The application of rhetorical structure theory to interactive news program generation from digital archives

C.A. Lindley, J.R. Davis, F. Nack, L.W. Rutledge

Information Systems (INS)

INS-R0101 January 31, 2001

Report INS-R0101
ISSN 1386-3681

CWI
P.O. Box 94079
1090 GB Amsterdam
The Netherlands

The Application of Rhetorical Structure Theory to Interactive News Program Generation from Digital Archives

Craig Lindley[1]
CSRIO Mathematical and Information Sciences
Locked Bag, 17, North Ryde, NSW 1670, Australia
Email: CraigLindley@cmis.csrio.au
Jim Davis[2], Frank Nack, Lloyd Rutledge
CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
Email: {Frank.Nack, Lloyd.Rutledg}@cmis.csrio.au

ABSTRACT
Rhetorical structure theory (RST) provides a model of textual function based upon rhetoric. Initially developed as a model of text coherence, RST has been used extensively in text generation research, and has more recently been proposed as a basis for multimedia presentation generation. This paper investigates the use of RST for generating video presentations having a rhetorical form, using models of the rhetorical roles of video components, together with rules for selecting components for presentation on the basis of their rhetorical functions. An RST model can provide a predefined link structure providing viewers with options for obtaining and dynamically modifying rhetorically coherent video presentations from video archives and databases. The use of an RST analysis for interactive presentation generation may provide a more powerful rhetorical device than conventional linear video presentation. Conversely, making alternative RST analyses of the same video data available to users can have the effect of encouraging closer and more independent viewer analysis of the material, and discourage taking any particular rhetorical presentation at face value.

1998 ACM Computing Classification System: H.5.2, I.3.4, I.3.8
Keywords and Phrases: interactive presentation generation, rhetorical structure theory, news, video, video archive
Note: The work was carried out under the ”ToKen200” and “Dynamo” projects

1. INTRODUCTION

Advances in digital video technology over the past decade have resulted in the widespread adoption of computer based non-linear editing systems for film and video post production. However, digital technology also allows the production artefact to be redefined, so that a fixed linear video stream is no longer the only form in which video data can be produced and distributed. In particular, it is possible to develop interactive video systems in which a linear video presentation is generated dynamically and adaptively from an underlying database having no predetermined linear structure in order to create a presentation tuned to the needs and interests of a particular viewer. Dynamic video presentation generation provides strategies for the use of video data that may otherwise be discarded during conventional linear video production (for example, the shooting ratio, or ratio of shot to used material, can be as high as 100 to 1 for documentary productions). Dynamic generation also allows presentation generation to benefit from the ongoing accumulation of topical video data within databases and archives, as well as providing techniques for the integration of historical archive data and stock footage material.

Previous research has demonstrated the automated selection of video clips from a video database into meaningful sequences for presentation to viewers. Sequencing in these cases has been based upon either a narrative model of video form [1, 2, 3], or a categorical model [4, 5]. Different theorists focus on different qualities that characterise narrative (see Stam et al, [6]). Basically though, narrative is about telling a story, and hence involves a system of causally interrelated events, actions, and situations. Commercial dramatic films are narrative films, although narrative

[1] This author works now at Starlab, Belgium (craig@starlab.net).
[2] This author works now for CourseNet Systems, USA (jrd3@alum.mit.edu)

organisation also appears throughout many forms of documentary. Narrative is concerned with the creation of a pattern of cause-effect relationships among the diegetic events, actions, and situations depicted by a film. Narrative in its broadest sense has been the goal of numerous research projects dealing with diverse media, from text to interactive 3D systems (see, for example, Mateas and Sengers [7, 8]). Research concerned with the construction of narrative video sequences by the selection and ordering of clips from a video database has tended to use a narrow interpretation of narrative in the sense of continuity-edited depictions of causally interconnected actions and events. Narrative in this sense is only one of a number of forms for the organization of filmic material. *Categorical* films are another form that uses semantic categories as a basis for syntactic organisation, typically basing each segment of the film on one category or subcategory [9]. Common examples of categorical films include lifestyle and gardening programs, travelogues, and sporting programs. The highest level of syntactic structure for news programming often has a categorical form, with categories such as "headline news", "international news", "local news", "sports", and "weather". Bordwell and Thomson [9] also identify *rhetorical* films that present an argument and lay out evidence to support it. [3] Common examples of rhetorical films are television commercials, but this form may also be expected in news and other forms of documentary production.

Each of these forms represents a different syntactic structure for film sequences. In most real films, the forms apply at multiple levels of film structure, a given film sequence may involve multiple forms at the same level, and multiple forms may occur at different levels. Hence, while a given form may be used to structure a video at a given level of decomposition, the elements being conjoined at that level may have an internal formal organisation that can be the same as or different from the form of that level. While the overall syntactic structure of a film or video can therefore be highly complex, research in automated video sequencing has tended to adopt one primary formal model as a basis for selecting and sequencing predefined video components. Previous research has concentrated upon narrative and categorical forms. In this paper we propose and investigate the use of rhetorical structure theory (RST) as a basis for the generation of video presentations having a rhetorical form. In particular, we describe how RST can model important aspects of the internal structure and semantics of a news story, and how that model can be used for automated, interactive, and adaptive news program presentation generation.

2. RHETORICAL STRUCTURE THEORY (RST)

Rhetorical Structure Theory (RST) is a relational theory of text structure originally devised for the analysis and synthesis of coherent texts [10]. RST relations are functional, modelling the rhetorical functions of meanings expressed within units of text. As such, a relation holds between two non-overlapping spans of text, one being referred to as the nucleus and the other as the satellite. A relation may include constraints upon both the nucleus and the satellite, and a relation functions to produce a specific effect within a reader. A set of RST relations is listed on Figure 1.

Nucleus-Satellite Relations		
Evidence	Justify	Antithesis
Concession	Circumstance	Solutionhood
Elaboration	Background	Enablement
Motivation	Volitional Cause	Non-Volitional Cause
Volitional Result	Non-Volitional Result	Purpose
Condition	Otherwise	Interpretation
Evaluation	Restatement	Summary
Multi-Nuclear Relations		
Sequence	Contrast	Joint

Figure 1: A listing of RST relations

[3] The additional categories of associational and abstract forms identified by Bordwell and Thomson [9] are not dealt with in this paper. Lindley [5] discusses these forms and analyses their relationship to categorical, narrative and rhetorical forms.

Relation definition does not constrain the linear order of spans. If the locus of effect of a relation is the nucleus, the satellite is inessential to the core meaning of a text. Relations are hierarchically structured, with the overall hierarchical structure of a text providing its holistic integrity and coherence. At the lowest level, spans map onto text units generally corresponding with clauses. The authors of RST recognize the ambiguity of RST analyses, arising partly from the ambiguous nature of text, and also from the nature of an RST analysis as a plausible interpretation.

3. RST ANALYSIS OF NEWS VIDEOS

RST has been used for a variety of purposes, from text generation (see http://www.sil.org/linguistics/rst/tgen.htm) to multimedia presentation generation [11, 12]. Noel [13] has used RST to analyse news broadcasts. Automated segmentation of news video has been addressed by the Informedia project (www.informedia.cs.cmu.edu), supporting ranked retrieval by recency or relevance. The FRANK project [14] has also demonstrated virtual video presentation generation by performing a text-based ranked retrieval on transcripts of news and currents affairs video data. These approaches may be useful when the object of interest is a specific item at the level of indexing of the retrieval and presentation system. However, rather than simply presenting a ranked list of atomic video segments, a more structured presentation can draw upon principles of montage and a higher level models of video form to create more specific and interesting productions having a typical video program structure. Here we investigate the use of RST to provide a set of relations that can be used to structure interactive video presentations having a rhetorical form as described above.

An RST analysis is a plausible analysis, and Mann et al [10] suggest that multiple RST analyses of the same text are part of the explication of how the text is informative. If presentation generation is based upon a single analysis, it must therefore exclude part of the potential informational function of the text. This can be used to impose a specific authorial view upon presentations based upon a single RST model (for pedagogical, rhetorical, or expressive purposes). Alternatively a system might include multiple RST analyses from which different presentations may be generated, which may be suited to different purposes, or may function as a method allowing users to explore the polysemy of the underlying video database. In this discussion we concentrate upon the use of a single RST model for video presentation generation. It cannot be assumed that RST can provide a plausible analysis of coherence for all forms, styles, and genres of video, and the question of which forms it can effectively and validly be applied to is currently an open one. In this paper we concentrate upon news programs, and examine the use of RST as a plausible account of the (rhetorical) coherence of a news program, and hence of the coherence of an automatically generated video news presentation. We describe how RST can provide a model for the semantics of video components that can be used to generate a rhetorically coherent presentation. That is, RST can model the coherence of a presentation, and also model the rhetorical potential of video components as a basis for algorithmically selecting components for insertion into a rhetorically coherent presentation.

Applying RST to a conventional linear video program raises the question of cross-media relations: does a span correspond to a meaningful subsequence of the composite video stream, or can separate spans be identified for the image and audio tracks, for different layers of the video image and/or different audio tracks, or for different spatial and/or semiotic components of the image (including captions and headlines)? RST analysis of text requires the division of text into units (segments) that provide the primitive elements from which spans can be constructed. For video, units might include multiple media modes, or may be limited to individual media types. Decomposition of video into more primitive media, spatial and temporal subcomponents may support more flexible and adaptive presentation generation based upon the resulting analysis, but this also complicates the presentation generation task. Such a decomposition could be specified in terms of abstract presentation characteristics such as screen coordinates and time codes. An alternative strategy might be to identify units or spans with the semantics of the video stream, such as meaningful objects and/or events that are represented within the video (image or sound tracks). Units and spans referring to higher level semantics will require a method of mapping from content-oriented descriptors to the data representation of the video file in order to support manipulation of the data, and could, for example, use an MPEG-7 description of the media content.

The simplest approach is to assume that the integrated video media will not be decomposed across modes, and identify units and spans with different linear subsequences of the multimodal video stream. This is a valid approach for systems that generate presentations from already composited video data. In systems in which the different image and sound layers are represented separately (e.g. within an MPEG-4 file format, see http://drogo.cselt.stet.it/ufv/leonardo/mpeg/standards/mpeg-4/mpeg-4.htm), it is preferable to develop techniques for adaptively combining the individual media objects and streams, supporting a greater range of adaptation and expressive

semantics[4]. However, assuming a composited video image, an RST analysis requires the identification of video subsequences corresponding to units and spans, and the interrelationship of spans according to their functional roles within RST relations. The RST analysis is based upon a definition of basic units, which will be linear segments of video. Since an RST analysis is a plausible analysis[5], it cannot be assumed that units can be defined independently of the RST analyses within which they are to play a role, although units or spans may correspond to intuitively obvious subdivisions of a video presentation, such as segments or stories within a news program.

As an example of an RST analysis applied to news programming, we analysed BBC television news programs at 1 pm, 6 pm, and 9 pm on one day during the week of 3 April 2000. The News at 1 program consisted of the following high level parts:

Headlines:
- Ethiopian famine
- Building collapse in Hull
- Microsoft legal challenge
- Bad weather in the UK

Stories:
1. Ethiopian famine
2. Building collapse in Hull
3. Microsoft legal challenge
4. Barclay's bank
5. Red tape and local businesses
6. Toddler ate ecstasy pill
7. Race driver loses appeal
8. Health risks of cell phones

Preview of stories to come:
- London Mayoral Elections
- Mission to save MIR

News Logo Segment

Stories (continued):
9. Bad weather in the UK
10. Human rights inquiry re Northern Ireland
11. Resignation of Japanese Cabinet after Collapse of Prime Minister
12. London Mayoral Elections
13. British Housing Policy
14. Mission to save MIR

Recap of headline about the Famine in Ethiopia

End of News Bulletin

Weather Forecast

The News at 6 had two different headlines (replacing the building collapse and Microsoft stories with the stories about the health risks of cell phones and a new story about a holiday company takeover), added new stories about deporting beggars and the state of school buses, and dropped the stories about the toddler eating ecstasy, the mission to save MIR, and the Japanese cabinet resignation. Similarly, the News at 9 dropped some stories, reinstated some stories from the News at 1, and added new stories. Hence the total set of stories for the day (i.e. 20 stories in total) was larger than the set presented during any single news broadcast (14 in the example described above). The headlines for the News at 9 included none of the headlines from the News at 1 and the News at 6, and had dropped three of the earlier headline stories altogether. In general, the different news presentations present different subsets of the total set of stories, with the individual stories being treated with variable length across the different programs.

[4] In this case, strategies for image compositing converge to some extent with strategies for hypermedia presentation generation of the kind discussed by Rutledge et al (2000b).

[5] In general an RST analysis is a kind of subtextual, or at least connotative, reading of the target text, in the sense described by Srinivasan et al (1999). As such its validity lies somewhere between the normative and the idiosyncratic.

Due to the different themes and subject matter of individual news stories, it appears to be most appropriate to model the high level coherence of a news program using the multinuclear *sequence* relation. The sequence relation reflects the sequential order of presentation of stories within a single news broadcast, as represented in the numerical order of stories in the News at 1 example described above. Headlines and previews may have a *summary* relation to the more detailed stories that they refer to. In an interactive news system, however, there is no particular need to preserve the sequential order of a given presentation, so the total set of stories may be interrelated by the multinuclear *joint* relation, which might include all stories presented on a particular day, or over any other arbitrary period of time.

To consider the detailed RST analysis of a particular news story, we concentrate upon the story about the famine in Ethiopia. This story was covered in all three news programs, but with a different anchor person each time, and variable content. In particular, the content of the field report had evolved between the News at 1 and News at 6 reports on one hand and the News at 9 on the other. The News at 9 also included a report about the background of the famine, including the previous Ethiopian famine in 1984. The News at 9 also included considerably more detail about Ethiopian accusations of poor international and British response to the famine, together with the British response to those accusations. The shot list for some of this material is segmented as shown below:

S43: Ethiopia has accused the international community of being far too slow to react to its warning of looming famine in the region.

S44: The country's foreign minister said that rich countries were waiting to see pictures of skeletons before answering appeals for aid.

S26: There were delays and there are logistic problems on the side of the [Ethiopian] government to deliver even what they have in hand

S27: and for the last month of March there has not been any food distribution to the displaced people as well as to the affected people at all.

S49: Unless international assistance arrives quickly, the world is likely to witness more and more such scenes of tragedy in Ethiopia.

S50: The Ethiopian ambassador here tonight criticized Britain for not spending enough on aid for his country.

S51: The amounts promised are not enough to bring safety to his dying people.

S57: Since then [1984] an early warning system has been in place.

S60: The British government says that lessons have been learned and they will not let the catastrophe happen again.

S61: Clair Short claims that it was worse in 1984 since they had a dictatorship then and the information needed for the world to act wasn't available.

S62: Britain will increase its aid.

The resulting RST analysis is shown on Figure 2. Unlike the example presented by Mann et al [10], RST relations in this case do not hold only between contiguous spans within the original video material, but can apply to spans separated by material having different rhetorical relationships and roles.

Performing an RST analysis on a news program highlights the interpretative nature of the analysis. For example, the BBC material presents the Ethiopian ambassador's criticism of Britain for not spending enough on aid for his country (S50) as well as presenting the British government's claim that it will not let the catastrophe happen again (S60), together with Clair Short's statement that Britain will increase its aid (S62). S60 stands in an antithetical relationship to S50. RST requires one of these statements to be nominated as a satellite and one as a nucleus, with the nuclear segment gaining in support from the antithetical relation to the satellite. Hence if the Ethiopian government in this context has low credibility, it should be given the satellite role in order to enhance the credibility of the British government statement. However, if the British government has low credibility, the opposite rhetorical direction holds, and its statement is the satellite.

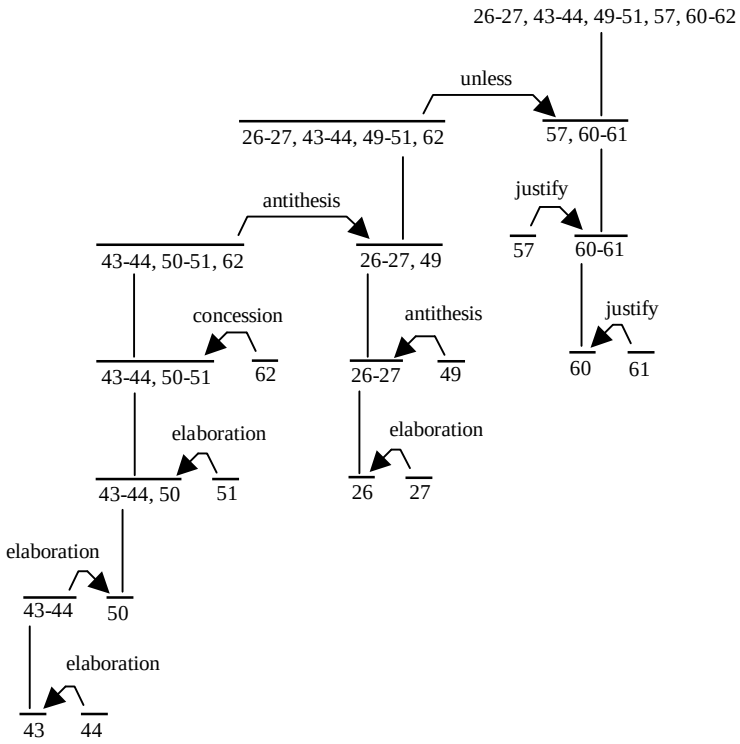

Figure 2: Segmentation of the components of news videos

The interpretation of segment S62 can have a critical role in the overall message read into the program and represented in the RST analysis. S62 can be interpreted as the volitional effect of the Ethiopian ambassador's claim that Britain should do more (S50), suggesting that antithetical criticism of the Ethiopian government (S26) has little credibility (since S26 potentially renders further aid futile). In this case, S50 should be the nucleus and S60 should be the satellite. Alternatively, S62 can be regarded as a concession to the concerns of the Ethiopian government for the sake of its relationship to Britain. In this case, S26 retains credibility, S60 should be the nucleus and S50 should be the satellite. This latter interpretation is the one represented on Figure 2, conveying the overall message of Britain responding appropriately according to the information that it has, with the Ethiopian government carrying at least some of the responsibility for ongoing suffering. This is not an unquestionable interpretation of the material, emphasising the nature of the RST analysis itself as an authored artefact.

4. Application of Rhetorical structure Theory to Automated Video Sequence Generation

Non-linear video systems allow the video data to be stored in multiple files, and units or spans can be identified with complete files or subsequences within files; in this case, the storage order of the video data does not need to correspond to the order of a particular video presentation. A span can be classified by rhetorical functional role, and related to another span by an RST relation. This unit can then function rhetorically within a higher level RST relation. For overall coherence of the interactive video production in RST terms, all of the video spans must ultimately be interconnected within the hierarchical structure of RST relations. The RST model of video components within a database constitutes a plausible account of the rhetorical coherence of the whole database. The RST model is an authored artefact, representing a kind of hyperlink structure over the video data. The automated generation of a linear video presentation from this structure involves an algorithm that will traverse the hyperlink structure and identify a subset of video data for

presentation that will satisfy some goal or specification (provided, for example, by the user/viewer). This process is analogous to the pruning process for generating hypermedia presentations described by Rutledge et al [12]; the main difference is that pruning of hypermedia leaves a tree structure for traversal by a user, while video presentation synthesis requires a tree traversal that will generate a single linear video presentation for the user. An RST model of video semantics supports both interactive, dynamic pruning and sequential video presentation generation, as described below. The video database and data structures used for the generation of a presentation represent a combinatorial space of possible presentations that may be much larger than any actual presentation generated for a particular viewer.

In the example considered in this paper, the set of BBC news broadcasts constitutes the video database. Material repeated within the different news presentations does not have to be duplicated within the database, and each individual broadcast corresponds with a particular path through the database. As stated above, the different stories within the overall database can be represented as independent RST nuclei, and the nuclei can be interrelated by the RST *joint* relation. Each story can be subdivided into segments, as demonstrated above for the story about the Ethiopian famine. At a simplistic level, the segments within a story could be modelled as independent nuclei and interrelated by the RST *sequence* relation to represent their presentation order in the original television broadcasts. RST appears to be unfalsifiable in this sense, since any text or media production could be modelled as a set of nuclei interrelated by joint or sequence relations. However, these are not terribly useful relations; while they support a simple semantics for user selections and presentation generation [11, 12], they are not sufficient for providing criteria for creating presentations having significant variations in semantics. A richer system of RST relations supports more adaptive presentation generation, provides more powerful rhetorical functions within the mechanics of presentation generation, and provides richer options for user interaction.

A simple method of generating variable linear video presentations from such a hyperlink structure is to present RST functions as categories that a viewer can select or deselect prior to or during a video presentation. For example, the analysis presented above uses the *unless*, *justify*, *antithesis*, *concession*, and *elaboration* relations. A full traversal of the tree structure presented in the example can be used to identify the leaf nodes corresponding to individual segments that may be selected for assembly into a presentation incorporating all of the modelled video material. A simpler presentation could be assembled by a similar traversal but excluding the segments linked by elaboration relations, or any other specific relation type or set of relation types. The exclusion of specific relation types will modify the semantics of a presentation.

In addition to simple pruning, the tree structure can be used to extract more specific forms of information about the video material, and support a form of discourse with the user of the database system. For example, analysis of the directions of the RST relations in the tree shows that segment S60, by the above analysis, is the primary nucleus of the news story. This information can be used to answer queries such as “What is the main point that the BBC is trying to make about the famine in Ethiopia?” In answer the user will be shown the segment S60 asserting that the British government says that lessons have been learned and they will not let the catastrophe happen again. If the user then wishes to know “Why is this so?”, the *justify* links to S60 can be traced to find the segments that state how in 1984, the time of the previous famine, the dictatorship in Ethiopia prevented an international response by withholding information, and how an early warning system has been in place since then. From that point, it is also possible to ask what may prevent the British government from acting as it states that it will. It is then possible to algorithmically trace the *unless* link to the subtree detailing internal Ethiopian problems with aid reaching the famine victims.

Links from a particular segment provide a basis for presenting specific query options to system users. For example, an *unless* link answers the question “What might prevent this?”. A *justify* link answers a “Why?” question. An *antithesis* link answers the question “Are there any arguments against this?”. A *concession* link answers the question “Are there any points to concede?”, while an *elaboration* link can satisfy a request for “More detail”. Presenting the viewer/user with this kind of dialog system provides a highly interactive model for viewing video material. It also imposes a strong rhetorical function on the material, since the system contextualises specific segments of the material with a very specific rhetorical interpretation that may be much more ambiguous when presented in the context of a conventional linear news program. Control of the query options presented to users also amounts to the presentation of leading questions that may suppress the users’ more independent criticisms and analyses of the material.

The RST structure of a video database can also be used to derive information about the relative importance of specific segments according to a particular analysis. For example, four segments in the example above have direct or indirect (i.e. via span) RST satellite relations to segment S50 before S50 itself functions as a satellite. Segment S26 has 2 atomic satellites, and the cluster of satellites around and including S50, for a total of 7 direct and indirect satellites. Segment S60 has 10 direct and indirect satellites. Segment S43 has one satellite, and the remaining segments have no satellites. The number of satellites attached directly or indirectly to a segment can be used as a heuristic for the

importance of the segment. The RST analysis for the example above therefore suggests an order in decreasing importance of: S60, S26, S50, S43, then the rest of the segments. Since the number of satellites associated with S50 is significantly higher than S43, the segments S60, S26, and S50 emerge as significantly more important than the rest, and can therefore be used as a summary of the major points of the story. Also, the relative ranking of S60, S26, and S50 suggests both an increasing order of detail and a decreasing order of priority in summarising the main points of the story, supporting the presentation of summaries of varying detail.

As well as the number of satellites linked to a segment, the distance of a segment measured in satellite links from the primary nucleus can also be used as a heuristic for the importance of a segment. For example, segment S61 in the example above is only one satellite link away from the primary nucleus, S60, while segment S44 is four satellite links away from S60. Hence, according to this RST analysis, Clair Short's justification for stating that disaster will not be allowed to happen again (i.e. that more information is now available due to an improved political situation in Ethiopia) is heuristically identified as being more important than the Ethiopian foreign minister's claim that rich countries are waiting to see pictures of skeletons before answering appeals for aid.

These functions clearly demonstrate the usefulness of an RST analysis for interactive video presentation generation. However, RST alone cannot determine the presentation order of the segments in a generated video presentation. In fact, different traversals of the tree structured RST model, driven either algorithmically or by user selections of relationship types and dialog options, can result in many different presentation orders for segments, and the dynamic juxtaposition of different segments for different presentations. It is therefore desirable that segments should have an internal structure and content that is compatible with a range of different temporal juxtapositions. That is, the interactive video system requires a *rhetorics of arrival and departure* [16]. The rhetorics of arrival and departure are the cues that make links between hypermedia components meaningful and coherent from the perspective of the viewer traversing the links. In the case of interactive video, this amounts to the need for the conjunction of video sequences into a single longer sequence to be meaningful, and for inter-sequence transitions to contribute to, and not detract from, that meaning. In this case it may be more appropriate to refer to a *rhetoric of montage*, referring to the system of semiotic codes used to ensure that a transition between video sequences is meaningful and coherent within the context of the production as a whole. Much of the difficulty of defining a narrative presentation generation system of the kind developed by Nack (1996) is in defining rich enough rules for the preservation of the continuity of action between cuts, and ensuring that discontinuities convey intended meanings. Detailed rules are required for lower level assembly, such as creating sequences having narrative continuity through the conjunction of short single shots. However, if the content units are longer (extending over numerous shots), they can be more self-contained. The rhetoric of montage can then be addressed by careful manual construction of the opening and closing ends of components, to ensure that transitions between components in automatically assembled sequences are appropriately marked and cued. If a component is as large as a complete story in a news program, this may be straightforward, using the normal conventions in news programming for closing the subject of one sequence at the end of that sequence and then introducing the next subject as the beginning section of its sequence. However, if an automatic system supports variability within a story, guaranteeing an effective rhetoric of montage may be problematic. If an interactive video system uses segments defined within a predefined linear video program (such as a news broadcast), the options for creating an effective rhetoric of montage may be constrained to the careful selection of segment boundaries. If the video database is intended to support interaction from the outset, creating an effective rhetoric of montage may impact upon the preparation of video material prior to entry in the database, and possibly during the original pre-production and production of the video data.

5. Supplementing Rhetorical structure Theory For Interactive Video Sequence Generation

At the highest level of modelling a news database as a joint relation between nuclei representing different stories, RST obviously needs supplementation with a content representation scheme that can indicate subject matter and bibliographical material such as the sources and originating dates of the video contents of the database. For a given subject, RST relations may be useful for identifying subsets of content for presentation, but within any given subset RST may provide only limited support for algorithmically deciding how much of the subset to present or in what order to present it. For example, in the analysis above there are a number of elaboration relations. A user might specify the inclusion of elaboration material, but this could amount to a longer presentation that the user really wants. The user can be provided with the option of terminating part of a presentation at any time. In this case, the link distance heuristic described above might be used to try to ensure that important material is presented first. However, for material having a similar link distance, RST itself provides no indications of the desirable presentation order. In these circumstances,

other strategies are clearly required to supplement the functionality provided by RST models. Ongoing research is required to investigate the applicability of narrative or associative/categorical sequencing techniques to provide some of the required supplementary functions, or to develop new methods compatible with the use of RST as a model of rhetorical semantics.

The research described in this paper has been based upon the explicit modelling of RST relations between particular video segments to provide a static hyperlink structure over a video database. Previous research in dynamic and interactive video sequence generation [1, 2, 3, 4, 5] has demonstrated dynamic link generation within video databases. RST relations could potentially function in a dynamically linked system, supporting the same interaction approaches described in this paper. To achieve this, the authoring process would have to include the association of the functional roles within RST relations with content descriptions, and provide content descriptions for the video segments within the video database. Instead of hard linking relations with segments, linking could then be achieved dynamically by matching content descriptions. Ongoing research is required to establish the viability of this approach to dynamic rhetorical video presentation generation.

6. CONCLUSION

The research reported in this paper clearly indicates that RST can provide a strong foundation for interactive and adaptive presentation generation of rhetorically intended news video. RST provides a basis for interactive video presentation generation that may have a stronger rhetorical force than linear video productions. While this may be desirable from the viewpoint of the author of an RST model, it may not be in the best interests of the user of the interactive video system in terms of gaining a deep understanding of the subtleties and ambiguities of an issue. This is because when an RST model is used as a basis for presentation generation, it acts to manifest a rhetorical function which may be only one of many possible potential rhetorical functions that the video data has. This effect may be countered by providing alternative RST analyses of a common video database. The availability of multiple analyses may have the opposite effect to the provision of a single analysis, highlighting the variations of interpretation that are possible, and emphasising how limited a single interpretation can be. This presents the user with the dilemma of resolving conflicting interpretations, encouraging their own analysis and closer study of the material available within the database, and discouraging the acceptance of any particular interpretation at face value.

REFERENCES

1. Davis M. (1994): "Knowledge Representation for Video", *Proceedings of the 12th National Conference on Artificial Intelligence*, AAAI, MIT Press, pp. 120-127.
2. Nack F. (1996): *AUTEUR: The Application of Video Semantics and Theme Representation for Automated Film Editing*. Ph.D. Thesis, Lancaster University, UK.
3. Nack F. and Parkes A. (1997): The Application of Video Semantics and Theme Representation in Automated Video Editing. *Multimedia Tools and Applications*, [Ed: Zhang, H.], Vol. 4, No. 1, pp. 57 - 83.
4. Davenport G. and Murtaugh M. (1995): "ConText: Towards the Evolving Documentary" Proceedings, ACM Multimedia, San Francisco, California, Nov. 5-11.
5. Lindley C. A. (2000): "A Video Annotation Methodology for Interactive Video Sequence Generation", *BCS Computer Graphics & Displays Group Conference on Digital Content Creation*, Bradford, UK, 12-13 April 2000.
6. Stam R., Burgoyne R., and Flitterman-Lewis S. (1992): *New Vocabularies in Film Semiotics: Structuralism, Post-Structuralism and Beyond*, Routledge.
7. Mateas M. and Sengers P. (1999): "Introduction to NI Symposium", AAAI 1999 Fall Symposium on Narrative Intelligence, http://www.cs.cmu.edu/~michaelm/narrative.html.
8. Sengers P. and Mateas M. (1999): "Narrative Intelligence", AAAI 1999 Fall Symposium on Narrative Intelligence, Massachusets. Available at: http://www.cs.cmu.edu/~michaelm/NISchedule.html.
9. Bordwell D. and Thompson K. (1997): *Film Art: An Introduction*, 5th edn., McGraw-Hill
10. Mann W. C., Matthiesen C. M. I. M. and Thompson S. A. (1989): "Rhetorical Structure Theory and Text Analysis", *Information Sciences Institute Research Report*, ISI/RR-89-242, November.
11. Lloyd Rutledge, Brian Bailey, Jacco van Ossenbruggen, Lynda Hardman, and Joost Geurts (2000): Generating Presentation Constraints from Rhetorical Structure In: Proceedings of the 11th ACM conference on Hypertext and Hypermedia (pages 19-28), May 30 - June 3, 2000, San Antonio, Texas, USA.
12. Lloyd Rutledge, Jim Davis, Jacco van Ossenbruggen, and Lynda Hardman. (2000): Inter-dimensional Hypermedia Communicative Devices for Rhetorical Structure In: Proceedings of International Conference on Multimedia Modeling 2000 (MMM00), November 13-15, 2000, Nagano, Japan,
13. Noel D. (1986): "Towards a Functional Characterisation of the News of the BBC World News Service", Antwerp, Belgium. Antwerp Papers in Linguistics, No. 49.
14. Lindley C. A., Simpson-Young B., and Srinivasan U. (1998): "The FRAMES Project: Reuse of Video Information Via the World Wide Web", Workshop on Reuse of Web Information (14 April), held in conjunction with WWW7, Brisbane, Australia, 1998.

15. Srinivasan U., Lindley C., Simpson-Young B. (1999): "A Multi-model framework for Video Information Systems", "Semantic Issues in Multimedia Systems", *8th IFIP 2.6 Working Conference on Database Semantics (DS-8)*, Jan 5-8 1999, Rotorua, New Zealand.

16. Landow G. P. (1991): "The Rhetoric of Hypermedia: Some Rules for Authors", *Hypermedia and Literary Studies*, Delany P. and Landow G. P. Eds., The MIT Press, 81-104.