

The Role of Narrative in Understanding Digital Video: An Exploratory Analysis

Todd Wilkens, Anthony Hughes, Barbara M. Wildemuth, Gary Marchionini

Open Video Project, Interaction Design Lab, School of Information and Library Science,
CB #3360, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3360

Narrative is perhaps the oldest and most widely used form for organizing information and human experience, thus, it is not surprising that there is a significant body of research concerning narrative and its importance to comprehension and understanding. One important outcome of this research is the concept of *narrative intelligence*, the human tendency to fit experience into narrative form. This research is extremely relevant to information seeking in general and sense-making¹ in particular. This paper outlines the basic principles and research supporting the concept of narrative intelligence and its applicability to the ways in which people make sense of digital video. We explore relevant theory and research in sense-making, surrogates, narrative, and narrative intelligence and then present the preliminary results of two research studies. The first clarifies and operationalizes the concept of narrative as it relates to video. The second demonstrates how narrativity can have significant effects on information seeking and sense-making in digital video. Results from these studies have implications for how syntactic form can be used as a means of indexing digital video.

Introduction

Stories provide a simple and versatile form for relaying information, allowing people to communicate across time (e.g. from one generation to the next) and space (e.g. from town to town or country to country). Even as our communication media have changed, the story, or narrative, form has persisted. This is because a narrative is no mere collection of facts. Rather, it has special qualities related to cause and effect, time, and space, the combination of which is quite powerful when it comes to

human comprehension. People use narrative as a way of organizing and making sense of experience. The presence or absence of narrative can play a significant role in comprehension of many kinds of information including spoken words, text, images, or moving pictures. This has bearing on almost any discussion of information seeking but is especially relevant when speaking of video and film. Videos typically convey information through visual (still and moving), textual, and audio media. Compared with text or audio alone, video is a much closer approximation of how humans experience the real world². As such, it is especially important that narrative be taken into account when thinking about information seeking or designing information systems that include video.

This paper lays out preliminary research regarding how narrative plays a role in understanding digital video. We begin by outlining the basics of sense-making in digital video and explain the importance of surrogates for this process. We then describe the basic principles and research concerning narrative and the concept of narrative intelligence. We present the results of two research studies. The first clarifies and operationalizes the concept of narrative as it relates to video. The second demonstrates how narrativity can have significant effects on information seeking and sense-making in digital video. This research was conducted as part of the Open Video project (www.open-video.org), an open source digital video repository that serves as a test bed for video research, including user studies and evaluations of interface prototypes for digital video applications.

SENSE-MAKING, SURROGATES, AND DIGITAL VIDEO

Information seekers have strategies, both conscious and unconscious, that they use when exploring information systems. A crucial part of those strategies is sense-making.

¹ Throughout the paper we are referring to the general process of sense-making (i.e. how people make sense of and construct meaning in their worlds) and not to the Sense-Making methodology of Dervin (1992).

² One might argue that virtual reality and other immersive media surpass video in this regard. However, these technologies are still experimental and not widely available. When it comes to *mass* media, film and video are by far the most realistic.

When using an information system, people must regularly make sense of information objects, categorization schemes, and interface mechanisms. This process can be especially difficult with digital video repositories because video incorporates multiple channels (i.e. visual and audio) simultaneously and this combination of audio and visual information leads to large file sizes and long viewing times. To address these problems, many modern video retrieval systems have begun using surrogates to stand in for video objects. A surrogate, sometimes called an abstract, is a compact representation that shares major attributes with the object it represents. By removing some audio and visual information, a surrogate can have significantly reduced file size and viewing time compared to the full video. However, the removal of information also means a surrogate does not precisely represent the full video. A surrogate, then, is not meant to replace the full video but merely to aid in the selection process. It does this by enabling a user to determine the overall meaning or gist of the video. If the surrogate supports gist determination well, the user is able to make accurate relevance judgments or selection decisions without watching the full video, thus saving time. Determining the gist of a video is clearly an important part of making sense of that video. As we will see, narrative plays a central role in this sense-making.

NARRATIVE AND NARRATIVE INTELLIGENCE

Narrative is arguably the most overused and under-defined word in all of academic writing. As such, using the term inevitably involves a risk of misinterpretation or impassioned – even hostile – disagreement. Be that as it may, the term is central to this discussion and so our hope is to use a precise yet general enough definition so as to avoid confusion and major disagreement. We define narrative as a chain of events related by cause and effect occurring in time and space and involving some agency. The definition has two important aspects. First, based on a definition from Bordwell and Thompson (1997), narrative is a chain of events related by cause and effect occurring in time and space. However, not just any such chain of events is a narrative. A chemical reaction fits that description but it is not a narrative. The same is true of earthquakes, weather patterns, or the rising and setting of the sun. What these situations lack is someone or something acting with intention. This forms the basis for the second part of the definition, based on Bruner (1991), which says narrative involves agency or intentionality. This agency is found in many different forms. Most often it is the characters or the narrator who are the source of intentionality in a narrative. These agents move the chain of causes and effects along, acting as primary causes or

dealing with effects. Thus, cause-effect and agency work together to make the narrative.

Narrative provides an ideal form for capturing how humans perceive and understand their world because it acts as a means for organizing and interpreting cause/effect, agency, time, and space. Imagine trying to describe the events of your day without these concepts. You could list the things you did but, without times, locations, reasons for why you did them or explanations of how one event led to the next, the description would lack any real meaning. It would be difficult for anyone to make sense of your activities. That is because cause/effect, agency, time and space are fundamental to human communication and understanding. Humans begin telling and listening to stories at a very early age and continue doing so throughout their lives. There is compelling evidence that narrative comprehension is one of the earliest mental powers of young children as well as one of the most widely used forms for organizing human experience (Nelson, 1989; Bruner, 1990). It is important to remember that narrative does not exist in the world. It is a human mental construct. *Narrative intelligence* is the term used to describe the human ability to fit experience into narrative form as a means of making sense of that experience (Blair & Meyer, 1997; Bruner, 1991). When applied to mediums of communication (such as video), narrative and narrative intelligence are related to media literacy, which is a person's ability to access, analyze, evaluate, and produce media (Aufderheide, 1992). The work reported here relates most closely to the first two abilities, accessing and analyzing media. Given narrative's fundamental nature, it is not surprising that it can be found in almost all areas of human communication. One of the first challenges in any exploration of narrative is explicitly determining what is *not* narrative.

NARRATIVE AND NON-NARRATIVE FORMS

Discussions of narrative in digital video generally focus on syntactic form. Syntactic form refers to the way scenes and events in a video are organized. For example, narrative is one syntactic form that uses cause-effect and agency as a basis for organizing scenes and events. Lindley and Nack (2000) provide an alternative, non-narrative film form based on Bordwell and Thompson (1997). "*Categorical* films use subjects or categories as the basis for their syntactic organization, typically basing each segment of the film on one category or subcategory." ³

³ Lindley and Nack also offer a second non-narrative form, the *rhetorical* form. However, this form is less about syntactic structure than about content. Rhetorical videos can be syntactically organized using either narrative or categorical

(Lindley & Nack, 2000, 113) Examples of the categorical form are travelogues, gardening programs, and sporting programs. Narrative videos are different from categorical videos because they create new meanings through sequential association of segments. Each segment leads to the next. In categorical videos, the segments stand alone and could be presented in any order without affecting the overall meaning of the film. Figure 1 depicts the distinction between narrative and categorical forms with the arrow representing required continuity/dependence and the ampersand representing simple conjunction.



Figure 1 Syntactic forms

In reality, few videos fall neatly into these categories. It is difficult to find videos, aside from avant-garde or experimental art films, that are completely devoid of narrative form. Yet, people can watch videos and feel comfortable saying that some are narrative and some are not. For example, *Huckleberry Finn* will generally be seen as more narrative than *The Reproductive Habits of the Sea Turtle*. In an earlier study not directly related to narrative (Wildemuth, et al, 2002), we found that participants were fairly consistent in spontaneously describing the videos they viewed as either “stories” (narrative) or “documentaries” (categorical). But what is making this distinction possible? While our definition of narrative suggests that this distinction should be closely related to cause/effect and agency, we felt that empirical investigation of this definition was warranted.

STUDY 1: PERCEPTION OF NARRATIVITY

What makes a film narrative? By definition, the answer is its organization by cause/effect relationships and the presence of agency. But it is also important to remember that narrative is a cognitive construct that relies on subjective interpretation. Thus, the real question is not what makes a film narrative but rather, what leads people to *perceive* a film as narrative? The distinction is subtle but important. It means that the only way to truly test a system for categorizing films is to test those categorizations against people’s perception of narrativity. Therefore, we

conducted a study to explore the effects of cause/effect and characters/agency on perception of narrativity in video.

Study materials. The study stimuli were 20 one-minute video clips taken from larger videos, half black and white and half color. Audio was removed to help isolate the visual channel. The video clips were categorized based on whether they contained characters (defined as agents who maintain camera focus across scenes) or cause/effect relations between scenes. Two members of the research team categorized them independently. The categorizations were then compared for consistency and the four inconsistencies were resolved through discussion. There were five videos with cause/effect and characters/agents, five with characters/agents only, five with cause/effect only, and five with neither.

Study procedure. The study was conducted using a kiosk at CHI 2002 (the annual meeting of the Association for Computing Machinery Special Interest Group on Computer-Human Interaction, ACM SIGCHI). Participants watched the 20 videos in random order and rated each on a scale of 1 to 5, 1 being “less narrative” and 5 being “more narrative”. Because the narrativity scale went from 1 to 5, we decided that a video would have to be rated over 3, the midpoint on the scale, to be counted as *narrative*. Any video clip rated 3 or below would not have sufficient narrativity and would therefore be classified as *non-narrative*. There were 20 conference attendees who volunteered to participate. The kiosk contained a computer and display and the stimuli were delivered via an automatic script developed for the study. All data were collected automatically. Because the study was performed in a public place with no human facilitator, participants were not obliged to view and rate all 20 videos; a total of 195 ratings of the 20 videos were collected. Videos received between 5 and 20 ratings each.

Results. The results indicate that both the presence of characters and the presence of cause/effect relationships had positive effects on ratings of narrativity. When characters were present, the mean rating was 3.3, compared with a mean rating of 2.4 on those videos that did not have characters/agents ($F=16.70$ with 1df, $p<0.0001$). When there were cause/effect relationships, the mean rating was 3.3, compared with a mean rating of 2.2 on those videos that did not have cause/effect relationships ($F=33.83$ with 1 df, $p<0.0001$). More importantly, there was a statistically significant interaction between these two variables ($F=13.44$ with 1 df, $p=0.0003$). When both cause/effect relationships and characters were present, the mean rating (3.9) was much higher than any of the other three possibilities (characters only, cause/effect only, or neither; see Figure 2).

techniques. Thus, we have chosen not to use this form in our categorization scheme.

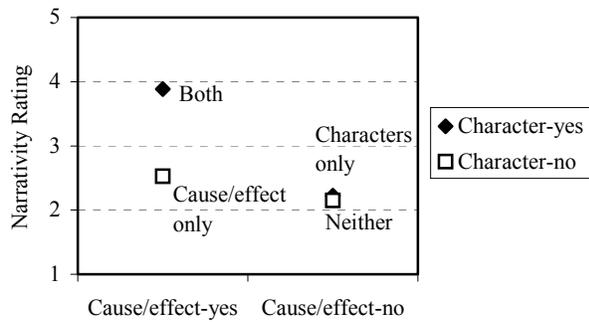


Figure 2. Interaction between presence of character(s) and presence of cause/effect relationships

Discussion. The presence of either cause/effect relationships or characters/agency resulted in ratings above the criterion score of 3, suggesting that either characteristic can lead to perception of narrativity. However, each of these main effects resulted in ratings that just barely met the criterion to be labeled as *narrative*, while the presence of *both* characteristics resulted in ratings of 3.9—noticeably higher. Thus, we conclude that only videos or video clips containing both characteristics can be dependably classified as *narrative*. This has become the guideline by which we classify all videos used in our studies as narrative or categorical.

It is worth noting that, even when both characteristics were present, the mean narrativity rating still did not attain the maximum possible rating of 5. This relatively low rating might suggest that there are other video characteristics besides cause/effect and characters that influence the perception of narrativity. While this may be true to some degree, we believe there is another explanation for the low ratings. It is important to remember that these video clips were only one-minute pieces taken from larger videos. Even those containing cause/effect and characters were not whole narratives nor were they meant to be. Additionally, all video clips had their audio removed so they were communicating only part of their content. It is not so surprising, then, to find that few of these video clips were given high narrativity ratings. This study was suggestive and stimulated considerable discussion in our subsequent project meetings and in a subsequent symposium on understanding video held in Fall 2002. Based on these results and discussions, we decided to revisit a previous study from the Open Video project to explore what effects narrativity may have had on people’s ability to determine the gist of a video from viewing its surrogate.

STUDY 2: FAST FORWARD SURROGATES

The primary purpose of the original study (Wildemuth et al., 2003) was to test the effects of different speeds of a fast forward surrogate on performance of recognition and gist determination tasks (only the gist determination tasks are discussed here). This study was not specifically focused on the effects of narrative or syntactic form. However, of the four videos selected as stimulus material for the study, two were narrative and two were categorical⁴ (using the operational guidelines derived in Study 1). Thus, this study provided an ideal opportunity to explore the effects of narrativity on gist determination performance.

Study procedures. Each of the 45 participants in this study interacted with four fast forward surrogates for each of the four videos. For each surrogate the subjects were asked to complete, among others, two measures of performance on linguistic gist determination (one multiple choice and one free text) and one on visual gist determination (incorporating the visual style of the video as well as its topicality).⁵ The multiple choice gist determination measure gives the user five candidate gist descriptions written by members of the research team and asks him/her to select the one that best describes the video represented by the surrogate. The free-text version of the measure asks the user to “write a brief summary of the video”. Once the study participants generated these gist descriptions, they were scored on correctness/accuracy and level of detail on each of two dimensions (objects/events and higher-level perspective), with a maximum possible score of 8. Two members of the research team independently scored the 180 gist statements, and their scores were strongly correlated ($r=0.76$, $p<0.0001$), indicating a satisfactory level of reliability. The visual gist determination measure asked the study participants to view 12 keyframes, none of which had been seen in the video surrogate, and to select those that “belong” to the video represented by the surrogate. Of the 12 choices, six were correct; the maximum possible score was 12, if all the correct frames were selected and all the incorrect frames were not selected. After the study was run, results were analyzed using syntactic form (narrative/categorical) as an independent variable to see if the form of the videos was affecting performance.

Results. Syntactic form did have significant effects on performance in free text gist determination and multiple choice gist determination. Study participants said they could determine the topic of a film organized by categories much easier than a narrative film. This was supported by

⁴ As noted before, narrative and categorical are two standard syntactic forms for film and video. Due to the lack of avant-garde or experimental films, we have decided to categorize all of our non-narrative films as categorical.

⁵ Yang et al. (2003) provide a full description of these measures.

performance data where participants had lower multiple choice gist comprehension scores for narrative videos (23% correct) than categorical videos (69% correct) (chi-square = 37.5829 with 1df, $p < 0.0001$). The scores on the free text gist determination measure indicated no statistically significant difference between the surrogates of narrative videos and those of categorical videos ($F = 0.02$ with 1df, $p = 0.8800$). However, there was a statistically significant interaction between syntactic form and surrogate speed ($p = 0.0199$; see Figure 3), indicating an advantage for narrative video surrogates at the higher display speeds and for the categorical video surrogates at lower speeds.

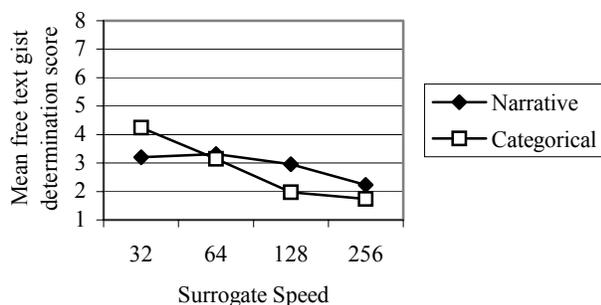


Figure 3. Interaction between syntactic form and surrogate speed on gist determination (free text)

Results of the visual gist determination measure indicated that the narrative videos had a slight advantage, with a mean score of 8.7, compared with the mean score of 8.2 for categorical videos ($F = 3.81$ with 1, 177 df, $p = 0.0525$). These results show how a video's syntactic form (narrative vs. categorical) can have significant effects on people's performance in determining the gist of a digital video from viewing a surrogate of that video. Additional studies using different videos are needed to more fully understand how video characteristics affect performance as people interact with the video surrogates.

Discussion. While several of the results for Study 2 are statistically significant, taken together, they do not present a clear picture of how narrativity is contributing to or inhibiting sense-making. In fact, these results raise some significant questions. Research on narrative intelligence, as well as our own work, has established that narrative should be an aid to sense-making and comprehension. This is supported by the participants' performance on the multiple choice and visual gist measures. But why, then, were participants more successful (as indicated by our full text measure) in determining the gist of categorical videos?

Bordwell's (1985) concept of the "canonical" story form provides the basis for one possible explanation. Bordwell holds that the majority of narratives, especially in film, have the same five stages:

- 1) Introduction to setting and characters,
- 2) Explanation of state of affairs,
- 3) Complicating action,
- 4) Ensuing events,
- 5) Outcome, and
- 6) Ending.

When viewers see one or more of these steps, it is a cue that the information they are receiving is in narrative form. This can be a tremendous aid for making sense of incoming information. However, Bordwell also reports that "distortions in comprehension and recall tend to occur at points when the events perceived by the viewer violate or ambiguate this ideal scenario" (1985, 35). This suggests an explanation for why people find gist determination for surrogates of highly narrative videos difficult. When a video is made into a surrogate, information is lost. For example, when a video is sped up as a fast forward, the viewer will not be able to perceive every scene or image in the video. Despite this lost information, there are generally enough cues for the viewer to recognize the video as narrative or categorical. When scenes or sections of the video are missing, it is nearly impossible for viewers to comprehend the cause/effect relationships or fit the information into the canonical form. They cannot "fill in the blanks". The familiarity of the narrative form makes the missing pieces more obvious. The viewer is very aware that they do not have the full picture and thus their confidence in their understanding is reduced. This can also help explain why the one-minute video clips from Study 1 received relatively low narrativity ratings. They were far too short to adequately represent the canonical story form. Categorical videos do not have this sequential relationship nor do they use the canonical form. Thus, they do not suffer the same drawbacks when made into surrogates.

The reasons for the other two effects are less clear. Both relate to the free text gist determination measure. There was no significant main effect of syntactic form on performance but there was a significant effect from the interaction of syntactic form and surrogate speed. The discussion above would suggest that narrative videos would lead to lower performance than categorical videos at higher speeds. Instead, the opposite was true (see Figure 3). One possible explanation for this effect is that the free text measure is addressing a different aspect of sense-making than the other measures. Future studies will explore this issue further.

DISCUSSION

The work reported here is exploratory and we have yet to perform more than preliminary studies. In particular, our research to date has focused exclusively on surrogates or small segments rather than on full videos. While this cannot present a complete picture of the role of narrative in digital video, working with these small pieces provides a means for exploring those aspects of video that are both *necessary* and *sufficient* for perception of narrative. Despite their exploratory nature, these results indicate that narrativity is a specific, measurable and significant part of how people understand digital video and that it deserves further study

Study 1 provides empirical support for our definition of narrative form as well as for our operationalization of that definition. The study results show that cause/effect relations between sequential scenes and the presence of characters/agency are two major contributors to the perception of narrative form. Film theorists have outlined several other aspects of films that contribute to narrativity such as evidence of a narrator and types of exposition (i.e., how the setting/theme is established). These could also be playing a role in people's perceptions, especially in relation to surrogates. For example, the removal of scenes and other information to make a surrogate must surely affect how the setting is established and therefore the perception of narrativity. We must reiterate that a discussion of narrative in videos is really a discussion of the *perception* of narrative in videos. Narrative is a cognitive construct and as such it requires active participation on the part of the viewer. Given our current results, Study 1 bears repeating in a controlled, laboratory environment. In particular, we would like to increase the number of participants and include qualitative techniques (think-aloud protocols and interviews) to get a better understanding of what leads people to perceive narrativity in videos. Our main question from this study remains, what are the minimal requirements for perception of narrativity? Because our research is focused on people's interactions with surrogates, we are particularly interested in the minimal time requirements (e.g., is 10 seconds enough? 20 seconds? 30 seconds?) and the effects of different types of surrogates (e.g., can people determine narrativity from key frames? from slideshows?).

Our analysis of Study 2 has definite implications for how people make sense of digital video information. In particular, it suggests a need for more research into the ways in which video surrogates disrupt the canonical story form, and the effects of such disruption on people's efforts to determine the gist of the video represented by the surrogate. In addition to the role of the canonical story form on perception of narrativity, it is important to investigate other possible factors and the role of the

viewer's confidence in his or her understanding of the video.

Despite the questions that remain, the results reported here have implications for the design of video retrieval systems. In particular, they suggest that syntactic form can be a useful index for digital video. While the combination of cause/effect relationships and characters/agency provided the clearest cues for people's perception of narrativity, it seems likely that only the detection of the presence of characters/agency can be easily automated. There are currently many techniques for face detection and recognition in video. If these techniques were combined with scene detection algorithms, it would be possible to detect when a given face was present across multiple scenes. This would provide a means to index videos for the presence of narrative. Given the importance of narrativity to information seeking and sense-making, having these automated tools would allow information systems to tailor their surrogates and interface mechanisms to suit the needs of users more effectively.

Improving the effectiveness of video information systems is becoming more and more important as technology for video production and digitization becomes more widely available thereby spurring a tremendous increase in the amount of digital video in online repositories. More and more people will be using these repositories to meet their information needs. It is therefore vital that we, as designers and developers of these repositories, recognize and account for how people seek and make sense of this information. Syntactic form in general and narrative in particular are integral parts of the process through which people make sense of video. We hope that the work reported here will help to spur discussion and further research on narrative and sense-making in digital video.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under Grant No. IIS 0099638.

REFERENCES

- Aufderheide, P. (1992). *A report of the National Leadership Conference on Media Literacy*. Queenstown, Maryland, December 7-9, 1992.
- Blair, D., & Meyer, T. (1997). Tools for an Interactive Virtual Cinema. R. Trapp, & P. Petta (Ed.), *Creating Personalities for Synthetic Actors: Toward Autonomous Personality Agents*. Berlin: Springer Verlag.
- Bordwell, D., & Thompson, K. (1997). *Film Art: an introduction* (5th ed.). New York: McGraw-Hill.

Bruner, J. (1990). *Acts of Meaning*. Cambridge, Mass : Harvard University Press.

Bruner, J. (1991). The Narrative Construction of Reality. *Critical Inquiry*, 18(1).

Dervin, B. (1992). From the mind's eye of the user: The Sense-Making qualitative-quantitative methodology. J. D. Glazier, & R. R. Powell (Eds), *Qualitative research in information management* (pp. 61-84). New York: Libraries Unlimited.

Lindley, C. A., & Nack, F. (2000). Hybrid narrative and categorical strategies for interactive and dynamic video presentation generation. *The New Review of Hypermedia and Multimedia* (Vol. 6). London: Taylor Graham.

Nelson, K. (1989). *Narratives from the Crib*. Cambridge, Mass: Harvard University Press.

Wildemuth, B. M., Marchionini, G., Yang, M., Geisler, G., Wilkens, T., Hughes, A., & Gruss, R. (2003). How fast is too fast? Evaluating fast forward surrogates for digital video. Paper submitted to Joint Conference on Digital Libraries, 2003 (under review).

Yang, M., Wildemuth, B. M., Marchionini, G., Wilkens, T., Geisler, G., Hughes, A., Gruss, R., & Webster, C. (2003). Measuring User Performance During Interactions with Digital Video Collections. Paper submitted to the Annual Meeting of the American Society of Information Science & Technology, 2003 (in press, in this volume).