



UNIVERSITÉ
TOULOUSE III
PAUL SABATIER



Université
de Toulouse



Présentation d'un article scientifique DOHM

Chris Ibrahim
Yoann Fleytoux

Incremental Dialog Processing in a Task-Oriented Dialog

- INTERSPEECH 2014

Auteurs: Fabrizio Ghigi, Maxine Eskenazi, M. Ines Torres, Sungjin Lee

À l'attention de Isabelle Ferrané

3ASRI - 18/12/2017

Plan

Contexte et objectif

Analyse du système précédent

Incremental Dialog Processing Strategy

Expérience

Résultats

Conclusion

Contexte et objectif

But du papier : montrer l'efficacité de l'IDP appliqué à un système réel fournissant des informations sur les horaires de transports et les itinéraires, dans le but d'améliorer le taux de succès d'une tâche, ce qui n'a pas encore été fait auparavant.

Analyse du système précédent

- Analyse d'un corpus de 100 dialogues d'utilisateurs et de leurs transcriptions (Octobre 2008 à Septembre 2009) pour établir des statistiques sur les énoncés longs et déduire les causes d'échecs.

Analyse du système précédent

Tableau 1 : Comparaison entre les dialogues dont les énoncés longs sont tous reconnus et ceux mal reconnus

	Correctly	Incorrect
Avg. Length (sec) (for Long Utterance)	4.41	5.01
Avg. Length (mots) (for Long Utterance)	4.96	5.72
Avg. #Values in parse(sec) (for Long Utterance)	2.61	2.67
Avg. Turn number (per Dialog)	13.33	13.21
#Long Utt. (Per Dialog)	1.49	2.21

La seule différence majeure est le nombre de longs énoncés.

Analyse du système précédent

Tableau 2 : Taux de succès de la tâche et sa longueur moyenne

	DLU (Dialogs with Long Utterances)	DILU (Dialogs with some Incorrect Long Utterances)	DCLU (Dialogs with only Correct Long Utterances)
Task Success	59.00%	45.68%	78.95%
Avg.Length	40.28	42.43	31.1

DCLU (Dialogue composé de longs énoncés tous corrects) avec un taux de succès élevés: les comprendre influence grandement la réussite de la tâche.

Analyse du système précédent

Tableau 3: Position des unités sémantiques dans l'énoncé basé sur les temps d'occurrence des résultats partiels

Correct slot position	
Beginning	30.16%
Middle	28.74%
Ending	41.09%

Près de 60% de l'information totale de l'énoncé est contenue dans les deux premières parties.

Analyse du système précédent

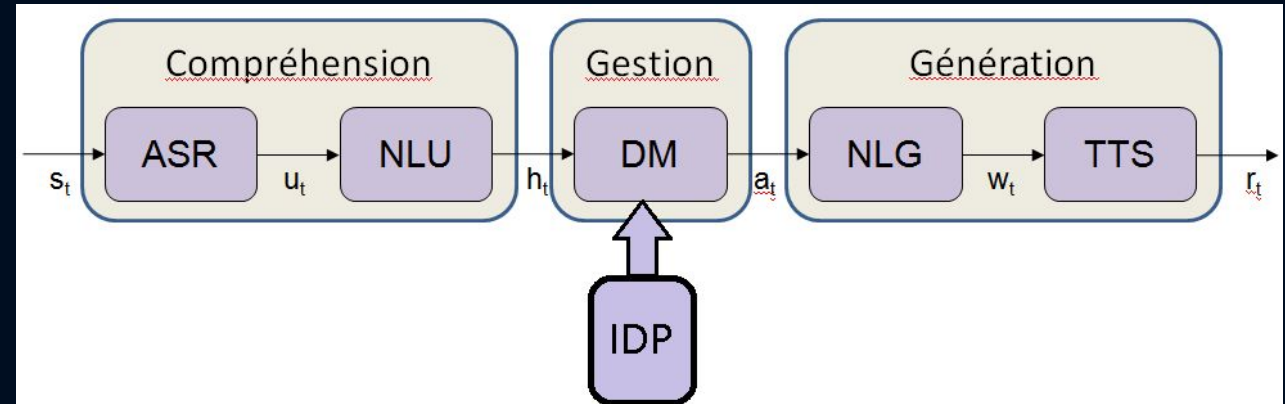
Tableau 4: Répartition des unités sémantiques erronés dans chaque partie l'énoncé long

Incorrect slot position	
Beginning	36.60%
Middle	47.62%
Ending	40.33%

Les informations les plus exploitables sont obtenues au début de l'énoncé.

Incremental Dialog Processing Strategy

- Paramètres de décision :
 - N derniers partiels égaux
 - durée d'un énoncé
 - score de confiance
 - durée de la pause finale



- Processus de décision :
 - enregistrement du temps d'arrivée du premier résultat partiel
 - Vérification de l'égalité des N derniers partiels, mesure de la durée de l'énoncé jusqu'à puis récupération des valeurs du score de confiance et de la durée de la pause suivante.
 - Interruption décidé si, les N derniers partiels ont le même *parse*, la durée de l'énoncé, celle de la pause finale, et le score de confiance dépassent un certain seuil.
- Selon les valeurs utilisées de ces paramètres, le système peut être plus ou moins agressif ou plus ou moins réactif.

Incremental Dialog Processing Strategy

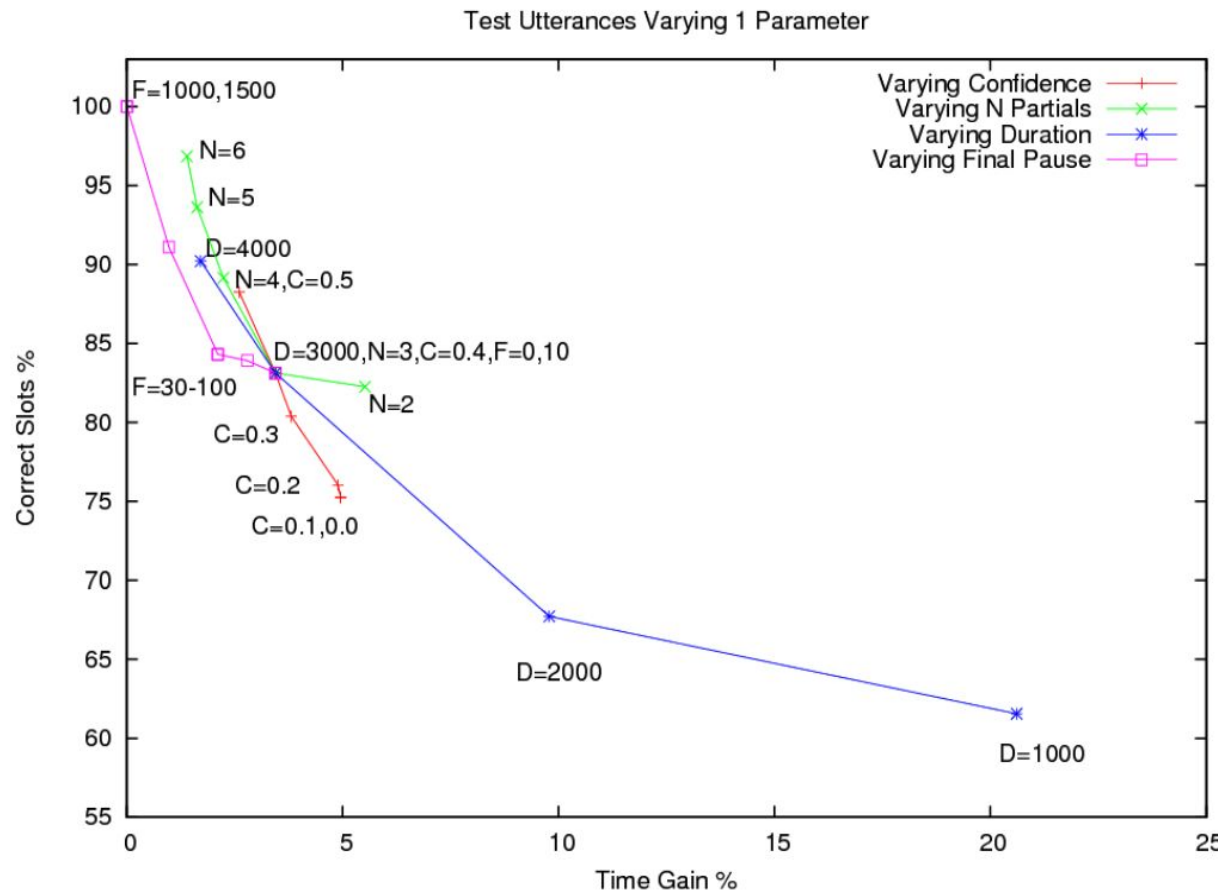


Figure 3: How the percentage of correct slots and time gain vary.

Courbe 1 : taux de reconnaissance des unités sémantiques et gains de temps associés lorsque l'on fait varier différents paramètres.

On voit qu'en changeant les différents paramètres, le gain de temps en pourcentage augmente mais diminuant le taux de reconnaissance des unités sémantiques.

Incremental Dialog Processing Strategy

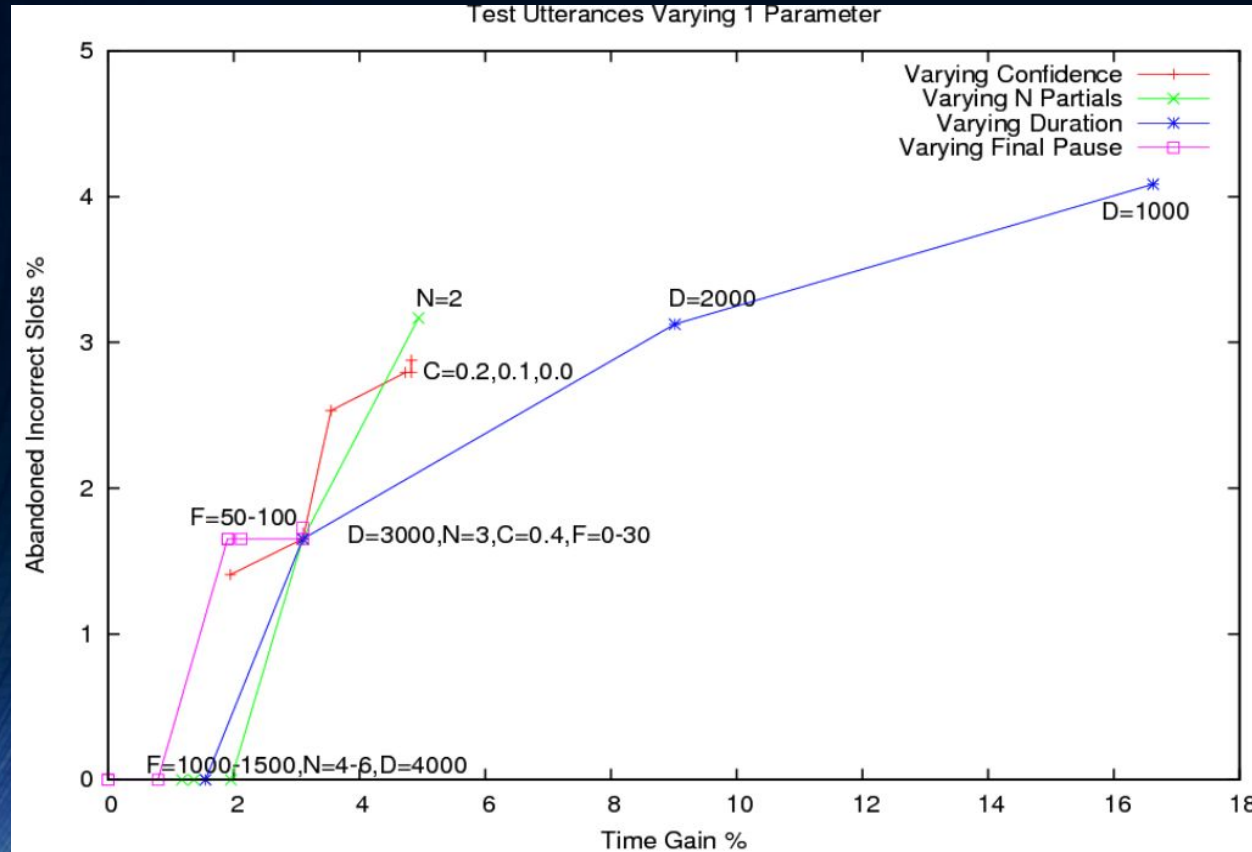


Figure 4: How the percentage of abandoned incorrect slots and time gain vary.

Courbe 2 : taux d'unités sémantiques mal-reconnues abandonnés et du gain de temps associés lorsque l'on fait varier différents paramètres.

On voit qu'au contraire ici qu'en changeant les différents paramètres, le gain de temps en pourcentage augmente avec le taux d'unités sémantique abandonnés mal reconnus.

Expérience

- Test réel sur des utilisateurs avec un système actif fournissant des informations sur les horaires de transports et les itinéraires un mois entier (Octobre 2013) puis comparaison avec résultat de l'ancien système (Septembre 2013).
- 7 métriques de comparaison :
 - ODSR (Observed Dialog Success Rate) ou taux de réussite de la tâche
 - ADL (Average Dialogue Length) ou durée moyenne d'un dialogue
 - ANLU (Average Number of Long Utterances) ou Nombre moyen d'énoncés longs
 - LUALU[N] (Long Utterance(s) after Log Utterance) ou N énoncé(s) long(s) après un énoncé long avec N=1,2 et 3

Résultats

Metric/Period	No IDP	With IDP
ODSR(Observed Dialog Success Rate)	66.67%	74.34%
ADL(Average Dialog Length)	18.76	21.15
ADD(Average Dialog Duration)	153.99	162.68
ANLU(Average Number Of Long Utterances)	2.93	2.21
LUALU[N](Long Utterances After Long Utterance) 1	25%	18%
LUALU[N](Long Utterances After Long Utterance) 2	6%	2%
LUALU[N](Long Utterances After Long Utterance) 3	2%	0%

Tableau 5: Comparaison des résultats avec et sans IDP

Légère augmentation de la durée de dialogue compensée par l'amélioration du taux de succès des dialogues.

Conclusion

- Certaines données ambiguës (manque d'unité, valeurs se référant à des choses différentes).
- Beaucoup de critères et métriques.
- Des résultats positifs et propositions de travaux futurs intéressants (mais pas encore publiés)
- Reprises de notions vues durant le cours.



Merci de votre attention

DES QUESTIONS ?

Références