

Vers une architecture générique de système de dialogue oral homme-machine

Hoá NGUYEN

Laboratoire CLIPS - IMAG – Université Joseph Fourier
385, rue de la Bibliothèque - B.P. 53 - 38041 Grenoble Cedex 9
Ngoc-Hoa.Nguyen@imag.fr

Mots-clefs – Keywords

Dialogue homme-machine, gestion du dialogue, gestion de la tâche, compréhension, interprétation.

Human-machine dialogue system, dialogue management, task management, comprehension, interpretation.

Résumé – Abstract

Cet article présente une architecture générique de système de dialogue oral homme-machine. Premièrement, nous abordons quelques problèmes soulevés par la généralité des systèmes de dialogue homme-machine. Nous décrivons ensuite dans ce cadre quelques systèmes récents et typiques. Nous présentons finalement une architecture générique pour concevoir/construire des systèmes de dialogue oral homme-machine.

In this paper, we present a general architecture for the dialog person-machine system that enables human interacts with a machine via speech's modality. First, we approach some notions of dialog system, the role of the speech in application, their interests and their limitations. Then, we present some recent, critical dialog person-machine systems and finally, we describe our architecture proposed for designing and constructing the general dialog systems.

1 Système de dialogue homme-machine

L'idée de remplacer l'interlocuteur humain par une machine est apparue depuis les années soixante où de nombreux systèmes informatiques ont été conçus pour cet objectif, avec pour certains, des résultats intéressants. Il s'avère que les technologies vocales n'ont pas seulement pénétré le domaine des applications professionnelles, mais bien celui du grand public, même

si cela reste encore fort discret. L'intérêt de l'interaction vocale dans un service est tout d'abord qu'il s'agit du mode le plus naturel et le plus courant de la communication entre les humains. Elle apporte également la rapidité, l'efficacité et l'extension du champ d'action du système. De plus, la parole permet d'avoir accès à des données distantes et rend la possibilité d'établir des échanges de niveau sémantique complexe, de manipuler des notions abstraites (Caelen, 1996). Par contre, du côté des inconvénients, les interactions vocales subissent actuellement les limitations technologiques comme média de transmission, puissance de calcul nécessaire pour une réponse en temps réel, mémoire nécessaire pour de grands vocabulaires... D'autre part, l'ambiguïté de la langue pénalise l'interprétation du message et réduit donc la performance du système vocal.

Nous distinguons les deux notions: "*système vocal*" et "*système de dialogue oral homme-machine*". Un système vocal est un service utilisant la parole comme une modalité d'entrée pour réaliser/accomplir certaines tâches dédiées. Dans ce type d'application, l'utilisateur peut dialoguer avec l'application en utilisant seulement les mots clés, des phrases courtes, simples et connues a priori. Par opposition, un système de dialogue homme-machine est un système qui est capable d'interagir de manière plus naturelle avec l'homme pour accomplir des tâches plus générales et plus complexes. Dans les deux cas cependant on se heurte au même problème concernant la portabilité d'une tâche à un autre : la **généricité**.

Il est évident que ce problème représente un verrou important à lever pour un développement en vraie grandeur de systèmes de dialogue homme-machine à large diffusion. Dans cet article nous présentons un premier travail de spécification d'une architecture générique effectué dans le projet PVE (Portail Vocal d'Entreprise) financé par le RNRT (Réseau National de Recherche en Télécommunications).

2 Etat de l'art

La dernière décennie a vu le développement d'un grand nombre de systèmes de dialogue homme-machine en langue naturelle. Ces systèmes permettent à l'utilisateur d'interagir avec la machine pour trouver une information ou pour résoudre un problème concret. Cette section présente quelques systèmes récents, dans ce domaine. Il est évident que la liste n'est pas exhaustive.

- COALA (COAdaptation Langagière pour l'Apprentissage) (Lehuen, 1997) est un système d'aide documentaire dans une médiathèque municipale. L'utilisateur peut communiquer au système en français pour la recherche documentaire en utilisant le modèle hypothético-expérimental (un modèle qui vise à introduire un maximum de souplesses dans la gestion de l'interaction.). COALA pose des hypothèses au fur et à mesure du dialogue, et les valide ou les infirme selon la suite des événements. Ainsi la machine apprend à chaque nouvelle interaction avec l'utilisateur.
- HALPIN (Hyperdialogue avec un Agent en Langage Proche de l'Interaction Naturelle) est un système de recherche d'information utilisant l'hyperdialogue¹ en

¹ Un dialogue homme-machine, coopératif et finalisé, dans un environnement hypertextuel. L'écrit, l'oral, voire le geste, peuvent être utilisés dans l'hyperdialogue

français. L'utilisateur dialogue avec le système HALPIN : il énonce sa demande de recherche des documents oralement et la machine lui répond en donnant le résultat de recherche sur le même mode (Rouillard, 2000).

- CMUCommutator (www.speech.cs.cmu.edu/Communicator/) est un système de dialogue provenant du laboratoire de la parole de l'université de Carnegie Mellon. Il a pour objectif de permettre à l'utilisateur de planifier un voyage en avion, de louer une voiture, de réserver une chambre en utilisant un serveur vocal et le réseau téléphonique (Rudnicky, Thayer, 1999). Dans ce système, on propose une architecture de gestion de dialogue basée sur un agenda qui gère les topiques concernant l'accomplissement de tâche (Xu, Rudnicky, 2000).

Le point faible commun de ces systèmes en particulier et des systèmes actuels de dialogue en général est la dépendance de la tâche. Cela signifie que ces systèmes ne sont conçus et développés que pour résoudre certaines tâches qui sont très restreintes et fermées. Ils ne peuvent pas couvrir un ensemble de tâches et l'ajout d'une tâche à ces systèmes peut conduire à reconstruire un nouveau système, en bénéficiant seulement d'un savoir-faire antérieur mais non de briques logicielles de base.

La généricité d'un système de dialogue a toujours été un des problèmes majeurs soulevés de nombreux chercheurs non seulement dans le domaine du dialogue mais également dans les autres domaines comme le traitement automatique de la langue. Notre ambition est de proposer une architecture de système de dialogue homme-machine plus générique, et relativement portable. Il est évident que ces systèmes ne pourront jamais être indépendants de la tâche mais peuvent comporter un gestionnaire de tâches autonome dont ils seront moins dépendants.

3 Architecture du système

Dans cette partie, nous décrivons une architecture qui facilite non seulement la conception mais également la construction d'un système de dialogue. Elle est générique, et pourra donc être instanciée. Cette architecture sépare les modules de traitement de données de manière générique/instanciée. Ainsi au niveau de la reconnaissance de la parole nous distinguons les données générales des données particulières de la tâche (vocabulaire spécialisé, noms propres par exemple). De même, au niveau du traitement du langage, nous distinguons le module de compréhension sémantique et d'interprétation pragmatique. Enfin, au niveau du dialogue nous séparons ce qui relève de la gestion proprement dite du dialogue (tours de parole, buts dialogiques, stratégies) de ce qui relève de la gestion de la tâche (plans, objets de la tâche, contexte d'interaction)

Dans cette architecture, le flux de données n'est pas linéaire :

- Le locuteur prononce un énoncé oral.
- Le système de reconnaissance produit une chaîne orthographique candidate.

- La chaîne orthographique est ensuite prise en charge par un composant linguistique (dit de compréhension) qui fournit un schéma sémantique représentant le sens littéral de l'énoncé.

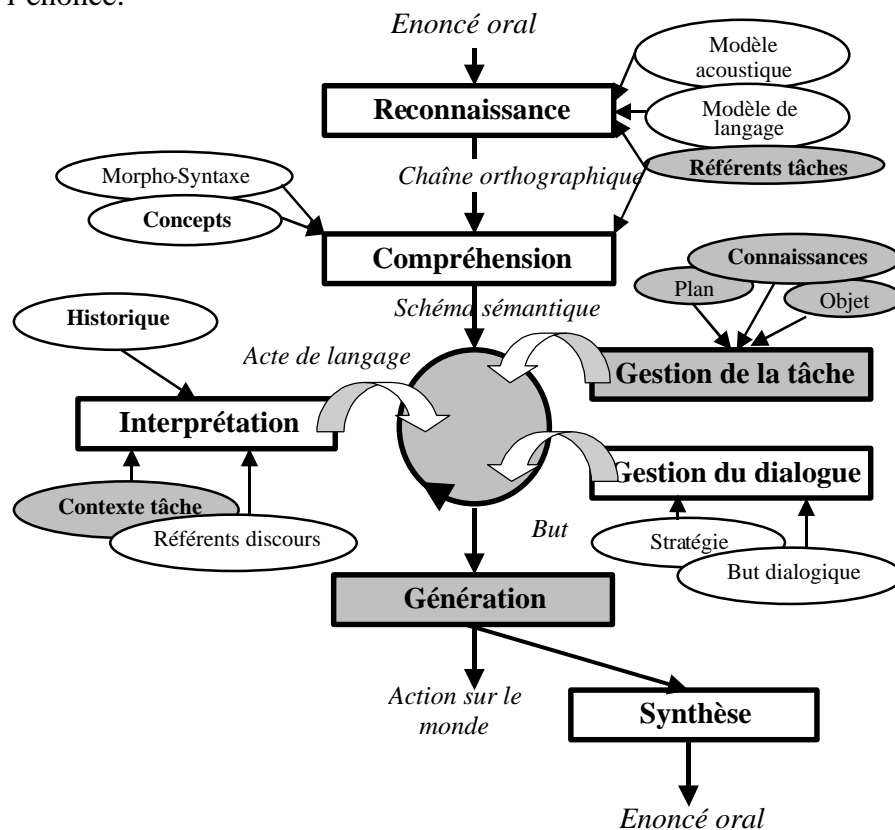


Figure 1 : Architecture générique d'un système de dialogue homme-machine, les parties grisées correspondent aux connaissances ou traitements dépendants de l'application

- On entre ensuite dans un processus itératif contrôlé par un mécanisme de gestion de dialogue. Le schéma sémantique s'enrichit au fur et à mesure des apports des autres modules pour devenir un acte de langage, en particulier par le module d'interprétation et par le module de gestion de la tâche. Cet acte de langage est pris en charge par le contrôleur de dialogue qui le soumet sous forme de but au gestionnaire de tâches.
- La réponse de ce dernier conduit à formuler une réaction sous forme de changement dans le "*monde de la tâche*" et/ou de réponse vocale synthétisée. En cas d'échec dans la résolution d'un plan ou d'une tâche, le contrôle est redonné au contrôleur de dialogue qui décide de la stratégie à tenir. Il peut alors redistribuer les rôles et les traitements à l'un des modules.

Nous décrivons les composants de cette architecture en détail ci-dessous.

3.1 Reconnaissance de la parole

Le module de reconnaissance automatique de la parole prend un énoncé oral en entrée et fournit une chaîne orthographique en sortie. Il comporte évidemment des ressources

acoustiques et linguistiques comme (a) un modèle acoustique de la langue, (b) un modèle statistique de langage et (c) un modèle spécifique de la langue de l'application (par exemple noms propres, entités et expressions du domaine) qui dépend du contexte de l'application, c'est-à-dire que ce module reste encore une partie dépendant de la tâche. On appelle ces entités spécifiques, "référents de la tâche".

Aujourd'hui, il existe quelques systèmes de reconnaissance de la parole en français ainsi qu'en anglais comme Dragon NaturallySpeaking (www.dragonsys.com/), SpeechWorks (www.speechworks.com/), ViaVoice d'IBM (www-3.ibm.com/software/speech/), Nuance (www.nuance.com/) qui proposent de telles possibilités.

3.2 Compréhension sémantique

Le module dit de "compréhension" prend la chaîne orthographique en entrée et fournit un schéma sémantique en sortie. Il utilise des outils généraux concernant la morphologie, la syntaxe et la sémantique lexicale comme données et algorithmes. La pragmatique n'est encore pas traitée à ce niveau. La performance du système de dialogue dépend évidemment de la qualité de la compréhension et donc elle doit être robuste pour ne pas être trop sensible aux erreurs de reconnaissance, c'est-à-dire en ayant une bonne anticipation/prédiction de ce qui est dit par l'utilisateur. Pour cette raison, on peut utiliser une approche descendante par analyse de concepts. Les concepts sont des sortes de réseaux sémantiques issus d'un thésaurus.

3.3 Interprétation pragmatique

La sortie du module de compréhension est l'entrée du module d'interprétation. Il a pour but de déterminer les actes de langage (Austin 1970, Searle 1986, Vanderveken 1990) et le contenu propositionnel de l'acte (Caelen 1997) en sortie. Ce module travaille sur des représentations du discours de type DRT- "Discourse Representation Theory" (Kamp 1981, Reyle 1993). Il utilise comme données, l'historique mémorisé au cours du dialogue, le contexte de la tâche (représenté sous-forme de thèmes à structure de graphe) et des algorithmes de calcul de la référence, notamment pour les ellipses et les anaphores (basés sur la DRT).

3.4 Gestion du dialogue

Ce module peut être considéré comme le contrôleur principal du système de dialogue. Il transmet les messages aux autres modules, décide de la stratégie adéquate, gère les buts et prépare les données pour le générateur d'énoncés. Pour réaliser toutes ces tâches, il est équipé de règles de calcul des stratégies et des règles de gestion des buts dialogiques (par exemple à l'aide de piles ou de files d'attente) (Caelen 1997). Les comportements de la machine sont déterminés selon la stratégie de dialogue du moment, le but actuel à résoudre et le plan de la tâche (séquence d'actions élémentaires pour la tâche). Les buts en cours de traitement sont ensuite transmis au gestionnaire de tâches. Les résultats sont récupérés, analysés, et selon le cas (incomplétude, échec ou succès) conduisent ou non à une nouvelle itération du processus.

3.5 Gestion de la tâche

La gestion de la tâche peut être considérée comme un planificateur classique. Il prend en entrée le but à résoudre et développe un plan pour l'atteindre. En cas d'échec il en donne les raisons et propose des solutions de repli. En cas d'incomplétude ou d'imprécision dans la formulation du but en entrée, il indique les paramètres en cause (par exemple lorsque les pré-conditions ou les contraintes ne sont pas satisfaites). Dans des applications fermées on peut représenter explicitement les tâches et sous-tâches à l'aide d'un arbre hiérarchisé. Si les tâches peuvent évoluer au cours du temps (tâches dynamiques), il est préférable d'adopter un modèle dit implicite.

3.6 Génération

Le module de génération prend en entrée les consignes du module de gestion de dialogue et fournit en sortie une action (suite d'instructions) exécutable et/ou produit une chaîne orthographique textuelle. Ce module peut se réduire à sa plus simple expression (collage de segments d'énoncés) s'il existe un adaptateur de noyau fonctionnel qui se charge de la transformation de la représentation du message en une chaîne.

3.7 Synthèse de la parole

Dans cette architecture, on constate que ce module est le plus simple par rapport aux autres. Il prend naturellement en entrée une chaîne orthographique textuelle et fournit en sortie un signal sonore. Actuellement, les synthétiseurs commercialisés sont d'une relative bonne qualité, ils peuvent être utilisés à profit dans un système de dialogue homme-machine. Mais la plupart du temps on préfère un générateur de phrases à trous si le nombre de types de réponses est limité.

4 Résultats attendus

Notre architecture sera validée dans le projet PVE². Ce projet vise le développement d'un modèle de génération d'interface de dialogue vocal pour un portail vocal d'entreprise. Plus précisément, il a pour but d'analyser, d'étudier et de formaliser un modèle générique de dialogue homme-machine vocal, dans l'optique de proposer des solutions technologiques adaptées aux nécessités d'un accès au système d'information de l'entreprise compatible avec la mobilité (au sens de la circulation) du personnel dans l'entreprise et hors ses murs. Les éléments fonctionnels prioritaires d'un portail vocal d'entreprise sont l'interrogation du répertoire du personnel, de l'agenda d'un groupe d'utilisateurs et le suivi du courrier électronique personnel. Ces fonctions doivent être activables de manière intégrée afin de permettre un dialogue utile et performant pour un utilisateur accédant au service par téléphone.

² Pour en savoir plus, consultez sur la page Web www.telecom.gouv.fr/rnrt/projets/res_01_5.htm

5 Conclusion

Nous avons présenté une architecture pour concevoir et construire des systèmes de dialogue homme-machine dans l'optique de couvrir un ensemble de tâches. Dans cette architecture, le modèle de dialogue est séparé complètement par rapport au modèle de la tâche. Cela apporte aux systèmes construits basés sur cette architecture à la fois la robustesse et la souplesse (par exemple l'ajout d'une nouvelle tâche à un tel système n'affecte pas le contrôleur de dialogue). L'implémentation d'un système de dialogue oral homme-machine dans le cadre du projet PVE est en cours. Le développement de ses modules est progressif et obéit à une logique de travail en étape. Etant achevée, cette architecture sera soumise à une évaluation en vue de tester et de diagnostiquer les apports de ses modules, en particulier le module de gestion de dialogue sur lequel nous nous focaliserons dans la continuité de notre thèse.

Références

Austin, J.L (1970), *Quand dire c'est faire*, Paris, Edition du Seuil.

Caelen J., Néel F.(1996), *Nouvelles Interfaces Homme-Machine*, ARAG0 série XVIII.

Caelen J.(1997), *Interaction Verbale*, Editions CEPADUES.

Coulon D., Kayser D. (1986), Informatique et langage naturel : Présentation générale des méthodes d'interprétation des textes écrits, *Revue TSI : Technique et Science Informatique*, Vol. 5. n°2.

Kamp H. (1981), *A Theory of truth and Semantic representation*, Amsterdam: Mathematical Centre, J. Groenendijk, T. Janssen and M. Stokhof Eds.

Kamp H., Reyle U.(1993), *From Discourse to Logic*, Kluwer Academic Publishers.

Lehuen J.(1997), *Un modèle de dialogue dynamique et générique intégrant l'acquisition de sa compétence linguistique - Le système COALA*, thèse du laboratoire d'Informatique, Université du Maine.

Rouillard J. (2000), *Hyperdialogue sur Internet – Le système HALPIN*, thèse de l'Université Joseph Fourier, CLIPS-IMAG.

Rudnicky A.I., Thayer E., Constantinides P., Tchou C., Shern R., Lenzo K., Xu W., Oh A.(1999), Creating Natural Dialogs In The Carnegie Mellon Communicator System, *Proceedings of Eurospeech*, n°4, 1531-1534.

Searle J.R., Vanderveken D.(1985), *Foundations of illocutionary Logic*, Cambridge University Press.

Vanderveken D. (1990), *Meaning and Speech Acts*, Volumes 1 and 2, Cambridge University Press.

Xu W., Rudnicky A.(2000), Task-based dialog management using an agenda. *ANLP/NAACL Workshop on Conversational Systems*, pp. 42-47.

