

Gestionnaire de dialogue pour un système d'informations à reconnaissance vocale

Sophie Rosset, Lori Lamel
LIMSI-CNRS, 91403 Orsay Cédex
rosset,lamel@limsi.fr

Résumé - Abstract

Dans cet article, nous présentons un gestionnaire de dialogue pour un système de demande d'informations à reconnaissance vocale. Le gestionnaire de dialogue dispose de différentes sources de connaissance, des connaissances *statiques* et des connaissances *dynamiques*. Ces connaissances sont gérées et utilisées par le gestionnaire de dialogue *via* des stratégies. Elles sont mises en œuvre et organisées en fonction des objectifs concernant le système de dialogue et en fonction des choix ergonomiques que nous avons retenus. Le gestionnaire de dialogue utilise un modèle de dialogue fondé sur la détermination de phases et un modèle de la tâche dynamique. Il augmente les possibilités d'adaptation de la stratégie en fonction des historiques et de l'état du dialogue. Ce gestionnaire de dialogue, implémenté et évalué lors de la dernière campagne d'évaluation du projet LE-3 ARISE, a permis une amélioration du taux de succès de dialogue (de 53% à 85%).

Mots-clés : dialogue oral homme-machine, gestionnaire de dialogue, modèle de dialogue, modèle de la tâche.

In this paper we describe the dialog manager of a spoken language dialog system for information retrieval. The dialog manager maintains both static and dynamic knowledge sources, which are used according to the dialog strategies. These strategies have been developed so as to obtain the overall system objectives while taking into consideration the desired ergonomic choices. Dialog management is based on a dialog model which divides the interaction into phases and a dynamic model of the task. This allows the dialog strategy to be adapted as a function of the history and the dialog state. The dialog manager was evaluated in the context of the LE3-ARISE project, where in the last user tests the dialog success improved from 53% to 85%.

Keywords : spoken language dialog, dialog management, dialog model, task model.

1 Introduction

Les principaux modules d'un système de dialogue oral homme-machine sont le système de reconnaissance, l'analyseur de langue naturelle et le gestionnaire de dialogue qui inclut l'accès à la base de données, le générateur de réponses en langue naturelle et le système de synthèse vocale. Nous présentons dans cet article notre gestionnaire de dialogue et son fonctionnement

en prenant pour exemple l'application ARISE (serveur vocal d'informations ferroviaires). Les connaissances qu'il utilise sont également décrites brièvement. Notre objectif était, pour le système de dialogue, de permettre la navigation (changement de thème), la négociation, d'autoriser une liberté d'expression la plus grande possible à l'utilisateur, de détecter et gérer les erreurs. Un autre objectif était également de permettre l'aboutissement de la demande en un minimum de temps avec un nombre d'échanges minimisé. Il nous fallait également tenir compte du fait que différents utilisateurs impliquent également différentes façon de guider les utilisateurs de la part du système. Celui-ci doit donc être *adaptatif*. Le gestionnaire de dialogue occupe une place centrale et a pour objet de gérer l'utilisation des différentes connaissances afin de permettre aux différents modules qui le composent de procéder aux analyses dont ils sont chargés. Afin de remplir les objectifs liés à la souplesse d'utilisation et d'expression, nous avons écarté, pour le modèle de dialogue, les approches orientées-plans (cf. (Guyomard et al. 1993)). Ces approches tendent à considérer le dialogue comme une suite d'actions simples et prévisibles. Or le dialogue, dans la mesure où il procède d'une co-construction et d'une collaboration entre une machine (pour un système de DOHM) et un humain, ne peut être que partiellement planifié, les utilisateurs étant susceptibles de ne pas suivre le plan fixé *a priori*. Notre conception du dialogue et de la gestion de dialogue se rapproche davantage de celle proposée par (Luzzati 1989). Par contre, cette approche a été utilisée pour le modèle de la tâche.

2 Les connaissances

Le gestionnaire de dialogue utilise différentes sources de connaissances. La distinction entre connaissances statiques et connaissances dynamiques est classique dans le domaine du DOHM. Les connaissances statiques sont composées des connaissances linguistiques et pragmatiques qui sont utilisées lors l'interprétation et la génération. Il s'agit d'un ensemble fini de connaissances, qui n'augmente pas au cours du dialogue. Elles regroupent notamment les modélisations du domaine, de la tâche et du dialogue ainsi que des connaissances linguistiques et pragmatiques. Une *représentation du domaine* est élaborée sous forme de réseau sémantique. Cette représentation permet de constituer un lexique comprenant des entrées propres à l'application et des entrées générales. Ces entrées sont des objets ou des modificateurs d'objets correspondant à un ensemble de mots ou d'expressions de la langue (par exemple, le modifieur **après* correspond à *plus tard, après, trop tôt...*, l'objet **carrissimo* correspond aux mots ou expressions tels que *carrissimo, réduction carrissimo, carte jeune, carré jeune, moins de 25 ans...*). Ces objets et modificateurs sont proches des segments conceptuels tels que les définit et utilise (Pérennou 1996). À partir de cette représentation du domaine, les différentes fonctionnalités (c'est-à-dire actions dans le cadre de l'application considérée, par exemple, pour ARISE la réservation, l'achat d'un billet..) sont définies. Ces fonctionnalités s'appliquent à un ensemble d'objets. Chaque fonctionnalité peut être structurée en sous-fonctionnalités. Les sous-fonctionnalités, de même que les objets, ne sont pas affiliées à une fonctionnalité donnée mais à l'ensemble de l'application. Le *modèle de la tâche* est constitué de l'ensemble de ces fonctionnalités et sous-fonctionnalités et des liens reliant les différentes (sous-)fonctionnalités et les objets du domaines. Il se caractérise par une représentation sur trois niveaux. Chaque niveau correspond à un groupe d'actions possibles dans le cadre de l'interaction (négociation, navigation et nouvelle demande). Le *modèle de dialogue* utilisé est fondé sur la détection de phases. Chaque dialogue est constitué d'une succession de phases qui peuvent être enchâssées les unes dans les autres. Une phase correspond à l'action qui est en cours dans une (sous-)fonctionnalité donnée (p. ex. la négociation dans la recherche d'un horaire). La phase peut avoir pour origine l'utilisateur ou le gestionnaire de dia-

Gestionnaire de dialogue pour un système d'informations à reconnaissance vocale

logue. Nous distinguons les phases d'*acquisition* (acquisition d'éléments), de *négociation*, de *navigation*, de *post-acceptation* (acceptation ou refus d'une proposition) et de *méta-traitement* (repérage et traitements des erreurs). Ces phases sont des états à partir desquels le gestionnaire peut prendre des décisions, et qui vont agir sur les noeuds de transitions du modèle de la tâche. L'identification de ces phases permet au gestionnaire de dialogue de désambigüiser des énoncés en se fondant sur le modèle et surtout l'état de la tâche (par exemple, l'énoncé *carrissimo* sera interprété comme une *question* ou une *contrainte* selon que la phase identifiée sera *acquisition* ou *négociation* ou encore *navigation*).

Les connaissances dynamiques sont composées d'un ensemble de connaissances qui sont modifiées et augmentées au cours du dialogue. Il s'agit des historiques du dialogue (historique des paires d'énoncés système-utilisateur, historique des fonctionnalités ouvertes...), d'un historique des évaluations de l'état du dialogue et du parcours dans la tâche. Certains sont donc liés à l'interaction et aux échanges et d'autres sont plus spécifiques à la tâche et à l'état du dialogue. L'*état de la tâche* peut être comparé à une sorte de pointeur sur le réseau de fonctionnalités. Il regroupe l'ensemble des chemins parcourus, probables et improbables. Il permet une certaine prédictivité pour l'interprétation des énoncés à venir. L'*état du dialogue* est caractérisé par un score que le gestionnaire de dialogue attribue pour *qualifier* le déroulement de l'interaction. Il est évalué localement (pour chaque fonctionnalité en cours) et globalement (pour l'ensemble de l'interaction en cours). Pour cela, le gestionnaire de dialogue utilise différents indices qui sont le *temps effectif* (le nombre d'échanges effectués à l'instant considéré), un *temps maximal théorique* (correspondant au nombre d'échanges maximal nécessaire pour obtenir une information et différent selon la fonctionnalité envisagée) et enfin une représentation du déroulement du dialogue à l'aide de deux valeurs (une valeur P sur un axe *progressif* instanciée à chaque fois que le dialogue progresse dans le "bon" sens et une valeur A sur un axe *accidentel* qui est instancié à chaque fois qu'une erreur est repérée). Cette représentation est proche de celle proposée par (Luzzati 1989) et utilisée par (Lehuen 1997) et (Lemeunier 2000). Ces connaissances dynamiques aident le gestionnaire de dialogue à s'adapter dynamiquement à la situation en cours.

3 Le fonctionnement

Dans cette section nous présentons le fonctionnement du gestionnaire de dialogue et l'utilisation qu'il fait des différentes connaissances. Nous décrivons les différents traitements que subit un énoncé, une fois celui-ci transformé en suite de mots par le système de reconnaissance, jusqu'au schéma sémantique que le gestionnaire de dialogue transmet au module de génération. La figure 1 illustre les différents traitements effectués par le gestionnaire de dialogue.

Le module de reconnaissance envoie une suite de mots au module d'interprétation littérale qui effectue une analyse hors contexte de cette suite de mots et produit un schéma sémantique. Le gestionnaire de dialogue traite ce schéma sémantique et l'incorpore au schéma dialogique. Le traitement est effectué en plusieurs étapes, par différents sous-modules, et abouti à la génération d'un nouveau schéma sémantique pour le module de génération de réponse en langue naturelle.

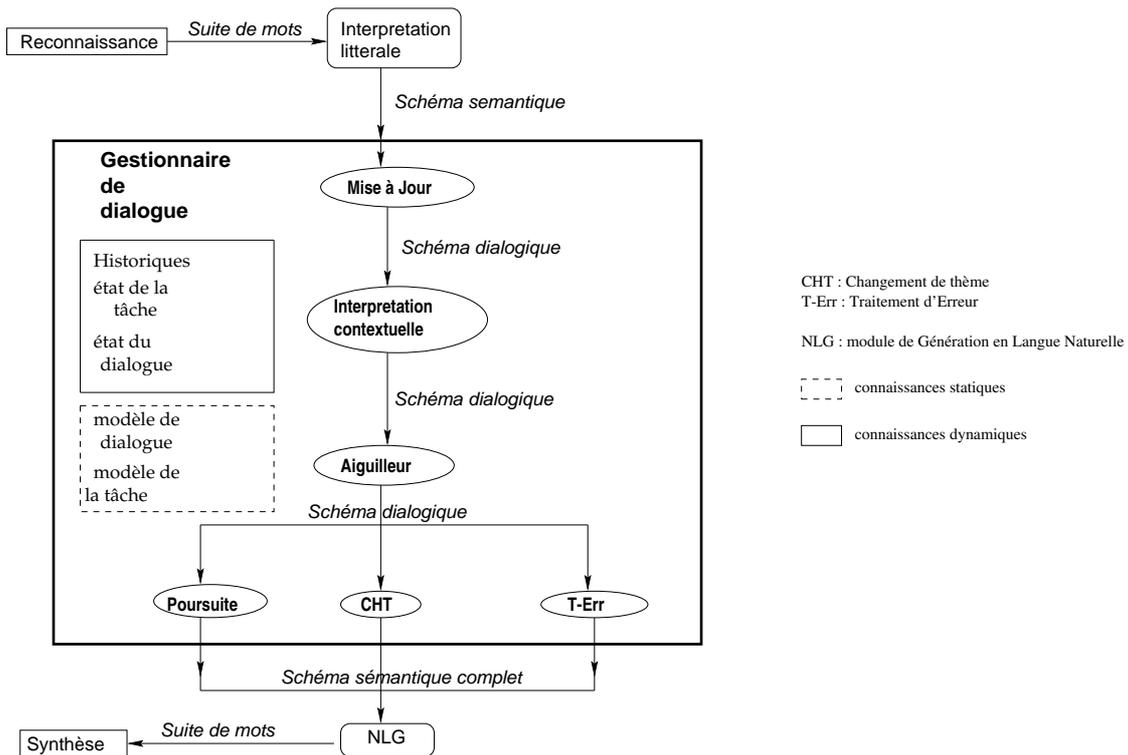


Figure 1: Étapes de traitement

3.1 Interprétation littérale

L'interprétation littérale est effectuée en deux étapes distinctes : une première permettant une analyse morphologique, syntaxique et sémantique de l'énoncé. Cette analyse aboutit à la réécriture sous la forme d'une suite d'objets et de relations entre ces objets. De cette façon, on aboutit à une représentation symbolique de l'énoncé. Cette représentation est ensuite interprétée littéralement. Il s'agit d'une interprétation partielle de l'énoncé qui sera ensuite réinterprétée par le gestionnaire de dialogue selon le contexte. Cette forme d'analyse a l'avantage d'être robuste et de permettre ensuite un traitement indépendant de la langue.

3.2 Interprétation et action dialogique

Le gestionnaire de dialogue procède à l'analyse des énoncés en plusieurs étapes. Il produit et gère un schéma dialogique qui contient des indicateurs de dialogue (temps, nombre d'échanges, état de la tâche), des schémas sémantiques et des informations provenant des différents historiques. Lors du dialogue, chaque énoncé de l'utilisateur passe par différentes étapes d'interprétation, chacune associée à des actions spécifiques, effectuées par un module de traitement spécifique. La première étape permet la **mise à jour** des différents marqueurs, indiquant l'état du dialogue, le type de demande, la fonctionnalité courante, le nombre de demandes ou encore le nombre de concepts renvoyés par l'interprétation littérale. Le schéma dialogique (schéma contenant toutes les informations ayant traités au dialogue en cours) est ainsi complété. Le rôle du module d'**interprétation contextuelle** est de réinterpréter le schéma sémantique reçu de

Gestionnaire de dialogue pour un système d'informations à reconnaissance vocale

l'interprétation littérale en fonction du déroulement du dialogue (de l'instant de l'énoncé sur l'axe de déroulement), de l'état du dialogue, de connaissances de type *sens commun* et du modèle de la tâche. Cette réinterprétation du schéma sémantique est fondée sur un ensemble de règles d'inférences et de normalisation. Cette étape permet de recalibrer le schéma dialogique sur le dialogue en cours, de corriger certaines erreurs d'interprétation littérale et/ou de reconnaissance et de clarifier d'éventuelles ambiguïtés. Le **module de décision** (ou aiguilleur) reçoit le schéma dialogique fourni par l'interprétation contextuelle. Il décide s'il y a un changement de tâche (navigation), une erreur (c'est-à-dire une contradiction entre l'historique du dialogue et l'interprétation contextuelle et/ou le modèle de la tâche) ou si le dialogue suit son cours. Pour cela, il s'aide du modèle de la tâche, des historiques et de l'état du dialogue. La phase (acquisition, post-acceptation, négociation, navigation ou méta-communication) dans laquelle se trouve le dialogue prend là toute son importance. Selon la phase précédente, une nouvelle phase sera instanciée. Cette nouvelle phase est une des conditions permettant par la suite de décider du passage à une autre tâche. Selon la décision prise, il envoie le schéma réinterprété et augmenté à l'un des trois modules suivants :

1. Poursuite : ce module gère la poursuite du dialogue lorsque aucune interruption n'est détectée. C'est le modèle de la tâche qui prend le relais.
2. Navigateur : c'est le module qui gère la navigation (changement de thème). La tâche courante est fermée (ou mise en attente, le cas échéant) et la nouvelle tâche activée selon le modèle de la tâche.
3. Correcteur : ce module gère le traitement des erreurs. Si une erreur est détectée, la stratégie pour la traiter dépend de la phase de dialogue dans laquelle se trouve le système (acquisition ou post-acceptation), de la tâche courante et des historiques. C'est à ce moment-là que des erreurs systématiques du système de reconnaissance peuvent être prises en compte.

Suite à ces différents traitements, un schéma sémantique est généré et envoyé au module de génération en langue naturelle.

4 Résultats et conclusions

Nos objectifs étaient d'obtenir un système d'interrogation orale d'une base de données (celle de la SNCF) permettant d'obtenir un échange aussi naturel que possible, imposant le minimum de contraintes à l'utilisateur. Afin d'atteindre ces objectifs, le système devait permettre les navigations entre et dans le cadre des différentes fonctionnalités définies de l'application et des négociations pour chacune des catégories d'informations demandées et des propositions du système. Nous avons donc opté pour un modèle de la tâche très souple et un modèle de dialogue dynamique. Pour y parvenir, nous avons structuré aussi précisément et clairement le domaine que possible et organisé les différents éléments sur trois niveaux, par regroupement successifs, tout en maintenant des liens entre les différents éléments d'un même niveau. *A priori*, le gestionnaire de dialogue autorise tous les chemins. Toutefois, il se fonde sur un chemin *le plus probable*. Ce chemin *le plus probable* varie au cours du dialogue selon le *score caractéristique de l'état du dialogue*, l'évaluation du chemin déjà parcouru et la *phase* de dialogue active. L'état du dialogue est évalué en fonction du modèle de la tâche, de ce qui s'est déjà produit (les historiques) et du nombre d'échanges. Ceci nous a permis non seulement d'assouplir le dialogue

mais également de le rendre plus prédictif et donc plus robuste face aux erreurs. Les différentes connaissances décrites ici existent dans le système soit sous la forme de lexiques (notamment la représentation du domaine) soit sous la forme de règles d'inférences.

Un corpus de plus de 10000 appels a été utilisé pour développer et tester le gestionnaire de dialogue et les différents modules présentés ici. Le système a été régulièrement évalué par la SNCF (partenaire du projet ARISE) entre Novembre 1997 et Novembre 1998. Les sujets ont été recrutés par une compagnie indépendante. Ces évaluations ont montré une amélioration des performances globales du système de dialogue (de 16% de réussite complète¹ pour le premier système à 85.2% pour le dernier système). Des évaluations plus spécifiques (Rosset et al. 2000) ont montré des améliorations sur la gestion des demandes de navigation et sur le traitement des erreurs. Les réussites en terme de passage d'une tâche à une autre ont été mesurées. Le gain global a été de 10% entre deux versions (en mai 1998: 82% de passages demandés ont réussi et 92% en novembre 1998). Il nous fallait également vérifier si le système ne passait pas à une autre tâche alors qu'aucune demande n'était faite. Entre mai 1998 et novembre 1998 le nombre de passages de tâches effectués alors que non demandés a diminué (de 40% des passages effectués qui n'étaient pas demandés en mai 1998 à 21% en novembre 1998). Pour vérifier la robustesse du système face à certaines catégories d'erreurs (erreurs de reconnaissance) et sa capacité à les repérer et les traiter, nous avons mis en rapport le taux de succès de dialogue et le taux d'erreurs de reconnaissance. Ce taux est un taux d'erreurs sur mots transcrits orthographiquement et manuellement. Toutes les erreurs n'impliquent pas nécessairement une erreur d'interprétation. Nous avons constaté qu'à un taux d'erreurs de reconnaissance identique, le taux de réussite de dialogue a augmenté entre les deux versions du système (par exemple, pour 30% d'erreurs de reconnaissance, nous avions en 1997 63% de succès dans le dialogue et 73% en 1998).

Références

- Guyomard M., Nerzic P. et Siroux J. (1993), Plans, métaplans et dialogue, Actes de la 4^{ème} école d'été sur les traitements des langues naturelles, Lannion, France.
- Lehuen J. (1997), Un modèle de dialogue dynamique et générique intégrant l'acquisition de sa compétence linguistique : le système COALA, Thèse de doctorat de l'Université de Caen.
- Lemeunier T. (2000), L'intentionnalité communicative dans le dialogue homme-machine en langue naturelle, Thèse de doctorat de l'université du Maine.
- Luzzati D (1989), Recherches sur le dialogue homme-machine: modèles linguistiques et traitements automatiques, Thèse d'état, Paris III.
- Pérennou G. (1996), Compréhension du dialogue oral - le rôle du lexique dans l'approche par segments conceptuels, *Lexique et communication parlée*, GDR - PRC, 169-178.
- Rosset S., Lamel L. et Bennacef S. (2000), Stratégies pour un système de dialogue oral homme machine, Actes 23^{èmes} Journées d'Étude sur la Parole, 329-332, Aussois, France.

¹L'ensemble du dialogue est une réussite : toutes les demandes, quelles qu'elles soient, ont abouti.