The discourse structure of video games

Wildfeuer, Janina; Stamenković, Dušan

Published in:
Language & Communication

DOI:
10.1016/j.langcom.2021.11.005

Language & Communication 82 (2022) 28–51

Contents lists available at ScienceDirect

Language & Communication

journal homepage: www.elsevier.com/locate/langcom

The discourse structure of video games: A multimodal discourse semantics approach to game tutorials

Janina Wildfeuer [a,*], Dušan Stamenković [b]

[a] University of Groningen, Faculty of Arts, Department of Communication and Information Studies, Groningen, the Netherlands
[b] University of Niš, Faculty of Philosophy, Department of English & Language Cognition Laboratory, Niš, Serbia

ARTICLE INFO

Article history:

Keywords:
Video games
Multimodality
Discourse
Inferences
Formal discourse semantics

ABSTRACT

The article proposes a multimodal discourse semantics approach to the analysis of video game tutorials that provides a discourse pragmatic analysis of the game canvases in these tutorials. The study mainly builds on linguistic approaches to formal dynamic discourse semantics that have already been successfully applied to other multimodal artefacts. The article will showcase the application of the resulting 'logic of multimodal discourse interpretation' to two specific cases of video game tutorials. This will outline particular discourse relations holding between events and segments in the tutorials as distinctive features of this video game genre and show the discursive patterns of these instructions.

1. Introduction

The study of video games has experienced a trend towards engaging digital games as multimodal discourse [e.g., Aarseth, 2014b; Ensslin, 2012; Ensslin and Balteiro, 2019; Gee, 2014; Popović and Stamenković, 2019; Stamenković et al., 2017], and the attempts to describe interactive audio-visual gameplay by drawing on our understanding of linguistic discourse are still increasing. Even publications that are not devoted to multimodality acknowledge the fact that the discourse of video games is shaped by several aspects, including language, design and play [see Paul, 2012]. The idea that not only the verbal interaction between players in games (in general), but instead all parts of the audio-visual (and sometimes haptic) material fulfil communicative purposes similar to verbal discourse is today mainly contextualized in the realm of visual and multimodal studies. These examine how different expressive resources combine for communication, involving language, static or moving images, sound, music, haptic elements, and more. The expanding body of multimodal research has analysed diverse media artefacts, performances, and interactions, including film (Bateman and Schmidt, 2012; Wildfeuer, 2014), comics and graphic narratives (Bateman and Wildfeuer, 2014b; Wildfeuer, 2019), sometimes in interaction with video games (Ng, 2020), experimental literature (Gibbons, 2011), digital technologies (Jewitt, 2013), interactions (Norris, 2004), as well as movement in space (McMurtrie, 2016).

Among these, approaches that are fully oriented towards video games have, in most cases, arisen relatively recently, i.e., at the end of the last decade. The multimodal analysis of video games then deals with the question of how different semiotic resources that are used for the design of the interface and the game itself construct meaning in their intersemiotic interplay. The focus of description is not only on verbal elements such as text inserts, a voice-over or the verbal interaction between players during the game, but also on all elements that are, as expressive forms, contributing in one way or another to the

* Corresponding author.
E-mail address: j.wildfeuer@rug.nl (J. Wildfeuer).

https://doi.org/10.1016/j.langcom.2021.11.005

meaning construction. These would include moving images or scenes that construct a setting and feature characters, but also certain colours or sounds, music, visual effects and animations, interactive elements to navigate the game, etc. They all take on certain roles and function in the complex interplay of these elements. A multimodal analysis aims at identifying these different roles and functions and at examining how they guide and influence the interpretation and experience of the game (Wildfeuer and Stamenković, 2020).

One important first attempt to apply a multimodal approach to video games can be found in David Machin and Theo van Leeuwen's *Global Media Discourse: A Critical Introduction* [Machin and Van Leeuwen, 2007: 74–104]. In 2009, Kevin Leander and Lalitha Vasudevan analyse, in a contribution on multimodality and 'mobile culture', the online game *Star Wars Galaxies* (LucasArts and Sony Online Entertainment, 2003) Ivarsson et al. (2009), take a closer look at interactions in interactive games. Astrid Ensslin devoted one of the chapters of *The Language of Gaming* to gaming and multimodality [Ensslin, 2012: 117–141]. Most of these earlier approaches to video games drew on social semiotics and multimodal discourse analysis (MDA), and this line of research has been active in previous years [e.g., Pérez-Latorre et al., 2017; Pérez-Latorre and Oliva, 2019]. There have also been approaches starting from a more general semiotic perspective, but also dealing with issues related to society [see, e.g., Hawreliak and Lemieux, 2020]. On the other hand, scholars belonging to the movement of ludo-hermeneutics oppose the idea that video games can be a channel for conveying ideologies [see Aarseth, 2014a]. Video games have often been the subject of smaller analyses or overview presentations [see also Masso, 2009] and most of the present studies in the field of empirical multimodal analysis of video games are based on small corpora [e.g., Stamenković and Wildfeuer, 2021; Stamenković, 2022]. Extensive disputes are scarce and first monographs have emerged recently [see, e.g., Toh, 2018; Hawreliak, 2018]. From the perspective of video game studies, there have also been calls that video game scholarship has to go further outside their usual circles and set towards other disciplines [see Bogost, 2015]; the present approach is one possible way of extending video game research.

The present paper joins these ranks by proposing an initial model of multimodal discourse pragmatics that accounts for the specifics of information communicated in and through video games – and in particular a specific text type within these games, i.e., tutorials. By providing a framework for the analysis of both meaning-making units as well as the inferential steps needed to understand and interpret these units, we aim to analyse the initial stage of communication between a video game and the players and to show how the latter are taught to play the game. Our approach is first and foremost fundamentally semantic and tackles the important and foundational aspect of comprehending multimodal artefacts, i.e., what recipients, or players, understand in their interpretation and processing of the video games. The analysis we provide in the following is a reconstruction of this interpretation process from a model-theoretical perspective by us as analysts on the basis of semantic and pragmatic frameworks; we do not take into account empirical data (such as demographic information, for example). By focusing on two example analyses of tutorial segments in the games *GTA V* and *Batman: Arkham Knight* and by providing a very fine-grained analysis of these segments, we offer a case study of initially testing our methodological approach.

This methodological approach is particularly based on linguistically-motivated formal accounts to discourse structure that examine the dynamically unfolding of a narrative storyline or argumentative structure – and, as we assume, are equally capable to describe the dynamic processes of the gameplay. Analysing these structures with regard to the inferences that are drawn by recipients/players in the interpretation process has already proven to be fruitful for the analysis of other multimodal phenomena, such as films (Bateman and Schmidt, 2012; Wildfeuer, 2014) and comics and graphic narratives (Bateman and Wildfeuer, 2014b; Wildfeuer, 2019). The approach follows the long-lasting tradition of considering (audio-)visual artefacts 'as language', as discussed for film already in the 1960s and 1970s [see, e.g., Metz, 1964, 1974] and still being explored for comics and graphic novels [see, e.g., Bateman and Wildfeuer, 2014b; Cohn, 2013, 2018; Miodrag, 2013], for example. As we have highlighted several times before [see, e.g., Bateman and Wildfeuer, 2014a, 2014b], the direct application or even equation of linguistic units (such as morphological or syntactical units) with visual or audio-visual units is not feasible and only invokes inappropriate levels of abstraction. At the same time, we have successfully proven that

> "significant similarities between verbal and visual communicative artefacts can be located at the more abstract levels of *discourse*. It is then only at these higher levels of abstraction that insights from linguistic models can be beneficially applied" [Bateman and Wildfeuer, 2014b: 181].

In this paper, we will continue to show this application by testing the transferability and applicability of the discourse analytical approaches we have been developing in the last couple of years to video games in general and video game tutorials in particular. For this, we will first introduce the particular discourse analytical framework of the *logic of multimodal discourse interpretation* in Section 2 and explain how it can be used to analyse processes of meaning construction in the dynamic unfolding of a video game. After a short overview of our example artefacts from video tutorials and their main tasks in video games in Section 3, we will then analyse the beginning sequences of the two different video games in Section 4. Finally, we will shortly discuss the findings of this analysis and give an outlook for future studies with this framework in Section 5.

2. Multimodal discourse interpretation in video games

The framework we use to analyse video games and their multimodal meaning construction is developed further from the *logic of multimodal discourse interpretation*, which has been provided in several approaches to the analysis of multimodal artefacts [cf., e.g., Bateman and Wildfeuer, 2014b; Wildfeuer, 2014; Wildfeuer, 2019; Wildfeuer et al., 2015]. This framework builds on substantial achievements in formal discourse analysis [see Asher and Lascarides, 2003; Hobbs, 1990; Kamp and Reyle, 1993] that again refer to several distinct logics to make visible the inferential work of the recipient to analyse verbal

and non-verbal discourses. The main aim of formal discourse semantics in particular is to formally realize the processes of inferring and to specify the interpretation of discourses on the basis of several logics. This inferential model of communication enables the qualitative reconstruction of the interpretation by hypothesizing about the recipients' inferences in their perception and experience, but without taking actual user data into account.[1] The importance and applicability of such a logics-based approach to formally model multimodal discourse has been shown to a different extent over the last two decades, including applications to gesture (Lascarides and Stone, 2009) and emojis (Grosz et al., 2021). All these accounts, including our own, aim at a detailed description of the semantics of multimodal artefacts or performances, either by providing a formal representation of specific units and their semiotic modes or by analysing the (mostly dynamic) discourse structure in which they are embedded or which they construct.

In comparison to the multimodal phenomena that have so far been approached from this perspective, video games have yet another degree of specificity that is crucial for our approach here. Namely, although, similar to some films, they aim at creating increasingly coherent worlds of fiction, this fiction is frequently hard to predict and control, as it can be incoherent, ambiguous and optional [see Juul, 2002]. This is one of the reasons why we decided to start with the tutorial segments of narrative video games – in these, the level of predictability seems to be slightly higher.

In order to analyse both the interplay of the semiotic modes and elements as well as their coherence and structure in these tutorial segments, our framework operates on both these levels of multimodal comprehension [see also Bucher, 2011]: first, the level of identification and arrangement of the meaning-making entities and, second, the level of coherence and structure. These two levels of interpretation are included in the logic of multimodal discourse interpretation with the help of two separate logics [cf. Bateman and Wildfeuer, 2014b; Wildfeuer, 2014]. We will describe both levels and logics in further detail in the following subsections.

2.1. Analysing semantics: the logic of information content

On the first level of our framework, on the level of the so-called *logic of information content*, logical forms of the discourse are constructed. These logical forms describe the semantic content of the interacting modalities by displaying the various events (e.g., of a film's diegesis or comic's narrative) and analysing the semiotic resources at play in terms of their *intersemiosis* that accounts for the diverse meanings to be identified by the recipient [see O'Halloran et al., 2013; Royce, 2007; Unsworth, 2007; Wildfeuer, 2012].

As we have argued in several of our previous works [see, e.g., Bateman and Wildfeuer, 2014b; Stamenković et al., 2017; Wildfeuer, 2014, 2017], this identification is always a process of abductive and defeasible reasoning about the best interpretation. The notion of abduction as introduced by Peirce as a basic logical process [cf. Peirce, 1931–1958; Wirth, 2000] helps in finding hypotheses and at the same time proving and verifying these hypotheses on the basis of certain logical principles. An important factor is also the recipient's assumed cognitive capacity and their ability to logically combine the information available in the discourse with further sources of information. As a result of this identification and combination process, the logical forms give details of the reconstruction of this reasoning process and identify the necessary units and referents involved. We provide a first example of such a logical form for video game discourse in Fig. 1 whose formal details we explain further below.

e_{π_1} = threaten
CHARACTERS AND OBJECTS: [v] represented participant: player 1 (a) [v] represented participant: male character (b) [v] represented participant: male character (c) [v] represented participant: female character (d) [v] object: gun (e) [a] "Get down there! – Alright, everybody pays attention, no one gets hurt." (f)
SETTING: [v] inside a building (g)
TECHNICAL FEATURES: [v] dynamically changing camera perspective: MS & CUs (h)
EVENTUALITY: $a, d \mid\sim$ **threaten** (e_{π_1})

Fig. 1. Example of a logical form described for the first *GTA V* scene and the narrative event happening in this scene.

[1] This also means that no demographic data or information about the players' culture, social context, or their experience in playing games are taken into consideration. In this way, our theoretical approach does not account for individual interpretations by players.

As described in further detail in Wildfeuer (2014, 2019), the *logic of information content* provides a formal language for analytically constructing a semantic representation of a film's event or a comics' panel – and, as we now assume, also a video game's sequence – in logical forms. These logical forms thus display an abstract formulation of what is happening or shown on a comic page or in a film or video game sequence. They describe the various entities occurring in these extracts and the dependencies and conditions holding between them. Their formal style is initially based on the discourse representation structures that Hans Kamp and Uwe Reyle (1993) introduced for verbal discourse in their Discourse Representation Theory (DRT) and extensions of this notation in so-called segmented discourse representation structures as developed by Nicholas Asher and Alex Lascarides (2003). Further details of descriptions and categories used within multimodal analysis (such as, for example, the identification of represented participants according to Kress and van Leeuwen, 2006) have been added. Building on their formal depictions, the representation of the logical forms as developed within the logic of multimodal discourse interpretation allows for a precise description and specification of all elements involved in the meaning-making process and thus leads to a reconstruction of the semantics of the respective units.

The box given in Fig. 1 displays an abstract formulation of what is happening in the first few seconds of the game *GTA V*. The first line of the box gives the inferred *eventuality*, the 'verbalization' of the content that is identified in the scene. This eventuality is the minimal semantic unit of description for the reconstruction process and is described as e_{π_n}. It represents a hypothesis of the recipient's inference process in making sense of all the audio-visual and verbal resources combined in a multimodal artefact. This might generally be a narrative event depicted, the description of a state or significant object, or the instruction to proceed with a specific task. Following Zacks et al. (2007), the delineation and segmentation of these eventualities and the narrative or instructive events are based on associations of "segment[s] of time at a given location that [are] conceived by an observer to have a beginning and an end" [Zacks et al., 2007: 273].[2] More details of how to build these eventualities with the help of event segmentation in film, for example, or on the basis of more static units such as panels are given in Wildfeuer (2014, 2019). We assume that similar processes such as the segmentation of several narrative events in a video game sequence or the transition from one setting to another setting with filmic cuts or other camera techniques will lead to the inference of a certain eventuality in video game sequence. Therefore, we take the eventuality and its representation in a logical form as the starting point for every examination of a video game scene that aims at describing the multimodal composition of all activities and processes in this scene.

This multimodal composition is more clearly described in the various other lines of the box which list elements identified and inferred from the different contributions of the semiotic resources. Each of these resources is listed according to their visual [v], auditory [a] or verbal-textual [t] characteristics and labeled with a variable in parentheses in order to better identify and compare the units. Again, the recipient's assumed world knowledge comes into play here, allowing to identify units in the game as representations of characters or screens or controllers as part of the games mechanics. All these resources contribute to the meaning-making process as so-called discourse referents. As an extension to the descriptions in Wildfeuer (2014) for film and in Bateman and Wildfeuer (2014b) for comic, the particular logic of information content has been further developed in Wildfeuer (2019) with different additional categories such as the SETTING and TECHNICAL FEATURES. Descriptions for these features use terminology from film and game studies to describe details of the moving images or instruction notes and allow a more fine-grained characterisation of the different semiotic elements that play a role for the meaning-making process. In the example logical form above, we describe the specific details of the camera perspective(s) from which the (in this case) narrative scene is filmed. In general, these features might have very different meaning potentials, and we assume that most of the diverse functions and details of the video game interface will be described in this part of the box.

The final line of the box and thus the semantic representation makes explicit which referents and features are actual sources of evidence for the interpretation of the eventuality. In the example form, it is simply the discourse referents labeled with *a* and *d* that allow for the inference of *threaten*. The logical operator in this line, $|\sim$, indicates that this inference is a defeasible consequence relation drawn on the basis of the recipient's world knowledge. Again, details of this inference process and the underlying logical framework are given in Wildfeuer (2014), Bateman and Wildfeuer (2014b), Wildfeuer (2019). Most of this work is resulting from an extension of the SDRT framework and its notations to the multimodal context, which is why the boxes are diverging from the well-known DRT representations used for verbal language. This means that whereas in the case of language, it is compositional semantics derived for clauses that provides the content of such boxed discourse structure representations, in the case of the audio-visual material we derive a similarly structured representation on the basis of what is being attended to visually and auditorily. For further detail see the discussion in Bateman and Wildfeuer (2014b). The final line thus makes visible how recipients would normally interpret the combination of all semiotic resources: as an event in which the player and several other characters threaten a fourth character (which might be identified as female).

In the first step of our analysis in Section 4, we will build logical forms of the semiotic entities at work in order to describe the semantics of the game scene in more detail. As shown in the example logical form, we will on the one hand focus on narrative events and eventualities that construct the game's story and unfolding narrative. These eventualities will mostly be built on the characters and objects in the scene, the setting, and some typical technical features such as camera perspectives,

[2] These associations can be found due to "a number of physical, visual and auditory cues that are not available from narrative language, including quantitative information about the movement of actors, objects, changes in object contact relations, facial expressions, and environmental sounds" [Zacks et al., 2009: 318].

animations, etc. which will in many cases be very similar to filmic techniques. In addition, we will describe more instructive eventualities which focus on the instructional and interactive elements of a scene. These will then contain important information about how to play the game or how to process specific steps – details that are only partially connected to the narrative.

We shall note here that the description of what we find in the analysed materiality will sometimes contain both parts of the gameworld and elements that do not belong to the gameworld (instructions and information imposed onto the gameworld). We by no means claim that these non-gameworld elements are incorporated into the gameworld, but we include them in the logical forms as they communicate important information towards the player and they represent part of the semantics of the game scenes. We consider them part of the communicative situation, which is why they need to be described.

2.2. Analysing discourse structure: the glue logic

On the second level of our framework, on the level of coherence and structure, the logical forms constructed out of the semiotic entities are related to each other in order to form a coherent discourse structure. By constructing discourse relations between the eventualities that follow each other in the linear order of watching a film, reading a comic, or playing a game, dynamically unfolding discourse structures are built that are in most of our analyses so far narrative (as in narrative feature films and comics and graphic novels). On the basis of the original framework provided for verbal narrative and exploratory discourses [cf. Asher and Lascarides, 2003], its fruitful applications to many different non-verbal discourses (see above), and some exploratory work on argumentative structures (as non-narrative structures) [cf. Wildfeuer, 2017; Wildfeuer and Pollaroli, 2018a, 2018b], we assume, however, that it is likewise possible to describe *instructive* structures that are typically used in video games and especially in video game tutorials. With this, we aim at reconstructing the recipients' inferential reasoning about the coherence and connectedness of segments of the game.

For the detailed reconstruction of these structures, the second logic, the so-called *glue logic* for gluing together the discourse segments [cf. Asher and Lascarides, 2003: 185], provides a set of discourse relations with detailed definitions for these relations and the conditions that have to be fulfilled within the context in order to interpret these relations. Among these conditions, there are meaning postulates which are hard constraints that need to be met from context knowledge for a discourse relation to obtain. They follow the general form $\varphi_{R(\alpha,\beta)} \Rightarrow conditions(\alpha, \beta)$, which expresses the monotonic and thus non-abductive logic of information content [Asher and Lascarides, 2003: 159]. On the left side of this rule, it says that a particular discourse relation R is added to the current discourse structure between the segments labeled α and β. If this relation holds, the conditions on the righthand side are needed to follow by regular, non-defeasible material implication [cf. Bateman and Wildfeuer, 2014b]. A second condition is expressed in the respective default axiom for each relation, which specifies which discourse relations may apply given specified properties of the discourse elements being related. The default axiom follows the schema $(?(\alpha, \beta, \lambda) \wedge someprinciples) > R(\alpha, \beta, \lambda)$ where $?(\alpha, \beta, \lambda)$ indicates an underspecified discourse relation holding between segments α and β in the context of the discourse structure labeled λ. *R* is the specified abduced discourse relation and $>$ is a defeasible implication, which can typically be read as "if …then normally…". *some principles* represents the conditions that have to hold in order to give evidence for the relation. This rule is thus abductive because the identification of a discourse relation can always be overridden if more information becomes available.

On the basis of the set of film discourse relations defined by Wildfeuer (2014) and the one for comics and graphic narratives by Bateman and Wildfeuer (2014b), we provide an initial set of video game discourse relations and their corresponding formal definitions in terms of meaning postulates and default axioms in Table 1. We will explain the details of these rules and conditions as necessary in our analyses below.

Table 1
The initial set of video game discourse relations with meaning postulates and default axioms for each relation.

Relation	**Meaning Postulate** ($\varphi_{R(\alpha,\beta)} \Rightarrow conditions(\alpha, \beta)$) **Default Axiom** ($(?(\alpha, \beta, \lambda) \wedge some\ principles) > R(\alpha, \beta, \lambda)$)
Narration	$\varphi_{Narration(\alpha,\beta)} \Rightarrow overlap(prestate(e_\beta), poststate(e_\alpha))$ $(?(\alpha, \beta, \lambda) \wedge occasion(\alpha, \beta)) > Narration(\alpha, \beta, \lambda)$
Elaboration	$\varphi_{Elaboration(\alpha,\beta)} \Rightarrow contains\ (e_\alpha, e_\beta)$ $(?(\alpha, \beta, \lambda) \wedge specification_D(\beta, \alpha)) > Elaboration(\alpha, \beta, \lambda)$
Explanation	$\varphi_{Explanation(\alpha,\beta)} \Rightarrow before(e_\alpha, e_\beta)$ $(?(\alpha, \beta, \lambda) \wedge cause_D(\beta, \alpha)) > Explanation(\alpha, \beta, \lambda)$
Result	$\varphi_{Result(\alpha,\beta)} \Rightarrow after(e_\alpha, e_\beta)$ $\varphi_{Result(\alpha,\beta)} \Rightarrow cause(e_\alpha, e_\beta)$
Background	$\varphi_{Background(\alpha,\beta)} \Rightarrow overlap(e_\beta, e_\alpha)$ $(?(\alpha, \beta, \lambda) \wedge circumstantial\ information) > Background(\alpha, \beta, \lambda)$
Parallel	$\varphi_{Parallel(\alpha,\beta)} \Rightarrow \Box(K_\alpha \sim K_\beta)$ $(?(\alpha, \beta, \lambda) \wedge semantic\ similarity(\alpha, \beta)) > Parallel(\alpha, \beta, \lambda)$
Contrast	$\varphi_{Contrast(\alpha,\beta)} \Rightarrow \Box(K_\alpha \sim K_\beta)$ $?(\alpha,\beta,\lambda) \wedge semantic\ dissimilarity(\alpha,\beta)) > Contrast(\alpha,\beta,\lambda)$

With this set, we do not particularly distinguish between narrative and instructive relations. Instead, we assume that most of these relations can be used to construct different discourse structures and that it is rather the specific combination of several of these relations that then identifies the particular type of structure. It can be hypothesized, for example, that instructive structures more often contain causal relations such as an *Explanation* for certain processes or describe a *Background* to certain details of the gameworld. We will demonstrate the identification of these relations in the example extracts in Section 4, but the set needs to be subsequently investigated empirically with a larger amount of data which will then also allow to say more about the specific arrangement of narrative and instructive structures.

Reasoning about discourse relations is again a matter of defeasible reasoning in order to construct maximally coherent discourse structures. The process of discourse update, i.e., constructing the final discourse structure of the text, is therefore accompanied by the so-called Principle of Maximise Discourse Coherence [MDC, Asher and Lascarides, 2003], which binds together the discourse segments by finding the most preferable structure. Consequently, a pragmatically preferred discourse structure contains those relations that can be interpreted within the discourse because of the clear meaning postulates and default axioms that are available for each relation and which operate within a simple and counterbalanced interpretation.

Within this interpretation, a particular relation is usually more preferred than another relation because of the actual content and the information available for the recipient in the particular moment of the game, i.e., that content and information that is analytically reconstructed in the first step of our analysis.[3] The inference of the preferred relation then realises the best update of the structure in that it displays its coherence with the relation chosen as the most preferred one that maximises the coherence of the overall structure. We will explain the process of discourse update in further detail in our analyses below.

3. Data: video game tutorials and instructions

In our analysis, we focus on video game tutorials or instructional beginnings, which represent the initial stage of communication between a video game and a player in a majority of video games. During this stage, the game introduces the player to the gameplay (and sometimes the story), which will include teaching them the rules and mechanics, as well as the main commands, controller or pointer movements and basic actions. These instructions are written or spoken (or both), given as a voice-over or by a real teaching figure in the game. In games that have a narrative, they are given along with, i.e., parallel to, the initial stages of the narrative (*in medias res*), which means that they are integrated into the game itself and as such more implicit. Examples of such games belong to the *Metal Gear Solid* series, the *Assassin's Creed* series, most of the *Fallout* series, or the *Grand Theft Auto* series.

There are also video games in which tutorials are more explicit – they represent a separate part of the game, they are done in a small training-oriented gameworld and sometimes they are even optional. Examples of games with separate tutorials include *Cuphead*, *Mortal Kombat X* and a range of strategy and management games. Sometimes they are given as a prologue to the main story (e.g., *NieR: Automata*, *Hitman 2* or *Fallout 3*). Verbal instructional components are frequently combined with interaction cues, with the use of graphic design and employment of natural cues from the gameworld. There are also video games in which instructions that are expected to be found in the beginning tend to be extended until the end of the game – a good recent example is *Grand Theft Auto V* (Rockstar North, 2013, see Stamenković and Wildfeuer, 2021). In the game, the player is provided with instructions which pertain to the interaction with the gameworld and with the controller, which is likely a result of the fact that the game keeps introducing new features until the end.

Tutorials are meant to provide an interactive learning experience and this is their main function regardless of their position in the game. If they are placed outside the main storyline, players usually feel that the tutorial level is not part of the real game and tend to skip it. This is why many games are trying to camouflage it within the gameworld [Rouse, 2005, p. 129]. More and more games are avoiding complex textual instructions, as learning by doing is more efficient than learning by reading any tutorial text. According to Hodent [Hodent, 2018, p. 40], when one acts in the game and accomplishes a task, it requires a deeper level of processing in the working memory. On the other hand, pressing a button to acknowledge reading a tutorial text does not require a deep level of processing, and is thus not as effective. Hodent [Hodent, 2018, p. 48] adds that in developing tutorials, i.e., determining "the onboarding plan", game makers have to think about these instructional elements as very important, as learning how to play a game is a substantial part of the playing experience.

Along with being substantial, tutorials can provide a fertile ground for analysing discourse structures, as they potentially encompass rich communicative situations, given the fact that games usually have a well-defined and directed agenda of what needs to be delivered to the player in the process of training. This is why we chose them as the starting point in our analysis – they are likely to reveal complexities that might not be visible later in the game. A drawback of this focus is the fact that tutorials usually entail a lower level of interactivity than the rest of the game due to their relatively high density of instructive information that needs to be processed by the players.

[3] This reconstruction might then also include information about preceding events or the story as a whole. However, it usually does not include – on this level of analysis – more individual expectations or anticipations by recipients that are not given in the material itself.

4. Analysis

The focus of our analysis at this stage is on video games that have similarities with artefacts that have been analysed using the approach presented above, and having in mind that the original framework aimed at analysing verbal discourse and narrative and exploratory texts. We have thus chosen *Grand Theft Auto V* and *Batman: Arkham Knight* (Rocksteady Studios, 2015) as video games that are narrative and filmic, partly scripted, and which have cutscenes. In both video games, the tutorial appears as integrated in the main game and the initial instructions start appearing shortly after the very beginning of the game. They are presented in stages, and in regard to what is needed at particular moments in each game. We are aware that analysing other, less filmic games would yield results that are potentially quite different from those we will present here, and at the same time quite different from analysing films or comics.

In the coding procedure, the authors initially collected material from both the PC and the Playstation version of *Grand Theft Auto V* and *Batman: Arkham Knight* as well as YouTube recordings of both and produced screenshots of the scenes to be analysed in further detail. In a second step, they manually coded the events for both games (using a LaTeX document) on the basis of a template for the logical form as provided in Wildfeuer (2019) and given in Fig. 2. The construction and coding of the discursive structure of events happened on the basis of the YouTube recordings, as the process required moving backward and forward several times. At the same time, using the same recording allowed for stability and consistency for both coders. Every author processed this step independently by constructing logical forms and producing graphical representations in LATEX. While Author 1 is used to work with this framework for quite some time now, Author 2 learned most of the coding steps for this project. In the final step, therefore, the authors compared each other's interpretations of various elements of events and their relations, agreed on the final forms and structures to be used, and resolved all outstanding issues in discussion. Most of these were minor disagreements in how to exactly describe the eventuality or which technical features to add to the boxes.

e_{π_1} = threaten
Characters and Objects: [v] represented participant: player 1 (a) [v] represented participant: male character (b) [v] represented participant: male character (c) [v] represented participant: female character (d) [v] object: gun (e) [a] "Get down there! – Alright, everybody pays attention, no one gets hurt." (f)
Setting: [v] inside a building (g)
Technical Features: [v] dynamically changing camera perspective: MS & CUs (h)
Eventuality: $a, d \mid\sim$ **threaten** (e_{π_1})

e_{π_2} = command	$e_{\pi_{2a}}$ = inform
Characters and Objects: [v] represented participant: player 1 (a) [v] represented participant: male character (b) [v] represented participant: male character (c) [v] represented participant: male character (i) [v] object: door (j) [a] "Open the door or they'll get worse than hurt! – Hey, hey, ah! Come on!" (k)	Characters and Objects: [v] object: radar/map (w)
Setting: [v] inside a building (g)	Setting: [v] inside a building (g)
Technical Features: [v] dynamically changing camera perspective: MLS & CU (l)	Technical Features: [t] instruction note: "The radar shows your position" (x)
Eventuality: $a, i, k \mid\sim$ **command** (e_{π_2})	Eventuality: $w, x \mid\sim$ **inform** ($e_{\pi_{2a}}$)

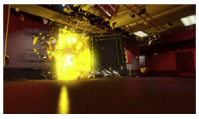
Fig. 2. The logical forms for the first two narrative events/eventualities in the extract from *GTA V* as well as the subordinated instructive eventuality ($\boldsymbol{e_{\pi_{2a}}}$); notation style adapted from Wildfeuer (2014, 2019).

4.1. Analysis: GTA V

As a first example, we take the opening scenes from the game *GTA V*, i.e., the game's prologue scenes, and analyse both the logical forms of multimodal meaning-making as well as the unfolding discourse structure of these scenes. For our analysis, we played the game and then worked with one of the recordings of the game that is available on YouTube, more particularly the

Singleplayer Walkthrough by TmarTn, https://www.youtube.com/watch?v=TOxuNbXrO28&ab|_channel=TmarTn, starting at 00:00:27 of the video (after an initial introduction by the player). Table 2 gives a rough transcription of the first six minutes, showing the main events and actions and featuring some notes on narrative elements or the setting, instructive elements, as well as interactive elements.

Table 2
Transcription of the first 6 min of the game recording of *GTA V*.

#	shot	narrative elements	instructive elements	interactive elements
1.		insert: spatiotemporal information	//	//
2.		robbery scene: characters & setting	//	//
3.		robbery scene: characters & setting	radar/map indicating positions of characters; instruction note(s): “The radar shows your position within the world”	//
4.		robbery scene: characters & setting	radar/map; instruction note(s): “Press X to bring up the phone”	phone in right corner
5.		robbery scene: characters & setting	//	//
6.		robbery scene (with filter): characters & setting	radar/map; instruction note: “…Use R to select a character, then release X to switch”	character switch menu
7.		shooting scene (in snow): characters & setting; insert in right corner	radar/map (with filter); instruction: “Hold LT to aim at enemies while in cover, then shoot with RT”	ammunition information, stars indicating police interest
8.		driving scene: characters & setting	//	//

The transcription of the prologue makes visible that besides the various narrative elements, there are also many important elements that are rather of instructive character or lead to a specific interaction with the game’s interface. For example, instruction notes are usually shown in the top left corner of the screen and inform or instruct the player in written text which keys to use and what to do next (see row 3). In row 4 of the table, a phone is shown which plays a crucial role for the unfolding of the narrative, since its use causes the explosion of the strong room and the possibility to

steal the money from there. The phone thus explicitly displays an interactive feature that requests the player's action to which the game reacts.

In general, we can identify several main events of a story, such as the robbery scene inside the building and a shooting scene in the snow, outside the building, which is ended by a driving scene that later on mainly shows the player and characters inside a car. These events clearly form a narrative structure that tells the dynamically evolving story in a very film-like atmosphere.

Several other instructive or interactive elements that appear in the prologue are added to these narrative events. For example, a map or radar is inserted which shows, as an additional instruction note explains, where in the gameworld the characters and player(s) are. Written details such as the information about the ammunition or instructions in the scheme of "Press X to do Y" do not immediately add to the narrative event, but rather present information about the interface or specific game processes on another description level. Some of these instructive elements are related to each other and work together in their sequential order; others are completely independent, but might be related to details in the gameworld.

4.1.1. GTA V: building logical forms

By first describing the logical forms resulting from these initial descriptions, we are able to formalize the details of the information content of this scene. Fig. 2 therefore presents the result of defeasibly reasoning about this information content and building discourse representations of what is shown and given in the first part of the scene. The eventualities each represent the events displayed in each row of Table 2. With this, we focus on the main events and the accompanying

e_{π_3} = lock up	$e_{\pi_{3a}}$ = instruct
Characters and Objects: [*v*] represented participant: player 1 (a) [*v*] represented participant: male character (b) [*v*] represented participant: male character (c) [*v*] represented participant: female character (d) [*v*] represented participant: male character (i) [*v*] represented participant: male character (m) [*v*] represented participant: male character (n) [*v*] represented participant: male character (o) [*v*] object: room (j) [*a*] "Hurry! – In the back! Come on!" (p)	Characters and Objects: [*v*] object: radar/map (w)
Setting: [*v*] inside a building (g)	Setting: [*v*] inside a building (g)
Technical Features: [*v*] dynamic moving camera in medium shot (q)	Technical Features: [*t*] instruction note: "Press X to bring up the phone." (y)
Eventuality: $a-c, d-o, j \mid\sim$ **lock up** (e_{π_3})	Eventuality: $y \mid\sim$ **instruct** ($e_{\pi_{3a}}$)

e_{π_4} = call	$e_{\pi_{4a}}$ = instruct
Characters and Objects: [*v*] represented participant: player 1 (a) [*v*] represented participant: male character (b) [*v*] represented participant: male character (c) [*a*] "All set. Phone it in." – "I'm calling it" (r) [*v*] object: phone (s)	Characters and Objects: [*v*] object: radar/map (z) [*v*] object: phone (a)
Setting: [*v*] inside a building (g)	Setting: [*v*] inside a building (g)
Technical Features: [*v*] dynamic moving camera in medium shot (q)	Technical Features: [*t*] instruction note: "Press A to view the contact list... (b)
Eventuality: $b, r, s \mid\sim$ **call** (e_{π_4})	Eventuality: $z, a \mid\sim$ **instruct** ($e_{\pi_{4a}}$)

e_{π_5} = explode
Characters and Objects: [*v*] object: detonator (t)
Setting: [*v*] strongroom (u)
Technical Features: [*v*] dynamically changing camera perspective: CU & MLS (v)
Eventuality: $t \mid\sim$ **explode** (e_{π_5})

Fig. 3. The logical forms for the next three narrative events and their subordinated instructive eventualities in the scene from *GTA V*.

important instructions to keep the game running. It would in fact be possible to describe even more and smaller events such as 'run through the building' or 'open the door of the car'. We forego this description in order to zoom in on the main instructions that guide the player through the game.

The first eventuality, e_{π_1} = *threaten*, as well as the second one, e_{π_2} = *command* are narrative eventualities that form the beginning of an unfolding storyline in the game in which the player as well as several other characters are involved in a robbery that can be constructed out of events such as threatening or commanding the characters in the building. As becomes visible in each second row of the two boxes, all characters, objects as well as the (most important) voice tracks are listed. The last line of the box then gives details about which of these specific discourse referents lead to the inference of the respective eventuality. In e_{π_1}, it is the player (a) who threatens one other character (d); in e_{π_2}, it is the player (a) who commands another character (i) to open the door.

This second eventuality is accompanied by another subordinated eventuality, $e_{\pi_{2a}}$ = *inform*, which is related to the narrative eventuality e_{π_2} in that the additional object, the radar or map, is added to the audio-visual display as a further level of description. It is important to note that this object is not part of the gameworld and can only be seen by the player looking at the interface (and for example not by the characters in the story); however, it displays the characters and their positions in the gameworld and is therefore strongly connected to the narrative eventuality. The object itself clearly is informative.

The instruction note that is displayed in the same eventuality has a similar function, namely to inform the player about (specific details of) the object. Since this note is almost purely textual and does not construct any other character or object, we see it more as a specific technical feature of the eventuality, similar to verbal inserts in filmic scenes or speech bubbles or captions in comic panels [see for further details Wildfeuer, 2014, 2019]. In $e_{\pi_{2a}}$, the instruction note explains the details of the map. We will see below that similar notes may also clearly instruct the player to press certain keys, for example.

In Figs. 3, 4 and 5, more narrative eventualities as well as subordinated instructive eventualities are constructed. While the narrative eventualities on the left list all characters involved in the scene as well as several objects that play a role in the storyline, the instructive eventualities all feature the radar/map and several distinct instruction notes. Depending on whether these notes rather inform or clearly instruct the player, the verbalization of the eventuality changes. The analysis in Section 4.1.2 will show how these eventualities relate to each other and construct a coherent discourse structure (see Fig. 6).

A further interesting case of specific interactive elements is, for example, given in $e_{\pi_{6a}}$. The scene offers a camera switch in order to show the actions and events from the perspective of another character and take over their actions. For this, a specific menu is shown in the bottom right corner that shows which character can be chosen for the camera switch. The instruction

e_{π_6} = detain	$e_{\pi_{6a}}$ = switch
Characters and Objects: [*v*] represented participant: player 1 (*a*) [*v*] represented participant: male character (*b*) [*v*] represented participant: male character (*w*) [*v*] object: gun (*j*) [*a*] "Give it up. – I got him!" (*x*)	Characters and Objects: [*v*] object: radar/map (*z*) [*v*] object: camera switch menu (*a*) [*v*] represented participant: player 1 (*b*) [*v*] represented participant: player 2 (*c*)
Setting: [*v*] inside a building (*g*)	Setting: [*v*] inside a building (*g*)
Technical Features: [*v*] dynamic moving camera zooming in (*y*)	Technical Features: [*t*] instruction note: "To switch hold to show the available characters" (*d*) [*v*] dynamic lighting, filters (*e*)
Eventuality: $w, a \mid\sim$ **detain** (e_{π_6})	Eventuality: $b, c \mid\sim$ **switch** ($e_{\pi_{6a}}$)
e_{π_7} = shoot	$e_{\pi_{7a}}$ = instruct
Characters and Objects: [*v*] represented participant: player 2 (*c*) [*v*] represented participant: male character (*b*) [*v*] represented participant: male character (*f*) [*v*] represented participant: cop (*g*) [*v*] represented participant: cop (*h*) [*v*] represented participant: cop (*i*) [*v*] object: guns (*j*)	Characters and Objects: [*v*] object: radar/map (*l*)
Setting: [*v*] outside the building (*k*)	Setting: [*v*] outside the building (*m*)
Technical Features: [*v*] dynamic moving camera, zooming (*l*) [*t*] insert: North Yankton (*m*)	Technical Features: [*t*] instruction note: "Hold LT to aim at enemies while in cover, then shoot with RT" (*n*) [*v*] munition info (*o*)
Eventuality: $b - f, g - i \mid\sim$ **shoot** (e_{π_7})	Eventuality: $l, n, o \mid\sim$ **instruct** ($e_{\pi_{7a}}$)

Fig. 4. The logical forms for the remaining narrative events and their subordinated instructive eventualities in the scene from *GTA V*.

e_{π_8} = drive
Characters and Objects: [v] object: car (n) [v] represented participant: player 2 (c) [v] represented participant: male character (b) [v] represented participant: male character (f)
Setting: [v] street covered in snow (u)
Technical Features: [v] dynamic camera: long shots, zooming in on characters (v)
Eventuality: $n \mid\sim$ **drive** (e_{π_8})

Fig. 5. Final event in the scene from *GTA V*.

note on the top left corner says: "To SWITCH hold 'down' to show the available characters." The lighting and filters used in this scene change when the 'down' key is held, so that the whole action is stopped for a moment. If the player processes these instructions and starts the camera switch, the instruction note changes to "While holding 'down' use R to select a character, then release 'down' to SWITCH." As soon as the player processes these steps, the action proceeds from the perspective of the new character. This of course also has consequences for the unfolding of the narrative, and our analysis of the discourse relations holding between these eventualities in Section 4.1.2 will show more details. Furthermore, the roles of the represented participants in the remaining logical forms and the dependencies between these participants and the other discourse referents need to be updated as well.

The scene we analyse here then ends with a simple narrative event without any further instructions in which the characters have entered a car and this car then drives away. The camera switches between long shots with the car on the street and medium shots and close ups of the characters and players in the car.

4.1.2. GTA V: relating the logical forms to each other

In the second step of the analysis, the logical forms built to display the interplay of the resources and referents are now related to each other with the help of discourse relations in order to construct the unfolding discourse structure. In Fig. 6, we first show a graphical representation of this discourse structure with the embedding and subordination of various substructures, each labeled π', π'', etc.

While the narrative structure continuously unfolds and is usually based on discourse relations such as *Narration* or *Result*, several instructional elements each form a further subordinated eventuality which, as we saw in Section 4.1.1, is always labeled as π_{n_a}. As becomes visible, these subordinated eventualities always hold an *Elaboration*-relation to the superordinated narrative event which is displayed with an arrow to the right. Some of these subordinated eventualities, such as $e_{\pi_{4_a}}$ and $e_{\pi_{5_a}}$ also hold a relation to the following narrative event, because they clearly influence or, even stronger, cause the further unfolding of the story. We will now explain in further detail how these discourse relations are inferred on the basis of the formal framework described above.

According to Asher and Lascarides (2003: 462), *Narration*-relations always hold between events that "occur in the sequence in which they are described". Furthermore, these events have to temporally overlap in that the end of the first eventuality is directly succeeded by the beginning of the second eventuality and that both eventualities are arranged in a temporal and spatial continuity. The meaning postulate for *Narration* that has been developed for multimodal discourse is therefore

$$\varphi_{Narration(\alpha,\beta)} \Rightarrow overlap(prestate(e_\beta), poststate(e_\alpha))$$

A crucial further factor for the inference of this relation is that the first eventuality occasions the second one in a typical 'natural-event sequence'. This is expressed in the default axiom for *Narration*:

$$(?(\alpha,\beta,\lambda) \wedge occasion(\alpha,\beta)) > Narration(\alpha,\beta,\lambda)$$

As soon as both conditions, which are in fact rather weak, are fulfilled, a *Narration*-relation can be inferred. This is the case for all situations in the game with two (or more) sequential events in the gameworld where the characters take some action. For instance, the eventualities e_{π_1} and e_{π_2} generally overlap in that the event of threatening is directly followed by the event of commanding. Since these two events also occasion each other, the inference of a *Narration*-relation is possible and feasible, since no other condition is fulfilled.

In other situations and under other conditions, however, the default axiom for *Narration* and especially the condition *occasion* can very quickly be overwritten by another condition, such as a *cause*, for example. As soon as it is possible to infer a causal relationship between the two eventualities, *Narration* is no longer the preferred discourse relations. Especially when the temporal conditions also slightly change and the eventualities no longer explicitly overlap, but rather follow each other in a temporal sequence, it is more likely that a *Result*-relation is inferred. This is for example the case between e_{π_4} and e_{π_5} where the action of dialling a certain number on a phone is causing the explosion of the strongroom and the latter eventuality thus

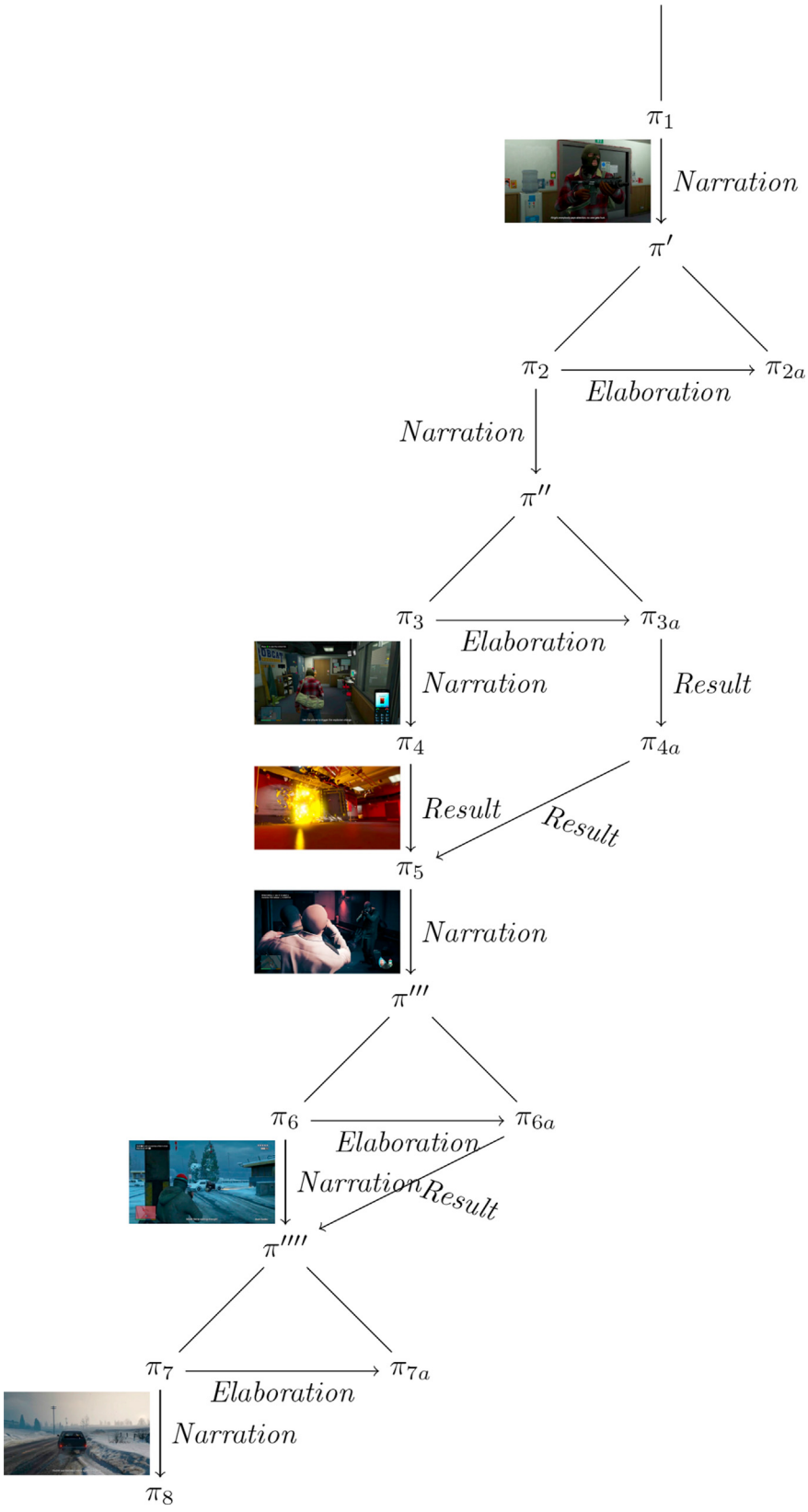


Fig. 6. Graphical representation of the discourse structure of the first extract from *GTA V*.

displays the *Result* of the preceding eventuality. Although the two events are in a spatiotemporal consequence, which would also be decisive for a *Narration*-relation, the effects for *Result* are much stronger.

A special case is given with the eventualities that contain elements of the interface such as the map or an instruction note. These elements are temporally included in the narrative eventuality which means that the latter contains the instructive eventuality. With this, they fulfill the meaning postulate for *Elaboration*:

$$\varphi_{Elaboration(\alpha,\beta)} \Rightarrow contains\ (e_\alpha, e_\beta)$$

At the same time, these elements represent a certain specification of the content and especially the interface that is clearly situated on another level of description. According to Asher & Lascarides (Asher and Lascarides, 2003; Bogost, 2015), the semantic function of this is a change of granularity of the description which is clearly given in the general difference between the gameworld depicted in the narrative events as well as the interface depicted in the more instructive eventualities. The specification given by elements such as a map or the respective instruction notes thus fulfill the default axiom for *Elaboration*:

$$(?(\alpha, \beta, \lambda) \wedge specification_D(\beta, \alpha)) > Elaboration(\alpha, \beta, \lambda)$$

and therefore enable the inference of an *Elaboration*-relation. We formally demonstrate this in Fig. 7 which contains both eventualities as well as a further line at the end of each of these eventualities that gives the temporal conditions holding for each logical form.

π_2, π_{2_a}

π_2 :	e_{π_2} = command
	CHARACTERS AND OBJECTS: [v] repr.part.: player 1 (a) [v] repr.part.: male character (b) [v] repr.part.: male character (c) [v] repr.part.: male character (i) [v] object: door (j) [a] "Open the door ... (k)
	SETTING: [v] inside a building (g)
	TECHNICAL FEATURES: [v] camera: MLS & CU (l)
	EVENTUALITY: $a, i, k \mid\sim$ **command** (e_{π_2}) $holds(e_{\pi_2}, t_1)$

π_{2_a} :	$e_{\pi_{2a}}$ = inform
	CHARACTERS AND OBJECTS: [v] object: radar/map (w)
	SETTING: [v] inside a building (g)
	TECHNICAL FEATURES: [t] instruction note: "The radar shows your position" (x)
	EVENTUALITY: $w, x \mid\sim$ **inform** ($e_{\pi_{2_a}}$) $holds(e_{\pi_{2_a}}, t_2)$ $contains(t_1, t_2)$

$Elaboration(\pi_2, \pi_{2_a})$

Fig. 7. Logical form of the discourse structure for the eventualities e_{π_2} and $e_{\pi_{2_a}}$

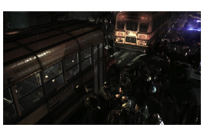
Since all instructive eventualities represent some sort of *specification* and are usually part of or included in the narrative event, they all hold an *Elaboration*-relation to this narrative event and with this construct a subordinated discourse structure. Each of these relations could be displayed similarly to the logical form given in Fig. 7. More examples of these and more complex discourse structures are given in Wildfeuer (2014).

4.2. Analysis: Batman: Arkham Knight

For our second example, we played and watched the initial stages of the video game *Batman: Arkham Knight* and, just like in the previous game, analysed both the logical forms of multimodal meaning-making and the unfolding discourse structure of the scenes encompassed in roughly the first eight minutes of the game. The YouTube channel lzuniy (https://www.youtube.com/watch?v=hqF0r4aTEAs) contains the section we analysed. Table 3 gives a transcription of the analysed initial part of the game and shows the main events, transitions, and actions and includes narrative elements, the setting and instructive/interactive elements.

As compared to *GTA V*, *Batman: Arkham Knight* in its initial stages is more intensively scripted, so there is a more diverse body of scenes that resemble film beginnings, but there are spots where narrative elements are interrupted by instructive and interactive elements, which is why the structure of the table resembles the one for *GTA V*. In both games, the tutorial is

Table 3
Transcription of the first 8 min of the game recording of *Batman: Arkham Knight*.

#	shot	narrative elements	instructive elements	interactive elements
1.		crematory furnace introduction: furnace mechanism pieces shown	//	//
2.		cremation scene: Joker in a furnace	instruction note(s): Incinerate (Space)	//
3.		Gotham introduction scene: various characters in rainy Gotham, first person narration	//	//
4.		diner conversation scene: a policeman (first person/player) and a waitress in Pauli's Diner	//	//
5.		smoking customer confrontation scene: a policeman (first person/player), a concerned customer and a smoking customer in Pauli's Diner	instruction note(s): (W)-(A)-(S)-(D) – To Move; (Space) – Confront Smoking Customer	//
6.		diner fear gas fight: a policeman (first person/player) and a group of skeleton-like characters	//	//
7.		fear gas attack report: Gotham citizens watching news report in the street	//	//
8.		Scarecrow warning: Scarecrow claims responsibility for the attack and warns Gotham citizens	//	//
9.		evacuation scene: Gotham citizens in the street with evacuation buses	//	//
10.		street chaos scene: criminals in Gotham streets vandalizing cars and shops	//	//

(*continued on next page*)

Table 3 (*continued*)

#	shot	narrative elements	instructive elements	interactive elements
11.		Batman introduction scene: Batman standing on a tall structure towering above Gotham	//	//
12.		gliding scene: Batman starting to glide towards Gotham buildings	instruction note(s): Glide (Space); compass line showing locations of interest; target detail/mission info	location of interest distance (553 m)
13.		landed scene:Batman on the GCPD building	instruction note(s): Glide (Space), Climb Down (LCtrl)+ mouse icon; compass line showing locations of interest; target detail/mission info	location of interest distance (60 m); local surveillance broadcast indicator and range

integrated in/intertwined with the narrative. The usual location of controller-related instructions is the bottom right part of the screen, but in the scene portrayed in the fifth row, these move to the centre of the screen. In the non-scripted part of the game, the middle right part of the screen is a placeholder for information and there is a dynamic compass line in the upper middle part of the screen.

We can differentiate between three major types of scenes in the transcription – most of them are entirely scripted, i.e., they represent interrelated cutscenes that are essentially film-like. The second type includes the scenes 2 and 5, where the player interacts with the game in a very basic manner – they are asked to press a button, or move from place to place, but they are tied to the scripted part of the game. The third type includes rows 12 and 13 and it reflects what we can call 'free play' – in this part of the game, players do have mission goals and tasks, but they are allowed to move freely, which is also possible in *GTA V* after the prologue we describe ends and the main part of the game starts. The initial scenes in *Batman: Arkham Knight* can also be called a (short) prologue, as the Joker incineration scene portrayed in rows 1 and 2 happens nine months prior to the main game events.

The free play part of the game allows for imposing markers, numbers, and other types of information onto the gameworld, so that one can see that the number showing distance from the next important point is written directly above the location the player needs to reach, which is one of the elements that make this part of the game different from the cutscenes. As this part of the game is not scripted, it can result in a different discourse structure for each player, since, in open-world video games, they are able to choose the order in which they will complete main and side missions. so that one can only analyse different instances of gameplay.

4.2.1. Batman: Arkham Knight: building logical forms

In describing the logical forms identified in our second example, we will mostly focus on the similarities with the first example (*GTA V*) and various specificities we detected that seemed to be unique to *Batman: Arkham Knight*. The figures will all represent defeasibly reasoning about the content and the discourse structure of what we see, hear, and read in the opening part of *Batman: Arkham Knight*. As both games are filmic and narrative, the eventualities will have compositions that will not be excessively different from other instances of formalization of a narrative discourse, yet they will have certain peculiarities, especially as we are nearing the free play part of the game.

After a rather unusual eventuality e_{π_1} = *introduce*, in which the gloomy crematorium scenery is accompanied by Sinatra's "I've Got You Under My Skin", we reach the next step in narration, the eventuality e_{π_2} = *incinerate*, where the game gives the player its first (though very basic) task and makes her/him the agent in the scene (see Fig. 8). The agency is transferred through the instruction note that reads "Incinerate (Space)", and this is represented in an accompanying subordinated eventuality $e_{\pi_{2a}}$ = *instruct*, similar to several eventualities we encountered in *GTA V*. The instruction note is tied to the narrative, but none of the instructions we encounter belong to the gameworld. With the following two eventualities, e_{π_3} = *introduce* and e_{π_4} = *order dinner*, we encounter the following two steps in the narrative (Fig. 9). In e_{π_3} we learn that there is a nine-month gap between the incineration scene and what follows. In the eventuality e_{π_4}, the perspective shifts into the next character the player will control, Officer Owens. The player takes control as we reach the eventuality e_{π_5} = *confront*, in which they are asked to go and confront a strange character smoking in a diner booth (Fig. 10). During this scene, the player is instructed on how to move the character and to perform basic actions. This is represented by the subordinated eventuality $e_{\pi_{5a}}$ = *instruct*. The following six eventualities, from e_{π_6} = *kill* to $e_{\pi_{11}}$ = *introduce* (represented in Figs. 10–12) follow the

e_{π_1} = introduce
CHARACTERS AND OBJECTS: [v] represented participant: unknown character/dead Joker (a) [v] object: metal pieces of a furnace (b) [v] object: cogs (c) [v] object: lying surface (d) [v] object: blazes (e) [v] object: withered flower (f) [v] object: furnace (g)
SETTING: [v] inside a crematorium (h)
TECHNICAL FEATURES: [v] dynamically changing camera perspective: CUs & MS (i) [v] music: "I've Got You Under My Skin" by Frank Sinatra (j)
EVENTUALITY: $g \mid\sim$ **introduce** (e_{π_1})

e_{π_2} = incinerate	$e_{\pi_{2a}}$ = instruct
CHARACTERS AND OBJECTS: [v] represented participant: dead Joker (a) [v] represented participant: player as agency (k) [v] object: furnace (g)	CHARACTERS AND OBJECTS: [v] (–)
SETTING: [v] inside a crematorium (h)	SETTING: [v] inside a crematorium (h)
TECHNICAL FEATURES: [v] top-down CU on dead Joker (l)	TECHNICAL FEATURES: [v] instruction note: "Incinerate (Space)" (m)
EVENTUALITY: $k, a \mid\sim$ **incinerate** (e_{π_2})	EVENTUALITY: $m \mid\sim$ **instruct** ($e_{\pi_{2a}}$)

Fig. 8. The logical forms for the first two narrative events/eventualities in the excerpt from *Batman: Arkham Knight* as well as the subordinated instructive eventuality ($e_{\pi_{2a}}$)

unfolding of a mostly filmic narrative, scene by scene, and this is described in a way comparable to what can be found in a formalised approach to typical feature films.

The final introduction to the character of Batman shifts the perspective to Batman's own and after $e_{\pi_{11}}$, the player takes control of Batman as the main character and enters the part of the game that depends on how well they will play the game. Within this part, players essentially need to complete missions and follow the instruction, but are at the same time allowed to choose their path and roam freely in between missions. This leads us to two levels where eventualities are organized in a structure that was not found in the prologue to *GTA V* (Fig. 13). Namely, the eventuality $e_{\pi_{12}}$ = *reach*, the first one that happens in a 'live' game environment, is accompanied by two subordinated eventualities, ($e_{\pi_{12a}}$) and ($e_{\pi_{12b}}$), one of which represents the instructive layer imposed onto $e_{\pi_{12}}$, while the other describes the informative elements contained on the canvas at the same time at which $e_{\pi_{12}}$ unfolds. There are two types of instruction notes we can see here – one of them relates to the keyboard controls and their use ('Glide (Space)'), and the other is bound to the gameworld and defines the point the player needs to reach in order to come closer to completing the first mission ('Meet Commissioner Gordon by the…'). The subordinated eventuality related to providing information encompasses details on distance and target and gives us a green attention marker that appears on the linear compass, which allows the player to have a proper orientation. The last eventuality in the described sequence, $e_{\pi_{13}}$ = *reach*, accompanied by ($e_{\pi_{13a}}$) and ($e_{\pi_{13b}}$) (see Fig. 14), functions in a very similar way, but its instructions feature a drawing of a controller (computer mouse) and its informative component includes local surveillance details, imposed as text and a graph onto the main screen and remaining outside of the gameworld (thus communicating with the game player only).

4.2.2. Batman: Arkham Knight: relating the logical forms to each other

The second step of the analysis again aims at relating the logical forms to each other with the help of discourse relations. This gives us the unfolding discourse structure of the beginning of *Batman: Arkham Knight* presented graphically in Figs. 15 and 16. The figure again presents a graphical representation of this discourse structure with embedding and subordination, which appears slightly more complex, mostly due to the fact that it encompasses free play at its end. Just like with *GTA V* the narrative structure in *Batman: Arkham Knight* is mostly based on two discourse relations: *Narration* or *Result*. Along with these, we also encounter several instructional and informative elements. These subordinated eventualities are labeled using π_{n_a} (and in some cases π_{n_b}). For the same reasons described in Section 4.1.2, subordinated eventualities hold the relation of

e_{π_3} = introduce
Characters and Objects: [v] represented participants: various people in the streets (n) [v] represented participant: narrator (o) [v] represented participant: a police officer/Officer Owens (p) [v] object: Pauli's Diner (q) [v] objects: various vehicles (r) [v] object: a kiosk (s) [v] object: Halloween party poster (t) [v] objects: jack-o'-lanterns (u) [v] Credits appearing (v) [a] "–Nine months ago Joker was cremated... I was just waiting for someone to pull the trigger." (w)
Setting: [v] rainy streets of Gotham (x)
Technical Features: [v] dynamically changing camera perspective and moving camera: WS, FS, craning, dollying, different angles; (y)
Eventuality: $x, q \mid\sim$ **introduce** (e_{π_3})
e_{π_4} = order dinner
Characters and Objects: [v] represented participant: Officer Owens (p) [v] represented participant: a waitress (z) [v] represented participants: several background customers (a) [v] object: a mug of coffee (b) [v] object: a cash register (c) [v] Credits appearing (d) [a] So, what'll it be? The usual? – No, you know what, make it waffles. (e)
Setting: [v] inside Pauli's Diner (q)
Technical Features: [v] POV with MS and FS on the waitress (f)
Eventuality: $p, z \mid\sim$ **order dinner** (e_{π_4})

Fig. 9. The logical forms for the third and the fourth narrative event/eventuality in the excerpt from *Batman: Arkham Knight*.

Elaboration to the superordinated narrative event in each instance. The instructive subordinated eventualities (namely, $e_{\pi_{2a}}$, $e_{\pi_{5a}}$ and $e_{\pi_{12a}}$) hold a *Result* relation to the following narrative event, not only because they influence, but also because they directly cause the following event – performing an action in line with the instructions given allows for the enaction of the next narrative event.

Even though the temporal gap between different described events varies and ranges from nine months to partial overlaps, they do occur in the described sequence, which is why we have opted for *Narration* relations in most of them. In those instances that involved a strong sense of causality (like in the relation between e_{π_8} = *threaten* and e_{π_9} = *evacuate*, where Gotham citizens escape the city *because* Scarecrow threatens them) and in relations that involved instructive elements ($e_{\pi_{2a}}$, $e_{\pi_{5a}}$ and $e_{\pi_{12a}}$), the *Result* relation seemed a more precise description. Finally, the eventualities that contain elements of the interface such as markers, instruction notes, compass, distance meters, are contained in the narrative eventuality, although

e_{π_5} = confront	$e_{\pi_{5a}}$ = instruct
Characters and Objects: [v] represented participant: Officer Owens (p) [v] represented participant: a concerned customer (g) [v] represented participant: a smoking customer (h) [v] represented participants: several customers (i) [v] object: a police hat (j) [v] object: a cigarette (k) [v] Credits appearing (l) [a] "– Hey I'm sorry to interrupt your dinner, officer. – Excuse me, sir. There's no smoking in here. (m)	Characters and Objects: [v] (–)
Setting: [v] inside Pauli's Diner (q)	Setting: [v] inside Pauli's Diner (q)
Technical Features: [v] POV with MS and FS on the customers (n)	Technical Features: [v] instruction notes: "(W)-(A)-(S)-(D) – To Move" (o) [v] instruction notes: "(Space) – Confront Smoking Customer)" (p)
Eventuality: $a, c \mid\sim$ **confront** (e_{π_5})	Eventuality: $o, p \mid\sim$ **instruct** ($e_{\pi_{5a}}$)

e_{π_6} = kill
Characters and Objects: [v] represented participant: Officer Owens (p) [v] represented participant: a smoking customer/skeleton (h) [v] represented participants: several skeletons and victims (r) [v] object: a jack-o'-lantern (s) [v] object: a pistol (t) [a] "– No." (u)
Setting: [v] inside Pauli's Diner (q)
Technical Features: [v] POV with MS and FS on skeletons/people (v)
Eventuality: $h, p \mid\sim$ **kill** (e_{π_6})

Fig. 10. The logical forms for the eventualities $\mathbf{e}_{\pi_5}$, $\mathbf{e}_{\pi_{5a}}$ and $\mathbf{e}_{\pi_6}$ in the excerpt from *Batman: Arkham Knight*.

they are not part of the gameworld. Similarly to *GTA V*, they fulfill the meaning postulate for *Elaboration*, as we have given them in Section 4.1.2.

A specificity that occurs only in *Batman: Arkham Knight* (and not in the analysed part of *GTA V*) can be found at the very bottom of the graphical representation in Fig. 16. Namely, there are two additional descriptive layers imposed onto the free gameplay narrative layer. As the instructive elements that present $e_{\pi_{12a}}$ are not only more relevant to gameplay, but also far more likely to be experienced than the informative elements found in $e_{\pi_{12b}}$, we have opted for the sequence in which the direction will be from $e_{\pi_{12}}$ to $e_{\pi_{12a}}$, and only then to $e_{\pi_{12b}}$. The same applies to $e_{\pi_{13}}$, $e_{\pi_{13a}}$ and $e_{\pi_{13b}}$.

5. Conclusions and further research directions

Our analysis has shown that the tutorials in the two video games exhibit similarly complex discourse structures pertaining to instructions and gameplay information. It can be assumed that this is mostly due to the fact that they belong to the same genre and type. The instructions in both games are usually subordinated to the main narrative structure, but the subordination always actively leads back to the main structure and with this directly allows for the continuation of the narrative.

At the same time, we have witnessed in both cases a rather similar use of discourse relations: (1) the *Elaboration*-relation for subordinated instructions and information segments that unfold in parallel to the narrative, and (2) the *Result*-relation for (re-)embedding the subordinated structures into the main narrative structure. This particular relation forms a recurring structure when it comes to the free play part of the game, as there is a pattern in which performing certain actions in accordance with the provided information and instructions leads us to the next bit of each game. It is very likely that within these two games, as well as in most games belonging to the same genre, we can predict that the recursion involving the *Result*-relation will be a distinctive feature, as it allows for progression that makes the game go forward. These initial observations build an important basis, as we think, for further evaluations of the notion of *procedurality* or *procedural rhetoric* in order to highlight the persuasive capacities of digital games as computational artefacts (Bogost, 2007; Murray, 2012;

e_{π_7} = inform
Characters and Objects: [v] represented participants: Gotham citizens (w) [v] represented participant: a smoking customer/skeleton (x) [v] objects: screens with news report on attack (y) [a] "Suspected chemical weapon attack in Gotham diner" (z)
Setting: [v] Gotham streets (a)
Technical Features: [v] MS and CUs on people (b)
Eventuality: $y, w \mid\sim$ **inform** (e_{π_7})

e_{π_8} = threaten
Characters and Objects: [v] represented participants: Gotham citizens (w) [v] represented participant: Scarecrow on a big screen (c) [v] objects: vehicles (d) [a] "–This demonstration used just 5 ounces of my latest toxin. – Gotham, this is your only warning." (e)
Setting: [v] Gotham streets (a)
Technical Features: [v] dynamically changing camera perspective: CUs & MS (f) [v] noise in Scarecrow's broadcast (g)
Eventuality: $c, w \mid\sim$ **threaten** (e_{π_8})

e_{π_9} = evacuate
Characters and Objects: [v] represented participants: Gotham citizens (w) [v] represented participant: narrator (o) [v] objects: evacuation buses (h) [v] objects: police vehicles (i) [v] objects: other vehicles (j) [v] object: a helicopter (k) [a] "– Scarecrow's threat worked..." (l)
Setting: [v] Gotham streets (a)
Technical Features: [v] dynamically changing camera perspective and moving camera: MS, FS, high-angle, aerial (m)
Eventuality: $h, w \mid\sim$ **evacuate** (e_{π_9})

Fig. 11. The logical forms for the eventualities $\mathbf{e}_{\pi_7}$, $\mathbf{e}_{\pi_8}$ and $\mathbf{e}_{\pi_9}$ in the excerpt from *Batman: Arkham Knight*.

$e_{\pi_{10}}$ = take over
CHARACTERS AND OBJECTS: [v] represented participant: narrator (o) [v] represented participant: a dead policeman (p) [v] represented participant: Penguin (q) [v] represented participant: Two-Face (r) [v] represented participants: various criminals/vandals (s) [v] objects: vehicles (t) [v] objects: birds (u) [v] Credits appearing (v) [a] "– The only people left on the streets..." are the sort that enjoy the chaos. (w)
SETTING: [v] Gotham streets (a)
TECHNICAL FEATURES: [v] dynamically changing camera perspective and moving camera: MS, FS, high-angle, aerial (x)
EVENTUALITY: $r, a \mid\sim$ **take over** ($e_{\pi_{10}}$)
$e_{\pi_{11}}$ = introduce
CHARACTERS AND OBJECTS: [v] represented participant: narrator (o) [v] represented participant: Batman (y) [v] object: Bat-Signal (z) [v] Credits appearing (a) [a] "– Tonight, Gotham's relying on one man..." to save us all... (b)
SETTING: [v] above Gotham streets (c)
TECHNICAL FEATURES: [v] dynamically changing camera perspective and moving camera: wide angle, tracking (d) [v] music: heroic tune (e)
EVENTUALITY: $o, y \mid\sim$ **introduce** ($e_{\pi_{11}}$)

Fig. 12. The logical forms for the eventualities $e_{\pi_{10}}$ and $e_{\pi_{11}}$ in the excerpt from *Batman: Arkham Knight*.

Hawreliak, 2018).[4] The fine-grained analysis of these (and potentially other) causal relations on a material level make visible the mechanics of the rule-based system of a game by looking into both semiotic and interactive dimensions [cf. Murray, 2012].

Also, as long as a video game involves informative pieces of interface and instructions, we can assume that an *Elaboration*-relation holds between the main narrative and these elements. In our opinion, this also invites a closer look into notions of games in relation to learning [e.g., Gee, 2003] and how knowledge is build through embodied experiences and semiotic domains. Further analyses are likely to include other types of discourse relations and other types of interpretations centred around the concepts of game mechanics and will enable a more thorough discussion of the semiotic affordances particular to video games.

[4] We thank the anonymous reviewer for pointing this out even clearer.

$e_{\pi_{12}}$ = reach
Characters and Objects: [v] represented participant: Batman (y) [v] object: Bat-Signal (z) [v] object: linear compass (f) [v] object: GCPD building (g) [v] object: a statue (h) [v] objects: various Gotham buildings (i) [v] objects: zeppelins (j) [v] objects: cranes (k)
Setting: [v] above Gotham streets (c)
Technical Features: [v] third-person Batman-bound view (l)
Eventuality: $y, z \mid\sim$ **reach** ($e_{\pi_{12}}$)

$e_{\pi_{12a}}$ = instruct	$e_{\pi_{12b}}$ = inform
Characters and Objects: [v] linear compass (f)	Characters and Objects: [v] linear compass (f)
Setting: [v] above Gotham streets (c)	Setting: [v] above Gotham streets (c)
Technical Features: [v] instruction notes: "Glide (Space)" (m) [v] instruction notes: "Meet Commissioner Gordon by the..." (n)	Technical Features: [v] distance: 553m (o) [v] green attention marker (p) [v] target detail (q)
Eventuality: $m, n \mid\sim$ **instruct** ($e_{\pi_{12a}}$)	Eventuality: $o, p, q \mid\sim$ **inform** ($e_{\pi_{12b}}$)

Fig. 13. The logical forms for the twelfth narrative events/eventuality in the excerpt from *Batman: Arkham Knight* and the subordinated instructive and informative eventualities ($\mathbf{e}_{\pi_{12a}}$ and $\mathbf{e}_{\pi_{12b}}$).

$e_{\pi_{13}}$ = reach
Characters and Objects: [v] represented participant: Batman (y) [v] object: linear compass (f) [v] objects: rooftop metal structures (r) [v] object: platform with Bat-Signal source (s) [a] "– Control, this is Falcon 5..." (t)
Setting: [v] GCPD rooftop (u)
Technical Features: [v] third-person Batman-bound view (l)
Eventuality: $y, s \mid\sim$ **reach** ($e_{\pi_{13}}$)

$e_{\pi_{13a}}$ = instruct	$e_{\pi_{13b}}$ = inform
Characters and Objects: [v] linear compass (f)	Characters and Objects: [v] linear compass (f)
Setting: [v] GCPD rooftop (u)	Setting: [v] GCPD rooftop (u)
Technical Features: [v] instruction notes: "Glide (Space)" (v) [v] instruction notes: "Climb Down (LCtrl) + Mouse drawing" (w)	Technical Features: [v] distance: 60m (x) [v] green attention marker (p) [v] local surveillance (y) [v] range: 214m (z)
Eventuality: $v, w \mid\sim$ **instruct** ($e_{\pi_{13a}}$)	Eventuality: $x, p, y, z \mid\sim$ **inform** ($e_{\pi_{13b}}$)

Fig. 14. The logical forms for the thirteenth narrative events/eventuality in the excerpt from *Batman: Arkham Knight* and the subordinated instructive and informative eventualities ($\mathbf{e}_{\pi_{13a}}$ and $\mathbf{e}_{\pi_{13b}}$).

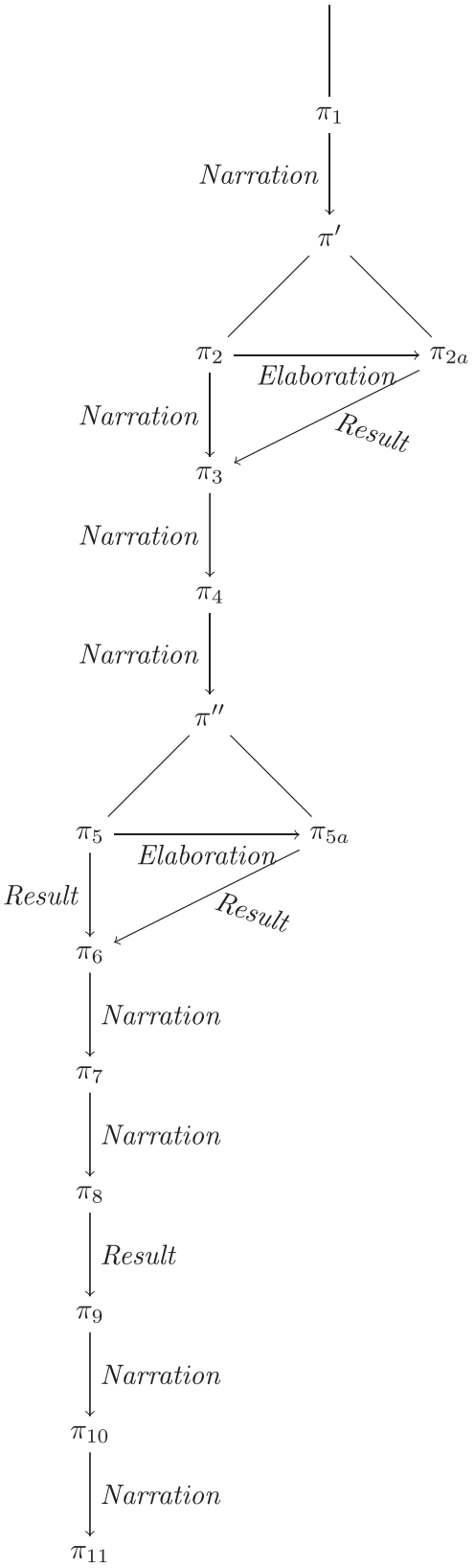


Fig. 15. First part of the graphical representation of the discourse structure of the extract from *Batman: Arkham Knight*.

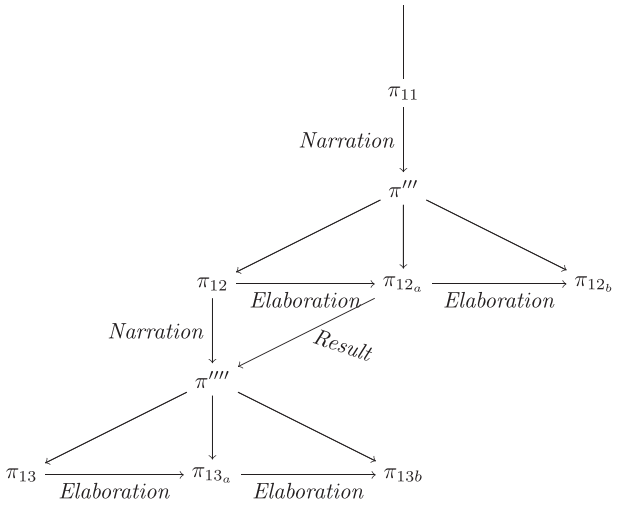


Fig. 16. Second part of the graphical representation of the discourse structure of the extract from *Batman: Arkham Knight*.

However, in order to confirm such assumptions and to apply the knowledge gained from our initial analyses of the specific type of tutorials, the present approach requires a lot more empirical power. These should not only be more analyses of tutorial scenes from other games, but also and in particular more analyses from regular and more interactive scenes from within the analysed games and other games within the same genre. This would allow us identify all kinds of events and structures that profile the genre of action-adventure open-world video games, though glimpses of these are visible in this study. Furthermore, the approach should be applied to analysing video games that belong to genres that are increasingly dissimilar to the artefacts that have so far been analysed using this framework, including puzzle games, as well as managerial and strategy games. Given the original application of the framework to spoken dialogues and conversation, another interesting direction would be to analyse video game conversations that involve player characters and non-player characters, as this form of dialogue can break players' immersion [see Rennick and Roberts, 2021].

We therefore understand our approach documented in this paper as an initial modeling of an analytical framework whose testing with two case studies builds a basis both for further empirical analysis as well as theoretical and methodological development. So far, we have shown that our approach provides important first aspects of a much needed framework for tackling all meaning-making entities in video game discourse(s) as well as exploring the experienced level of coherence and structure within this discourse. The specific affordances and particularities of video games as *interactive* multimodal artefacts, however, need a more critical examination of the specifics of this interaction which goes beyond the basic semantic understanding of the communicative functions of instructing and informing players in a game, and this represents an important research direction.

Acknowledgments

Preparation of this paper was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia grant to Dušan Stamenković (Contract No. 451-03-9/2021-14/200165).

References

Aarseth, E., 2014a. Ludology. In: Wolf, M.J.P., Perron, B. (Eds.), The Routledge Companion to Video Game Studies. Routledge, London, pp. 185–189.
Aarseth, E., 2014b. Ontology. In: Wolf, M.J.P., Perron, B. (Eds.), The Routledge Companion to Video Game Studies. Routledge, London, pp. 484–492.
Asher, N., Lascarides, A., 2003. Logics of Conversation. Cambridge University Press, Cambridge.
Bateman, J.A., Schmidt, K.-H., 2012. Multimodal Film Analysis: How Films Mean. Routledge, London and New York.
Bateman, J.A., Wildfeuer, J., 2014a. Defining units of analysis for the systematic analysis of comics: a discourse-based approach. Stud. Comics 5 (2), 371–401. https://doi.org/10.1386/stic.5.2.373_1. http://www.intellectbooks.co.uk/journals/view-Article,id=18991/.
Bateman, J.A., Wildfeuer, J., 2014b. A multimodal theory of visual narrative. J. Pragmat. 74, 180–208. https://doi.org/10.1016/j.pragma.2014.10.001.
Bogost, I., 2007. Persuasive Games. The Expressive Power of Video Games. The MIT Press, Cambridge, MA.
Bogost, I., 2015. How to Talk about Videogames. University of Minnesota Press, Minneapolis.
Bucher, H.-J., 2011. Multimodales Verstehen oder Rezeption als Interaktion. Theoretische und empirische Grundlagen einer systematischen Analyse der Multimodalität. In: Dieckmannshenke, H., Klemm, M., Stöckl, H. (Eds.), Bildlinguistik. Theorien - Methoden - Fallbeispiele. Erich Schmidt Verlag, Berlin, pp. 123–156.
Cohn, N., 2013. The Visual Language of Comics: Introduction to the Structure and Cognition of Sequential Images. Bloomsbury, London.
Cohn, N., 2018. In defense of a "grammar" in the visual language of comics. J. Pragmat. 127, 1–19. https://doi.org/10.1016/j.pragma.2018.01.002.
Ensslin, A., 2012. The Language of Gaming. Palgrave Macmllan, Basingstoke.
Ensslin, A., Balteiro, I. (Eds.), 2019. Approaches to Videogame Discourse: Lexis, Interaction, Textuality. Bloomsbury Academic, London.
Gee, J.P., 2003. What Video Games Have to Teach Us about Learning and Literacy. Palgrave Macmillan, New York.
Gee, J.P., 2014. Unified Discourse Analysis: Language, Reality, Virtual Worlds and Video Games. Routledge, London and New York.
Gibbons, A., 2011. Multimodality, Cognition, and Experimental Literature Routledge Studies in Multimodality. Routledge, London.

Discourse anaphoricity and first-person indexicality in emoji resolution. In: Grosz, P., Martí, L., Pearson, H., Sudo, Y., Zobel, S. (Eds.), Proceedings of Sinn und Bedeutung, 25. University of Konstanz, pp. 340–357. https://doi.org/10.18148/sub/2021.v25i0.941.
Hawreliak, J., 2018. Multimodal Semiotics and Rhetoric in Videogames. Routledge, New York/London. https://doi.org/10.4324/9781315159492.
Hawreliak, J., Lemieux, A., 2020. The semiotics of social justice: a multimodal approach to examining social justice issues in videogames. Discourse: Stud. Cult. Polit. Educ. 41 (5), 723–739. https://doi.org/10.1080/01596306.2020.1769936.
Hobbs, J.R., 1990. Literature and Cognition. Lecture Notes.
Hodent, C., 2018. The Gamer's Brain: How Neuroscience and Ux Can Impact Video Game Design. CRC Press, Boca Raton, FL. https://doi.org/10.1201/9781315154725.
Ivarsson, J., Linderoth, J., Säljö, R., 2009. Representations in practices. a socio-cultural approach to multimodality in reasoning. In: Jewitt, C. (Ed.), The Routledge Handbook of Multimodality. Routledge, London and New York, pp. 201–212.
Jewitt, C., 2013. Multimodal methods for researching digital technologies. In: Price, S., Jewitt, C., Brown, B. (Eds.), The Sage Handbook of Digital Technology Research. SAGE Publications, London, pp. 250–265.
Juul, J., 2002. Half-real: Video Games between Real Rules and Fictional Worlds. The MIT Press, Cambridge, MA.
Kamp, H., Reyle, U., 1993. From Discourse to Logic. Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory. Kluwer Academic Publishers, Amsterdam.
Kress, G., van Leeuwen, T., 2006. Reading Images: the Grammar of Visual Design, second ed. Routledge, London and New York.
Lascarides, A., Stone, M., 2009. A formal semantic analysis of gesture. J. Semant. 26, 393–449. https://doi.org/10.1093/jos/ffp004.
Machin, D., Van Leeuwen, T., 2007. Global Media Discourse: A Critical Introduction. Routledge, London and New York.
Masso, I.C., 2009. Developing a methodology for corpus-based computer game studies. J. Gaming Virt. World 1 (2), 143–169. https://doi.org/10.1386/jgvw.1.2.143/7.
McMurtrie, R., 2016. The Semiotics of Movement in Space: A User's Perspective. Routledge, London and New York.
Metz, C., 1964. Le cinéma: langue ou langage? Communications 4, 52–90.
Metz, C., 1974. Film Language: a Semiotics of the Cinema. Oxford University Press and Chicago University Press, Oxford and Chicago. Translated by Michael Taylor.
Miodrag, H., 2013. Comics and Language: Reimagining Critical Discourse on the Form. University Press of Mississippi, Jackson, Miss.
Murray, J.H., 2012. Inventing the Medium. Principles of Interaction Design as a Cultural Practice. The MIT Press, Cambridge, MA.
Ng, C., 2020. Interfacing comics and games: a socio-affective multimodal approach. In: Rauscher, A., Stein, D., Thon, J.-N. (Eds.), Comics and Videogames: from Hybrid Medialities to Transmedia Expansions. Routledge, London and New York, pp. 29–44.
Norris, S., 2004. Analyzing Multimodal Interaction. A Methodological Framework. Routledge, London and New York.
O'Halloran, K., Marissa, K.L.E., Podlasov, A., Tan, S., 2013. Multimodal digital semiotics: the interaction of language with other resources. Text Talk 33 (4–5), 665–690. https://doi.org/10.1515/text-2013-0030.
Paul, C.A., 2012. Wordplay and the Discourse of Video Games: Analyzing Words, Design, and Play. Routledge, London and New York.
Peirce, C.S., 1931–1958. In: Collected Papers of Charles Sanders Peirce, vols. 1–6. Harvard University Press, Cambridge, MA, pp. 1931–1935 edited by Charles Hartshorne and Paul Weiss; Vols. 7–8, 1958, edited by Arthur W. Burks.
Pérez-Latorre, O., Oliva, M., 2019. Video games, dystopia, and neoliberalism: the case of bioshock infinite. Game. Cult. 14 (7–8), 781–800. https://doi.org/10.1177/1555412017727226.
Pérez-Latorre, O., Oliva, M., Besalú, R., 2017. Videogame analysis: a social-semiotic approach. Soc. Semiotic. 27 (5), 586–603. https://doi.org/10.1080/10350330.2016.1191146.
Popović, M., Stamenković, D., 2019. Critical discourse analysis, rules and persuasion in video games. In: Lopičić, V., Ilić, B.M. (Eds.), Jezik, književnost, teorija. Filozofski fakultet, Niš, pp. 313–325.
Rennick, S., Roberts, S., 2021. Improving video game conversations with trope-informed design. Game Stud. 21 (3). http://gamestudies.org/2103/articles/rennick_roberts?s=09.
Rockstar North, 2013. Grand Theft Auto V. Rockstar Games, New York.
Rocksteady Studios, 2015. Batman: Arkham Knight. Warner Bros, Burbank.
Rouse, R., 2005. Game Design: Theory & Practice, second ed. Wordware Publishing, Inc, Plano, TX. Illustrations by Steve Ogden.
Royce, T.D., 2007. Intersemiotic complementarity: a framework for multimodal discourse analysis. In: Royce, T.D., Bowcher, W.L. (Eds.), New Directions in the Analysis of Multimodal Discourse. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 63–110.
Stamenković, D., 2022. The stylistic journey of a video game: a diachronic approach to multimodality in the football manager series. In: Werner, V., Schubert, C. (Eds.), Stylistic Approaches to Pop Culture. Routledge, London and New York.
Stamenković, D., Jaćević, M., Wildfeuer, J., 2017. The persuasive aims of *Metal Gear Solid*: a discourse theoretical approach to the study of argumentation in video games. Discourse, Context and Media 15, 11–23. https://doi.org/10.1016/j.dcm.2016.12.002.
Stamenković, D., Wildfeuer, J., 2021. An empirical multimodal approach to open-world video games: A case study of *Grand Theft Auto V*. In: Pflaeging, J., Wildfeuer, J., Bateman, J.A. (Eds.), Empirical Multimodality Research: Methods, Applications, Implications. De Gruyter, Berlin, pp. 259–279.
Toh, W., 2018. A Multimodal Approach to Video Games and the Player Experience. Routledge, London and New York.
Unsworth, L., 2007. Image/text relations and intersemiosis: towards multimodal text description for multiliteracies education. In: Barbara, L., Sardinha, T.B. (Eds.), Proceedings of the 33rd International Systemic Functional Congress (33rd ISFC). Pontifícia Universidade Católica De Sao Paulo (PUCSP), Sao Paulo, Brazil, pp. 1165–1205. In: http://www4.pucsp.br/isfc/proceedings/Artigos%20pdf/59pl_unsworth_1165a1205.pdf. Online publication at. http://www.pucsp.br/isfc.
Wildfeuer, J., 2012. Intersemiosis in film: towards a new organisation of semiotic resources in multimodal filmic text. Multimodal Commun. 1 (3), 233–304. https://doi.org/10.1515/mc-2012-0016.
Wildfeuer, J., 2014. Film Discourse Interpretation. Towards a New Paradigm for Multimodal Film Analysis. Routledge, London and New York.
Wildfeuer, J., 2017. It's all about logics?! analyzing the rhetorical structure of multimodal filmic text. Semiotica 220, 95–121. https://doi.org/10.1515/sem-2015-0139.
Wildfeuer, J., 2019. The inferential semantics of comics: panels and their meaning. Poetics Today 40 (2), 215–234. https://doi.org/10.1215/03335372-7298522.
Wildfeuer, J., Pollaroli, C., 2018a. Seeing the untold. Multimodal argumentation in movie trailers. In: Tseronis, A., Forceville, C.J. (Eds.), Multimodal Argumentation and Rhetoric in Media Genres. John Benjamins, Amsterdam, pp. 189–216.
Wildfeuer, J., Pollaroli, C., 2018b. When context changes: the need for a dynamic notion of context in multimodal argumentation. Int. Rev. Pragmat. 10, 179–197. https://doi.org/10.1163/18773109-01002003.
Wildfeuer, J., Schnell, M.W., Schulz, C., 2015. Talking about dying and death: on new discursive constructions of a formerly postulated taboo. Discourse and Society 26 (3), 366–390. https://doi.org/10.1177/0957926514564739.
Wildfeuer, J., Stamenković, D., 2020. Multimodale Forschungsperspektiven auf Computerspiele. In: Engelns, M., Voßkamp, P. (Eds.), OBST 96/2020. Sprechende Pixel – Computerspielphilologie in Schule und Hochschule. Universitätsverlag Rhein-Ruhr, Duisburg, pp. 7–28.
Wirth, U., 2000. Die Welt als Zeichen und Hypothese. Perspektiven des semiotischen Pragmatismus von Charles Sanders Peirce. Suhrkamp, Frankfurt a.M.
Zacks, J.M., Speer, N.K., Reynolds, J.R., 2009. Segmentation in reading and film comprehension. Am. Psychol. Assoc. 138 (2), 307–327. https://doi.org/10.1037/a0015305.
Zacks, J.M., Speer, N.K., Swallow, K.M., Braver, T.S., Reynolds, J.R., 2007. Event perception: a mind-brain perspective. Psychol. Bull. 133 (2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273.