# After Dialog Went Pervasive: Separating Dialog Behavior Modeling and Task Modeling

**Amanda J. Stent**
AT&T Labs - Research
Florham Park, NJ 07932, USA
stent@research.att.com

**Dialog Goes Pervasive** Until recently, many dialog systems were *information retrieval* systems. For example, using a telephone-based interactive response system a US-based user can find flights from United (1-800-UNITED-1), get movie schedules (1-800-777-FILM), or get bus information (Black et al., 2011). These systems save companies money and help users access information 24/7. However, the interaction between user and system is tightly constrained. For the most part, each system only deals with one domain, so the task models are typically flat slot-filling models (Allen et al., 2001b). Also, the dialogs are very structured, with system initiative and short user responses, giving limited scope to study important phenomena such as coreference.

Smart phones and other mobile devices make possible *pervasive* human-computer spoken dialog. For example, the Vlingo system lets users do web searches (information retrieval), but also connects calls, opens other apps, and permits voice dictation of emails or social media updates[1]. Siri can also help users make reservations and schedule meetings[2].

These new dialog systems are different from traditional ones in several ways; they are *multi-task, asynchronous, can involve rich context modeling*, and have *side effects in the "real world"*:

*Multi-task* – The system interacts with the user to accomplish a series of (possibly related) tasks. For example, a user might use the system to order a book and then say *schedule it for book club* - a different task (e.g. requiring different backend DB lookups) but related to the previous one by the book information. Multi-task interaction increases the difficulty of interpretation and task inference, and so requires new kinds of dialog model (e.g. (Lison, 2011)).

*Asynchronous* – the user may give the system a command (e.g. *Add Hunger Games with Mary for 3 pm*), and the system may follow up on that command an hour later, after considerable intervening dialog (e.g. *Mary texted you about the Hunger Games*). Because the dialog is multi-task, it is more free-flowing, with less clear start and end points but more opportunities for adaptation and personalization.

*Rich context modeling* – Mobile devices come with numerous sensors useful for collecting non-linguistic context (e.g. GPS, camera, web browser actions), while the semi-continuous nature of the interaction permits collection of rich linguistic context. So far, dialog systems have used this context only in limited ways (e.g. speech recognizer personalization). However, the opportunities for modeling human interaction behavior, including multi-modal interaction, are tremendous.

*Side effects "in the real world"* – the system (with input from the user) can cause changes in the state of the world (e.g. emails get sent, hotel rooms get booked). This increases the importance of grounding and agreement in the interaction. But it enables new kinds of evaluation, for example based on the number of successfully completed subtasks over time, or on comparing the efficacy of alternative system behaviors with the same user.

**Dialog Challenges and Task Challenges** The implications for research on dialog systems are clear. It is unsustainable to reimplement dialog behaviors for each new task, or limit the use of context to the

[1] www.vlingo.com

[2] http://www.apple.com/iphone/features/siri.html

*NAACL-HLT 2012 Workshop on Future directions and needs in the Spoken Dialog Community: Tools and Data*, pages 17–18,
Montréal, Canada, June 7, 2012. ©2012 Association for Computational Linguistics

most basic semantic representations. As the field moves forward, *dialog behavior modeling will be increasingly separated from task modeling* (Allen et al., 2001a; Allen et al., 2001b). Research on dialog modeling will focus on dialog *layers*, task-independent dialog behaviors such as (incremental) turn-taking, grounding, and coreference that involve both participants. Research on task modeling can focus on the design of task models that are agnostic to the types or forms of interaction that will use them, on general models for interactive problem-solving (Blaylock and Allen, 2005), and on rapid acquisition and adaptation of task models (Jung et al., 2009).

Within this space, there can be two types of (collaborative or competitive) "dialog challenge":

*Dialog layer-focused* – Participants focus on models for a particular dialog behavior, such as turn-taking, grounding, alignment, or coreference. Implementations cover both the interpretation and the generation aspects of the behavior. Evaluation may be based on a comparison of the implemented behaviors to human language behaviors (e.g. for turn-taking, inter-turn silence, turn-final and turn-initial prosodic cues), and/or on user error rates and satisfaction scores. An initial dialog layer-focused challenge could be on turn-taking (Baumann and Schlangen, 2011; Selfridge and Heeman, 2010).

*Task modeling focused* – This type of challenge will move from modeling individual tasks, to automatic acquisition and use of task models for interactive tasks in dialog systems. Future challenges of this type would build on this by incorporating (in order): (a) tasks other than information retrieval (e.g. survey tasks (Stent et al., 2008)); (b) task completion (tasks with subtasks that have side effects, e.g. purchasing a ticket after looking up a route); (c) task adaptation (during development, participants work with one task, and during evaluation, participants work with a different but related task); and (d) multi-task modeling. Participating systems could learn by doing (Jung et al., 2009), via user simulation (Rieser and Lemon, 2011), from corpora (Bangalore and Stent, 2009), or from scripts or other abstract task representations (Barbosa et al., 2011).

**Tools for the Community** It has never been easier (with a little Web programming) to rapidly prototype dialog systems as mobile apps, or to use them to collect data. To enable researchers to focus on dialog- and task-modeling rather than component development, AT&T is happy to offer its AT&T WATSON$^{SM}$ speech recognizer and Natural Voices$^{TM}$ text-to-speech synthesis engine in the cloud, through its Speech Mashup platform (Di Fabbrizio et al., 2009), to participants in dialog challenges. The Speech Mashup supports rich logging of both linguistic and non-linguistic context, and is freely available at *http://service.research.att.com/smm*.

## References

J. F. Allen, G. Ferguson, and A. Stent. 2001a. An architecture for more realistic conversational systems. In *Proceedings of IUI*.

J. F. Allen et al. 2001b. Towards conversational human-computer interaction. *AI Magazine*, 22(4):27–37.

S. Bangalore and A. Stent. 2009. Incremental parsing models for dialog task structure. In *Proceedings of EACL*.

L. Barbosa et al. 2011. SpeechForms - from web to speech and back. In *Proceedings of Interspeech*.

T. Baumann and D. Schlangen. 2011. Predicting the micro-timing of user input for an incremental spoken dialogue system that completes a user's ongoing turn. In *Proceedings of SIGDIAL*.

A. W. Black et al. 2011. Spoken dialog challenge 2010: comparison of live and control test results. In *Proceedings of SIGDIAL*.

N. Blaylock and J. F. Allen. 2005. A collaborative problem-solving model of dialogue. In *Proceedings of SIGDIAL*.

G. Di Fabbrizio, T. Okken, and J. Wilpon. 2009. A speech mashup framework for multimodal mobile services. In *Proceedings of ICMI-MLMI*.

H. Jung et al. 2009. Going beyond PBD: A play-by-play and mixed-initiative approach. In *Proceedings of the CHI Workshop on End User Programming for the Web*.

P. Lison. 2011. Multi-policy dialogue management. In *Proceedings of SIGDIAL*.

V. Rieser and O. Lemon. 2011. Learning and evaluation of dialogue strategies for new applications: Empirical methods for optimization from small data sets. *Computational Linguistics*, 37(1):153–196.

E. Selfridge and P. Heeman. 2010. Importance-driven turn-bidding for spoken dialogue systems. In *Proceedings of ACL*.

A. Stent, S. Stenchikova, and M. Marge. 2006. Dialog systems for surveys: The Rate-A-Course system. In *Proceedings of SLT*.