREL, «Recherche tolérante d'informations dans un système d'aide vocal pour une standardiste malvoyante», *in: Actes ERGO-IA '94*, p. 203–214, Biarritz, octobre 1994.
- [32] F. DUERMAEL, B. GAIFFE, J.-M. PIERREL, «Un moteur à deux temps pour référer aux actions», *in: Actes ERGO-IA '94*, p. 512–523, Biarritz, Octobre 1994.
- [33] D. FINCK, «Application des notions d'objet et d'agent à la traduction automatique», *in: Actes du congrès AFCET-RFIA (AFCET-AFIA)*, Paris, janvier 1994.
- [34] D. FINCK, «Comprendre un ordre pour pouvoir l'exécuter», *in: Actes Congrès ERGO-IA '94*, p. 501–511, Biarritz, Octobre 1994.
- [35] B. GAIFFE, A. REBOUL, L. ROMARY, «Les NP définis : anaphore, anaphore associative et cohérence», *in: Actes du Colloque Relations anaphoriques et incohérence*, Université d'Anvers, décembre 1994.
- [36] B. GAIFFE, A. REBOUL, L. ROMARY, «Références et gestion du dialogue», *in: Actes Congrès TALN'94*, p. 134–151, Marseille, Avril 1994.
- [37] R. HAJISLAM, Y. ANGLADE, J.-C. JUNQUA, J.-M. PIERREL, «Etude acoustique du réflexe Lombard en vue de la reconnaissance de la pa-

- role produite en milieu bruité», *in: Actes 3ème Congrès d'Acoustique*, Toulouse, mai 1994.
- [38] Y. LAPRIE, M.-O. BERGER, «A New Paradigm for Reliable Automatic Formant Tracking», *in: Proceedings of ICASSP'94*, 2, p. 201–204, Adelaide (Australia), avril 1994.
- [39] Y. LAPRIE, L. MERCIER, «Un environnement logiciel pour un atelier phonétique», *in: Actes XXèmes Journées d'Etude sur la Parole*, p. 209–214, Trégastel, Juin 1994.
- [40] J.-F. MARI, D. FOHR, Y. ANGLADE, J.-C. JUNQUA, «Hidden Markov Models and Selectively Trained Neural Networks for Connected Confusable Word Recognition», *in: Proceedings International Conference on Spoken Language Processing*, p. S26.11, Yokohama (Japan), September 1994.
- [41] V. PAGEL, N. CARBONELL, Y. LAPRIE, «Perception de l'accent et des marques prosodiques structurelles en français: premiers résultats», *in: Actes Séminaire GDR PRC Communication Homme-Machine*, Nancy, Mars 1994.
- [42] X. POUTEAU, L. ROMARY, J.-M. PIERREL, «Voix, geste et multimodalité: quand dire c'est faire faire», *in: Actes Congrès ERGO-IA '94*, p. 491–500, Biarritz, Octobre 1994.
- [43] A. REBOUL, «I and you: personal identity and simple identity», *in: Acts of the 1rst International Colloquium on Deixis: Time, space and identity*, Lexington, Kentucky, 4-6 décembre 1994.
- [44] A. REBOUL, «(In)cohérence et anaphore: mythes et réalités», *in: Actes du Colloque Relations anaphoriques et incohérence*, Université d'Anvers, décembre 1994.
- [45] A. REBOUL, «What, if anything, is accessibility? A relevance-oriented criticism of Ariel's Accessibility theory of referring expressions», *in: Actsof the 6th International Conference on Functional Grammar*, York, 22-26 août 1994.
- [46] L. ROMARY, N. BELLALEM, D. SCHANG, «Positioning objects in a graphical environment: Reference and Gesture», *in: Actes du premier colloque international sur la déixis - Time, Space and Identity*, University of Kentucky, décembre 1994.
- [47] L. ROMARY, «L'interprétation de ici dans des énoncés de positionnement», *in: Actes Workshop "le dialogue homme-robot en langage naturel: problèmes psychologiques"*, Université de Caen, octobre 1993.
- [48] L. ROMARY, «Sens et action, ou comment aménager son salon», *in: Actes Congrès TALN'94*, p. 165–174, Marseille, Avril 1994.

- [49] D. SCHANG, L. ROMARY, «Frames, A Unified Model for the Representation of Reference and Space in a Man-Machine Dialogue», *in : Proceedings 3rd International Conference on Spoken Language Processing*, Yokohama (Japan), 1994.
- [50] D. SCHANG, L. ROMARY, «Framing the World: Towards a Localized Spatial Reasoning», *in : Proceedings 3rd International Conference on the Cognitive Science of Natural Language Processing*, Dublin (Ireland), 1994.
- [51] D. SCHANG, «Modélisation de la référence et du positionnement spatial dans un système de dialogue Homme-Machine», *in : Actes Secondes Rencontres Nationales des Jeunes Chercheurs en Intelligence Artificielle*, Marseille, 1994.
- [52] Y. TOUSSAINT, «From Natural Specification to Formal Specification : Linguistic Engineering for Software Specification», *in : Workshop on Artificial Intelligence and Software Engineering, International Conference on Software Engineering, Sorrento, Italie*, mai 1994.

### Rapports de recherche et publications internes

- [53] N. CARBONELL, L. CHAPELIER, C. FAY, J.-M. PIERREL, A. ROUSSANALY, V. SAINT-DIZIER, «Corpus d'interactions multimodales homme-machine à partir du logiciel GOCAD», *Rapport interne*, Centre de Recherche en Informatique de Nancy, Vandœuvre-lès-Nancy, 1994.
- [54] N. CARBONELL, L. CHAPELIER, C. FAY, J.-M. PIERREL, A. ROUSSANALY, V. SAINT-DIZIER, «Recueil d'un corpus d'interactions multimodales homme/machine», *Rapport interne*, Centre de Recherche en Informatique de Nancy, Vandœuvre-lès-Nancy, 1994.
- [55] N. CARBONELL, «Éléments de définition d'un langage multimodal acceptable par l'utilisateur», *Rapport interne*, Centre de Recherche en Informatique de Nancy, Vandœuvre-lès-Nancy, 1993.
- [56] X. POUTEAU, B. BACCONNET, «Co-référence voix+geste dans un dialogue multi-modal: vers une interaction naturelle», *Rapport interne*, Centre de Recherche en Informatique de Nancy, Vandœuvre-lès-Nancy, 1993.
- [57] THOMSON-CSF/SDC, CERMA, CRIN, «Dialogue homme-machine intelligent : définition d'une grammaire multimodale», *rapport de fin de contrat DRET G1 91/419 n° Thomson-CSF/SDC/DTS/PTI/194/93*, Thomson-CSF/SDC and CERMA and CRIN, novembre 1993.

## Divers

- [58] V. MORLOT, «Paramétrisation d'un système de dialogue par l'application», Diplôme d'Ingénieur CNAM, Juillet 1994.
- [59] V. SAINT-DIZIER, « Une problématique psycho-sociale pour la conception et l'implémentation d'un système d'aide au logiciel GOCAD », mémoire de DEA de l'Université de Nancy II, Juillet 1994.

## 8 Abstract

The main purpose of our group - which is affiliated both to the CNRS and INRIA - is to provide advances in the definition and development of human-computer dialogue systems accepting spoken or written natural language, in order to achieve robust and reliable communication.

We therefore deal with problems ranging from acoustic-phonetic decoding to the interpretation of natural language utterances and dialogue management.

For this purpose, we use Pattern Recognition and Artificial Intelligence techniques, with a three-fold goal:

- developing basic studies on the fundamental mechanisms associated with speech, language and dialogue understanding;
- confronting these models with the results obtained in related research domains, such as linguistics (phonetics and pragmatics) or cognitive psychology (ergonomics). For instance, our participation to one of the Cognisciences-Est projects (Cognitive Science in the East of France) gives us the opportunity to develop such a pluridisciplinary research approach;
- the definition of applications dedicated to efficient process control, user assistance or information retrieval.

**Main areas of research** Our research thus deals with three complementary aspects which can be summarized by the three following words Recognition-Understanding-Dialogue:

- **speech recognition** or artificial speech perception. Our aim is to integrate global and analytical methods based on phonetical knowledge :

- recognition of difficult vocabularies (isolated or connected words) by means of mixed models associating Dynamic Time Warping, Hidden Markov Models and formal Neural Networks ;
- multispeaker acoustic-phonetic decoding based upon explicit knowledge and involving various types of speech units, such as phonemes or triphones ;
- enhancement of the recognition robustness through the study of acoustic models robust to noise and the Lombard effect ;
- **the study of linguistic phenomena** which is necessary for the recognition of continuous speech, as well as for the understanding and pragmatic interpretation of natural language utterances; we are currently tackling the following subjects :
  - utterance recognition and lexical matching ;
  - syntactic and semantic parsing based on unification grammars ;
  - reference interpretation in discourse ;
  - inference mechanisms and representations needed for the interpretation of utterances, texts or dialogues, with a specific attention towards time and space references ;
  - terminology : from term extraction to conceptual representation
  - the definition of computational linguistic methodologies and tools in order to define and evaluate quality criteria for technical documents. We are currently studying software specifications written in natural language.
- **the implementation of human-computer dialogue systems** dedicated to standard interactive applications; our research activity in this area is focused on :
  - the system overall architecture and dialogue strategies ;
  - the dialogue structure.

On the whole, in-as-much as they involve speech, natural language, human-computer interaction, our basic research activities are deeply pluridisciplinary. Thanks to the development of experimental dialogue

systems, we also get the opportunity to test the models we have designed and to contribute to the definition of future user interfaces.

## Table des matières

<b>1</b>	<b>Composition de l'équipe</b>	<b>1</b>
<b>2</b>	<b>Présentation du projet</b>	<b>3</b>
<b>3</b>	<b>Actions de recherche</b>	<b>4</b>
3.1	Reconnaissance de la parole . . . . .	4
3.1.1	Reconnaissance de mots isolés et enchaînés . . . . .	5
3.1.2	Décodage acoustico-phonétique . . . . .	7
3.2	Traitement de la langue . . . . .	12
3.2.1	Référence . . . . .	12
3.3	Terminologie et textes techniques . . . . .	17
3.3.1	Lexique et syntaxe . . . . .	20
3.4	Dialogue et applications . . . . .	21
3.4.1	Modélisation de l'utilisateur . . . . .	21
3.4.2	Architectures logicielles pour le dialogue . . . . .	22
<b>4</b>	<b>Actions industrielles</b>	<b>25</b>
4.1	Projet Esprit Roars . . . . .	25
4.2	Intégration dialogue-reconnaissance de parole (DRET, Thomson ASM) . . . . .	26
4.3	Collaboration avec Thomson LCR dans le cadre d'une convention CIFRE . . . . .	27
4.4	Intégration dialogue et reconnaissance vocale pour avions d'armes (DRET, Sextant avionique) . . . . .	28
4.5	Tobie-Sol . . . . .	28
<b>5</b>	<b>Actions nationales et internationales</b>	<b>28</b>
5.1	Projet Lingua . . . . .	28
5.2	Action Régionale scientifique Cognisciences Grand-Est PIR-CNRS Cognisciences . . . . .	29
5.3	Projet inter-PRC DALI . . . . .	30

Rapport d'activité INRIA 1994 — Annexe technique

5.4	Projet PRC-Sciences Cognitives . . . . .	31
5.5	Projet régional IRMA . . . . .	31
5.6	Relations avec la communauté nationale . . . . .	32
5.7	Relations internationales . . . . .	32
<b>6</b>	<b>Diffusion des résultats</b>	<b>33</b>
6.1	Diffusion de produits . . . . .	33
6.2	Colloques, congrès, revues . . . . .	34
6.3	Actions d'enseignement . . . . .	35
<b>7</b>	<b>Publications</b>	<b>36</b>
<b>8</b>	<b>Abstract</b>	<b>42</b>