

VIAN-DH: Computational Multimodal Conversation Analysis Software and Infrastructure

Authors : Teodora Vukovic, Christoph Hottiger, Noah Bubenhofer

Abstract : The development of VIAN-DH aims at bridging two linguistic approaches: conversation analysis/interactional linguistics (IL), so far a dominantly qualitative field, and computational/corpus linguistics and its quantitative and automated methods. Contemporary IL investigates the systematic organization of conversations and interactions composed of speech, gaze, gestures, and body positioning, among others. These highly integrated multimodal behaviour is analysed based on video data aimed at uncovering so called "multimodal gestalts", patterns of linguistic and embodied conduct that reoccur in specific sequential positions employed for specific purposes. Multimodal analyses (and other disciplines using videos) are so far dependent on time and resource intensive processes of manual transcription of each component from video materials. Automating these tasks requires advanced programming skills, which is often not in the scope of IL. Moreover, the use of different tools makes the integration and analysis of different formats challenging. Consequently, IL research often deals with relatively small samples of annotated data which are suitable for qualitative analysis but not enough for making generalized empirical claims derived quantitatively. VIAN-DH aims to create a workspace where many annotation layers required for the multimodal analysis of videos can be created, processed, and correlated in one platform. VIAN-DH will provide a graphical interface that operates state-of-the-art tools for automating parts of the data processing. The integration of tools that already exist in computational linguistics and computer vision, facilitates data processing for researchers lacking programming skills, speeds up the overall research process, and enables the processing of large amounts of data. The main features to be introduced are automatic speech recognition for the transcription of language, automatic image recognition for extraction of gestures and other visual cues, as well as grammatical annotation for adding morphological and syntactic information to the verbal content. In the ongoing instance of VIAN-DH, we focus on gesture extraction (pointing gestures, in particular), making use of existing models created for sign language and adapting them for this specific purpose. In order to view and search the data, VIAN-DH will provide a unified format and enable the import of the main existing formats of annotated video data and the export to other formats used in the field, while integrating different data source formats in a way that they can be combined in research. VIAN-DH will adapt querying methods from corpus linguistics to enable parallel search of many annotation levels, combining token-level and chronological search for various types of data. VIAN-DH strives to bring crucial and potentially revolutionary innovation to the field of IL, (that can also extend to other fields using video materials). It will allow the processing of large amounts of data automatically and, the implementation of quantitative analyses, combining it with the qualitative approach. It will facilitate the investigation of correlations between linguistic patterns (lexical or grammatical) with conversational aspects (turn-taking or gestures). Users will be able to automatically transcribe and annotate visual, spoken and grammatical information from videos, and to correlate those different levels and perform queries and analyses.

Keywords : multimodal analysis, corpus linguistics, computational linguistics, image recognition, speech recognition

Conference Title : ICCLCD 2023 : International Conference on Computational Linguistics and Corpus Design

Conference Location : Singapore, Singapore

Conference Dates : July 03-04, 2023