

Corpus oral de renseignements touristiques

S. Rosset, L. Lamel, S. Bennacef, L. Devillers, J.L. Gauvain

Groupe Traitement du langage parlé
LIMSI-CNRS, BP 133 91403 Orsay cedex, FRANCE
{rosset, lamel, bennacef, devil, gauvain}@limsi.fr

Résumé

Cet article décrit les activités du LIMSI concernant l'acquisition de corpus de dialogue homme-machine pour une tâche de renseignements touristiques dans le contexte de l'action B2 de l'AUPELF: Dialogue oral. Notre expérience dans le développement de systèmes de compréhension pour des tâches de renseignements sur les tarifs et les horaires de trains ou d'avions (ATIS et Mask) nous a appris qu'il était avantageux de placer les sujets devant un système aussi tôt que possible. Ceci permet en effet d'améliorer itérativement le système en terme de performances et de fonctionnalités en se basant sur les interactions entre les sujets et le système.

Un premier module de compréhension du langage parlé a été développé (analyse par grammaire de cas avec un système de dialogue et un système de génération de réponses restreints pour la première phase). Ce module a été utilisé pour acquérir un premier corpus d'énoncés écrits. Ces énoncés ont permis de constituer un premier lexique et un premier modèle de langage d'un système complet de dialogue qui est actuellement utilisé pour acquérir un corpus oral.

Introduction

La conception de systèmes de dialogue oral homme-machine dans le cadre d'applications spécifiques, doit répondre autant que possible aux besoins et aux comportements des utilisateurs potentiels. L'étude de corpus correspondant à l'application se révèle alors être un excellent guide à condition que ces corpus soient aussi représentatifs que possible des différents types de dialogues entre de réels utilisateurs et un système utilisé.

Deux approches principales sont utilisées pour l'acquisition de corpus, la technique du Magicien d'Oz et l'utilisation d'un système prototype [2].

Ces deux approches d'acquisition de corpus ont été étudiées dans le cadre du projet ESPRIT Mask¹ [3]. Alors que les tests Magiciens d'Oz se sont avérés être particulièrement intéressants pour les développements et les tests successifs d'interfaces, ils ne permettaient pas l'acquisition de grandes quantités de données de parole significatives pour l'application. En nous basant sur cette expérience, nous avons utilisé la deuxième méthode pour acquérir le corpus destiné au développement et

¹ Demande d'informations sur les transports SNCF entre plus de 600 gares en France, telles que les horaires, tarifs, réductions, prestations et les réservations.

à l'évaluation des systèmes de compréhension de l'action B2 de l'AUPELF.

L'acquisition du corpus est faite à partir d'un système de dialogue construit de manière itérative. Un système initial a été développé à partir de la connaissance de la tâche et des informations linguistiques et dialogiques recueillies dans des applications similaires. Les données recueillies avec ce premier système ont été analysées afin d'améliorer et d'augmenter les capacités du système et d'obtenir une meilleure couverture de la tâche. Une partie de ces données peut être utilisée pour l'évaluation du système, et ce aux différentes étapes de sa conception.

On peut distinguer plusieurs étapes nécessaires à l'enregistrement d'un corpus de dialogues homme-machine :

- la construction de modèles acoustiques et de langage adaptés à l'application pour le système de reconnaissance vocale,
- le développement d'une grammaire minimale de compréhension et d'interprétation des énoncés des utilisateurs pour la mise en place d'un module de compréhension,
- la mise en place d'un module de dialogue préliminaire permettant une interaction homme-machine,
- l'intégration de la base de données d'informations au système de dialogue,
- la transcription et l'étiquetage des données enregistrées (transcription orthographique, classification des énoncés, annotations des dialogues),
- l'organisation et l'archivage des données.

Description de la tâche (cadre B2)

Le cadre d'application retenu dans l'ARC B2 est le renseignement touristique autour de la gare Saint-Lazare à Paris. Le système informe sur les moyens de transport dans le quartier ainsi que sur les hôtels, restaurants magasins et cinémas des environs de la gare, les monuments et musées à Paris et les services en gare.

Les informations disponibles portent sur :

- hôtels (prix, adresse, prestations, catégories, trajet)
- restaurants (prix, adresse, prestations, types, trajet)

objet: hôtel
nom: astotel
tarif: de 150FF à 450FF
horaire: 6h-23h
prestation: chambre double (simple) avec douche (bain)
categorie: 3
type: -
description: ...
adresse: 29, rue Caumartin, 01-47-42-95-95
trajet: prendre sortie Rome, 1ère a droite.

Figure 1: Exemple de définition d'un objet de la base d'information disponible.

- monuments (adresse, horaires, descriptif, trajet)
- musées (adresse, horaires, prix, descriptif, trajet)
- magasins (adresse, horaires, prix, descriptif, trajet)
- cinémas (adresse, horaires, prix, descriptif, trajet)
- services en gare (information, café, pharmacie...)

N'ayant pas accès à une base de données contenant ce type d'informations, il a été nécessaire d'en construire une. Celle-ci contient entre autres des informations sur les tarifs, horaires, prestations, trajet d'une dizaine de chaque catégorie d'informations traitées. La Figure 1 montre un exemple du contenu et de l'organisation de la base.

Les services en gare sont traités différemment, ils renvoient à un plan de la gare. Des informations spécifiques à ces services seront fournies ultérieurement.

Les scénarii proposés aux sujets portent sur les différents objets contenus dans la base et sur leurs attributs. Il s'agit de de demandes d'informations sur :

- la localisation des services de la gare,
- les hôtels et restaurants environnants en fonction de leurs tarifs et de leurs commodités,
- le trajet à faire pour aller d'un lieu à un autre dans la gare ou bien de la gare à un lieu extérieur,
- les horaires et jours d'ouverture des hôtels, musées, restaurants, magasins et cinémas.

Description générale du système

La Figure 2 contient le synoptique du système de dialogue. Les principaux composants sont le module de reconnaissance, l'analyseur sémantique, le gestionnaire de dialogue, le module de recherche d'informations (pour l'accès à la base de données), et le générateur de réponses en langage naturel.

Le logiciel de reconnaissance traite le signal en temps réel sur un processeur standard de type RISC. Des modèles statistiques sont utilisés au niveau acoustique et linguistique. Les modèles acoustiques indépendants du locuteur ont été estimés sur des enregistrements d'un grand nombre de locuteurs. Des modèles de phones contextuels sont utilisés pour tenir compte des variantes phonologiques observées dans des contextes phonétiques différents. Les

modèles de langages statistiques de type n -gramme ont été estimés sur les transcriptions des phrases collectées. Le lexique de reconnaissance contenant environ 1000 mots est représenté phonémiquement avec un ensemble de 35 phonèmes.

La sortie du système de reconnaissance est transmise au module de compréhension qui analyse l'énoncé au moyen d'une grammaire de cas pour en extraire le sens [1]. Le principal travail lors du développement du module de compréhension est l'écriture des règles de la grammaire de cas qui comprend également la définition des concepts significatifs pour la tâche et les mots-clés associés. Le rôle du module de dialogue est de guider l'utilisateur afin qu'il fournisse l'information nécessaire à la construction de requêtes pour le gestionnaire de la base de données. Les réponses en langage naturel sont générées à partir du schéma sémantique et de l'information extraite de la base de données. Actuellement, le mode de présentation des informations est assez simplifié. Notre objectif à terme est de fournir des réponses en utilisant différents modes (tableaux, images, photos, vidéos...) selon le contexte et le contenu de l'information.

Le corpus homme-machine

Afin de permettre la construction du modèle de langage, un premier corpus de plus de 1000 phrases écrites a été recueilli. La Tableau 1 montre l'état du corpus en terme de taille de lexique et de nombre de dialogues réalisés.

#sujets	9
#dialogues	320
#phrases	1149
#total de mots	8393
#mots distincts	825

Table 1: Etat du corpus écrit

Ce premier corpus a permis la construction du modèle de langage nécessaire à la mise en place d'un premier système de reconnaissance.

Il a ainsi été possible de commencer les enregistrements avec des sujets.

Protocole d'enregistrement des données

Les sujets qui utilisent le système prototype ne reçoivent pas d'explications quant aux possibilités de celui-ci. L'expérimentateur leur explique le fonctionnement du système (appuyer sur une touche pour que la demande soit enregistrée par exemple) et leur présente des scénarii à réaliser. En milieu de session, il est demandé aux sujets de réaliser un scénario libre. Aucune contrainte de formulation ne leur est imposée. Ceci permet de recueillir des corpus riches d'un point de vue linguistique, dialogique et comportemental face à une machine en situation réelle de dialogue.

La Figure 3 montre des exemples de scénarii proposés aux sujets.

Chaque locuteur résoud environ 5 scénarii. Tous les dialogues enregistrés sont transcrits orthographiquement. La

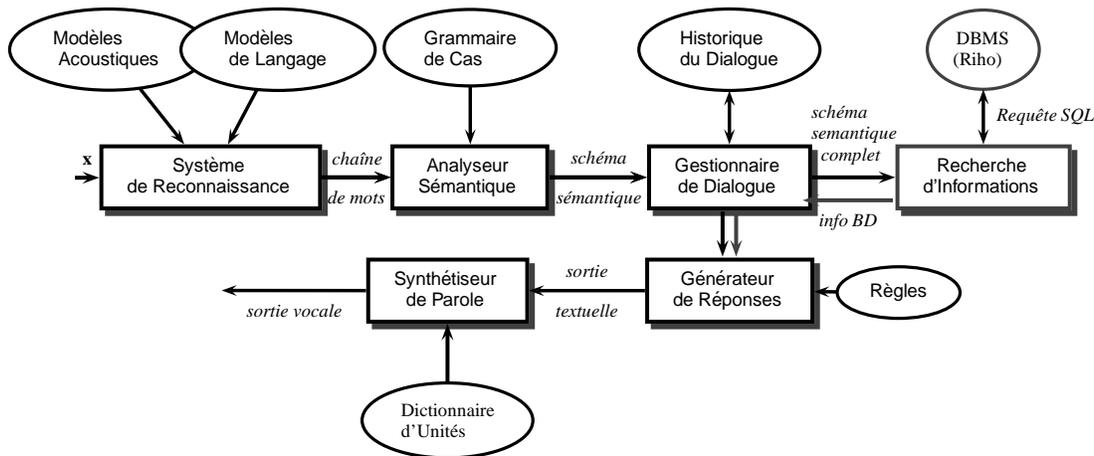


Figure 2: Système de dialogue d'informations à partir d'un énoncé oral. (x est le signal vocal et w est la suite de mots délivrée par le système de reconnaissance).

A- Vous restez 3 jours à Paris. Vous connaissez mal la ville et souhaitez trouver un hôtel près de la gare.
B- Vous disposez de 2 heures environ avant votre train. Vous souhaitez déjeuner avant de le prendre. Vous préférez la cuisine indienne.
C- Vous souhaitez visiter le musée du Louvre. Le musée est-il ouvert ? Quel est le meilleur moyen pour y aller.

Figure 3: Exemples de scénarii proposés.

durée de chaque dialogue, le nombre d'échanges par dialogues, les temps de réaction du système et du locuteur sont notés.

Nous avons enregistré environ 400 phrases avec 13 sujets. Le Tableau 2 montre l'état du corpus.

#sujets	13
#dialogues	82
#phrases	402
#total de mots	3926
#mots distincts	581
#mots réels	558

Table 2: Etat du corpus oral

Le nombre moyen de mots par phrase est d'environ 9. La plus courte comporte 1 mot et la plus longue 42. Ces 402 phrases s'intègrent dans 82 dialogues ce qui représente environ 5 phrases par dialogue.

Conclusion

Dans cet article nous avons décrit notre démarche pour l'acquisition de corpus de dialogues homme-machine dans le cadre de l'action B2 de l'AUPELF.

Les principales étapes préalables à la constitution de ce corpus sont :

- la définition du domaine d'application et des fonctionnalités de base du système,
- la définition des concepts et le développement d'un premier système de compréhension,

- la définition du vocabulaire de base du système et la construction du premier modèle de langage à partir d'énoncés écrits.

Un premier système a été mis en place et a permis de collecter environ 400 phrases de 13 sujets. Ces données sont actuellement analysées et vont nous permettre d'étendre les fonctionnalités du système et en particulier d'ajouter de nouveaux concepts aux modules de compréhension et de génération.

Références

- [1] S. Bennacef, H. Bonneau-Maynard, J.L. Gauvain, L. Lamel, W. Minker, "A Spoken Language System For Information Retrieval," *ICSLP'94*.
- [2] L. Lamel, S. Rosset, S. Bennacef, H. Bonneau-Maynard, L. Devillers, J.L. Gauvain, "Development of Spoken Language Corpora for Travel Information" *EUROSPEECH'95*.
- [3] A. Life, I. Salter, J.N. Temem, F. Bernard, S. Rosset, S. Bennacef, L. Lamel, "Data Collection for the Mask Kiosk: WOz vs Prototype System" *ICSLP'96*.