TALN 2004, Session Poster, Fès, 19-21 avril 2004

Prédiction d'actes et attentes en dialogue : expérience avec un assistant virtuel simulé

Yannick Fouquet

Laboratoire CLIPS-IMAG – Université Joseph Fourier
B.P. 53, 38041 Grenoble cedex 9, France
Yannick.Fouquet@imag.fr

Résumé – Abstract

Dans cet article, nous présentons une plate-forme de test et de recueil de dialogue oral homme-machine. Dans son architecture générale, des magiciens d'Oz simulent la compréhension des énoncés des utilisateurs et le contrôle du dialogue. Puis, nous comparons, dans un tel corpus, la prédiction statistique d'acte de dialogue avec les attentes du locuteur.

This paper presents a platform for testing and building human-computer spoken dialog system. In the general architecture of the platform, understanding and dialog management are simulated. Thus, comparison between statistic act prediction and expectation will be made. First results obtained show a credible way of capturing and annotate spoken dialogs.

Keywords – Mots Clés

Dialogue, attentes, magicien d'Oz, pragmatique, analyse, statistique
Dialog, expectations, magicien d'Oz, pragmatics, analysis, statistics

1 Introduction

Les attentes du locuteur semblent une alternative intéressante à la simple prédiction d'actes de dialogue dans le cadre du dialogue oral homme-homme (DOHH) (Fouquet, 2002). Nous étudions ici ces attentes en dialogue oral homme-machine (DHM) finalisé dans le domaine d'un Portail Vocal d'Entreprise (PVE). Nous présentons un système de DHM dans lequel s'intègrent les attentes des interactants. Une expérimentation à base de magicien d'Oz a permis de collecter des données linguistiques. Nous abordons la méthodologie adoptée puis l'architecture générale du système dans laquelle la compréhension et le contrôleur de dialogue sont simulés. Nous présentons ensuite la comparaison entre la prédiction statistique d'acte de dialogue et notre approche orientée attentes du locuteur dans le cadre du DHM.

2 Méthodologie

Le développement du système de DHM oral PVE suit la méthodologie développée par (Rouillard, Caelen, 1998) et (Rouillard, 2000). Plus de 800 dialogues inter-humains entre des secrétaires et leurs interlocuteurs ont été enregistrés. Les dialogues les plus représentatifs des tâches habituellement dévolues à des secrétaires ont été repérés puis analysés afin de mettre en évidence leurs particularités et leurs parties génériques. 3 tâches très fréquentes ont été soulignées : joindre une personne, prendre un rendez-vous et réserver une salle. 3 autres tâches, considérées comme importantes pour un assistant virtuel, ont été rajoutées : gérer un agenda partagé, recevoir une information et envoyer un document. Avec la transcription de 44 dialogues, choisis pour leur adéquation avec les 6 tâches sélectionnées, nous avons défini le vocabulaire et des classes propres à l'application. Un premier modèle de dialogue, issu de l'analyse des transcriptions, offre 122 phrases typiques que le système devra être capable de dire. Des variantes ont aussi été introduites pour obtenir des dialogues plus naturels. En outre, les phrases contiennent 12 paramètres dépendants de la tâche : les titre, nom et coordonnées de l'utilisateur, la date et l'heure. L'analyse de tels dialogues étant insuffisante pour créer un système de DHM, une plate-forme magicien d'Oz, développée pour simuler le comportement du système (contrôle du dialogue et de la tâche), a permis d'élaborer un corpus de DHM oral.

3 Attentes

Fondée sur la théorie des actes de langage de (Austin, 1962) et (Searle, 1972) et sur celle des intentions, la théorie des attentes (Fouquet, 2002) met l'accent sur la compréhension par la machine des attentes de l'utilisateur au delà de la seule compréhension de l'énoncé en cours. Il existe un grand nombre de manières de laisser transparaître une attente et les attentes sont souvent non marquées linguistiquement. Les déterminer apporte des indices tant au niveau de l'interprétation de l'énoncé qu'au niveau du contrôle du dialogue.

Les attentes sont une **liste de réponses** que l'utilisateur est susceptible d'attendre lorsqu'il formule un énoncé. La notation, dérivée de (Greimas, 1966) et (Vandervecken, 1990), considère 8 actes (cf. Figure 1). L'acte de dialogue a pour forme : $\mathbf{F^{fs}_{U}(p)}$. F^{fs} est l'acte de langage. $_U$ est l'identifiant de l'utilisateur ($_A$, l'assistant pour la machine). (p) est une représentation logico-sémantique du contenu propositionnel.

	Acte assumé	Acte délégué
Acte actionnel	F : action	F^f : demande d'action
Acte communicationnel	F^s : information	F^{fs} : demande d'information
Acte engageant actionnel	F^d : engagement	F^{fd} : demande d'engagement
Acte engageant statif	F^p : possibilité, invite	F^{fp} : demande de possibilité

Figure 1 : Notre classification en actes de langages

4 Architecture générale du système

Le système de dialogue est réalisé autour d'une architecture client-serveur avec 4 « serveurs »: la reconnaissance vocale, la synthèse text-to-speech et deux autres, simulés par magiciens d'Oz, l'interprétation en actes et le contrôle du dialogue (Figure 2) .

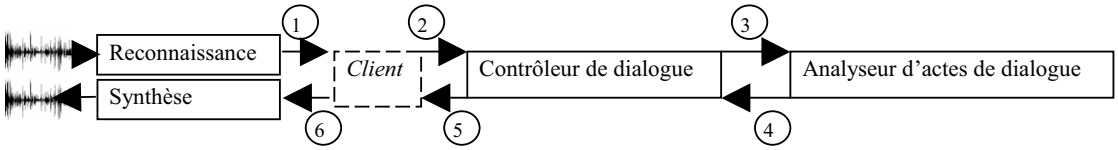

Figure 2 : Architecture du système de dialogue

Le client reçoit (1) l'**hypothèse** de reconnaissance d'un serveur de reconnaissance local et l'envoie (2) au contrôleur de dialogue. Le contrôleur de dialogue transmet (3) cette **hypothèse** à l'analyseur d'actes de dialogue qui envoie en retour (4) l'**acte de dialogue** correspondant à l'hypothèse reçue. Le contrôleur de dialogue choisit alors la **réponse** la plus appropriée et l'envoie (5) au client qui l'achemine (6) vers le serveur local de synthèse vocale.

Le **système de reconnaissance vocale** en parole continue utilise la boîte à outils Janus-III (Woszczyna et al., 1993). Le modèle acoustique dépend du contexte (Besacier et al., 2001). Pour le modèle de langage, le corpus d'apprentissage est issu des 44 dialogues obtenus en DOHH et d'une collecte de DHM réalisée via un premier système en magicien d'Oz résultant de leur analyse. Spécifique à la tâche, il contient 1110 phrases composées d'un vocabulaire de 1119 mots différents. Deux autres corpus ont été testés : un issu de Le Monde 97-01 et un collecté sur Internet. Nous les avons associés à notre vocabulaire enrichi des X mots les plus fréquents de Le Monde, X variant de 0 à 25000. 2 boîtes à outils ont été testées : CMU-SLM (www.speech.cs.cmu.edu/SLM_info.html) et SRI-LM (www.speech.sri.com/projects/srilm/). Le corpus d'évaluation est constitué de 63 phrases issues d'un pré-test du système durant lequel les sujets, via un micro-casque, devaient résoudre des tâches prédéfinies. Notre corpus spécifique croisé avec SRI-LM obtient les meilleures performances (39% d'erreurs de mot). Les magiciens n'ont demandé que 3 fois au sujet de répéter dans les 6 dialogues du pré-test.

L'**analyseur d'actes de langage** est simulé par magicien d'Oz pour annoter l'hypothèse de reconnaissance vocale en actes de dialogue. De l'identification de ces actes dépend l'ensemble du modèle. Le magicien annotateur reçoit la phrase reconnue par le système. Il choisit dans une liste l'acte le plus approprié. Pour ce choix, la première attente, déduite de l'acte précédent et qui maximise la probabilité de prédiction, est présélectionnée. Lorsque le magicien envoie l'acte correspondant à la phrase, les attentes sont mises à jour avec cette nouvelle occurrence. Les probabilités sont donc apprises dans le contexte de l'application.

Le **contrôleur de dialogue** est simulé par magicien d'Oz pour collecter et tester des dialogues homme-machine spécifiques aux tâches choisies. Le magicien contrôleur reçoit de l'utilisateur l'hypothèse de reconnaissance vocale et du premier magicien l'acte de dialogue correspondant à cette hypothèse. Il a alors les rôles de contrôleur de dialogue, contrôleur de tâche et générateur de réponse. Il choisit les paramètres adéquats dans l'interface de la tâche et sa réponse parmi les 122 possibles et les envoie à la synthèse du client. Si l'utilisateur dit « Pourrais-je parler à X ? », le magicien coche X dans l'annuaire et répond (i) « X n'est pas disponible » ou (ii) « Je vous passe X » en sélectionnant la phrase correspondante.

Le diagramme de l'architecture des dialogues de DOHH a permis de les découper en phases. Les 122 phrases ont alors été classées par tâche, phase et variante. La première attente déduite de l'acte reçu est soulignée. Pour chaque réponse, la stratégie correspondante (directive, réactive ou coopérative) est mise en valeur de sorte que le magicien peut choisir sa stratégie : réactive pour (i) ou (ii) ; directive pour « Avez-vous essayé son poste direct ? » par exemple.

La **synthèse** Text-to-speech, qui produit un signal audio à partir du texte, est implantée comme serveur qui reçoit le texte du système, crée le fichier signal correspondant et le joue. Elle doit se faire rapidement pour minimiser le temps d'attente de l'utilisateur. Le système Mbrola TTS (tcts.fpms.ac.be/synthesis/euler), de qualité de synthèse vocale suffisante pour nous, nécessite environ 30% du temps de la phrase pour produire le fichier son correspondant.

5 Prédiction d'actes statistique et prédiction d'attentes en DHM

Nous avons montré dans (Fouquet, 2002) l'intérêt de prendre en compte les attentes en DOHH sur un corpus de renseignement touristique (Besacier et al., 2001). Nous montrerons ici, suivant la même méthodologie, l'intérêt de ces attentes en DHM à travers un magicien d'Oz.

Nous adoptons une démarche stochastique pour prédire l'acte subséquent le plus probable à partir d'un historique plus ou moins grand. Nous utilisons les probabilités d'actes de dialogue n-grams, ce qui nous donne, à partir de la formule de probabilités conditionnelles de Bayes : $P(a_{n+1} \mid a_1,\ldots,a_n) = P(a_1,\ldots,a_n, a_{n+1}) / P(a_1,\ldots,a_n)$. Les probabilités sont alors estimées par des techniques de fréquence relative. Pour prédire le $i^{ème}$ acte de dialogue a_i, les *n-1* actes précédents déterminent le plus probable par la formule : $a_i = \text{argmax}_a\ P(a|a_{i-1},a_{i-2},\ldots,a_{i-n+1})$. Nous n'avons pas pu collecter un très grand nombre de données pour estimer correctement les probabilités. Nous traitons alors le cas de l'entrée non attendue et celui de la prédiction de plusieurs actes avec la même probabilité en utilisant le modèle (n-1)-gram et récursivement.

La prédiction d'attentes est dérivée de la prédiction d'actes. Dans l'historique, seuls nous intéressent les énoncés où l'interlocuteur pose une attente : les cas de demande d'action, d'information ou d'engagement. Les règles de gestion de ces attentes suivent la gestion des buts (Fouquet, 2002). Nous ne considérons pour l'expérience que l'attente la plus probable. Au moment de prédire un acte de A, nous allons donc vérifier que son historique contient un acte posé par B et portant une attente et comparer, si tel est le cas, l'acte prédit et celui posé.

6 Résultats

86 dialogues oraux homme-machine ont été collectés. Une nouvelle tâche (Ouverture), générique aux dialogues étudiés, a été introduite. Elle concerne la première phrase du dialogue et les transitions inter-tâches dans un même dialogue (« Voulez-vous autre chose ? »). Les dialogues sont alors répartis en 105 tâches et deux parties : apprentissage (env. 1500 actes, 65 dialogues et 80 tâches) et test (env. 400 actes, 21 dialogues et 25 tâches) (Figure 5) :

Tâches	Apprentissage	Test	Total
Standard / Redirection	16	5	21
Rendez-vous	9	3	12
Réservation de salle	12	3	15
Communication d'information	11	3	14
Gestion d'un agenda	15	5	20
Envoi de document	17	6	23
Ouverture	65	21	86

Figure 5 : Description du corpus

Plusieurs tailles d'historique ont été testées, des uni-grammes (simple répartition statistique) aux 8-grams (7 actes précédents déterminant le 8ième). L'ouverture, prédictible à 100% (l'agent commence par « Bonjour »), a été ôtée des statistiques mais sans amélioration significative.

Historique	Actes				Attentes			
	Assistant Virtuel		Utilisateur		Assistant Virtuel		Utilisateur	
	Taux	Nb	Taux	Nb	Taux	Nb	Taux	Nb
n=1	39.3	178	49.2	244	39.3	178	49.2	244
n=2	55.6	178	52.1	244	69.3	49	91.8	97
n=3	48.3	178	50.4	244	53.2	62	61.6	130
n=4	51.1	178	49.2	244	50.7	73	56.8	162
n=5	51.7	178	46.3	244	53.6	84	52.2	186
n=6	51.7	178	45.9	244	53.5	86	49.5	200
n=7	51.1	178	46.3	244	52.8	91	48.4	215
n=8	50.0	178	46.3	244	52.2	92	48.2	220

Figure 6 : Taux (%) de prédictions d'acte et de prédiction d'attentes et nombre à prédire.

Les résultats (Figure 6) de prédiction d'acte sont un peu moins fiables que ceux de Verbmobil (Reithinger, Maier, 1995) : 40,3% pour n=1, 59.6% pour n=2, 71,9% pour n=3. Cependant, leur corpus mono-tâche contient environ 7200 actes de dialogue. En outre, nos résultats les corroborent : un grand historique est inutile, le taux le meilleur étant souvent trouvé pour n=2.

Les résultats de prédiction d'attentes montrent une amélioration par rapport à la prédiction d'acte. Les énoncés portant une intention, même sans la déterminer, permettent de prédire plus facilement l'acte qui suivra. Une bonne gestion des attentes permet alors de contrôler le reste du dialogue. Dans ce cadre de DHM oral finalisé, la prise en compte des attentes de l'interlocuteur permet donc, comme en DOHH, d'améliorer le contrôle du dialogue.

L'acte de l'utilisateur est bien prédit par les attentes. Plutôt réactif, il se laisse guider par les questions de l'assistant virtuel et répond souvent aux attentes de ce dernier. La théorie des attentes peut donc apporter une aide significative à l'interprétation et à l'annotation. L'acte de l'assistant est moins bien prédit. S'il répond souvent directement à la 1ère attente de l'utilisateur en étant réactif ou coopératif, sa stratégie directive lui indique de répondre également à la 2ième attente la plus probable, la demande de précisions. L'utilisateur s'attend d'ailleurs fortement à cette dernière en début de dialogue car il sait qu'un certain nombre de paramètres manque pour accomplir la tâche pour laquelle il appelle. Tenir compte également de ces autres attentes semble alors important pour permettre un meilleur contrôle du dialogue.

7 Conclusions et perspectives

Nous avons présenté les modules développés pour l'annotation de parole et l'élaboration d'un corpus de 86 dialogues oraux homme-machine. Dans les modules simulés par magicien d'Oz, la théorie des attentes permet d'améliorer l'interprétation en actes de dialogue et le contrôle du dialogue en prédisant l'acte le plus attendu et ceux moins attendus mais non négligeables. Cette théorie offre aussi une aide à l'annotation par la prédiction de l'acte subséquent.

L'analyse statistique et linguistique des dialogues permettra le test et la mise à jour du système. Elle servira également à l'automatisation de l'interprétation (pour une tâche spécifique) et d'un contrôleur générique usant tous deux des attentes en dialogue.

Remerciements

Transcription, annotation, architecture des 44 dialogues réels, 122 énoncés du contrôleur, magiciennes : A.-C. Descalle, S. Hollard. Modèles de langage : B. Bigi. n-ième passe et relecture, magicienne : A. Vidal. Autres magiciens : Multicom. Merci de leur aide précieuse.

Références

AUSTIN J.L. (1962), *How To Do Things With Words*, Oxford, U.P.

BESACIER L., BLANCHON H., FOUQUET Y., GUILBAUD J.P., HELME S., MAZENOT S., MORARU D., VAUFREYDAZ D. (2001), Speech Translation for French in the NESPOLE ! European Project, Actes de *Eurospeech*.

FOUQUET Y. (2002), Une modèle de dialogue par les attentes du locuteur, Actes de *TALN*.

GREIMAS A.J. (1966), *Sémantique structurale*, Paris, Seuil.

REITHINGER N., MAIER E. (1995), Utilizing statistical dialogue act pro-cessing in verbmobil. Actes de 33rd Annual Meeting of the *Association for Computational Linguistics*, pp. 116-121.

ROUILLARD J. (2000), Hyperdialogue sur Internet ; le système Halpin, Thèse informatique.

ROUILLARD J., CAELEN, J. (1998), Etude Du Dialogue Homme-Machine En Langue Naturelle Sur Le Web Pour Une Recherche Documentaire, Deuxième Colloque International Sur *L'apprentissage Personne-Système*, Caps'98, Caen, Juillet 98.

SEARLE J.R (1972), *Les Actes De Langage : Essai De Philosophie Du Langage* (Trad. Française Par H. Pauchard), Paris, Hermann.

VANDERVEKEN D. (1990), *La logique illocutoire,* Bruxelles, Mandarga éd.

WOSZCZYNA M., COCCARO N., EISELE A., LAVIE A., MCNAIR A., POLZIN T., ROGINA I., ROSE C., SLOBODA T., TOMITA M., TSUTSUMI J., AOKI-WAIBEL N., WAIBEL A., WARD W. (1993), "Recent Advances in JANUS : A Speech Translation System". *Eurospeech*, Vol. 2, pp. 1295-1298.