An approach to modeling attentional information in human-machine interaction

Milan Gnjatović, Vlado Delić
Faculty of Technical Sciences, University of Novi Sad
Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia
milan.gnjatovic@alfanum.co.rs, vdelic@uns.ac.rs
http://www.ftn.uns.ac.rs/

One of the challenges in spoken human-machine interaction (HMI) is how a language interface should address various inherently present dialogue phenomena related to the users' language, such as syntactically very simple utterances, high frequency of ungrammaticalities, predominant use of anaphora over complex syntactic constructions, affected speech, meta-language, etc. In addition, an additional non-linguistic context shared between the user and the system is often involved in HMI (e.g., a graphical interface) and may influence the language of the user, e.g., causing predominant use of elliptical and minor utterances, context dependent utterances, etc. It is clear that problems that occur in spoken HMI are not related only to technical deficiencies such as inaccurate automatic speech recognition. Generally, it is not reasonable to expect that users will always behave "cooperatively" and that they will produce utterances that fall within the application's domain, scope and grammar. Forcing them to always produce "correct" utterances would significantly limit the naturalness of the interaction.

In this paper, we discuss and illustrate an approach to modeling attentional information in HMI that facilitates processing of more flexibly formulated users' utterances. The implementation of this model in the prototype dialogue systems was demonstrated to work well for different syntactic forms of users' commands: elliptical commands, verbose commands (i.e., the commands that were only partially recognized by the speech recognition module), and context dependent commands. Two dedicated tasks in our prototype systems are: manipulation with graphical entities represented on the display that includes spatial reasoning (in German and English), and searching for textual content in newspapers' and websites' articles (in Serbian). We point out advantages of the introduced model that can reduce the degree of miscommunication on the conversational level and on the intentional level. It should be also noted that our approach to modeling attentional information is not limited only to verbally uttered commands. It supports also non-verbal dialogue acts produced by the user (e.g., using a mouse or a keyboard, etc.) or by the system (e.g., performing an instructed command) in the course of the interaction.

An important question is to what extent this approach can be generalized. We discuss this question from two points of view: the linguistic point of view and the engineering point of view. The linguistic point of view considers the question to which types of dialogue can this approach be applied. With respect to the domain of the interaction, this approach covers the class of spoken dialogue systems that are intended to manage a subclass of task-oriented dialogues, i.e., dialogues that are primarily concentrated on a given task, where the state of the task is observable in the sense that it can be explicitly defined and evaluated regarding to how it corresponds to expected final states. The engineering point of view considers primarily implementation aspects. The proposed modeling method and algorithms are not *a priori* related to some specific predefined interaction domain. The introduced algorithms are independent of the given task structure and of the content of the phrasal lexicon. This gives a relatively high level of generalizability of the proposed model—the given task can be relatively easy redefined or extended.

References

[1] J. Alexandersson, R. Engel, M. Kipp, S. Koch, U. Küssner, N. Reithinger, and M. Stede. Modeling Negotiation Dialogs. In W. Wahlster, editor, Verbmobil: Foundations of Speech-to-Speech Translation, pages 441–451. Springer-Verlag Berlin Heidelberg, 2000.

[2] E. Bilange. An approach to oral dialogue modelling. In M. Taylor, F. Néel, and D. Bouwhuis, editors, The Structure of Multimodal Dialoque II, pages 189–205. John Benjamins Publishing Company Philadelphia/Amsterdam, 2000.

[3] D. Bohus and A. Rudnicky. Sorry, I Didn't Catch That! An Investigation of Non-Understanding Errors and Recovery Strategies. In L. Dybkjær and W. Minker, editors, Recent Trends in Discourse and Dialogue, volume 39 of Text, Speech and Language Technology, pages 123–154. Springer, 2008. ISBN 978-1-4020-6820-1.

[4] N. Campbell. On the Structure of Spoken Language. In Proceedings of the 3rd International Conference on Speech Prosody 2006, Dresden, Germany, 2006.

[5] M. Gnjatović. *Adaptive Dialogue Management in Human-Machine Interaction*. Verlag Dr. Hut, München, ISBN 978-3-86853-189-3, 2009.

[6] M. Gnjatović and D. Rösner. An approach to processing of user's commands in human-machine interaction. In *Proc. of the 3rd Language and Technology Conference (LTC'07)*, pages 152–156, Adam Mickiewicz University, Poznan, Poland, ISBN 978-83-7177-407-2, 2007.

[7] B. Grosz and C. Sidner. Attention, Intentions, and the Structure of Discourse. Computational Linguistics, 12(3):175–204, 1986.

[8] D. Luzzati. A Dynamic Dialogue Model for Human-Machine Communication. In M. Taylor, F. Néel, and D. Bouwhuis, editors, The Structure of Multimodal Dialoque II, pages 207–221. John Benjamins Publishing Company Philadelphia/Amsterdam, 2000.

[9] Y. Wilks, R. Catizone, and M. Turunen. Dialogue Management. Technical report, COMPANIONS Consortium: State of the Art Papers, 2006. URL http://www.companions-project.org/downloads/Companions_SoA2_Dialogue_Management.pdf. Public report.