

Directions for Research on Spoken Dialog Systems, Broadly Defined*

Nigel G. Ward

University of Texas at El Paso

El Paso, Texas 79968, USA

nigelward@acm.org

Abstract

To increase impact and accelerate progress, the spoken dialog systems research community should work on four shareable things that will also engage and support sister fields of science and engineering.

1 To Reach Out to the VoiceXML Community, a Commercial-Dialogs Corpus

Although many people are frustrated with the commercial dialog systems they use every day, spoken dialog systems research has been only sporadically relevant to these issues. Although service interactions are pervasive in everyday life, and can be rich and interesting, the vast majority of attempts to model and engineer them have attempted to optimize efficiency and surface-goal completion. The results are all around us, from crudely scripted up-selling attempts at fast food restaurants to stilted dialog systems that tediously elicit the pieces of information needed to complete a database query. One reason is that the research community has come to shun most practical dialog types, perhaps to avoid seeming old-fashioned or being tainted by low expectations, or perhaps due to a misperception that industry is addressing these issues. A resource that would help progress here would be a commercial-dialogs corpora that is shareable by all.

Personally, I would like this corpus to be one with a truly exemplary person in the service role, someone who puts customers at ease, develops rapport,

*This work was supported by NSF Award IIS-0914868.

brings humor and sparkle, and makes them want to call back. Having several thousand short dialogs where diverse customers call in to that person, and modeling how she handles them, would take us a long way to understanding responsive and adaptive behaviors. Even prototype systems built on such dialogs could help set the agenda for future generations of commercial dialog systems.

2 To Reach Out to the Applied Linguistics Communities, Dialog Analysis Tools

Although many people are fascinated by language and dialog, spoken dialog systems research has only sporadically tapped this enthusiasm. For example, researchers in the conversation analysis tradition and teachers of foreign languages, not to mention many undergraduates, love to explore patterns of dialog. However spoken dialog research so far has produced scant findings about language behavior that are interesting to and graspable by non-engineers.

Personally, I think the biggest opportunity here involves tools to support non-technical people in discovering things themselves. Even amateurs, such as high school science fair participants, should be able to satisfy curiosity or confirm hunches, and experience the joy of systematically examining dialog phenomena. Our community ought to be producing tools and toolsets that support the complete workflow in such inquiries, eclectically supporting tagging, searching, juxtaposing clips and so on, and supporting both perceptually-based analysis and quantitative analysis in an integrated way. In particular we need to go beyond in-lab solutions (Ward and Al Bayyari, 2006) to develop robust toolsets that

can be used effectively without months of training.

3 To Reach Out to the Psycholinguistics Community, Modeling-Related Goals

Although many people are curious about how communication feats are achieved daily by human minds, spoken dialog research has only sporadically raised questions of real scientific interest. The spoken dialog community ought to formulate one or two high-profile grand-challenge problems that would inspire and bring people together, either cooperatively or in competition. Rather than “dialog management” and systems-type problems, these should be framed as “dialog modeling” problems, to make it clear that they are true scientific problems, and formulated so that they can be addressed more empirically and/or more theoretically, without requiring researchers to work with end-to-end systems. Such purer formulations should also help focus on questions of the fundamental human perceptions and abilities involved here, and how they vary with age, personality, language and culture.

Personally I think the most central and dialog-specific issues in our field are those relating to interpersonal coordination. Topics here have been nibbled at, perhaps most saliently in the study of turn-taking phenomena. Possible grand challenges may relate to topics such as “dialog dynamics” and “prediction of the interlocutor’s actions,” but formulating these problems so that they are general, and yet relevant and tractable, has been difficult (Ward, 2010; Ward et al., 2010).

4 To Reach Out to the Speech Processing Community, More Open Models

First, although speech generation and speech synthesis researchers are currently looking for new challenges, beyond correctness and intelligibility, the dialog systems community has only sporadically offered them interesting goals. These systems need somehow to be able to express the richness of the attitudes, structures, and intentions people convey in dialog, in real time, and we ought to provide specifications for this. Personally I think that multi-dimensional vector-space models of dialog states, situations, and intentions have promise here, and that these can best be developed by bottom-up em-

pirical studies (Ward and Vega, 2012 submitteda), one of which suggests that the important dimensions of dialog include, at least, in rough order of importance: who has the floor, the activity level, topic aging and transition, turn taking, seeking vs. establishing grounding, empathy, and sympathy, lexical access and planning processes, dominance, confidence, affect and attitude, rhetorical structure and strategy, and indications of concentration and involvement.

Second, although research on emotion and other nonverbal aspects of speech is advancing, this has only sporadically been guided by the needs of dialog systems. We ought to be thinking more about how emotion, attitude, stance, and related dimensions of communication are used in dialog. Personally I think that empirical studies of prosody, again, can be informative.

Third, although speech recognition researchers are adding flexibility and incrementality, speech recognizers’ interactions with the dialog manager are still very limited. In particular, the role of the dialog model in telling the recognizer what words are likely to come next, that is, its role in language modeling, is still underdeveloped. Personally I think we need dialog models that track more aspects of the dialog, and do so continuously, and supply that information to the recognizer (Ward and Vega, 2012 submittedb).

References

- Nigel Ward and Yaffa Al Bayyari. 2006. A case study in the identification of prosodic cues to turn-taking: Back-channeling in Arabic. In *Interspeech 2006 Proceedings*.
- Nigel G. Ward and Alejandro Vega. 2012, submitteda. A bottom-up exploration of the dimensions of dialog state in spoken interaction. In *Sigdialog*.
- Nigel G. Ward and Alejandro Vega. 2012, submittedb. Towards empirical dialog-state modeling and its use in language modeling. In *Interspeech*.
- Nigel G. Ward, Olac Fuentes, and Alejandro Vega. 2010. Dialog prediction for a general model of turn-taking. In *Interspeech*.
- Nigel G. Ward. 2010. The challenge of modeling dialog dynamics. In *Workshop on Modeling Human Communication Dynamics, at Neural Information Processing Systems*.