A proposal of a framework for analysis of multimodal message coherence (FAMC). Live report coherence pilot study.

Agata Karaśkiewicz, Agnieszka Czoska

Institute of Linguistics, Adam Mickiewicz University in Poznan,
Institute of Psychology, Adam Mickiewicz University in Poznan
ul. Wieniawskiego 1, 61-712 Poznań

agata.karaskiewicz@wp.pl, aczoska@amu.edu.pl

Abstract

We present an application of Framework for Analysis of Multimodal Message Coherence (FAMC) to a pilot study on live report (LR) coherence impact on audience's memory and impressions. The study participants (6) were asked to watch LR with or without coherence manipulation and asked to recall key information and unusual elements from LR. FAMC is based on the Event Indexing Model (EIM) of message processing (Zwaan, Langston and Graesser 1995, Zwaan and Radvansky 1998, Czoska and Karaśkiewicz in print). The framework may enable assessment of local and global coherence of a multimodal message and provide basis for testing the impact of coherence manipulation on message understanding.

1 Live report

In live report (LR) a news story is reported from the authentic scene in real time: "here and now" (Snoeijer, De Vreese and Semetko 2002). LR standard structure consists of three phases. First the presenter provides the kernel details of the report (e.g. naming the place, time, character of the event: "we now go live to Kiev where a riot continues") and poses a question to the reporter. This is followed by the reporter's coverage, during which detailed information is provided verbally and visually (sometimes with additional video material, pictures or infographics). Usually, a short exchange between the reporter and the presenter is added close to the end of LR (Montgomery 2007). Finally a topic shift is announced by the presenter.

It is a semi-spontaneous message, with pre-planned structure. The reporters know the questions that will be asked by the presenter, have time to prepare themselves to give an account, interview participants and witnesses present at the scene also before the LR is aired, collect additional video material. However, what they say during the live airing is produced mostly on-line. Moreover, sometimes improvisation is needed due to technical problems, delays, occurrence of new information.

2 Event-Indexing Model

Event-Indexing Model (EIM) was proposed by Zwaan, Langston, Graesser and Radvansky (Zwaan, Langston and Graesser 1995, Zwaan and Radvansky 1998). It defines text as a set of *events* (presented sequentially) the processing of which consists of three stages:

- current model (processing of a current *event*),
- integrated model (the model of all the previous text which may be updated with information from the current model),
- complete model (created when the text processing is finished).

An *event* is defined as new verb or phrase in the text. From a multimodal message *events* may be extracted on the basis of the textual elements (for analysis of co-speech visual, behavioural or other paraverbal elements) but also according to other modalities, like gestures or movements as well as scenes or shots in video material (Zacks, Speer and Reynolds 2009, Kurby and Zacks 2012). Every *event* is indexed across five dimensions: time (1), space (2), protagonists and objects (3)[1], intentions (4) and causality (5). Construction of the situation model is called "updating" and may mean developing an integrated model (attaching a new *event*) or adjusting the index values that vary between adjacent *events* at the moment of integrating the model.

In FAMC we use terms video or audio *event* to refer to surface elements (shots, scenes, intonation phrases, speech parts). We also propose a term *mental event* to refer to a mental representation constituted of (in most cases) parallel *events* from different channels that may be integrated into a single *event* in the situation model (see Figure 1). This distinction was not explicitly mentioned in the works of EIM authors, and the term event referred there both to surface (message) as well as representation elements. We think the distinction is called for especially since there is a difference between index values establishment on the basis of surface cues and all the other ones, involving information processing and knowledge activation.

Within EIM message coherence is defined as identity of index values between *adjacent events*. To adjust the model to analysis of multimodal message and take into account all the modalities (or channels) employed in FAMC *adjacency* will be understood as simultaneous occurrence (coherence between two *events* from different modalities) or succession (between two subsequent *events* in the same or different modalities). In consequence within FAMC two different perspectives for coherence analysis may be distinguished: parallel and sequential (Czoska and Karaśkiewicz in print).

Analysis of message processing in terms of EIM is based on two hypotheses (Therriault, Rinck and Zwaan (2006):

- *the processing load hypothesis*: the more index values change between the integrated and the current model, the mode difficult is the updating – the integration process;
- *the memory organization hypothesis*: the more (out of five) index values for two *events* are similar, the more associated they should be in the situation model (although they may be far apart in the surface code).

The *processing load hypothesis* enables research on, for example, local and global text coherence and its impact on the model creation (Kurby and Zacks 2012, Radvansky and Copeland 2004), effects of index continuity and variance on message processing and perception of its parts (structure reconstruction) (Zacks, Speer and Reynolds 2009, Kurby and Zacks , 2012) or interaction between situation model creation and working memory (Brunyé and Taylor 2008). The *memory organization hypothesis* enables for example testing the impact of the surface code proximity and index values on concepts sorting (Zwaan, Langston and Graesser 1995) or predicting next event (Zacks, Kurby, Eisenberg and Haroutunian 2011) and (re)analysis of memory tasks results from research in different frameworks (Zwaan and Radvansky 1998). The model of text processing proposed by Zwaan et al. was at first tested only on texts, but was created as modality-independent and as such was used also for analysing films (Zacks, Speer and Reynolds 2009, Zacks, Kurby, Eisenberg and Haroutunian 2011, Czoska and Karaśkiewicz in print).

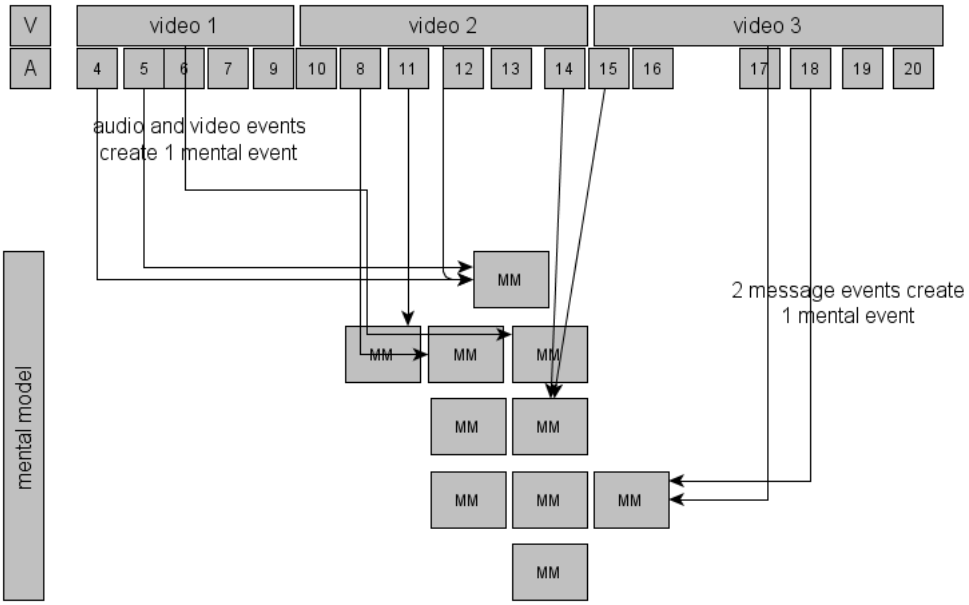
The simplest testable prediction based on the two *hypotheses* is that the lower the message coherence, the greater the processing load and lower the message understanding. What is more, the more disorganized (incoherent)the resulting situation model is, the worse the message content recall.

1. Determining index values in FAMC

Index value ascription is based on the cues from the surface code, audience's semantic knowledge, knowledge of the world and inferences based on combining the sources (Zwaan, Langston and Graesser 1995). Ascription on the basis of the surface code means that the information on time, space and other indexes was stated explicitly in the message. The integrated model serves as an immediate context and provides the most salient cues (Brunyé and Taylor 2008).

[1] Divided by Zacks et al. into two separate indexes: characters and objects; see for example: Zacks, Speer and Reynolds 2009.

Figure 1: Message situation model construction in FAMC. Two separate surface *event* lines are shown: audio (A) and video (V) channels. Numbers are ascribed only to surface events (shots and intonation phrases or sentences).



In some multimodal messages, values provided via the audio channel may be more explicit and salient (Graber 1990, Crigler, Just and Neuman 1994) providing the default values for the parallel video events or for the combined mental events, reconstructed by the audience (see Figure 2). The relations between simultaneous audio and video events such as overlap (all values overlapping), displacement or dichotomy (lack of overlapping values) (Meinhof 1994, Bednarek and Caple 2012) may influence this default ascription, with incoherences promoting updating. Moreover lack of coherence on local or global level should increase audience cognitive load according to *the processing load hypothesis*.

An important notion in analysing the impact of incoherence on message processing is audience's coherence standard. On the one hand, one may assume that it is going to be high for television broadcasts. Knowing, that such a message should be coherent and well-formed, the audience is prone to put a lot of effort into creating very coherent model and using every new information (Snoeijer, De Vreese and Semetko 2002). On the other hand, LR is a semi-spontaneous message, with global structure and topic planned in advance, but produced on-line. This may increase audience's acceptance for errors and incoherences, lowering the tendency to update the model whenever index value seems to be changed. In both cases multimodality of the message and the fact, that lot of information is provided verbally may also induce high coherence standards in parallel perspective with verbally given index values as default (Brunyé, Taylor, Rapp and Spiro (2006).

3 Pilot study

The main idea of the study was to compare audience's reactions on two LR on Poznań joining Fairtrade Towns Movement (FTM). One LR included several coherence modifications. Two conditions were compared: watching the original LR (OLR) and the modified one (MLR).

1. Stimuli

The structure of both LR was identical and consisted of two of the aforementioned phrases (the termination phrase was missing): (1) presenters introduction (backgroud of the event), (2) reporter's statement. The

reporter provided information about the date and place of the ceremony of granting Poznan the FTM certificate, short history of FTM, the rules of joining it and the event promoting the initiative held on Poznan Old Market Square. In both LR additional video material (e.g. showing the mayor collecting the certificate, the event on the Old Market Sguare), pictures (showing fair trade products, farms and people from Africa and South America) and infographics (slide presenting FTM rules) were used. Coherence of MLR stimulus was modified. Incoherent information was placed in:

- audio channel: In the last sentence of introduction false information is given concerning the place where the reporter was;
- audio channel: Twice reporter's intonation was modified (unnatural tone of voice, falling or rising, inadequate to the emotional content of the speech);
- video channel: Scene background was modified (in a corner of the screen a rainbow was added on the cloudy sky; see Figure 3);
- video channel: Inadequate pictures were shown (pictures of 2 happy African families on holiday were shown while the reporter talked about fair trade schools; see Figure 4); No other modification was added and especially there was no difference between the information stated verbally during the reporter's statement. Figures 2 and 3 show the elements visible during the presenter's introduction and the location of the reporter.

The LR were prepared by professional WTK (Wielkopolska Telewizja Kablowa, Great Poland Cable Television) team specially for the purpose of the study and was never aired.

2. Participants and procedure

There were 6 participants, 3 watched the OLR and 3 the MLR. All were students from the Institute of Linguistics (Adam Mickiewicz University in Poznan), women and Polish native speakers, aged 21-23. The experiment was conducted with each participant individually. They were all provided with an instruction read aloud by the experimenter, informing them that they will watch LR and there will be discussion afterwards which will be recorded. Participants watched one LR on a 11" screen computer with in-build loudspeakers. After watching LP participants answered 13 questions (asked by the experimenter present the whole time), part of which asked about the content of the LP and the participants' impressions (8 questions) and about the participant's background (4) and one control question about having previously seen the LR. The answers were recorded and later transcribed.

Figure 2: Studio introduction in the moment the information on reporters location was provided. In the additional video material a symbol of Poznan is shown (goats on Poznan Town Hall).

Figure 3: During all shots showing the reporter he stayed in Fryderyk Chopin Park in the back of Poznan Municipal Office. The right picture presents one modification of the video channel, inserting the rainbow.

3. Results

No participant has previously seen the LR. All participants correctly described the LR topic (Poznan joining FTM) and remembered its scene (Poznan).

Content questions There were no differences between the groups in answering question on the date of the event: 1 correct answer in each group (1th November 2013). There was a difference in answering the question on the place of ceremony (Municipal Office): 2 correct answers in OLR and 1 correct in MLR. The two MLR participants that misremembered the place based their incorrect and imprecise answers on video channel information.

The question on about the background, against which he reporter stayed (see Figure 3), was answered correctly and precisely by all OLR and 2 MLR participants. The question on the exact place where the reporter stayed was answered correctly by all the participants, but only 2 MLR participants answered it according to their background knowledge naming the place (others used visual information as in the question on the background). Asked how they know what is the scene of the LR participants from both groups referred to the information presented visual channel, not to the information given verbally during introduction. The participants referred also to their knowledge of similar events and familiarity with Poznan Old Town.

Impression questions The participants were asked whether they noticed anything unusual in the LR. Only one person in MLR mentioned the incoherent African families pictures (however she may have also been referring to the rest of picture material; see Figure 4), and another one mentioned "something wrong" with the sound. 1 person in each group mentioned, that using the English term "fair trade" was unusual and inappropriate, since there is a Polish one ("sprawiedliwy handel"). She also mentioned, that the term should at least be pronounced with appropriate English accent. No participant noticed the insertion of the rainbow (see Figure 3).

Background questions When asked what they expect from LR the participants mentioned emotional content, spontaneity and no additional video material or pictures. Answering that question two of them also mentioned, that they were prepared for a message hard to understand, chaotic, with more pronunciation errors that the presented one. The participants were also asked whether they watch television news (not at all or very rarely) and where they look for news (internet newspapers).

4 Conclusions and future work

The aim of this study was to analyse the impact of incoherent *events* insertion on LR understanding. The insertions were in both audio and video, were mostly parallel. The analysis was based on the frameworks of

Figure 4: Additional material: pictures on farming (OLR and MLR) showing happy families (MLR)

EIM and FAMC (which may be treated as EIM version adapted to multimodal message analysis, in which all modalities should add to the same content and the same situation model).

LR is a semi-spontaneous message, different from what is associated to be a typical, well-prepared and controlled television broadcast. LR reporters are allowed a few slips of the tongue. Even if their speech is previously planned it is impossible to direct it from beginning till the end, because of their certain spontaneous behaviours and reactions, which is also known by the audience. Because of that even the MLR that was prepared for the pilot experiment was acknowledged "correct" and natural by the participants. They applied low coherence standards while processing the message and created situation model that did not allowed detailed recall of the LR, but enabled understanding it on a general level.

The hypothesis that the video channel *events* provide the default index values is not supported by the results. First of all most errors in answering the content questions may be attributed to information provided in video channel. The fact that the participants misremembered the place of the ceremony may be due to a big amount of additional video material showing the celebration on Poznan Old Market Square, Poznan Town Hall and the goats. Moreover no MLR participant mentioned the misinformation in the introduction, although they all knew the reporter was not on the Old Market Square.

What remains to be tested is the hypothesis that the lack of coherence increases processing load. Although the pilot study provides some evidence that incoherent message influences recall (more errors in answering detailed content questions), a better-controlled study with more participants is needed to test the effect. What is more, an experiment measuring reaction time or monitoring message on-line processing should be proposed, since processing load increase may be indicated not only by errors, but also by increased reaction times or attentiveness (operationalized for example as gaze latency).

In further research more controlled manipulation of the coherence of the message might be provided by modification of the main index values across the LR: time, space, causality (Czoska and Karaśkiewicz in print). Especially greater control must be ensured when it comes to possible sources of index values, for example by minimising the amount of additional materials inserted. Moreover analysis of relations between those indexes values may enable research on the hierarchy of index values in LR processing and remembering.

References

Bednarek, M. and Caple, H. 2012. News discourse (Vol. 46). Continuum.

Brunyé, T. and Taylor, H. 2008. "Working memory in developing and applying mental models from spatial descriptions." Journal of Memory and Language, 58(3): 701–729.

Brunyé, T. T., Taylor, H. A., Rapp, D. N. and Spiro, A. B. 2006. "Learning procedures: The role of working memory in multimedia learning experiences." Applied Cognitive Psychology, 20(7): 917–940.

Crigler, A. N., Just, M. and Neuman, W. R. 1994. "Interpreting visual versus audio messages in television news." Journal of Communication, 44(4): 132–149.

Czoska, A. and Karaśkiewicz, A. (in print). "Analiza mechanizmów zachowania spójności w relacjach na żywo w ramach teorii modelu sytuacji - propozycja modelu spójności dla telewizyjnego przekazu reporterskiego." Investigationes Linguisticae(XXXI).

Graber, D. A. 1990. "Seeing is remembering: How visuals contribute to learning from television news." Journal of Communication, 40(3): 134–156.

Kurby, C. and Zacks, J. 2012. "Starting from scratch and building brick by brick in comprehension." Memory and Cognition: 1–15.

Meinhof, U. 1994. Double talk in news broadcasts. Media texts: Authors and readers: 212–223.

Montgomery, M. 2007. The discourse of broadcast news. Routledge.

Radvansky, G. and Copeland, D. 2004. "Working memory span and situation model processing." The American Journal of Psychology: 191–213.

Snoeijer, R., De Vreese, C. H. and Semetko, H. A. 2002. "Research note: The effects of live television reporting on recall and appreciation of political news." European Journal of Communication, 17(1): 85–101.

Therriault, D., Rinck, M. and Zwaan, R. 2006. "Assessing the influence of dimensional focus during situation model construction." Memory and Cognition, 34(1): 78–89.

Zacks, J., Kurby, C., Eisenberg, M. and Haroutunian, N. 2011. "Prediction error associated with the perceptual segmentation of naturalistic events." Journal of Cognitive Neuroscience, 23(12): 4057–4066.

Zacks, J., Speer, N. and Reynolds, J. 2009. "Segmentation in reading and film comprehension." Journal of Experimental Psychology: General, 138(2): 307.

Zwaan, R., Langston, M. and Graesser, A. 1995. "The construction of situation models in narrative comprehension: An event-indexing model." Psychological Science: 292–297.

Zwaan, R. and Radvansky, G. 1998. "Situation models in language comprehension and memory." Psychological Bulletin, 123(2): 162–185.