[Abstract] Available in MMSYM 23's Book of Abstracts. Here authors' copy.

Deleted: version

Event Chronography in Multimodal Data: a Method for Quantitative Analyses

Anaïs Murat, Trinity College Dublin
Maria Koutsombogera, Trinity College Dublin
Carl Vogel, Trinity College Dublin

Studying interactions in their complexity requires the exploration of different modalities and their temporal arrangement. Yet, distinct channels (e.g., vocal linguistic content, laughter, gesture, gaze, etc.) do not perfectly align in start or end times, which makes it difficult to count events in cross-modal comparisons. Consider the depiction of sequences of events in two modalities over time:

Answering the simple question "how many events of the lower sort accompany each event of the upper sort?" is not straightforward because the latter spans boundaries of the former. To overcome this difficulty, the literature tends to focus on longer events (such as dialogue episodes), or discontinuous events (e.g. turns with intervening intervals). But such an arrangement is not always relevant, and comparing shorter units of different types can be challenging. For instance, in Murat et al. (2022), overlaps between turns and mutual gazes (MG) (two relatively short annotation types) led longer MGs that spanned through several turns to be counted several times. This honest but necessarily skewed representation of the data limited the account of temporal relationships between the two modalities, notably in the exploration of the durational aspect of MG patterns.

Our ambition here is to contribute a method for studying temporality in multimodal corpora. We detail a new way of individuating instances according to chronologically arranged *events* – rather than time intervals – for quantitative purposes, and show a validating example.

The solution we present highlights the onset and offset of each annotation, and treats them as singular *events*. It then creates a table which chronologically orders these events. Typically, one column accounts for one annotation type, and each row accounts for one *event*: either the *onset* (identified by the letter "B", for "beginning") or the *offset* (identified by "E" for "end") of an annotation. When an *event* precipitates a new line, it also creates a "M" cell (for "middle") in any already started annotation of another type. "M"s can then be numbered to keep track of events occurring within an annotation. Therefore, a "BMMME" sequence reads

Deleted: version

"during this annotation, three *events* of another type occurred". Ultimately, the finalised table comprises a minimum of two columns (minimum two annotation types), and as many lines as events (onsets and offsets) present in the corpus.

With this so-called BME method, we investigate the Multisimo Corpus (Koutsombogera & Vogel, 2018) to illuminate two things: (1) the various possible relations between turns and MG (such as whether MGs tend to be included inside, between, or across turns), and (2) the relevance of event-based analyses over time-based analyses of duration.

Contingency table analysis of the resulting annotations enabled quantification of MG beginnings and endings to occur during middles of turns, suggesting a potential priority of the turn channel. Our second analysis of the MG's Bs, Ms, and Es compared to the amount of repetition by turn illustrates the relevance of event-based duration in addition to time-based duration and suggests that the length of a MG as counted in events is positively correlated to the amount of repetition occurring in between its onset and offset.

Finally, we argue the generality of our method. The strength of the BME method lies in its ability to handle in a similar manner isolated and overlapping annotations, as well as longer and shorter annotations. It pushes time-related matters (such as gaze fixation definition, or cross-annotator agreement time differences) to the background by shifting the focus from a -more or less- arbitrary time duration to a meaningful arrangement of data where the length of an annotation translates the number of other *events* occurring through it. In the future, we aim to use this method to study relationships among words, parts-of-speech and non-verbal aspects of interactions, such as gestures, gaze patterns, and laughter.

Keywords: multimodality; temporal analysis; interaction

Koutsombogera, M., & Vogel, C. (2018). Modeling Collaborative Multimodal Behavior in Group Dialogues: The MULTISIMO Corpus. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, 2945–2951.

Murat, A. C., Koutsombogera, M., & Vogel, C. (2022). Mutual Gaze and Linguistic Repetition in a Multimodal Corpus. *Proceedings of the Language Resources and Evaluation Conference*, 2771–2780.