Corpus-driven conversational agents: tools and resources for multimodal dialogue systems development

Maria Di Maro {Università degli Studi di Napoli 'Federico II' - maria.dimaro2@unina.it}

Introduction

The present work gives an overview of a PhD research on dialogue modelling for conversational agents, to be framed in the ongoing national project **CHROME**, whose aim is to define a methodology of collecting, analysing and modelling multimodal data in designing virtual agents serving in museums. The PhD research under consideration is, therefore, mainly concerned with building the interaction with the resulting virtual gatekeeper, which will guide museum visitors in the exploration of cultural contents. In more details, starting from an empirical study of conversational phenomena, especially in cultural heritage domains, common ways of expressing requests and curiosities by visitors, and strategies of communicating cultural contents by guides will be collected and analysed, along with semantic, syntactic and paralinguistic language-dependent strategies.

System Architecture

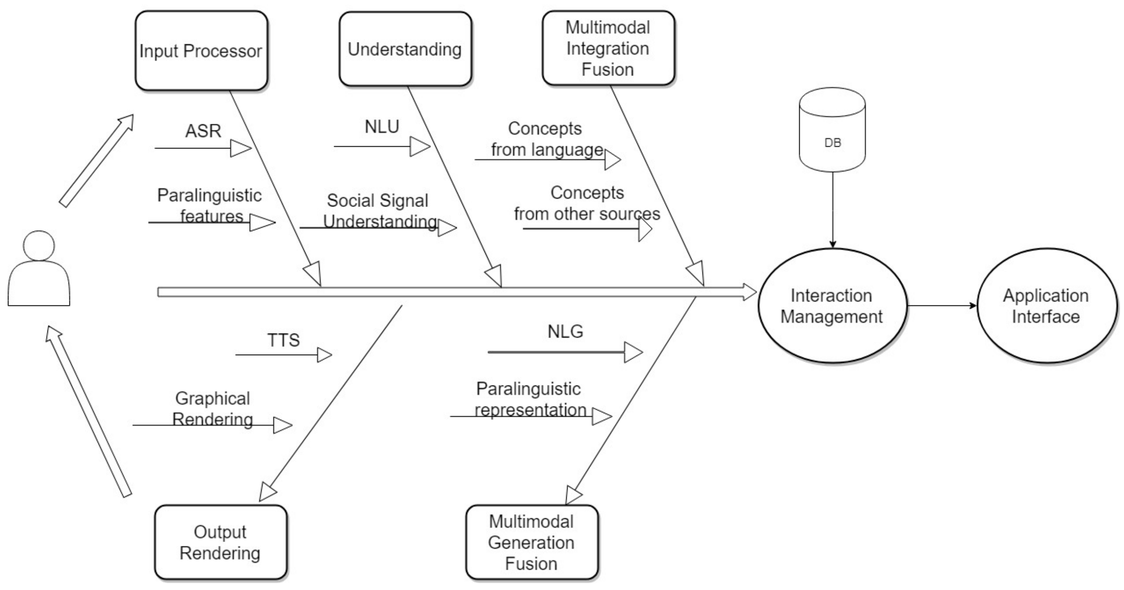


Figure 1: Multimodal Dialogue System Architecture

Annotations Fusion

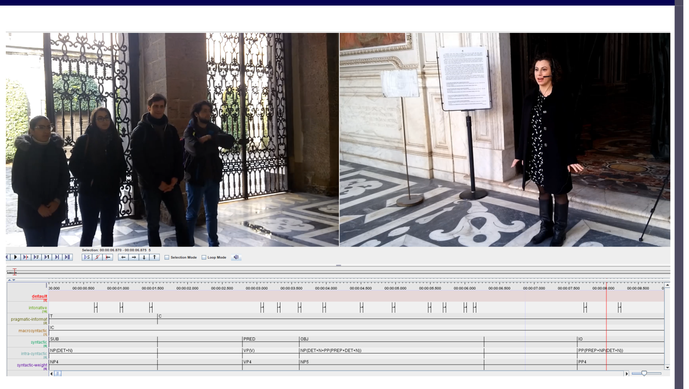


Figure 2: Screen-shot of the ELAN interface showing the fusion of different annotation levels for multimodal dialogue systems (see "CLARIN Tools and Resources" section)

CLARIN Tools and Resources

Input Processor: Automatic Speech Recognition for Phonetic features modelling and extraction

- Automatic Sementation and Alignment of audio data: Munich AUtomatic Segmentation System (MAUS).
- Features extraction: Praat.

Natural Language Understanding: Language Model

- Language modelling: Speech Recognition Grammars (SRGS).
- Semantic Networks: ItalWordNet.
- POS-Tagging: TreeTagger.
- Pragmatic annotation: Dialogue Act Mark-up Language (DiAML); EXMARaLDA.
- Corpora: SPOKES, CLIPS.

Multimodal fusion

- Multimodal annotations synchronisation: ELAN Software.

References

- Bunt, H., Alexandersson, J., Carletta, J., Choe, J. W., Fang, A. C., Hasida, K., Soria, C. (2010, May). Towards an ISO standard for dialogue act annotation. In Seventh conference on International Language Resources and Evaluation (LREC'10).
- Di Maro, M., Valentino, M., Riccio, A., and Origlia, A. (2017). Graph Databases for Designing High-Performance Speech Recognition Grammars. In IWCS 2017âĂŤ12th International Conference on Computational Semantic-sâĂŤShort papers.
- Kisler, T., Reichel U. D. and Schiel, F. (2017): Multilingual processing of speech via web services, Computer Speech Language, Volume 45, September 2017 (pp. 326âĂŞ347).
- Pezik, P. (2015, August). Spokes-a search and exploration service for conversational corpus data. In Selected Papers from the CLARIN 2014 Conference, October 24-25, 2014, Soesterberg, The Netherlands (No. 116, pp. 99-109). LinkÃűping University Electronic Press. Roventini, A., Alonge, A., Calzolari, N., Magnini, B., and Bertagna, F. (2000, May). ItalWordNet: A Large Semantic Database for Italian. In LREC.
- Savy, R., Cutugno. F. (2009). CLIPS. Diatopic, diamesic and diaphasic variations in spoken Italian, in in M. Mahlberg, V. GonzÃąlez-DÃŋaz, C. Smith (eds.), On-line Proceedings of 5th Corpus Linguistics Conference, (p 213).
- Schiel, F. (1999). Automatic Phonetic Transcription of Non-Prompted Speech. In Proc. of the ICPhS (pp. 607-610). Schmid, H., Baroni, M., Zanchetta, E., and Stein, A. (2007). The Enriched TreeTagger System. In Intelligenza Artificiale IV-2. (pp. 22-23).
- Schmidt, T., and Woerner, K. (2009). EXMARaLDAâĂŞCreating, analysing and sharing spoken language corpora for pragmatic research. Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA), 19(4), 565-582.
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H. (2006). ELAN: a professional framework for multimodality research. In Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006) (pp. 1556-1559).