

Concept Form Adaptation in Human-Computer Dialog

Svetlana Stoyanchev and Amanda Stent

Department of Computer Science

Stony Brook University

Stony Brook, NY 11794-4400, USA

svetastenchikova@gmail.com, amanda.stent@gmail.com

Abstract

In this work we examine user adaptation to a dialog system's choice of realization of task-related concepts. We analyze forms of the time concept in the *Let's Go!* spoken dialog system. We find that users adapt to the system's choice of time form. We also find that user adaptation is affected by perceived system adaptation. This means that dialog systems can guide users' word choice and can adapt their own recognition models to gain improved ASR accuracy.

1 Introduction

Considerable research has now demonstrated that human dialog partners exhibit lexical and syntactic *convergence*; that is, that in a human-human conversation the participants become more similar in their use of language over time (Brennan and Clark, 1996; Lockridge and Brennan, 2002; Pickering and others, 2000; Reitter et al., 2006). Several Wizard-of-Oz studies have also shown evidence of convergence in human-computer dialog (Branigan and others, 2003; Brennan, 1996; Gustafson and others, 1997).

In recent work, we examined user adaptation¹ to the system's choice of verb and preposition using the deployed *Let's Go!* spoken dialog system (Stoyanchev and Stent, 2009a). This was the first study to look at convergence with real users of a real dialog system and examined user adaptation to verbs and prepositions. The study described in this paper is a follow-on to our previous study.

¹In this paper, we use the term *adaptation* to indicate directional convergence, e.g. user adaptation to a system. We make no claims about the psycholinguistic models underlying this adaptation.

Here we look at user adaptation to the system's choice of realization of task-related concepts. In this paper, we: (1) Confirm our previous results showing that users adapt to the system's choice of words, using transcribed data rather than ASR output; (2) Show that adaptation can persist over time; and (3) Show that user adaptation is strengthened by perceived system adaptation.

In addition to providing further evidence of convergence in human-computer dialog, our results have implications for dialog system design. Currently, much dialog systems research is devoted to improving ASR accuracy, because this is a significant contributor to task success rates and to dialog length. One way to improve ASR accuracy is to use targeted language models. Since users adapt to the system's choices of realization for task-related concepts, we can predict the user's choice of realization and use this to adjust the ASR's language model, improving ASR accuracy specifically on concept words. Another way to improve ASR accuracy is to guide the user into using words that are likely to be recognized correctly (Hockey and others, 2003; Sheeder and Balogh, 2003; Tomko and Rosenfeld, 2006). Our results imply that if the designer of a dialog system wants to improve ASR accuracy, system prompts should be designed to use word choices that are more recognizable; and, when, possible, to be adaptive to the user's choice of form for task-related concepts.

2 System

We conducted our experiment using the *Let's Go!* deployed telephone-based spoken dialog system which provides information about bus routes in Pittsburgh (Raux and others, 2005). *Let's Go!* answers the phones at the transit authority's website outside of normal business hours. Its users are

SYS:	Welcome to the CMU Let's Go! Bus Information System. How may I help you?
USR:	I'd like to go to Forbes and Murray.
SYS:	To Forbes and Murray. Where are you leaving from?
USR:	Carnegie Mellon.
SYS:	From Carnegie Mellon. What time would you like to leave?
USR:	Ten thirty p.m.
SYS:	Leaving at ten thirty p. m.?
USR:	Yes
SYS:	There are five buses running between Carnegie Mellon and Forbes and Murray...

Figure 1: Sample dialog with *Let's Go!*

Time form	Example realizations	Frequency
TIME	four, five, six thirty...	31.1%
APM	four a.m., ten p.m., one fifteen p. m.	43.5%
POD	four in the morning/evening/afternoon/	4.6%
OCLOCK	five o'clock	16%
OTHER	four o'clock p. m., sixteen hundred hours	4.8%

Table 1: Time forms in users' utterances and their relative frequencies in one month of *Let's Go!* 2006 data.

naive callers from the general population with a real task to perform. In order to provide bus route information, *Let's Go!* elicits values for several task-related concepts: an optional bus route number, a departure place, a destination and a desired travel time. Each concept is explicitly confirmed. Figure 1 shows a sample dialog with the system.

In this work we investigate adaptation to the time concept because it has multiple different realizations, as shown in Table 1. This variability is not unique to time; however, it is the only task-related concept in *Let's Go!* that is not usually realized using named entities (which exhibit less variability).

3 Method

In order to study adaptation, we need to identify a *prime*, a point in the conversation where one partner introduces a realization. In *Let's Go!* the system always asks the user to specify a departure time. The user then typically says a time, which the system confirms (see Figure 1). We simulate an ASR error on the user's response to the system's time request, so that when the system confirms the departure time it confirms a time other than that recognized in the user's response. To make the system's error more realistic, the time in the simulated error is a time that is phonetically

close to the time (hour and minute) recognized in the user's response. The system's confirmation prompt is our *prime*.

The system runs in one of the three conditions: SYS_TIME, SYS_APM, or SYS_POD. In each condition it uses the corresponding time format (TIME, APM, or POD as shown in Table 1). TIME is the most frequent form in the 2006 *Let's Go!* corpus, but it is potentially ambiguous as it can mean either night or day. APM is the shortest unambiguous form. POD is longer and has a very low frequency in the 2006 *Let's Go!* corpus.²

We collected approximately 2000 dialogs with *Let's Go!* using this setup. We used the ASR output to identify dialogs where a time appears in the ASR output at least twice³. We manually transcribed 50 dialogs for each experimental condition. Some of these turned out not to contain mentions of time either before or after the system's time confirmation prompt, so we excluded them.

We examine whether the user adapts to the system's choice of form for realizing the time concept, both in the first time-containing post-confirmation utterance, and in the rest of the dialog (until the user hangs up or says "New query").

4 Results

In this section we first examine user adaptation to system's choice of time expression, and then look at how perceived system adaptation affects user adaptation.

4.1 User adaptation to system time form

If the user adapts to the system's time form, then we would expect to see a greater proportion of the system's time form in user utterances following the prime. We compare the proportion of three time forms (APM, TIME, and POD) in each system condition for 1) *Unprimed*, 2) *First After*, and 3) *All After* user's utterances, as shown in Table 2. *Unprimed* utterances are the user's time specification immediately prior to the prime (the system's confirmation prompt). *First After* utterances are user utterances immediately following the prime. *All After* utterances are all user utterances from the prime until the user either hangs up or says "New

²We would have liked to also include OCLOCK in the experiment. However, due to resource limitations we had to choose only three conditions.

³The most frequent user response to the system's request to specify a departure time is "Now"; we exclude these from our experiment.

Unprimed			
system/user	Usr:APM	Usr:TIME	Usr:POD
SYS_APM	25%	42%	8%
SYS_TIME	30%	52%	2%
SYS_POD	24%	49%	4%
First After			
system/user	Usr:APM	Usr:TIME	Usr:POD
SYS_APM	49%	29% ♠	2%
SYS_TIME	21% ♣	58%	0%
SYS_POD	29%	45%	5%
All After			
system/user	Usr:APM	Usr:TIME	Usr:POD
SYS_APM	63%	19% ♣	3%
SYS_TIME	21% ♣	50%	2%
SYS_POD	37% ♣	38%	4%

Table 2: Proportions of time forms in different system prompt conditions. The highest proportion among system conditions for each time form is highlighted. Occurrences of time forms other than the three examined time forms are excluded from this table. ♠ indicates a statistically significant difference from the highlighted value in the column ($p < .05$ with Bonferroni adjustment). ♣ indicates a statistically significant difference from the highlighted value in the column ($p < .01$ with Bonferroni adjustment).

query”. To test the statistical significance of our results we perform inference on proportions for a large sample.

APM There are no statistically significant differences in the proportions of Usr:APM⁴ forms in *Unprimed* utterances for the different system conditions. The proportion of Usr:APM forms in *First After* utterances is significantly higher in the SYS_APM condition than in the SYS_TIME condition ($p < .01$), although not significantly different than in the SYS_POD condition. The proportion of Usr:APM forms in the *All After* utterances is significantly higher in the SYS_APM condition than in both the SYS_TIME and the SYS_POD conditions ($p < .01$). We conclude that there is user adaptation to system time form in the SYS_APM condition.

TIME There are no statistically significant differences in the proportions of Usr:TIME forms in *Unprimed* utterances for the different system conditions. The proportions of Usr:TIME forms in the *First After* utterances in the SYS_TIME condition is significantly higher than that in the SYS_APM condition ($p < .01$), but not significantly higher than that in the SYS_POD condition. The same is true of Usr:TIME forms in the *All After* utter-

⁴Usr:time-form refers to the occurrence of the time-form in a user’s utterance.

condition	keep	adapt	switch	total
adaptive	81.8%	-	18.2%	33
non-adaptive	37.5%	29.1%	35.4%	48

Table 3: Proportions of user actions in *First After* confirmation utterances

ances. We conclude that there is user adaptation to system time form in the SYS_TIME condition.

POD We did not find statistically significant differences in Usr:POD forms for the different system conditions in either the *Unprimed*, *First After* or *All After* data. Because this is the *long unambiguous* form, users may have felt that it would not be recognized or that it would be inefficient to produce it.

Figure 2 illustrates the effect of user adaptation on time form for the SYS_APM and SYS_TIME conditions.

4.2 The effect of system adaptation on user adaptation

Sometimes the user happens to use the same form in their initial specification of time that the system uses in its confirmation prompt. This gives the illusion that the system is adapting its choice of time form to the user. We examined whether users’ perception of system adaptation affected user adaptation in *First After* confirmation utterances.

For this analysis we used only the dialogs in the SYS_APM and SYS_TIME conditions since the POD form is rare in the *Unprimed* utterances. We distinguish between three possible user actions following the system’s confirmation prompt: 1) *keep* - use the same form as in the unprimed utterance; 2) *adapt* - switch to the same form as in the system’s confirmation prompt; and 3) *switch* - switch to a different form than the one used in the system’s confirmation prompt or in the unprimed utterance.

Table 3 shows the proportions for each possible user action. In the *adaptive* condition users are twice as likely to keep the time form than in the *non-adaptive* condition (81.8% vs. 37.5%). This difference is statistically significant ($p < .001$). In the *non-adaptive* system condition users who change time form are slightly more likely to switch (35.4%) than to adapt (29.1%).

These results suggest that when the system does not adapt to the user, the user’s choice is unpredictable. However, if the system adapts to the user, the user is likely to keep the same form. This

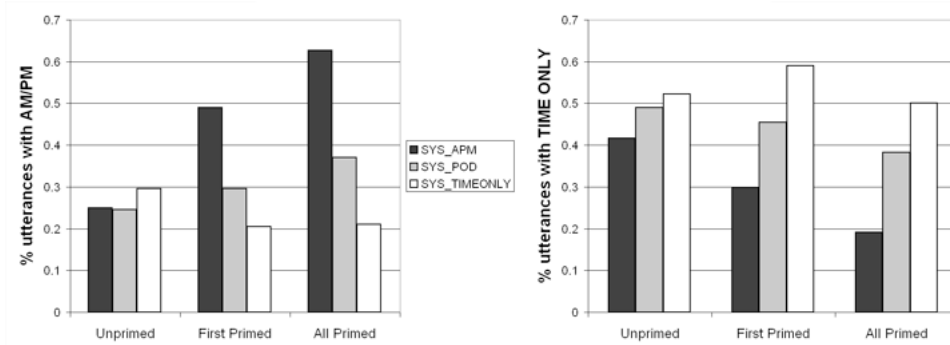


Figure 2: User Utterances with TIME_APM and TIME_ONLY.

means that if the system can adapt to the user when the user chooses a form that is more likely to be recognized correctly, that provides positive reinforcement, making the user more likely to use that felicitous form in the future. Furthermore, if the system does adapt to the user then it may be possible with high accuracy to predict the user's form for subsequent utterances, and to use this information to improve ASR accuracy for subsequent utterances (Stoyanchev and Stent, 2009b).

5 Conclusions and Future Work

In this paper, we analyzed user adaptation to a dialog system's choice of task-related concept forms. We showed that users do adapt to the system's word choices, and that users are more likely to adapt when the system appears to adapt to them. This information may help us guide users into more felicitous word choices, and/or modify the system to better recognize anticipated user word choices. In future work we plan to analyze the effect of ASR adaptation to user word choice on speech recognition performance in spoken dialog.

References

- H. Branigan et al. 2003. Syntactic alignment between computers and people: The role of belief about mental states. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society*.
- S. Brennan and H. Clark. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology*, 22(6):1482–1493.
- S. Brennan. 1996. Lexical entrainment in spontaneous dialog. In *Proceedings of ISSD*, pages 41–44.
- J. Gustafson et al. 1997. How do system questions influence lexical choices in user answers? In *Proceedings of Eurospeech*.
- B. Hockey et al. 2003. Targeted help for spoken dialogue systems: intelligent feedback improves naive users performance. In *Proceedings of EACL*.
- C. Lockridge and S. Brennan. 2002. Addressee's needs influence speakers' early syntactic choices. *Psychonomics Bulletin and Review*, 9:550–557.
- M. Pickering et al. 2000. Activation of syntactic priming during language production. *Journal of Psycholinguistic Research*, 29(2):205–216.
- A. Raux et al. 2005. Let's go public! taking a spoken dialog system to the real world. In *Proceedings of Eurospeech*.
- E. Reitter, J. Moore, and F. Keller. 2006. Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. In *Proceedings of CogSci*.
- T. Sheeder and J. Balogh. 2003. Say it like you mean it: Priming for structure in caller responses to a spoken dialog system. *International Journal of Speech and Technology*, 6:103–111.
- S. Stoyanchev and A. Stent. 2009a. Lexical and syntactic priming and their impact in deployed spoken dialog systems. In *Proceedings of NAACL*.
- S. Stoyanchev and A. Stent. 2009b. Predicting concept types in user corrections in dialog. In *Proceedings of the EACL Workshop on the Semantic Representation of Spoken Language*.
- S. Tomko and R. Rosenfeld. 2006. Shaping user input in speech graffiti: a first pass. In *Proceedings of CHI*.