*TALN 2005, Dourdan, 6–10 juin 2005*

# RITEL : dialogue homme-machine à domaine ouvert

Olivier Galibert, Gabriel Illouz, Sophie Rosset
LIMSI - CNRS
F-91403 Orsay Cedex
{galibert,gabrieli,rosset}@limsi.fr

**Mots-clefs :** dialogue homme machine, recherche d'information précise, corpus

**Keywords:** human machine dialog, question answering, information retrieval, corpus

**Résumé** L'objectif du projet RITEL est de réaliser un système de dialogue homme-machine permettant à un utilisateur de poser oralement des questions, et de dialoguer avec un système de recherche d'information généraliste (par exemple, chercher sur l'Internet "Qui est le Président du Sénat ?") et d'en étudier les potentialités. Actuellement, la plateforme RITEL permet de collecter des corpus de dialogue homme-machine. Les utilisateurs peuvent parfois obtenir une réponse, de type factuel (**Q :** qui est le président de la France ; **R :** Jacques Chirac.). Cet article présente brièvement la plateforme développée, le corpus collecté ainsi que les questions que soulèvent un tel système et quelques unes des premières solutions envisagées.

**Abstract** The project RITEL aims at integrating a spoken language dialog system and an open-domain question answering system to allow a human to ask a general question (f.i. "Who is currently presiding the Senate?") and refine his research interactively. As this point in time the RITEL platform is used to collect a new human-computer dialog corpus. The user can sometimes recieve factual answers (**Q :** who is the president of France ; **R :** Jacques Chirac). This paper briefly presents the current system, the collected corpus, the problems encountered by such a system and our first answers to these problems.

## Introduction

Les progrès réalisés ces dernières années tant en reconnaissance de la parole qu'en recherche d'information permettent d'envisager de nouvelles études. L'objectif du projet RITEL est de réaliser un système de dialogue homme-machine permettant à un utilisateur de poser oralement des questions, et de dialoguer avec un système de recherche d'information généraliste (par exemple, chercher sur l'Internet "Qui est le Président du Sénat ?") et d'en étudier les potentialités. Le terme de **système de dialogue** indique généralement un système permettant une interaction entre un humain et un système dans un cadre restreint (Glass J. R. et al., 2000). Toutefois, notamment dans le cadre des travaux sur les systèmes de question-réponse, le cadre tend à s'élargir. Un dialogue est une suite d'échanges entre interlocuteurs dans un contexte donné. Un système de dialogue homme-machine interprète les requêtes de l'utilisateur en fonction de la tâche à accomplir, de l'histoire du dialogue et du comportement de l'utilisateur. Son

objectif est de donner à l'utilisateur les informations recherchées tout en assurant une interaction efficace et naturelle. Actuellement, Les systèmes concernent des domaines restreints tels que l'information horaire de moyens de transports (trains, avions, cinéma) ou les informations touristiques, par exemple le projet européen Le3-Arise (horaires de train), le projet américain ATIS du DARPA Communicator (voyages en avions), et le projet français Technolangue MEDIA (informations touristiques). Ces projets ont donné lieu à des évaluations et ont permis d'en asseoir la faisabilité. Des modèles de gestion dynamique du dialogue et de génération adaptée ont été proposés. Même si ce qu'on entend par interaction naturelle est très variable d'un système à un autre (Villaneau J., 2003), ces systèmes permettent une interaction (orale) relativement naturelle : l'utilisateur peut à tout moment changer d'avis et revenir sur des choix exprimés, interrompre la réponse du système en prenant la parole, le système peut lui aussi changer de stratégie d'interaction en fonction des réactions de l'utilisateur. Un système de dialogue utilise des sources de connaissances diverses et complexes : connaissances acoustiques, phonétiques, lexicales, morphologiques, syntaxiques et sémantiques, pragmatiques, ainsi que des connaissances sur le dialogue, sur la tâche à réaliser et sur l'interlocuteur. En recherche d'information et extraction d'information, des progrès (Harabagiu S. et al., 2001) ont été motivés par des campagnes d'évaluations (américaine TREC, 98-2003, européenne CLEF, et nationale Equer/Technolangue, 2004). Ces systèmes sont limités à une question et une réponse. Des tentatives ont été faites pour des questions enchaînées, portant sur un même thème. Mais il ne s'agissait pas de dialogues, il n'y avait pas réellement d'interaction, il n'y avait pas de négociations possibles. Le projet RITEL s'appuie sur des progrès récents en reconnaissance de la parole conversationnelle et multi-locuteurs (Gauvain J.L., Lamel L., 2002) Un projet riche et complexe doit faire face à plusieurs points épineux. Les plus évidents sont : la reconnaissance de la parole qui doit être à grand vocabulaire et sur laquelle une contrainte temps réel s'applique, la gestion d'un dialogue en domaine ouvert, la communication et l'échange d'informations entre un système de question-réponse et le dialogue, la génération de la réponse. Cet article ne répond pas à toutes ces questions, seuls certains problèmes rencontrés et les réponses apportées seront présentés. Nous présentons un état des lieux du projet RITEL et décrivons le corpus collecté jusqu'à présent avec la plateforme. Nous concluons sur les perspectives de cette étude.

# 1 Dialogue oral et recherche d'information : intégration

Un système de dialogue oral homme-machine a pour objectif de donner à l'utilisateur l'information qu'il recherche en s'aidant de diverses sources de connaissances (statiques : connaissance du domaine, dynamique...), et le système de question-réponse de rechercher une réponse précise à une question. L'objectif du projet RITEL est l'intégration de ces différents systèmes en une plateforme unique.

## 1.1 Reconnaissance de la parole

Intégrer un système de reconnaissance de parole dans une telle plateforme suppose de mettre en place un système de reconnaissance à grand vocabulaire (taille de lexique de 65000 à 300000 mots), temps réel, multi-locuteur et fonctionnant sur un signal téléphonique. Dans de telles conditions aucun système de reconnaissance, au niveau de l'état de l'art, ne permet d'obtenir des performances nécessaires pour un système de dialogue (aux alentours de 20% d'erreur). Il faut donc envisager des techniques d'adaptations dynamiques des différents modèles. L'adaptation

Table 1: Exemples avec une question : *qui est le président des États-Unis*?

| *Réponses* | score | thème | contexte | *Réponses* | score | thème | contexte |
|---|---|---|---|---|---|---|---|
| G.W. Bush | 0.99 | pol | 2000- | élu au suffrage indirect | 0.6 | droit | - |
| Bill Clinton | 0.7 | pol | 93-2000 | né sur le sol américain | 0.6 | droit | - |
| Satan | 0.1 | pol_opi | 2005 | un pantin | 0.2 | pol_opi | - |
| Mère Theresa | 0.2 | pol_opi | 2002-04 | | | | |
| Dumbo Bush | 0.3 | pol_opi | 2004 | | | | |
| Bartlet | 0.6 | fic_série | 1999- | | | | |
| H. Ford | 0.5 | fic_film | 1997- | | | | |

dynamiques des modèles de langage est rendue possible par l'indexation en thème des énoncés utilisateur et des réponses du système de recherche d'information. L'adaptation des modèles acoustiques s'effectue dynamiquement et de manière de plus en plus poussée au fur et à mesure des échanges entre l'utilisateur et le système. De plus le gestionnaire de dialogue peut le cas échéant demander à l'utilisateur d'épeler certains mots de sa demande.

## 1.2 Recherche d'information

Le système d'information prend en entrée la sortie du système de reconnaissance. Il s'agit de parole libre et potentiellement erronée. Il est donc nécessaire d'adapter la communication par rapport à un moteur question-réponse classique, notamment l'analyse de la question ne peut se faire à l'aide de contraintes morpho-syntaxiques fortes mais plutôt d'une analyse syntaxico-sémantique plus lâche. Les étapes suivantes restent sensiblement les mêmes que dans un moteur question-réponse classique. Par contre, pour ce qui est du retour de l'information, celle-ci doit être adapté au dialogue. Il ne s'agit plus seulement de mettre les "meilleures" réponses en premier mais bien de permettre au dialogue d'aider l'utilisateur à choisir celles qui lui conviennent le mieux. Pour ce faire, les réponses sont constituées de listes indicées par un score de confiance ce qui aide le système de dialogue à prendre sa décision pour générer une réponse (ou non) à l'utilisateur (cf. tableau 1). Selon ce score la réponse pourra comporter une information informant l'utilisateur du degré de confiance qu'a le système dans sa réponse (ex. je crois que,...) Le nombre de document est lui aussi associé à une réponse. Pour un grand nombre de document retourné, deux stratégies sont possibles.

Le **Regroupement par thèmes** permet au dialogue de proposer différentes possibilités à l'utilisateur pour que celui-ci oriente sa recherche. Pour chacun de ces topics, un score de confiance sera attribué de façon à aider le dialogue à orienter au mieux l'utilisateur. (**R:** *J'ai plusieurs réponses possibles, 2 dans le domaine de la politique, 3 qui s'apparentent à des opinions et 2 concernant des fictions. Quelle est la thématique de votre recherche ?* )

La **Demande de précision** a lieu si le regroupement par thème n'est pas possible. Le système soit demande à l'utilisateur de préciser sa requête soit lui propose quelques exemples pour qu'il y réagisse. ( **R:** *J'ai des réponses avec des noms de personnes et d'autres sans. Par exemple, G.W. Bush est le ... ou une réponse de type définition comme élu au suffrage universel indirect. Que recherchez-vous précisément ?* ) Ainsi, contrairement à l'augmentation de la précision en utilisant le retour d'information en aveugle (*Blind relevance feedback*), nous avons ici un retour d'information éclairé par l'utilisateur.

# 2 Architecture générale du système actuel

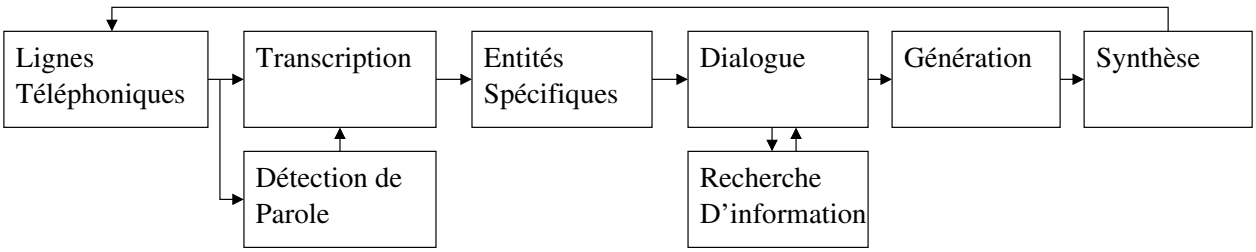


Figure 1: Organisation des composants de RITEL

Le système est composé d'un ensemble simple de composants communicants (figure 1). Le module de gestion de lignes téléphonique envoie le signal audio au détecteur de parole et au système de transcription. Le détecteur de parole contrôle l'activité de la transcription. La suite de mots obtenue est envoyée à la détection d'entités spécifiques puis au gestionnaire de dialogue. Celui-ci communique avec le composant de recherche d'informations pour décider d'une réponse à l'utilisateur. Le schéma sémantique obtenu est envoyé à la génération qui en fait une phrase que la synthèse transforme en un signal audio qui revient finalement au module téléphonique. Nous présentons brièvement le système de reconnaissance, le système d'analyse et le gestionnaire de dialogue dans les sections suivantes.

## 2.1 Détection de parole et système de reconnaissance

La détection de parole est basée sur notre segmenteur de conversations téléphoniques spontanées : des modèles multi-gaussiens modélisent le silence et la parole. Un décodage de viterbi classique segmente le signal. Le système de reconnaissance est un décodeur monopasse utilisant des modèles acoustiques triphones et un modèle de langage trigramme suivi d'un rescoring en consensus avec un modèle quadrigramme. Les modèles acoustiques sont ceux du système de dialogue Arise (Lamel et al., 2000). Vocabulaire (65000 mots, 1.7% OOV) et modèles de langage ont été obtenus par interpolation de modèles de diverses sources sur le développement. Elle sont le corpus d'entraînement RITEL, les textes de journaux de 1988 à nos jours, les transcriptions d'émissions de nouvelles radio- et télé-diffusées, des questions venant d'internet (ex.: Madwin.fr). La prochaine version intègrera les nouvelles transcriptions du corpus ainsi que des transcriptions de conversations téléphoniques spontanées.

## 2.2 Détection d'entités spécifiques

Un système de détection d'entités spécifiques est utilisé. Il prend en entrée la suite de mots envoyée par le système de reconnaissance. Les entités spécifiques se décomposent en trois classes: **Entités nommées** telles que les personnes (<pers>), les productions (<prod>), les marqueurs temporels (<time>), les organisations (<org>), les unités lexicales (<ul>) et les lieux (<loc>), les **Marqueurs syntaxiques de requête** (qui, où, quand, comment, combien...), les **Marqueurs sémantiques de requête** qui correspondent à des thèmes (littérature, géographie, histoire, société, cinéma, politique, économie...) et sous-thèmes (auteur, dirigeant, orthographe,

Table 2: Génération de réponses : à gauche, une réponse factuelle, à droite une incitation

| | |
|---|---|
| **S:** Bonjour et bienvenue sur le service Ritel du Limsi. Souhaitez-vous écouter le guide ? | **S:** Bonjour et bienvenue sur le service Ritel du Limsi. Souhaitez-vous écouter le guide ? |
| **U:** euh non merci je voudrais la capitale de la France | **U:** je voudrais des informations sur l' ONU |
| **S:** La capitale de la France est Paris. Quelle autre informations désirez-vous ? | **S:** Vous cherchez une information sur ONU. Pouvez-vous être plus précis ? |
| **U:** euh quel est son président | **U:** quels pays sont impliqués |
| **S:** Jacques-Chirac est le dirigeant de la France. Quelle autre informations désirez-vous ? | **S:** Votre question porte sur la géographie. Mais plus précisément ? Un pays ? Un ville ? |

population...). Toutes ces entités sont extraites et normalisées via un ensemble de règles écrites sous forme d'expressions régulières de mots. Des macros (règles locales) et des listes peuvent être utilisées pour la définitions des règles. Ces macros et listes, ainsi que des règles elles-mêmes, peuvent servir dans la définition de contexte. Les analyses sont séparées selon leur fonction : la première étape consiste à traiter les nombres. La deuxième étape consiste en un traitement lexical c'est à dire une normalisation des mots marqueurs syntaxiques de questions. La troisième étape enfin annote les thèmes, sous-thèmes et entités nommées. Ainsi un énoncé comme *qui a écrit le rouge et le noir en mille huit cent trente* a pour représentation *<subsubject> <Tauteur> <Tqui> qui </Tqui> a écrit </Tauteur> </subsubject> <prod> le rouge et le noir </prod> <time> <1830> mille huit cent trente </1830> </time>* .

## 2.3 Gestionnaire de dialogue

La première version du gestionnaire de dialogue a consisté exclusivement à inciter les sujets à parler le plus possible tout en permettant de conserver une interaction raisonnablement naturelle. La seconde version permet d'interroger une base de données. Le rôle du gestionnaire de dialogue est d'interpréter contextuellement le schéma sémantique que le système de détection d'entités spécifiques lui envoie ; récupérer les informations nécessaires à l'interrogation de la base de données ; générer les schémas sémantiques pour la génération ; choisir une stratégie de génération ; Le gestionnaire de dialogue commence par mettre à jour les différents marqueurs caractérisant le dialogue en cours (fonctionnalités en cours, générations précédentes, nombre de nouveaux éléments...) L'énoncé est alors traité. **l'interprétation contextuelle** génère un schéma sémantique en adéquation avec l'état du dialogue. Le **module de décision** réinterprète ce schéma selon un historique plus large en se fondant sur le modèle de la tâche et le modèle de dialogue. Si ce module "décide" que l'énoncé correspond à une possible recherche dans une base de données, celle-ci est effectuée. Sinon, la requête est traitée par le **module d'incitation** (cf. tableau 2). Ces deux modules génèrent enfin un schéma sémantique qui est envoyé au **module de génération** en langue naturelle.

Table 3: Corpus RITEL

| | | | | | |
|---|---|---|---|---|---|
| dialogues | 369 | mots distincts U | 1993 | Thèmes | 767 |
| durée totale U | 3h40 | moyen énoncés U / dial. | 9 | Sous-Thèmes | 701 |
| énoncés U | 3300 | durée moyenne U / dial. | 34s. | marqueurs syntaxiques | 3220 |
| mots U | 32634 | Entités Nommées | 8174 | | |

# 3 Corpus

Le corpus (Table 3) a été collecté entre septembre 2004 et janvier 2005. 13 personnes ont appelé le serveur. Ces personnes ont reçu chacune une liste différente d'environ 300 questions. Il leur était demandé de chercher à obtenir une information et d'utiliser pour cela les moyens qu'elles souhaitaient. Il leur était précisé qu'elles ne devaient pas lire les questions qu'elles avaient en exemple et qu'elles pouvaient en choisir d'autres plus proches de leurs intérêts.

# Conclusion - Perspectives

Actuellement la plateforme permet une interaction naturelle, quoique limitée, entre un utilisateur qui recherche des informations et le système. Elle a permis de collecter un corpus réaliste et riche d'un point de vue linguistique. Une recherche d'information (minimaliste) est possible, puisque une partie des dialogues a abouti à une réponse. Cette première étude nous permet de dégager les points sur lesquels nos travaux futurs vont porter : reconnaissance vocale temps-réel en flux sur vocabulaire large dynamiquement extensible et en domaine ouvert avec adaptation des modèles à l'interlocuteur tout au long de l'interaction, gestion de dialogue en domaine ouvert et gestion de l'information multi-niveau retourné par le système de recherche d'information, recherche d'information(classification/présentation des réponses suivant l'état du dialogue), types d'information nécessaire pour permettre au dialogue et aux autres modules d'évaluer leurs analyses et réponses suivant l'état du dialogue et des résultats de la recherche d'information, étude du coût des différentes stratégies pour le lancement de la recherche d'information (en continu ou sur décision du gestionnaire de dialogue), fonctionnement en parallèle de la recherche d'information et du dialogue, génération de la réponse, résumé automatique, etc...

# Références

EQueR, ELDA (2003) - Evaluation de systèmes de Question-Réponse, http://www.elda.org/article118.html

Gauvain J. L. et Lamel L. F. (2002), Systèmes de reconnaissance, de compréhension et de dialogue, *Reconnaissance de la parole Traitement automatique du langage parlé*, Hermes Lavoisier, J. Mariani.

Glass J. R. et al. (2000), Data collection and performance evaluation of spoken dialogue systems : the MIT experience, Actes de *ICSLP'00*, Pekin, Chine.

Grau B. (2005), Les systèmes de question-réponse, *M*éthodes avancées pour les systèmes de recherche d'informations, sous la direction de Madjid Ihadjadene, collection Traité des sciences et techniques de l'information, Hermes-science.

S. Harabagiu, et al., (2001), The Role of Lexico-Semantic Feedback in Open-Domain Textual Question-Answering, Actes de *Association for Computational Linguistics*

L. Lamel et al., (2000), The LIMSI ARISE System, *Speech Communication* Vol. 31(4):339-354.

Devillers L. et al. (2004), The French MEDIA/EVALDA Project: the Evaluation of the Understanding Capability of Spoken Language Dialogue Systems, *L*REC'04

J. Villaneau (2003), Contribution au traitement syntactico-pragmatique de la langue naturelle parlée : approche logique pour la compréhension de la parole, Thèse de Doctorat, Université de Bretagne Sud.