

Inattendus structurels et communication orale finalisée : influence de la tâche et du contexte interactif

Sabine Letellier-Zarshenas, Pascale Nicolas, Jérôme Goulian et Jean-Yves Antoine

VALORIA — Université de Bretagne Sud — EA 2593
IUP Vannes, rue Yves Mainguy, 56000 Vannes — Mél : {Nom.Prenom}@univ-ubs.fr

1. PROBLEMATIQUE

La Communication orale Homme-Machine (CHM orale) a atteint une maturité que traduit l'apparition de premiers systèmes opérationnels grand public (systèmes de réservation ferroviaire par exemple [1]). En dépit de ces avancées significatives, les systèmes actuels restent néanmoins toujours dédiés à une application très précise (CHM finalisée). La question de la portabilité d'une application à un autre n'étant pas encore surmontée [2], la conception de tout nouveau système nécessite le redéveloppement, par apprentissage automatique sur corpus, de modèles de langage, de compréhension et de dialogue spécifiques. N'étant pas encore en possession du système réel, on est dans l'obligation de recréer ces situations d'interaction en laboratoire afin d'obtenir des corpus qui soient aussi proches que possible de la réalité [3].

Deux approches sont alors envisageables. La première repose sur la technique du *bootstrap*. Elle consiste à construire rapidement un système primitif que l'on améliore progressivement par une succession de tests en situation réelle d'interaction Homme-Machine. Pour développer le premier prototype du système, on élabore à la main un corpus artificiel. La réussite de cette approche repose donc sur la qualité de cette première simulation. L'expérience montre que cet objectif est réalisable dans le cas d'applications très finalisées. L'extension de cette technique à des domaines applicatifs plus riches n'est cependant pas garantie.

La seconde approche consiste au contraire à procéder à une analyse préalable des usages afin de cerner précisément les pratiques des utilisateurs qui doivent être modélisées au sein du système [3]. On procède alors au recueil de corpus *pilotes* (interaction réelle Homme-Homme) ou de Magicien d'Oz (interaction réelle avec machine simulée). Cette analyse des usages au plus près de l'application réelle est essentielle. Diverses études différentielles ont en effet montré que le comportement langagier de l'utilisateur varie sensiblement en fonction du contexte d'interaction [4].

Dans cette optique, cet article présente une étude expérimentale portant sur l'usage des répétitions/corrections et des ellipses en CHM orale. Ces deux classes d'inattendus structurels, qui posent problème à l'analyse automatique, sont essentielles à la naturalité du langage oral spontané [5]. Ils sont étudiés ici dans trois situations interactives différentes. Les résultats que nous présentons dénotent l'extrême variabilité des usages d'une application à l'autre. Ils montrent en outre l'intérêt des études de corpus spécifiques en CHM orale.

2. PRESENTATION DE L'ETUDE

Nous avons choisi de confronter trois situations interactives différentes issues de deux corpus pilotes (Air France et Murol) et d'un corpus (Levelt) issu d'une expérience de magicien d'Oz :

- Le corpus **Air France** (AF), recueilli dans le cadre du projet DALI [6], réunit un ensemble de conversations téléphoniques entre un centre de réservation aérienne et différents clients. Le degré d'interactivité est celui de la conversation téléphonique. Le dialogue reste cependant contenu, l'hôtesse étant tenue à une certaine réserve. Ce corpus est composé de 35 dialogues représentant 1639 tours de parole.

- Le corpus **Murol** du laboratoire CLIPS-IMAG [7] réunit deux interlocuteurs simulant une conversation téléphonique entre un touriste et un syndicat d'initiative. Un scénario a été conçu afin de favoriser l'apparition de situations de négociation. Il s'agit donc d'un dialogue finalisé assez peu coopératif qui témoigne d'une interactivité très marquée. La partie analysée du corpus comporte trois dialogues représentant 583 tours de parole.
- Le corpus **Levelt**, de l'Institut de la Communication Parlée, correspond à une application de conception architecturale assistée par ordinateur. La tâche consiste à concevoir un plan d'intérieur par commande vocale. Elle est par conséquent très répétitive. Le système, simulé par magicien d'Oz, se contente de réagir aux directives du locuteur. Cette situation est donc très faiblement interactive. La partie analysée du corpus comporte 5 dialogues représentant 706 tours de parole.

Du point de vue de l'interaction, ces corpus peuvent se distinguer suivant plusieurs critères. Nous en avons retenus trois, présentés dans le tableau ci-dessous.

Tableau 1 — Caractérisation du matériel

	Taille (tours parole)	Degré de finalisation	Répétivité	Interactivité
AF	1639	très élevée	très faible	élevée
Murol	583	modérée	faible	très élevée
Levelt	706	très élevée	très élevée	très faible

L'étude que nous avons menée a consisté à mesurer la fréquence d'apparition des ellipses et corrections/répétitions dans les corpus présentés. A la différence de l'écrit, il est difficile de segmenter les productions orales en phrases [5]. Dans l'optique d'un traitement automatique, nous avons choisi le tour de parole comme unité de comptabilisation. Les fréquences d'apparition sont calculées en terme de pourcentage d'énoncés porteurs du phénomène considéré. La dispersion des données est analysée en terme d'écart-type de classes, chaque classe correspondant à un dialogue.

3. RESULTATS

Le tableau 2 ci-dessous présente les résultats expérimentaux obtenus.

Ellipses — Nous observons un taux relativement élevé d'ellipses dans les corpus Air France et Levelt, par opposition au corpus Murol. L'analyse en écart-type montre que ces résultats sont assez significatifs. Ils s'expliquent avant tout par les degrés de finalisation différents des corpus.

Tableau 2 — Fréquence d'apparition des ellipses et répétitions/corrections

	Ellipses			Répétitions / corrections		
	AF	Murol	Levelt	AF	Murol	Levelt
Fréquence	27,6%	9,4%	18,7%	9,3%	34,8%	13,7%
Ecart-Type	7,9%	5,2%	9,0%	6,6%	3,3%	10,2%

Dans le cas du corpus Air France, la tâche est très finalisée (réservation) et se résume pour le client à l'obtention d'un nombre de paramètres limités (prix, horaire, vol...). Ce sont ces mêmes informations ciblées qui sont fournies à l'hôtesse sur ordinateur. Ainsi, la tâche est simple et les intentions des interlocuteurs sont claires, ce qui facilite l'usage de structures elliptiques. En outre l'hôtesse reprend souvent une partie des informations afin de meubler l'attente. On observe aussi de nombreuses confirmations elliptiques chez les deux interlocuteurs.

Au caractère finalisé du corpus Levelt s'ajoute la nature très opérative et répétitive de la

tâche. On observe ainsi un usage croissant des structures elliptiques au fil du discours, l'utilisateur ne précisant finalement que les paramètres qui varient d'une action à l'autre.

En revanche, le corpus Murol, moins finalisé, donne lieu à des échanges sensiblement plus riches et variés. Les intentions des interlocuteurs évoluent fréquemment, ce qui explique ce taux moins élevé.

Répétitions/corrections — Ici, nous observons au contraire un taux élevé de ces phénomènes dans le corpus Murol par opposition aux deux autres. L'analyse en écart-type montre que ces résultats sont significatifs. Ils s'expliquent ici par le degré d'interactivité des différents corpus.

Dans le corpus Murol, on constate une très forte interactivité qui se traduit par des fréquentes interruptions. Le caractère peu coopératif du scénario, de même que la richesse de la tâche sont la cause de cette interactivité. Il en résulte des échanges très spontanés qui sont accompagnés des multiples répétitions et corrections observées.

A l'opposé, le corpus Air France se traduit par une interactivité plus faible du fait de la nature plus contenue du dialogue, alors qu'il n'y a pas d'interaction avec le magicien d'Oz dans le corpus Levelt. Il n'en reste pas moins que l'élocution reste spontanée et donc source d'hésitations ou d'erreurs, ce que traduit la fréquence d'apparition non négligeable de ces phénomènes.

4. CONCLUSION

Les résultats expérimentaux que nous avons présentés montrent l'intérêt d'une analyse préalable des usages réels propres à une application. Nous avons ainsi noté l'usage important des répétitions dans certains cadres applicatifs, tandis que d'autres se caractérisent en revanche par un usage fréquent des procédés elliptiques. Or, le traitement de ces phénomènes répond à des contraintes différentes. Ce genre d'information s'avère donc très utile au prototypage des systèmes de CHM orale, d'autant plus que nos résultats semblent répondre à des causes identifiables a priori (interactivité, degré de finalisation).

Dans deux cas, ces expériences ont été menées sur des corpus d'interaction homme-homme. Or, certains travaux ont mis en évidence des variations sensibles des usages entre l'interaction inter-individus et la CHM [4]. Cependant, ces modifications ne dénotent que les capacités d'adaptation au système de l'utilisateur moyen. Or, cet effort d'adaptation n'est pas acceptable par tous, ce qui peut avoir des conséquences sociales ou économiques. L'exemple de l'annuaire téléphonique sur Minitel est très instructif de ce point de vue. En effet, l'ajout au système d'interrogation initial, très contraignant, de capacité de traitement robuste du langage naturel s'est traduit par un accroissement élevé des consultations de ce service. Preuve, s'il en fallait, de l'importance d'une analyse des usages spécifiques à chaque application.

REFERENCES BIBLIOGRAPHIQUES

- [1] ARISE consortium, (1996), *Automatic Railway Information System for Europe*, deliverable LE3-4229.
- [2] L. Hirschman, (1998), *Language understanding evaluations : lessons learned from MUC and ATIS*, LREC'98, Granada, Espagne, 117-122.
- [3] J. Caelen *et al.*, (1997), *Les corpus pour l'évaluation du dialogue homme-machine*, JST-FRANCIL'97, Avignon, France, 215-222.
- [4] M.A. Morel *et al.*, (1989), *Analyse linguistique de corpus*, Publ. Sorbonne Nlle, Paris.
- [5] C. Blanche-Benveniste *et al.*, (1990), *"Le français parlé; études grammaticales"*, CNRS Editions, Paris, France
- [6] G. Sabah, (1994), *Projet DALI*, rapport d'activité GDR-PRC CHM, 71-88.
- [7] M. Bessac, G. Caelen, (1995), *Analyses pragmatiques, prosodiques et lexicales d'un corpus de dialogue oral homme-machine*, JADT'95, Roma, Italie, 363:370.