Projet CORDIAL

Communication multimodale personne-machine à composantes orales : méthodes et modèles

Localisation : Rennes

---

1 Composition de l'équipe

Responsable scientifique

Jacques Siroux [professeur, IUT]

Personnel UPRES A **6064**

Yolande Anglade [maître de conférences, IUT]
Olivier Boëffard [maître de conférences, Enssat]
Marc Guyomard [professeur, Enssat]
Mouloud Kharoune [maître de conférences, IUT]
Guy Mercier [professeur associé, Enssat]
Laurent Miclet [professeur, Enssat, (participe également au projet Aïda)]
Pierre Nerzic [maître de conférences, IUT]
Jean-Christophe Pettier [maître de conférences, Enssat]

Chercheurs doctorants

Jacques Chodorowski [bourse sur contrat Cnet]
Boris Cormons [bourse Cnet, (thèse soutenue en mars 1999)]
Hélène François [bourse sur contrat Cnet]

2 Présentation et objectifs généraux

La conception et la réalisation de systèmes informatiques, tels que par exemple des serveurs ou des gestionnaires d'information, destinés à des usagers professionnels ou occasionnels, doivent intégrer de façon explicite et intentionnelle le traitement de tous les aspects de la communication personne-machine. En effet, nombreux sont les exemples d'échecs (aux conséquences économiques souvent lourdes) de systèmes, échecs ou insuccès dûs à une conception superficielle de l'interface gérant les interactions entre les usagers et le système. La prise en compte des nombreuses facettes de la communication personne-machine nécessite différents modèles (de l'usager, du dialogue, de connaissances,...) à définir ou à adapter à partir de résultats existants.

Parmi les moyens de communication d'un système, la parole se révèle l'un des plus intéressants. Du point de vue de l'utilisateur, l'usage de la parole avec celle naturellement associée de la langue naturelle facilite la manipulation des informations fournies ou reçues du système. Des points de vue utilisateur et système, la parole s'est avérée être le plus performant des média lors de nombreuses expériences

effectuées en laboratoire ou en situation réelle. Ces caractéristiques justifient les recherches effectuées sur le traitement de la parole et la progression, lente mais régulière, du nombre de systèmes commercialisés et mis à la disposition du public. Cependant les résultats encore peu fiables (sans restriction d'environnement) des techniques de reconnaissance de la parole et les difficultés de compréhension du langage naturel en parole spontanée constituent des obstacles qui limitent la mise en place d'un plus grand nombre de systèmes oraux. Il apparaît alors intéressant de rechercher des méthodes et des moyens pour pallier les problèmes posés par la reconnaissance et la compréhension de la parole. Nous pensons que les capacités de dialogue du système, l'ajout de moyens de communication supplémentaires (multimodalité) et l'architecture même de systèmes forment une bonne partie de ces méthodes et moyens.

Dans ce domaine, nos objectifs et activités se déclinent naturellement sur deux dimensions complémentaires : théorique et pratique. Le point de vue théorique s'attache à comprendre et modéliser les fondements de l'interaction afin d'une part de prendre en compte le maximum de phénomènes interactionnels (qu'ils soient de nature cognitive ou comportementale) et, d'autre part, de faciliter la mise au point de nouveaux systèmes. L'un des premiers objectifs est de faire en sorte qu'à partir de l'acte d'énonciation des utilisateurs on puisse faire émerger la signification contextuelle complète de l'énoncé, c'est-à-dire extraire de l'énonciation non seulement les aspects structurels de l'énoncé mais aussi le sens (sémantique) ainsi que l'intention finale de l'utilisateur.

Les obstacles à ces ambitions sont nombreux : comme on l'a déjà signalé, la reconnaissance de la parole ne fournit pas de manière fiable les messages prononcés (de plus ceux-ci peuvent faire l'objet d'altérations lors de leur émission), la structure et le sens des messages sont soumis à des déviations par rapport aux normes généralement admises et enfin, fréquemment, les intentions véhiculées n'apparaissent pas explicitement dans les énoncés. Nous cherchons donc à pallier ces problèmes en utilisant différentes approches et techniques et en nous appuyant sur les concepts intégrateurs d'actes de langage et de planification. Ces concepts servent dès à présent de fondements pour modéliser quelques-uns des principes essentiels du dialogue tels que le suivi des activités des utilisateurs sous-jacentes à l'interaction ou encore la gestion des phases principales du dialogue. Il reste néanmoins plusieurs points à développer sur et autour de ces fondements.

Il est d'abord nécessaire d'enrichir certains aspects de la modélisation afin de traiter les *incidents* de dialogue dûs aux comportements et connaissances des interlocuteurs (utilisateur et système) qui ne sont pas toujours en harmonie ainsi que la gestion du dialogue qui ne dépend pas directement de la tâche à réaliser (gestion dite phatique). Si acte de langage et planification constituent des fondements scientifiques particulièrement appropriés et centraux pour la communication, il reste cependant à éclaircir les relations qu'ils entretiennent avec l'extérieur (*i.e.* l'application et les énonciations des utilisateurs). Il s'agit d'une part de déterminer les différents composants des actes de langage de l'énonciation dans le contexte de la reconnaissance de la parole. Pour cela, nous suivons deux approches ; la première est fondée sur les connaissances linguistiques classiques (lexicale, syntaxique, sémantique,...), la prosodie et la pragmatique (notion de présupposé). La seconde vise l'utilisation de techniques d'apprentissage automatique pour faire apprendre au système à partir de corpus de phrases l'extraction des éléments pertinents ; l'avantage escompté de cette approche réside dans le fait qu'il sera possible d'adapter plus facilement le système à de nouvelles applications.

D'autre part, les mécanismes généraux liés à la planification ne permettent pas de prendre en compte de manière satisfaisante les aspects de la coopération qui proviennent du mode opératoire de celle-ci (par exemple la substitution de paramètres ou encore la modification raisonnée de valeurs de paramètres pour atteindre l'information demandée). Le but est donc de proposer une modélisation de l'application qui autorisera une exploitation optimale de ses caractéristiques dans le cadre du dialogue. Nous avons montré que l'ajout d'une nouvelle modalité (le geste par l'intermédiaire d'un écran tactile) à un système de dialogue oral améliore la qualité de l'interaction en augmentant la compétence du système. Ce constat a été effectué sur un système prototype limité ; il reste à étudier de manière plus fine le comportement des utilisateurs face à un tel système. De façon plus précise, nous étudions les moyens utilisés pour référencer des éléments de l'application, que ceux-ci soient présents ou non dans le contexte visuel. Ces études qui doivent aboutir à une modélisation et à des propositions d'architecture de traitement, portent sur les aspects linguistiques ainsi que gestuels liés aux types des éléments désignés.

La dimension pratique recouvre trois préoccupations principales. Il s'agit en premier lieu de valoriser les activités de recherche plus fondamentales en intégrant les résultats dans des systèmes opérationnels. Cette intégration pose également des problèmes de conception et d'architecture de système. Il est en effet nécessaire de faire coexister de manière souple et efficace les modules construits sur des bases hétérogènes.

Une deuxième préoccupation concerne la promotion de l'oral et des technologies vocales. Nous visons ainsi de nouvelles applications (logiciels éducatifs par exemple) qui illustrent les avantages de l'usage de la voix. Ceci nécessite en amont des activités de recherche et développement.

Enfin, la mise au point de systèmes oraux permet aussi de viser deux buts complémentaires : le recueil de corpus réels d'interaction personne-machine qui sont nécessaires pour affiner les connaissances sur ce sujet, et l'évaluation de système qui est importante d'un point de vue technologique et qui ne peut être étudiée que si l'on dispose d'un système.

3 Fondements scientifiques

3.1 Panorama

Résumé : *Les activités se rattachent à quatre domaines complémentaires par leurs objets d'études et leurs méthodes. Le premier domaine, dialogue et modélisation, s'intéresse au code et à la structure de l'interaction ainsi qu'aux champs d'application des systèmes. Le second concerne la multimodalité (avec une emphase pour l'oral) et les prototypes de systèmes (architecture et évaluation). Le troisième porte sur les techniques d'apprentissage et leurs applications à l'analyse syntaxique. Enfin, le dernier domaine concerne la synthèse de parole (recherche et application).*

3.2 Dialogue et modélisation

Mots clés : acte de langage, planification, reconnaissance de plan.

Glossaire :

acte de langage dans la théorie des actes de langage fondée par Austin[Aus70] et développée par Searle[Sea82], le postulat de base affirme que l'émission d'un énoncé s'assimile à l'accomplissement d'actions qui modifient les états mentaux des interlocuteurs.

plan séquence d'actions destinées à réaliser un but intentionnel.

reconnaissance de plan reconnaître un plan à partir d'une séquence d'actions observées consiste à déterminer les relations qu'entretiennent ces actions afin de déterminer les buts et suites possibles du plan en cours.

Résumé : *Le projet utilise une modélisation fondée sur la notion de plan d'actes de langage. Cette modélisation prend en charge le cadre général de la communication et facilite la mise en œuvre informatique, mais ne résoud pas certains problèmes comme ceux de l'extraction des actes de langages à partir des énoncés observés, de l'intégration des différentes sources d'informations et d'une mauvaise communication entre les interlocuteurs.*

L'interaction personne-machine peut être considérée comme une succession d'actions particulières – les actes de langage [Aus70,Sea82] (nommés dans notre contexte : actes de dialogue) – qui portent la fonction de l'action dans le dialogue (exemple : une interrogation, un ordre,...) ainsi qu'un contenu propositionnel

[Aus70] J. AUSTIN, *Quand dire c'est faire,* Editions du seuil, Paris, 1970.
[Sea82] J. SEARLE, *Sens et expression,* Les éditions de minuit, 1982.
[Aus70] J. AUSTIN, *Quand dire c'est faire,* Editions du seuil, Paris, 1970.

(exemple : le thème de l'interrogation). Ces actes sont aussi caractérisés par leurs conditions d'utilisation qui concernent les états mentaux des participants à l'interaction (leurs intentions, connaissances et croyances). La modélisation informatique la plus pertinente est celle d'un opérateur de plan [All87,Lit85] dans lequel on peut faire figurer des préconditions et contraintes d'utilisation ainsi que l'effet de l'acte.

Par exemple, l'acte de demander à autrui l'exécution d'une action peut se modéliser sous la forme :

Requérir(Locuteur, Auditeur, Action(A))
précondition-intention : *Veut(Locuteur, Requérir(Locuteur, Auditeur, Action(A)))*
précondition-préparatoire : *Veut(Locuteur, Action(A))*
corps : *Croyance-mutuelle(Auditeur, Locuteur, Veut(Locuteur, Action(A)))*
effet: *Veut(Auditeur, Action(A))*

qui peut se paraphraser par : lorsqu'un agent veut que son auditeur réalise une action A, il peut employer l'action étiquetée *Requérir* qui consiste à établir un consensus entre les interlocuteurs pour exécuter A. La réalisation de ce consensus est confiée à une autre action non décrite ici. L'ensemble des actions nécessaires à la réalisation d'un but s'appelle un plan et cette approche fait l'hypothèse que chacun des interlocuteurs participe à la réalisation du plan de son interlocuteur.

Cette modélisation des actes de dialogue permet d'envisager plusieurs types de raisonnements automatiques nécessaires à la conduite d'un dialogue. Le premier concerne la compréhension contextuelle des énoncés de l'interlocuteur par un mécanisme appelé reconnaissance de plans. Cela consiste à reconstruire une partie du plan de l'interlocuteur ; cette partie, si elle est correctement identifiée, permet d'expliciter les motivations de l'interlocuteur et ses croyances. Un second traitement est destiné à calculer une réponse convenable, par un mécanisme de planification, qui tient compte, par la nature même de la modélisation, des informations déjà connues et des malentendus éventuels. Ce type de modélisation rend possible une mise en œuvre informatique dans certains cas simples, mais laisse encore à traiter des problèmes importants d'ordres divers.

Extraction des actes de dialogue

Le premier problème est celui du passage des énoncés prononcés par l'utilisateur à l'acte de dialogue. Ce passage n'est pas un simple problème de transcodage. Il est en effet nécessaire de prendre en compte de manière intégrée une grande variété de connaissances (états mentaux, présuppositions, prosodie,...) ainsi que des indices présents dans l'énonciation elle-même (structure syntaxique, éléments lexicaux). De plus, la forme de surface des énoncés oraux présente de nombreuses irrégularités (problèmes de performance) qui compliquent la tâche de la reconnaissance de parole ainsi que celles de la compréhension et de l'interprétation de l'énoncé.

Modèles de systèmes

Le second problème réside dans l'utilisation du formalisme de plan [3, 6] pour associer trois points de vue : celui de l'application, celui du dialogue «principal» (qui concerne les intentions de l'utilisateur vis-à-vis de l'application) et celui de la gestion du dialogue lui-même (méta dialogue et dialogue phatique). Des solutions partielles ont été proposées [Lit85], mais elles résistent mal à des applications de type manipulation d'informations (interrogation de bases de données) ou qui comportent plusieurs tâches en parallèle, ainsi qu'au traitement de certaines fonctions de gestion de communication. Une approche possible de résolution de ces problèmes peut être une modélisation multi-agents. En effet, ce cadre

[All87] J. ALLEN, *Natural Language Understanding*, Benjamin/Cummings Menlo Park, 1987.

[Lit85] D. J. LITMAN, *Plan Recognition and Discourse Analysis : An Integrated Approach for Understanding Dialogues*, thèse de doctorat, University of Rochester, TR 170, 1985.

[Lit85] D. J. LITMAN, *Plan Recognition and Discourse Analysis : An Integrated Approach for Understanding Dialogues*, thèse de doctorat, University of Rochester, TR 170, 1985.

conceptuel permet de combiner des modèles et contextes de dialogue a priori exclusifs afin d'accroître la couverture dialogique. La problématique se déplace ainsi en partie de la modélisation du dialogue vers la modélisation de l'intégration.

Erreurs de communication

Le troisième problème se pose fréquemment dans toute interaction : celui d'une mauvaise communication. Chacun des deux intervenants (*i.e.* l'usager humain et le système) peut en effet posséder des connaissances erronées sur l'application, sur les compétences de l'autre et sur les éléments du dialogue lui-même tels que les références employées pour désigner les objets mis en cause durant le dialogue. Une erreur concernant ces informations peut, à plus ou moins longue échéance, entraîner un échec, c'est-à-dire une impossibilité pour la machine de satisfaire l'interlocuteur. La détection et le traitement de ces erreurs nécessitent en amont une tâche de caractérisation, puis une modélisation dans le cadre de la planification.

Modélisation de l'application

L'application, dans un système interactif, doit se comporter comme un élément actif. Dans les systèmes actuels, la modélisation de l'application présente deux types de défaut majeurs : soit les modèles de tâche s'avèrent trop figés (plans dans les systèmes de transfert d'informations), contraignant ainsi trop fortement l'initiative de l'utilisateur, soit ils sont fondés sur des contraintes (comme dans les applications de CAO), ce qui permet une activité de l'utilisateur plus libre mais peut aussi entraîner un manque de coopérativité pour l'aider s'il ne connaît pas la suite d'actions à accomplir pour atteindre son but. Nous pensons que la modélisation de l'application doit comporter les éléments suivants : les données et leur ontologie, les connaissances sur l'utilisation des données (modes opératoires) et l'interface avec le reste du système. Enfin, cette modélisation doit être envisagée de façon à pouvoir changer facilement d'application.

3.3 Systèmes et multimodalité

Mots clés : multimodalité, référence, synthèse de parole.

Résumé : *Pour pallier certains des problèmes dûs à l'utilisation de la parole, nous étudions une modalité supplémentaire de communication, un écran tactile. Les problèmes à traiter concernent l'intégration des messages provenant des différents canaux, le traitement de la référence ainsi que l'évaluation des systèmes.*

L'utilisation des techniques vocales actuelles dans les systèmes interactifs a pour conséquence l'apparition de nouveaux problèmes et difficultés qui vont de la réalisation de logiciels complets (y compris la recherche de l'application) à la spécification d'architecture, en passant par l'amélioration de la synthèse de parole et l'introduction de la multimodalité.

La communication entre personnes est rarement monomodale : le geste et la parole sont souvent utilisés conjointement pour des raisons fonctionnelles (désignation d'éléments, fiabilité de la communication). Dans un environnement de parole, l'introduction d'une modalité supplémentaire – le geste par l'intermédiaire d'un écran tactile dans notre cas – est d'autant plus intéressante qu'elle permet de pallier les erreurs de reconnaissance de parole.

Cette adjonction fait surgir très rapidement de nouvelles difficultés. La première concerne la façon dont doivent être traitées les informations qui proviennent des différents canaux de communication : leur intégration doit elle se faire à un niveau syntaxique, sémantique ou pragmatique ? Quel type de modélisation faut-il utiliser ? Il existe encore peu de réponses satisfaisantes à ces questions. Nous avons choisi de nous appuyer sur les travaux de M. Maybury [May90] destinés à un contexte différent (production d'actes communicatifs en sortie d'un système). Maybury propose plusieurs niveaux d'actes de communication (avatars d'actes de langage), ce qui permet d'intégrer à chaque niveau des informations provenant de modalités différentes. Nous reprenons ce principe (qui est cohérent avec notre modélisation du dialogue) mais en reconnaissance d'actes : les modalités tactile et parole sont traitées séparément pour fournir des actes communicatifs qui sont ensuite fusionnés pour donner des actes de langage.

La seconde difficulté porte sur le traitement de la référence plus particulièrement dans le cadre de l'application choisie (interrogation d'une base de données géographiques et touristiques). La désignation des objets intéressants du dialogue s'effectue à l'aide du langage et du geste (pointé et tracé de zones) et tient compte du contexte applicatif (l'utilisateur peut suivre un contour d'un objet cartographique).

Les études dans ce domaine se font en linguistique et en intelligence artificielle (représentation des connaissances). Certains linguistes [Van86] proposent des études très fines sur les conditions d'utilisation (approche fonctionnelle) des prépositions utilisées dans la désignation des objets. Nous pensons qu'il s'agit là de résultats intéressants que nous avons adaptés pour notre traitement syntaxique des énoncés. Du côté de l'intelligence artificielle, plusieurs modélisations des relations spatiales ont été proposées. Nous utilisons celle suggérée par l'Irit (Toulouse) pour vérifier la cohérence sémantique des expressions référentielles dans le cadre de notre application. Ce modèle est fondé sur certaines caractéristiques (dimension, morphologie,...) des éléments qui régissent l'utilisation des termes linguistiques dans les expressions référentielles (par exemple, le mot bord ne peut être utilisé qu'associé à un objet possédant deux dimensions).

L'ambition de mettre sur le marché des systèmes de dialogue doit s'accompagner d'exigences sur la qualité de l'interaction. Il faut pouvoir évaluer et comparer, dans le cadre d'applications équivalentes, différents systèmes selon plusieurs points de vue (performances de la reconnaissance, efficacité du dialogue, capacités dialogiques et langagières,...) et éventuellement pour un même système, évaluer des approches différentes. Un certain nombre de métriques ont déjà été proposées[Sun93,CS94] (exemple : longueur du dialogue, nombre de tours de parole pour la récupération d'erreurs de reconnaissance,...) mais elles ne rendent pas compte de toutes les dimensions d'un système interactif. De nouvelles pistes sont actuellement envisagées dans la communauté scientifique : elles sont fondées (comme au laboratoire Clips de Grenoble) sur des aspects pragmatiques tels que la pertinence, ou bien sur un concept d'autoévaluation de système (pour tester ces différentes capacités) qui consiste à faire traiter par le système ou par une partie de celui-ci, des fragments de dialogue qui présentent les particularités à tester tout en fournissant tous les éléments contextuels nécessaires.

3.4 Apprentissage et traitement de l'information textuelle et sonore

Mots clés : apprentissage automatique, inférence grammaticale, analyse syntaxique, synthèse de la parole, base de données sonores.

[May90] M. MAYBURY, « Communicative Acts for Explanation Generation », *International journal of Man-machine studies 37(2)*, 1990, p. 135–172.

[Van86] C. VANDELOISE, *L'espace en français*, Éditions du seuil, Paris, 1986.

[Sun93] SUNDIAL, « SUNDIAL, Prototype performance evaluation report », *Deliverable n° D3WP8*, projet Sundial P2218, septembre 1993.

[CS94] A. COZANNET, J. SIROUX, « Strategies for Oral Dialogue Control », *in : Proceedings of International Conference on Spoken Language Processing (ICSLP) 94, 2*, p. 963–966, Yokohama, Japon, 1994.

Résumé : *Cette étude a pour objectif d'élaborer des techniques d'apprentissage pour améliorer les parties amont et aval du dialogue oral : le traitement de requêtes orales ou écrites et la synthèse de la parole.*

Apprentissage d'outils syntaxiques

Dans la partie amont d'un système de dialogue oral, la parole est traitée par un outil de reconnaissance qui produit en général un *treillis* de mots, c'est-à-dire le tableau des hypothèses lexicales entre deux instants de la phrase. Il faut ensuite utiliser une syntaxe pour en extraire une suite unique de mots, celle dont la vraisemblance à la fois acoustique et syntaxique est la plus forte.

Dans le cas d'une entrée écrite, on dispose d'une unique séquence de mots. Cependant, le fait que chacun de ces mots puisse en général appartenir à plusieurs catégories grammaticales oblige à une *désambiguïsation* par analyse syntaxique.

Cette analyse syntaxique est effectuée en général soit par un modèle formel fourni a priori par le concepteur du système, soit par un modèle statistique simple fondé sur l'enchaînement des classes grammaticales (*bi-gram, n-gram*), dont les paramètres sont fixés par apprentissage sur un corpus.

Il est intéressant de chercher à combiner ces deux approches, en extrayant du corpus d'apprentissage un ensemble de règles de grammaire (éventuellement probabilisées) : on profite alors des avantages de la seconde («coller» aux données d'apprentissage) et de la première (autoriser des dépendances à long terme qui reflètent une véritable structure).

Nous cherchons donc à apprendre des structures syntaxiques à partir d'exemples de phrases regroupées dans un corpus d'apprentissage. Nous employons actuellement deux techniques, l'une dans le cadre d'une application de dialogue par message écrits (Minitel et extensions), l'autre pour le dialogue oral (interrogation de l'Annuaire Général des Services).

La première méthode est une extension de la méthode DOP (*Data Oriented Parsing*) [Bod95]. Elle a été conçue pour les syntaxes *context-free* standard ; nous l'étudions pour les syntaxes LFG (*Lexical Functionnal Grammars*), qui sont composées de règles *context-free* accompagnées de contraintes écrites sous forme logique.

La seconde est l'emploi des techniques d'*inférence grammaticale* [2], permettant d'extraire d'un ensemble de phrases une grammaire régulière, éventuellement probabilisée. Une étape préliminaire d'apprentissage est nécessaire : extraire du corpus des exemples (ici une liste de phrases) un ensemble de catégories grammaticales. Ensuite, il faut adapter des méthodes qui ont été en général développées en dehors d'applications comme celle que nous avons à traiter.

Méthodologies de synthèse de parole

La partie aval d'un système de dialogue oral est constituée d'un générateur de texte produisant une suite de mots correspondant à l'énoncé du message à diffuser. Cet énoncé textuel est ensuite converti en énoncé oral au moyen d'un système de synthèse de la parole à partir du texte.

Dans ce cadre, on peut distinguer plusieurs pistes de recherche permettant d'une part de produire une parole de synthèse la plus naturelle possible et d'autre part de diversifier les styles de voix.

[Bod95] R. BOD, *Enriching Linguistics with Statistics: Performance Models of Natural Language*, thèse de doctorat, Institute for Logic, Language and Computation, University of Amsterdam, 1995, ftp://ftp.fwi.uva.nl/pub/theory/illc/dissertations/DS-95-14.text.ps.gz.

Un premier axe d'étude consiste à remettre en cause les hypothèses de traitement de la matière acoustique dans un système de synthèse de la parole. La plupart des systèmes actuels juxtaposent des unités acoustiques correspondant à des unités linguistiques bien définies (par exemple, des diphones). Pour créer cette matière sonore, nous proposons de rechercher, au fil de la synthèse, des segments acoustiques pris dans une base de parole continue.

Dans cette optique, il n'y a plus de notion d'unités définies a priori sur des critères linguistiques, mais une recherche constante de la meilleure unité à retenir. La recherche des segments se fait au moment de la synthèse en s'appuyant sur un modèle acoustique conduit par la chaîne phonétique du message à prononcer. Les paramètres de ce modèle sont obtenus par apprentissage automatique.

Un second axe d'étude cherche à réduire la distance qu'il y a entre le générateur de texte du système de dialogue et le signal à la sortie du système de synthèse. Nous proposons de traiter par le système de synthèse toutes les informations connues du système de dialogue comme par exemple les informations syntaxiques et grammaticales, les informations sémantiques, voire certaines informations pragmatiques. En tenant compte de ces données, il est alors possible de générer des façons de parler adaptées à certaines situations particulières du système de dialogue (information, répétition, insistance,...).

Enfin, un dernier axe d'étude consiste à diversifier les voix de synthèse tant sur le plan du timbre que sur le plan de l'énonciation en caractérisant ce problème comme un problème de conversion de voix. À partir d'une ou de plusieurs voix de synthèse de référence et à partir de la signature vocale d'une voix cible, les caractéristiques segmentales et prosodiques de la voix de référence sont transformées pour ressembler sur le plan de la perception à celles de la voix cible. Des travaux ont permis d'évaluer une technique de transformation du timbre de la voix par une modélisation probabiliste des espaces acoustiques de la voix de référence et de la voix cible.

l'apprentissage de la langue bretonne

3.5 Technologies vocales pour l'apprentissage de la langue bretonne

Mots clés : technologie de la parole, interface graphique, logiciels éducatifs, aide à l'apprentissage des langues, prosodie, enseignement de la langue bretonne, dictionnaires, évaluation, linguistique.

Résumé : *Cette étude concerne l'adaptation et la mise au point de techniques, développées en reconnaissance, en synthèse et en analyse de la parole, pour les appliquer à la langue bretonne et pour les intégrer dans des logiciels éducatifs d'aide à l'apprentissage et à l'enseignement du breton. La réalisation d'un système de synthèse pour la langue bretonne et son utilisation dans des applications pédagogiques, la conception et le développement d'exercices d'apprentissage de la prononciation (production, perception) et de la prosodie où les technologies de la parole peuvent apporter un complément primordial aux outils existants, constituent les objectifs essentiels de cette recherche.*

Les technologies de la parole (reconnaissance, synthèse, analyse, transformation, visualisation de la parole) ont atteint un niveau de maturité acceptable et sont maintenant utilisées dans de nombreux domaines d'application. Les ordinateurs peuvent entendre et parler, ce qui ouvre des perspectives prometteuses pour l'apprentissage des langues, assisté par ordinateur. L'introduction de ces techniques dans le domaine de l'aide à l'enseignement des langues a commencé, il y a une décennie environ, avec l'apparition des premiers Cédérom d'Auralog [Aur] utilisant des logiciels de reconnaissance de la parole ou encore avec la sortie de la technologie MacinTalk d'Apple utilisant la synthèse de la parole. Depuis lors, les recherches et les développements dans ce domaine progressent régulièrement ; les faits marquants les plus récents sont les produits commerciaux tels que "Tell me more" ou "Dyned" qui utilisent les techniques de reconnaissance pour évaluer la prononciation des élèves, mais aussi de nouveaux prototypes d'aide à l'apprentissage des langues incluant divers outils tels que la reconnaissance, la synthèse,

[Aur] « CD ROM language learning software based on speech recognition », `http://www.auralog.com`.

l'animation 3D de têtes parlantes artificielles [Ca98] ou encore des outils d'évaluation de la prosodie de l'apprenant (système Slim [Del] par exemple). On peut signaler également le projet européen Spell[H+93] et plus récemment l'existence de séminaires spécialisés.

Cependant les technologies de la parole ne sont pas encore optimales et pour les langues régionales ou minoritaires telles que le breton, pour lesquelles on ne dipose pas de corpus de parole spécialisés et correctement étiquetés à tous les niveaux linguistiques (condition nécessaire pour l'obtention de bons outils), il reste encore beaucoup à faire. Les techniques de reconnaissance de la parole sont insensibles à la fréquence fondamentale, à l'amplitude, aux détails de durée. Elles ne sont sensibles qu'aux aspects segmentaux (phonétiques) et non aux aspects suprasegmentaux et aux informations de nature prosodique qui sont très importants pour l'apprentissage d'une langue ; il faut donc modifier les algorithmes pour tenir compte de ces contraintes.

De plus ces systèmes sont destinés à des locuteurs natifs et les modèles sont adaptés à ce type de locuteurs. Les apprenants de langue sont par contre généralement non-natifs. Les technologies de la parole doivent donc être adaptées aux locuteurs non-natifs et à un parler homogène de la langue, pour fournir une évaluation adéquate de la performance de l'apprenant. Pour l'enseignement de la langue bretonne, il faut en même temps développer ou adapter les outils existants de traitement de la parole et construire des exercices ou des outils spécifiques bien adaptés au niveau des élèves. Nous avons donc décidé de procéder par étape, en nous concentrant d'un côté sur l'enseignement de la prosodie et en privilégiant les outils d'analyse et de visualisation de la parole qui permettent à l'élève d'imiter la voix du maître en mettant en relief les paramètres prosodiques et les courbes caractéristiques de l'élève et du maître ; l'autre axe principal de nos recherches concerne la synthèse de la parole et des mots avec comme objectifs la réalisation d'un dictionnaire parlant et l'adaptation du système de dictée par synthèse vocale à la langue bretonne.

Un des thèmes principaux de cette étude a pour cadre l'élaboration de la base d'unités acoustiques pour la synthèse par concaténation d'unités. Dans un premier temps, on a choisi comme unité le diphone mais à court terme il faudra passer à des unités plus longues, comme la syllabe et le mot, et choisir l'unité optimale, pour améliorer la qualité segmentale de la synthèse.

Le second thème de recherche concerne la conversion orthographique / phonémique, avec comme approches privilégiées, l'approche lexicale pour la synthèse des mots du dictionnaire et l'utilisation de règles de transcription pour la synthèse de phrases. Ces deux approches doivent être combinées et complétées par une analyse grammaticale.

Le troisième volet, lié au naturel et à la qualité de la synthèse concerne l'analyse et la modélisation prosodique. Il existe plusieurs techniques pour calculer les paramètres prosodiques et pour générer le contour intonatif de la phrase. L'approche choisie est une approche par règles dont le rôle est de calculer la durée des phonèmes et les valeurs de la fréquence fondamentale F_0.

4 Domaines d'applications

Résumé : *Les domaines d'application du projet sont nombreux. Ce sont essentiellement des domaines d'activité humaine réunissant plusieurs personnes, où la communication orale joue un rôle important. En particulier, on peut citer les activités telles que les services d'informations, de réservation par téléphone et également les industries de la langue avec en particulier l'enseignement assisté par ordinateur. La motivation du projet est de fournir*

[Ca98] R. COLE, AL., « Intelligent animated agents for interactive language training », *in : STILL*, p. 163–166, Marholmen, Sweden, 1998.

[Del] R. DELMONTE, « A prosodic module for self-learning activities », *in : MATISSE*, p. 129–132, University College, London.

[H+93] S. HILLER *et al.*, « SPELL, an automated system for computer-aided pronunciation teaching », *Speech Communication 13*, 1993, p. 463–473.

un substitut automatique à l'être humain dans cette tâche d'assistance lorsque le travail à accomplir est particulièrement contraignant.

Les domaines d'application spécifiques des techniques de traitement de la parole sont ceux où l'usage de l'oral apporte un confort d'utilisation incontestable (télématique vocale, domotique) ou bien s'avère indispensable (consultation de bases de données par téléphone, autres moyens de communications inopérants ou occupés). L'usage de la langue naturelle dans la communication orale permet aussi de placer les applications dans le domaine des industries de la langue. Les logiciels et produits que nous développons se situent dans ce domaine et plus particulièrement dans le sous-domaine des logiciels éducatifs. Les techniques et savoir-faire utilisés nous permettent d'envisager de nouvelles applications proches ; par exemple, le logiciel Ordictée (*cf 5.1*) pourra être utilisé pour l'apprentissage de la langue bretonne, ou bien encore une partie du logiciel de synthèse du breton sera utilisable pour une autre langue.

Les techniques d'apprentissage automatique peuvent être utilisées pour le développement d'interfaces homme-machine pour les applications utilisant le langage naturel. Dans ce type d'application, la variabilité des énoncés des utilisateurs ainsi que la nécessaire robustesse de la compréhension du système rendent pertinent le recours à des techniques d'ingénierie de conception à base d'apprentissage.

5 Logiciels

Résumé : *Nos activités ont permis la production de logiciels qui concernent deux types d'usages différents : didactique et support pour le développement. Pour le premier usage, deux logiciels ont été développés :* Ordictée, *pour l'apprentissage de l'orthographe et* Ar geriadur a gomz *qui est un dictionnaire vocal français-breton. Pour le second usage, nous avons produit* Epigram *qui est un ensemble de fonctions relatives à l'inférence grammaticale permettant le développement d'applications.*

5.1 Ordictée

Participant : Marc GUYOMARD[*correspondant*].

Ordictée est un logiciel destiné à l'apprentissage de l'orthographe utilisant la synthèse de parole. Il permet à l'élève de réaliser une dictée de manière autonome et dans un environnement non stressant. Le système joue le rôle de l'instituteur, il dicte le texte à l'élève puis corrige le texte tapé en présentant à l'élève les fautes commises. Le logiciel est constitué de trois modules : le module «élève» qui effectue la dictée proprement dite, le module «tuteur» qui permet à l'instituteur de créer ses propres dictées et le module «concepteur» qui permet la gestion de l'ensemble.

Le logiciel est enregistré à l'APP (Agence pour la Protection de Programmes) par l'université de Rennes 1 sous le numéro IDDN.FR.001.070006.01.S.P.1996.000.21100. Le guide d'installation et d'utilisation (version 1.1 1996) est disponible auprès de M. Guyomard.

5.2 Dictionnaire vocal

Participant : Guy MERCIER[*correspondant*].

Un dictionnaire vocal (Ar geriadur a gomz) français-breton et breton-français a été réalisé en partenariat avec la maison d'édition Tes de Saint-Brieuc et gravé sur Cédérom. C'est un dictionnaire bilingue multimédia intégrant texte, sons et images animées, à usage pédagogique (copyright CRDP de Bretagne/Tes).

La version 1.0 de ce Cédérom a été distribuée aux classes des écoles enseignant le breton ou en breton (5500 élèves en Bretagne) et a été commercialisée par l'association *Skol Vreizh* en décembre 1998.

Le logiciel de consultation appelé Gervogal permet non seulement de rechercher un mot, d'obtenir sa signification et sa traduction dans l'autre langue, mais encore de visualiser les catégories grammaticales, les transcriptions phonétiques, les variantes de prononciation proposées pour chaque entrée de la partie *breton-français*. On peut entendre chaque mot et ses différentes prononciations. Le logiciel offre en plus la possibilité de retrouver des mots ayant plusieurs variantes orthographiques et il peut proposer des listes de mots répondant à certains critères de sélection (débuts ou fins de mots identiques, par exemple). Ce logiciel est enregistré à l'APP par l'université de Rennes 1.

5.3 Epigram

Participant : Laurent MICLET [*correspondant*].

Afin d'accélérer et uniformiser le développement d'applications pour l'inférence grammaticale, nous avons développé un outil sous forme d'une bibliothèque de classes C++. Cette bibliothèque de fonctionnalités de haut niveau a été conçue selon deux principes : la généricité et l'indépendance de la stratégie d'implantation. La première permet à l'utilisateur de la bibliothèque d'adapter facilement des fonctions implantées à son problème, et de raccourcir ainsi le temps de programmation. La seconde laisse libre le choix du style de programmation (statique, dynamique) tout en fixant un cadre formel par le biais de mécanismes des classes abstraites.

Cet outil, appelé Epigram (Environnement de Programmation pour l'Inférence GRAMmaticale), a été développé au cours de l'année 1997. Actuellement il possède deux implémentations : dynamique — en tant qu'une surcouche de la bibliothèque C++ Leda — (bibliothèque de types de données et d'algorithmes de programmation combinatoire, développée à Max-Planck-Institut für Informatik, Saarbruck, Allemagne) et statique (développée par J. Chodorowski). Tous les programmes produits par l'équipe dans le cadre de recherches ayant pour thème l'inférence grammaticale sont désormais implantés dans *Epigram*. Le développement de cette bibliothèque est poursuivi en collaboration avec l'équipe de l'université de Saint-Étienne (Colin de la Higuera, Franck Thollard) ainsi qu'avec le projet Aida (Jacques Nicolas, François Coste) de l'Irisa.

6 Résultats nouveaux

6.1 Dialogue et modélisation

Résumé : *Deux études concernant les principes fondamentaux d'un module de gestion de dialogue ont été menées. La première concerne les mécanismes d'implication qui sont utilisés par les raisonnements sur les connaissances épistémiques et doxastiques du système. La seconde porte sur une reformulation d'un modèle de dialogue par plans par l'incorporation des postulats de traitement de phénomènes dialogiques supplémentaires. Enfin, quelques travaux ont été effectués sur le problème de l'évaluation de système de dialogue.*

Amélioration d'un moteur modal de dialogue

Participants : Mouloud KHAROUNE, Pierre NERZIC.

Nous avons commencé à étudier les améliorations qu'il est possible d'apporter au moteur de dialogue proposé par P. Bretier [Bre95], sur les idées de D. Sadek [Sad91]. Ce moteur effectue des raisonnements sur des formules prédicatives contenant des opérateurs modaux représentant par exemple les connaissances et les intentions des participants. Ce moteur est constitué de deux mécanismes : la résolution, comparable à celle des prédicats, est adaptée pour le traitement des opérateurs modaux et l'instanciation de schémas. Ces deux mécanismes ont pour objet de produire de nouvelles formules (résolvantes et instances de schémas) à partir d'un jeu initial (croyances sur le monde, le dialogue,...).

La réalisation que P. Bretier a proposée dans sa thèse souffre de plusieurs problèmes théoriques liés notamment au langage choisi pour l'implémentation, Prolog. Certaines unifications de termes modaux sont permises par Prolog mais sont interdites par la logique. Il semble également que, dans ces cas, les optimisations proposées par Bretier ne soient pas toujours valides. Notre équipe travaille sur une meilleure définition de ce mécanisme permettant de lui conserver ses qualités.

L'équipe a également formé le projet d'intégrer à ce moteur de dialogue les propositions de P. Nerzic sur la reconnaissance des plans invalides dans le cadre du modèle de H. Kautz [Kau91]. L'ensemble de ces travaux devrait permettre d'arriver à un système de dialogue bien fondé sur un plan théorique et possédant des qualités de robustesse avérées.

Reformulation du modèle par plans

Participants : Marc GUYOMARD, Jean-Christophe PETTIER.

Suite à une réflexion menée sur une extension multi-agents du modèle par plans de Litman, il est apparu nécessaire d'étendre et de remanier ce formalisme pour permettre la génération directe d'applications opérationnelles. En effet, des aspects dialogiques fondamentaux (référenciation, coopération, introspection) n'y sont pas traités explicitement et certains problèmes posés par le calcul sont éludés, ce qui nécessite des interventions spécifiques à chaque tâche. Notre approche vise à préciser le formalisme, y introduire des extensions permettant d'accroître la diversité des énoncés traités tout en garantissant un niveau minimal de consistance des informations.

Parmi les différentes facettes d'une approche par plans, la modélisation de l'action a été abordée prioritairement car celle-ci détermine pour une bonne part les possibilités de l'interaction dialogique. Une analyse détaillée des différentes conditions d'exécution dans le contexte spécifique du dialogue a révélé l'existence de situations variées sous les dénominations habituelles de préconditions et contraintes. Une typologie a alors été proposée afin que le concepteur d'applications ne puisse éluder les questions essentielles sur la nature des conditions et par conséquent soit amené à réfléchir sur la cohérence de la modélisation de l'application. Dans la même perspective de vérification statique, le formalisme proposé permet de s'assurer de la présence de l'information requise par l'exécution. Enfin, il distingue l'information nécessaire à la modélisation du suivi de la tâche de celle pertinente pour les échanges verbaux avec l'utilisateur. Globalement, ce formalisme dote chacune des rubriques de l'action d'une sémantique non équivoque.

[Bre95] P. BRETIER, *La communication orale coopérative : contribution à la modélisation logique et à la mise en oeuvre d'un agent rationnel dialoguant*, thèse de doctorat, Université de Paris Nord, 1995.

[Sad91] D. SADEK, *Attitudes mentales et interaction rationnelle : vers une théorie formelle de la communication*, thèse de doctorat, Université de Rennes 1, 1991.

[Kau91] H. KAUTZ, *Reasoning about plans*, Morgan Kaufmann, 1991, ch. A Formal Theory of Plan Recognition and its Implementation.

Évaluation de systèmes de dialogue

Participant : Jacques SIROUX.

Une première transcription des dialogues de test sur Géoral Tactile (*cf 6.2*) a été effectuée. Elle a permis de mettre en évidence quelques phénomènes de désignation que nous n'avions pas encore rencontrés et qui ont été pris en compte dans les études pour faire évoluer le système.

L'étude du corpus enregistré à Grenoble dans le cadre du contrat avec l'Aupelf-Uref n'a pas encore fourni de résultats intéressants. Une autre transcription, plus riche, va être effectuée au Valoria (Vannes) et nous permettra de faire de nouvelles observations.

En ce qui concerne l'évaluation des systèmes de dialogue, le groupe de travail (Irit, Clips, Limsi, Valoria) auquel nous participons a peu progressé du fait de problèmes administratifs. Nous avons mené une étude sur la méthode DCR (Demande, Contrôle, Résultat) proposée par le Clips et le Valoria. Cette méthode, prédictive et générique, s'intéresse à l'évaluation de la compréhension des systèmes et propose différents niveaux d'analyse. Nous avons montré qu'il était encore nécessaire de préciser plus finement certains points méthodologiques pour l'application de la méthode. D'autre part, l'extension de la méthode à la partie dialogique des systèmes reste encore à explorer.

6.2 Systèmes et multimodalité

Résumé : *Une partie des travaux a porté sur le traitement de la parole (reconnaissance, recherche dans le treillis lexical). L'étude des phénomènes de référence pour une version enrichie de Géoral Tactile a été menée. Enfin, nous avons poursuivi des activités de développement pour l'amélioration du logiciel Ordictée (prise en compte des fautes d'origine phonétique, suivi de la frappe).*

Géoral tactile et référence

Participants : Marc GUYOMARD, Jacques SIROUX.

Le changement en cours du reconnaisseur de parole (remplacement de la carte Media50 par le logiciel HTK) nous permet d'envisager de nouveaux développements d'envergure dans le système de dialogue *Géoral Tactile*[S+97]. L'accroissement du nombre de mots du vocabulaire donnera la possibilité aux utilisateurs de formuler des phrases linguistiquement plus complexes. Nous utilisons ce fait pour enrichir l'univers de l'application en ajoutant de nouveaux éléments sur la carte qui sert de support aux interrogations. Dans ce nouveau cadre, plusieurs points sont à étudier : une modélisation du contexte cartographique, les comportements linguistiques et gestuels des utilisateurs pour désigner les éléments sur la carte et enfin l'organisation même du système.

Dans un premier temps, une expérimentation a été menée afin de déterminer le comportement langagier des utilisateurs dans leur activité de désignation des éléments sur la carte. Un grand nombre de formes linguistiques ainsi que l'utilisation d'éléments construits (par exemple désignation d'un triangle à l'aide de points particuliers) ont été observés. Une nouvelle forme de geste (suivi d'une ligne) est également apparue.

[S+97] J. SIROUX *et al.,* « Multimodal References in Georal Tactile », *in : Proceedings of the workshop Referring Phenomena in a multimedia Context and their Computational Treatment, SIGMEDIA and ACL/EACL*, p. 39–44, Madrid, juillet 1997.

Nous proposons un modèle syntaxique pour analyser et filtrer les expressions référentielles dans les énoncés des utilisateurs. Ce modèle est fondé sur les travaux de Vandeloise[Van86] et d'A. Borillo[Bor88] qui prennent en considération les caractéristiques spatiales des éléments manipulés. Nous avons ensuite développé un modèle sémantique qui permet de filtrer plus finement les productions de l'analyseur syntaxique. Le modèle est dérivé de celui d'Aurnague [Aur93] qui utilise des propriétés caractéristiques des éléments (telles que la dimension, la consistance, la situation,...). Nous n'utilisons que trois caractéristiques (dimension, consistance et forme) mais de manière combinée, compte tenu des constructions linguistiques possibles.

Du point de vue cartographique, nous avons développé un nouveau modèle de données ainsi que des algorithmes de recherche plus adaptés aux éléments manipulés.

Enfin, le fait de traiter en plusieurs phases des énoncés plus complexes et celui de voir apparaître des objets qui ne sont pas présents dans la base de données ainsi que le nouveau geste à traiter nous ont amenés à faire évoluer l'architecture du système et la philosophie des traitements. Fondé sur la prééminence des activités gestuelles sur les activités orales (le contraire de ce qui se passe dans la version actuelle), le principe permet de vérifier de manière progressive et éventuellement de corriger les expressions linguistiques référentielles, de déterminer les référents potentiels sur la carte et de construire le cas échéant de nouveaux éléments dans la base. Certains des algorithmes ont été implantés et devront être testés quand le nouveau logiciel de reconnaissance de parole fonctionnera.

Ordictée

Participants : Marc GUYOMARD, Jacques SIROUX.

Ordictée est un logiciel permettant à un élève de réaliser une dictée de manière autonome. Le logiciel est constitué de trois modules : le module «élève» qui effectue la dictée proprement dite, le module «tuteur» qui permet à l'instituteur de créer ses propres dictées et le module «concepteur» qui permet la gestion de l'ensemble. L'ensemble des trois modules a été totalement réécrit afin d'en améliorer la robustesse et l'ergonomie.

Une campagne d'évaluation auprès des élèves d'une classe de l'enseignement primaire (comme pour la première version du système) est prévue.

l'information textuelle et sonore

6.3 Apprentissage et traitement de l'information textuelle et sonore

Résumé : *La méthode Data Oriented Parsing a été particulièrement étudiée pour une utilisation dans le cadre de la théorie* LFG*, notamment sur le traitement des ambiguïtés. Les études sur l'inférence de grammaire pour le dialogue oral ont été poursuivies sur deux axes : l'amélioration de l'efficacité des méthodes n-grams et l'exploitation des méthodes d'exploration de l'espace de recherche. Le travail sur la conversion de la voix a commencé cette année.*

[Van86] C. VANDELOISE, *L'espace en français*, Éditions du seuil, Paris, 1986.

[Bor88] A. BORILLO, « Le lexique de l'espace : les noms et les adjectifs de localisation interne », *Cahiers de grammaire 13*, 1988, p. 1–22.

[Aur93] M. AURNAGUE, *A unified processing of orientation for internal and external localization*, Groupe Langue, Raisonnement, Calcul, Toulouse, 1993.

Analyse syntaxique à base de corpus

Participants : Boris CORMONS, Laurent MICLET.

Ce sujet est traité dans le cadre de la thèse de B. Cormons, soutenue en Mars 1999. L'essentiel du travail a été effectué au Cnet, à Lannion.

Le but de cette thèse était l'élaboration d'un analyseur syntaxique robuste et à large couverture pour le formalisme des grammaires lexicales fonctionnelles. Il a été choisi d'étudier l'application de la méthode Dop (Data Oriented Parsing) aux grammaires fonctionnelles. En effet,

1. l'analyse ne se fait pas à l'aide d'une grammaire mais à l'aide du corpus lui-même et cette manière de faire semble propice à l'obtention d'une large couverture.
2. la méthode *Tree-Dop* donne d'excellents résultats en terme de *désambiguïsation* et on peut donc espérer obtenir des résultats de bonne qualité sur un autre type de corpus.
3. d'autre part, l'instanciation de Dop décrite dans ce travail permet de proposer une analyse raisonnable pour des phrases qui sont habituellement considérées comme incorrectes (fautes d'accord, par exemple), ce qui peut donc être vu comme de la *robustesse*.

Ce type d'approche consiste à définir un modèle probabiliste sur l'ensemble des analyses possibles et considérer que, pour une phrase donnée, l'analyse la plus pertinente est tout simplement la plus probable. Ce sont les *approches statistiques*. De nombreux travaux (voir par exemple Sekine & Grishman [SG95] et Charniak [Cha96a] [Cha96b]) ont été réalisés dans cette direction mais s'adressent principalement à des représentations non contextuelles. On peut toutefois citer Eisner [JE92] qui utilise un modèle probabiliste à historique (*history-based*) mais Bod [Bod95] a montré que ce genre de modèle est inférieur au modèle Dop. Abney [Abn97] propose l'utilisation d'un modèle à base de champs aléatoires et cette approche est reprise par Geman & Johnson et Eisele (en cours de publication).

Cependant, dans les deux cas, on considère que la grammaire est donnée et il s'agit de déterminer parmi les analyses proposées par la grammaire la plus pertinente. L'approche de cette thèse est beaucoup plus ambitieuse puisque nous ne disposons ni d'une grammaire ni d'un lexique et c'est en ceci qu'elle nous semble tout à fait novatrice.

Il a été démontré dans ce travail que le modèle probabiliste proposé par Bod & Kaplan [BK98] ne peut que donner de mauvais résultats sur un corpus de représentations de taille raisonnable et qu'une modification simple de leur modèle donne par contre forcément de bons résultats si le corpus est suffisamment grand. Cette condition est toutefois irréalisable en pratique ; le travail expérimental a consisté à déterminer si les résultats sont acceptables sur un corpus de taille modeste.

Inférence de grammaires pour le dialogue oral

Participants : Jacques CHODOROWSKI, Laurent MICLET.

Ce travail suit deux directions : le développement de nouvelles méthodes pour l'inférence grammaticale et l'adaptation des techniques existantes aux données réelles (fournies par le Cnet).

[SG95] S. SEKINE, R. GRISHMAN, « A Corpus-based Probabilistic Grammar with Only Two Non-terminals », *in : Fourth International Workshop on Parsing technologies*, 1995.

[Cha96a] E. CHARNIAK, « Tree-bank Grammars », *rapport de recherche*, Department of Computer Science, Brown University, Rhode Island, 1996, CS-96-02.

[Cha96b] E. CHARNIAK, « Tree-bank Grammars », *in : AAAI*, 1996.

[JE92] M. A. JONES, J. EISNER, « A Probabilistic Parser Applied to Software Testing Documents », *in : AAAI*, 1992.

[Bod95] R. BOD, *Enriching Linguistics with Statistics: Performance Models of Natural Language*, thèse de doctorat, Institute for Logic, Language and Computation, University of Amsterdam, 1995, ftp://ftp.fwi.uva.nl/pub/theory/illc/dissertations/DS-95-14.text.ps.gz.

[Abn97] S. ABNEY, « Stochastic Attribute-Value Grammars », *Computational Linguistics 23*, 4, décembre 1997.

[BK98] R. BOD, R. KAPLAN, « A probabilistic Corpus-Driven Model for Lexical-Functional Analysis », *in : Coling*, 1998.

Le premier axe de travaux de recherche s'est orienté vers une approche encore peu exploitée dans le domaine de l'inférence grammaticale. Il s'agit d'une méthode d'exploration de l'espace de recherche à partir de l'élément le plus général pour générer les solutions de plus en plus spécifiques. Cette méthode conduit normalement vers une explosion combinatoire ; le travail a consisté à trouver des critères qui limitent l'espace de recherche. Ce travail a été finalisé par l'implémentation d'un programme de construction par spécialisation du treillis des solutions. L'amélioration de ces travaux a fait l'objet d'étude théoriques et pratiques en collaboration avec le projet Aïda.

Le deuxième axe a ouvert deux voies d'expérimentation : d'une part pour l'amélioration de l'efficacité des méthodes *n-grams* et des *bi-grams* en particulier, d'autre part pour l'utilisation d'un algorithme d'inférence grammaticale ECGI *(Error Correcting Grammatical Inference)* avec les données mentionnées plus haut. Une première réflexion nous a amenés à considérer, dans les deux cas, les effets de la réduction de la taille du vocabulaire. Pour cela, nous nous sommes tournés vers les méthodes de classification des mots d'un lexique. Étant donné l'ampleur de la tâche et dans le souci de garder la cohérence de la démarche, nous n'avons pris en compte que les méthodes de classification automatique et plus particulièrement la classification automatique hiérarchique.

Au cours de la deuxième année du contrat, nous avons approfondi l'étude de l'algorithme ECGI et de son application dans la tâche réelle de reconnaissance vocale. Les résultats de ce travail ont été synthétisés dans la publication présentée à ICGI'98 [1]. Un travail sur l'évaluation de la qualité des modèles construits uniquement avec les exemples positifs a été également entamé pour valider nos travaux sur le plan théorique. Parallèlement, nous avons mis en place un outil d'étiquetage en vue de son application sur le corpus du Cnet. Cette méthode de classification introduit des connaissances extérieures (linguistiques) et sera ajoutée aux méthodes purement statistiques citées plus haut.

Pendant la troisième et dernière année du contrat (en cours) nous utilisons les résultats théoriques et pratiques obtenus au cours des années précedentes, en les appliquant dans le cadre précis du filtrage syntaxique des treillis accoustiques. Nous avons proposé un paramétrage de l'algorithme d'apprentissage des structures syntaxiques (ECGI) puis l'adaptation de ces structures (grammaires) aux données réelles, notamment le *lissage des grammaires* [2]. Actuellement, nous intégrons le résultat de la catégorisation des mots d'un corpus dans notre modèle afin de pouvoir comparer son efficacité à celle des modèles de langages utilisés couramment dans le processus de la reconnaissance vocale et qui subissent une approche équivalente. En parallèle, nous nous efforçons de réduire la taille des grammaires apprises afin de les rendre compatibles avec les contraintes temporelles du traitement de la langue parlée. Si la réduction du vocabulaire va dans ce sens, ce but peut être probablement atteint par la construction de plusieurs grammaires, plus petites et plus adéquates aux caractéristiques du corpus. Ces travaux vont clore l'ensemble de recherches menées dans le cadre du contrat Cnet.

Conversion de la voix

Participant : Olivier BOËFFARD.

L'utilisation de plus en plus importante des systèmes de synthèse de la parole pose le problème de la diversité des timbres de voix disponibles. Pour des systèmes de dialogue, il est en effet souhaitable de les distinguer par des voix bien distinctes.

Pour fabriquer une phrase de synthèse, une méthode de référence consiste à assembler des unités de parole élémentaires. Ces éléments de parole ont été préalablement enregistrés par un locuteur selon une méthodologie permettant d'atteindre une très bonne qualité de voix synthétique. Cette méthodologie reste cependant relativement lourde. L'alternative à l'enregistrement d'un locuteur consiste à conserver un locuteur de référence unique et à mettre en œuvre des techniques de modification du timbre de la voix et de certains paramètres prosodiques.

On dispose de l'enregistrement acoustique d'un ensemble de phrases couvrant à peu près tous les phones et transitions entre phones pour deux locuteurs A et B (une voix d'homme et une voix de femme). Une fraction des deux bases, le corpus d'apprentissage, sert à apprendre les paramètres des modèles de

transformation de la voix B vers la voix A. L'autre fraction sert à valider la capacité de transformation du modèle pour des phrases non présentes dans le corpus d'apprentissage.

Le modèle de transformation tient compte d'informations segmentales (timbre de la voix et onde glottique) ainsi que de certaines informations prosodiques (durée segmentale et le fondamental).

Nous avons développé une technique de transformation du timbre de la voix en modélisant les espaces acoustiques par des mélanges de gaussiennes. La conversion de voix consiste ensuite à transformer linéairement les paramètres de chacune des gaussiennes du mélange entre les locuteurs A et B. Nous avons montré que la représentation du signal de parole par des paires de lignes spectrales est mieux adaptée qu'une représentation pas des coefficients cepstraux filtrés sur une échelle Mel. Nous avons utilisé le critère Bic (Bayesian Information Criterion) comme critère d'arrêt dans l'estimation du nombre de classes modélisant chacun des espaces acoustiques (nombre de gaussiennes). Nous avons de plus montré que ce critère a aussi un sens dans l'espace euclidien sur lequel sont opérées les transformations linéaires. Le nombre minimal de gaussiennes selon le critère BIC est aussi celui qui minimise la distance euclidienne entre les vecteurs spectraux des locuteurs.

6.4 Intégration des technologies de la parole pour l'enseignement de la langue bretonne

Participants : Guy MERCIER, Jacques SIROUX.

Résumé : *Les activités de 1999 se déclinent selon trois axes : l'amélioration du dictionnaire vocal (amélioration au niveau de l'installation du logiciel, correction de certains défauts), la réalisation d'une première version d'un système de synthèse à partir du texte pour la langue bretonne (intégration des niveaux prosodiques) et l'utilisation des outils* TCL/TK *et Snack pour améliorer l'interface du correcteur de prosodie. Ces travaux de développement s'avèrent nécessaires pour l'obtention d'une plate-forme technologique suffisante pour des recherches plus fondamentales.*

Amélioration du dictionnaire vocal

La première version du cédérom «ar geriadur a gomz» a été distribuée dans les écoles (300 exemplaires) à la fin 1998 et au cours du premier semestre 1999. Ceci nous a permis de collecter oralement et par écrit les avis des enseignants et des élèves sur le plan technique et pédagogique. Ainsi, un certain nombre de bogues (chargement difficile du logiciel, arrêt intempestif,...) et de lacunes ont été signalés. Les principales améliorations ont porté sur la correction de ces "bogues" plus ou moins génants et sur l'amélioration du moteur de recherche de mots. Ainsi parmi les corrections réalisées, on peut citer le logiciel d'aide, le logiciel d'installation et l'insertion de la phonétique de la traduction bretonne de la version français-breton du dictionnaire.

La nouveauté principale concerne la recherche du mot à l'aide d'un «grep» qui nous permet d'intégrer sans difficultés, dans la recherche d'un mot, les variantes orthographiques, les mutations et les caractères «joker». Cette fonction «grep» est maintenant intégrée dans l'application sous la forme d'une DLL, ce qui permet au logiciel de fonctionner sur la plateforme Windows NT. D'autres modifications ayant pour effet de supprimer certains plantages liés à des cas de figures particuliers comme la présence de parenthèses dans l'orthographe des mots ont été introduites dans le logiciel ou dans le dictionnaire.

Grâce à cet ensemble d'améliorations, on peut envisager le gravage d'une nouvelle version du Cédérom, pour la fin de l'année.

Correcteur de prosodie

La première version initiée en 1998 a été entièrement rénovée. Ce logiciel a pour objectif principal l'apprentissage de la prononciation du breton et en particulier l'apprentissage de la prosodie. Le projet comprend deux parties : le correcteur de prosodie et l'interface. Cette année, l'accent a été mis sur l'interface et nous avons choisi le langage script TCL-TK qui est un outil facile à mettre en oeuvre et multiplateformes. Ce langage nous permet de faire communiquer facilement l'interface avec des programmes écrits en C et d'utiliser le module Snack du Laboratoire K.T.H. (Stockhölm) qui gère les fichiers sons et qui fournit un certain nombre d'outils de traitement de la parole.

Comme le logiciel «Ordictée», ce logiciel comporte trois modules de base.

1. Le module «concepteur» qui constitue la base de lancement de l'application et la boîte à outils des autres modules. Il se charge en particulier des travaux sur la parole dont «Snack» ne s'occupe pas, en particulier le calcul de l'énergie, la détection et le calcul du pitch (hauteur) du signal, ainsi que le diagnostic final sur la bonne ou la mauvaise prononciation de l'élève.
2. Le module «création d'exercices» servant à créer des exercices de prononciation. La version actuelle permet au tuteur (enseignant ou concepteur) d'enregistrer des phrases «modèles» ayant une prosodie correcte. Il permet d'afficher, à la demande, le signal, son spectre, la courbe d'évolution du pitch, et la courbe d'évolution de l'énergie. Il comporte des fonctions de «zoom» et d'écoute de portions de signal. Il comprend également un module d'étiquetage phonétique ou syllabique. Il sera ainsi possible de comparer et de visualiser les caractéristiques spectrales ou prosodiques (durée et intonation) des zones correspondant aux mêmes syllabes ou aux mêmes phonèmes du maître et de l'élève. Ce module crée des fichiers qui mémorisent respectivement le signal, les descriptions phonémiques, l'énergie et la courbe de pitch.
3. Le module «élève» constitue la partie visible pour l'utilisateur. Il permet de charger des exercices, de s'enregistrer et d'obtenir un diagnostic de la prononciation. Concrètement, il utilise les fichiers créés par le module «création d'exercices» ; il contient un fichier correspondant à l'enregistrement de l'utilisateur et doit réaliser la comparaison prosodique du signal «maître» et du signal «élève» et fournir un diagnostic. Il reste à écrire les programmes de comparaison prosodique et de diagnostic et à déterminer la forme définitive des interfaces et des exercices après discussion avec les utilisateurs (enseignants et élèves).

Synthèse de la langue bretonne à partir du texte

Après la synthèse vocale des mots réalisée en 1997 et intégrée en 1998 au dictionnaire vocal, la première version du logiciel de synthèse de phrases bretonnes est maintenant opérationnelle. Ce logiciel intègre un module de transcription phonétique du texte à partir d'un fichier de règles qui reste à compléter pour prendre en compte la coarticulation à la frontière entre les mots, un module de génération automatique des paramètres prosodiques (durée des phonèmes et des pauses et valeurs cibles de la fréquence fondamentale des phonèmes voisés) et le module de génération du signal, réalisant l'assemblage et le lissage des diphones de la phrase à synthétiser et les modifications de durée et d'intonation associées au texte à synthétiser.

Le travail réalisé cette année concerne principalement l'intégration de ces modules, la création d'une interface commune et l'écriture de l'ensemble des règles gérant la prosodie du breton. Ce logiciel est une première version et de nombreuses améliorations doivent y être apportées tant au niveau de la modélisation prosodique qu'au niveau de la transcription phonétique en l'intégrant par exemple au dictionnaire vocal. Cependant, on peut déjà envisager de l'intégrer au logiciel Ordictée pour en faire une version adaptée à l'enseignement de la langue bretonne.

7 Contrats industriels (nationaux, européens et internationaux)

7.1 Apprentissage

Un contrat entre l'université de Rennes 1 et France-Télécom (numéro Cnet 97 1B 004) «Inférence grammaticale pour l'apprentissage de la syntaxe en reconnaissance de la parole et dialogue oral» a été notifié le 16 janvier 1997 pour une durée de trois ans. Il comporte en particulier le financement de la thèse de Jacques Chodorowski. Les principaux objectifs sont rappelés dans le paragraphe 3.4.

Un contrat entre l'université de Rennes 1 et France-Télécom, *Synthèse flexible de la parole* a été notifié. Ce contrat comporte le financement de la thèse d'Hélène François. Les objectifs de cette thèse sont rappelés dans le paragraphe 6.3.

8 Actions régionales, nationales et internationales

8.1 Actions régionales

L'ensemble des travaux sur l'intégration des technologies de la parole, pour l'aide à l'apprentissage de la langue bretonne est menée en collaboration avec la maison d'édition Tes, dépendant du rectorat de l'académie et subventionnée par l'Etat, la Région et le Conseil général, et avec l'université de Rennes II. L'apport principal de Tes concerne les aspects pédagogiques et l'université de Rennes II apporte son savoir-faire en linguistique de la langue bretonne (phonétique, prosodie, lexiques).

Pendant l'année scolaire 1998-1999, le dictionnaire vocal «Ar geriadur a gomz» a été distribué gratuitement dans les établissements scolaires assurant un enseignement du breton. L'éditeur «Skol Vreizh» a assuré l'édition commerciale du Cédérom, disponible en librairie pour un prix de l'ordre de 225 F. Compte tenu des diverses remarques et des propositions d'amélioration du produit, une nouvelle version sera proposée l'année prochaine.

8.2 Actions nationales

Le contrat Cnet cité dans le paragraphe 7.1 comporte une collaboration avec l'université de Saint-Étienne. Le projet Aïda est également officiellement partie prenante à ce contrat, sans financement de thèse.

Les travaux de recherche décrits dans le paragraphe 6.3 dépendent d'un projet du RNRT (Réseau National de Recherche en Télécommunications) labelisé en juillet 99 et dénommé Syrus. Cinq partenaires participent à ce projet : Météo-France, France-Télécom, Elan Informatique, Irisa (université de Rennes I). L'Irisa est leader du projet.

8.3 Réseaux et groupes de travail internationaux

L'action d'évaluation de systèmes de dialogue est soutenue par une convention avec l'Aupelf-Uref (Arc n°X/7 1004) ; des laboratoires français et canadiens font partie de cette action de recherche concertée.

Le projet Cordial fait partie du réseau d'excellence européen Elsnet (linguistique informatique) ainsi que du réseau francophone Francil (linguistique).

9 Diffusion de résultats

9.1 Animation de la communauté scientifique

Jacques Siroux et Marc Guyomard ont fait partie des comités de programmes et des relecteurs des ateliers et congrès suivants : Méthodes hybrides TALN/TAP, "la langue naturelle dans l'interaction personne-machine", Recital99. Jacques Siroux fait partie du comité de lecture de la revue InCognito.

9.2 Enseignement universitaire

O. Boëffard enseigne le cours de synthèse de la parole dans le séminaire Parole du DEA Stir, Rennes 1.

M. Guyomard et J. Siroux enseignent dans l'option *Dialogue oral* du DEA informatique de l'Ifsic, Rennes 1.

M. Guyomard est responsable de l'option *Dialogue oral* du DEA informatique et du DESS Isa de l'Ifsic.

M. Guyomard, G. Mercier et J. Siroux enseignent le module *Communication homme-machine*, en troisième année (option LSI) de l'Enssat Lannion.

L. Miclet enseigne le cours de *Reconnaissance des Formes* dans le séminaire Parole du DEA Stir, Rennes 1, une partie du module *Classification et Apprentissage* dans le tronc commun du DEA Informatique de Rennes I, un bloc *Apprentissage Automatique* dans le DEA Miash (ENSTB et Paris IV) et un cours d'*Apprentissage Automatique* en 3ème année de l'Enssat.

9.3 Participation à des colloques, séminaires, invitations

O. Boëffard est conseiller scientifique pour le Cnet en synthèse de la parole à partir du texte.

B. Cormons a été invité quatre mois en fin d'année 1998 au centre de recherches Xerox, Palo Alto, par le Pr. R. Kaplan.

L. Miclet a participé au jury de thèse de B. Cormons, Université de Rennes, mars 1999, et à celui de H. Zaragoza (comme rapporteur), université Paris VI, juillet 1999.

L. Miclet a été invité à donner un tutoriel, en collaboration avec C. de la Higuera, au congrès International Conference on Machine Learning, juin 1999, à Bled (Slovénie). Titre : *Grammatical Inference : Learning Syntax from Sentences* [3].

J. Siroux et G. Mercier ont participé à la présentation officielle du Cédérom «ar geriadur a gomz», au président de la région bretagne, au recteur d'académie et à la presse, le 25 mars 1998.

Guy Mercier a été invité à présenter une communication au séminaire «Speech Technology applications in Call», lors du congrès Eurocall99, à Besançon (15-18 octobre 1999). Le sujet de la communication était : «Synthèse de la parole en breton - Didacticiels pour une langue minoritaire».

J. Siroux a participé au jury de thèse (comme rapporteur) de F. Wolf, université H. Poincaré de Nancy (janvier 1999).

9.4 Accueil d'étudiants et de stagiaires

Le projet a accueilli les étudiants suivants :

- Y. Aubry, licence Mime du Mans, synthèse du breton
- D. Bin, 3^{me} année Enssat (projet court), reconnaissance de la parole
- A.-S. Capelle, DEA Stir, conversion de la voix
- P. Kerbaul, 3^{me} année Enssat (projet court), synthèse du breton
- J. Labidurie, 3^{me} année Enssat (projet court), dictionnaire vocal
- D. Le Tacon, 3^{me} année Enssat (projet court), logiciel Ordictée
- G. Mocquard, 2^{me} année Diic, correcteur de prosodie
- P. Parnet, Ifsic, correcteur de prosodie
- D. Seddah, Maîtrise d'informatique linguistique, logiciel d'étiquetage

10 Bibliographie

Ouvrages et articles de référence de l'équipe

[1] P. DUPONT, L. MICLET, E. VIDAL, « What is the search space of the regular inference ? », *in : Grammatical Inference and Applications, Lecture notes in AI 862*, Springer Verlag, septembre 1994.

[2] P. DUPONT, L. MICLET, « L'inférence grammaticale régulière : fondements théoriques et principaux algorithmes », *rapport de recherche n° 3449*, INRIA, juillet 1998.

[3] M. GUYOMARD, P. NERZIC, J. SIROUX, « Plans, métaplans et dialogue », *rapport de recherche n° 1169*, Irisa, septembre 1998.

[4] M. GUYOMARD, J. SIROUX, *Suggestive and Corrective Answers: A Single Mechanism*, North Holland, Amsterdam, 1989.

[5] L. MICLET, *Méthodes Structurelles pour la Reconnaissance des Formes*, Eyrolles, 1986.

[6] P. NERZIC, M. GUYOMARD, J. SIROUX, « Reprise des échecs et erreurs dans le dialogue homme-machine », *Cahiers de linguistique sociale 21*, 1992, p. 35–46.

[7] P. NERZIC, *Erreurs et échecs dans le dialogue oral homme-machine, détection et réparation*, thèse, université de Rennes 1, janvier 1993.

[8] J. SIROUX, M. GUYOMARD, F. MULTON, C. RÉMONDEAU, « Oral and Gestural Activities of the users in the GÉORAL System », *in : Intelligence and Multimodality in Multimedia, Research and Applications*, John Lee (ed), AAAI Press, 1998.

Communications à des congrès, colloques, etc.

[1] J. CHODOROWSKI, L. MICLET, « Applying Grammar Inference in Learning a Language Model for Oral Dialogue », *in : International Colloquium on Grammatical Inference*, p. 102–113, Ames, Iowa, 1998. Springer-Verlag.

[2] J. CHODOROWSKI, L. MICLET, « Apprentissage et Evaluation de Modèles de Langage par des Techniques de Correction d'Erreurs », *in : Actes de 6e conférence annuelle sur le Traitement Automatique des Langues Naturelles*, p. 253–262, Cargèse, Corse, 1999.

[3] C. DE LA HIGUERA, L. MICLET, « Grammatical Inference : Learning Syntax From Sentences », *in : International Congres on Machine Learning*, Bled (Slovenia), 1999. http://www.univ-st-etienne.fr/eurise/gi/gi.html.

[4] G. MERCIER, M. GUYOMARD, J. SIROUX, « Synthèse de la parole en breton - Didacticiels pour une langue minoritaire », *in : Eurocall99*, p. 57–61, Besançon.