http://dx.doi.org/10.5007/2175-8026.2013n64p49

MULTIMODAL ANALYSIS OF FILM WITHIN THE GEM FRAMEWORK

John A. Bateman
University of Bremen

Abstract

In this paper, the predominantly visual framework developed for the analysis of static multimodal documents within the Genre and Multimodality project 'GeM' is considered as a foundation for treating non-static multimodal artifacts. The paper introduces the original framework and characterizes how it can be beneficially extended to work with the moving audio-moving image. Several illustrations of the new framework's application to narrative film are presented in order to show how it may provide stronger support for empirical investigations of artifacts of this kind.

Keywords: : layout, film, gem model, multimodality, analysis

1 Introduction: from static to dynamic documents

In Bateman (2008) a detailed model for the analysis of static multimodal documents was presented focusing on the mutually constraining influences of 'genre' and 'multimodality' — hence the

framework's name: GeM (*Genre and Multimodality*). Although linguistically inspired, the GeM model re-centered attention away from the language occurring in multimodal documents and towards the artifact as such *as a primarily visually* realized semiotic object. This paved the way for the empirical exploration of a broader range of distinguishable semiotic modes than previously considered (cf. Bateman, 2011). A natural question that this raises is the potential of the framework for considering non-static 'documents'. In this paper, we set out the position of one very common type of non-static multimodal document, the 'narrative film', when viewed through the lens offered by the GeM framework. We will see that, on the one hand, many properties of film fall naturally out of the resulting description and, on the other, that this view offers a rich site of integration for some quite distinct but nevertheless valuable methods for analyzing film. In addition, we shall also see why the continued use of the term 'document' is not only justified for dynamic artifacts such as film but also beneficial, in that it opens up a much needed source of further constraint when building detailed models of film and its interpretation.

2 The GeM framework

The approach to static documents considered here was originally deployed in order to clarify the notion of *genre* in the multimodal context. In earlier work (e.g., Bateman et al., 2001), we had noted that designing multimodal page-based artifacts within a model relying on communication goals or intentions still appeared to leave far too many 'design decisions' open. Our hypothesis was then that this variation was due to an insufficient consideration of the constraints brought about by the requirements that a document participate in a particular genre. The resulting *Genre and Multimodality* model (GeM: Delin et

al., 2002; Bateman, 2008), set about to provide an overarching scheme within which genre could be explored multimodally and by which the additional design constraints required for particular classes of documents could be empirically investigated and formally specified. The kinds of documents considered in this research project included traditional print newspapers, web-based newspapers, instruction manuals, and information booklets — particularly bird field guides.

One result of this work was a multi-layered analysis and annotation scheme by which any static multimodal document could be decomposed at several distinct levels of abstraction. Recurrent patterns at the different levels were then to be described in terms of mutual constraints, which, taken together, constituted proposals for the definition of individual or families of 'multimodal genres'. Whereas this work continues for static documents (e.g., Bateman et al., 2007; Thomas, 2009; Hiippala, 2011), our focus here will be to follow the implications of the model when we look at *dynamic* documents, in particular, narrative film. We do this for at least two reasons. First, there is nothing about the approach pursued in the GeM framework that is inherently restricted to static documents and so it is necessary to explore this further concretely in order to evaluate the framework. And second, a variety of problems are known from film studies, particularly concerning issues of reliable segmentation for analysis, that appear appropriate for treatment within a strongly structuring framework such as GeM.

A principal addition made by the GeM framework to genre and multimodal analysis was the full acceptance of the importance of artifact *materiality*. The significance of materiality for semiotic accounts has grown considerably in recent years and there are many distinct directions pursuing the consequences of materiality for meaning-making; this is also the case within multimodal semiotic

work such as that of Kress and van Leeuwen (2001) and Kress (2010), which we build on here. Accepting materiality within an account means that, when carrying out multimodal analysis, the physical properties of the artifacts under investigation must also be considered for their potential contributions to meaning-making — i.e., choice of material brings with it its own constraints and makes its own 'communicative' statements. Selection of a particular kind of handmade paper for a document, for instance, might indicate values such as selectivity, exclusivity and expense while also constraining the kinds of design decisions that may be made due to absorption qualities of the paper, its ability to successfully carry fonts of various sizes, and so on. The genres that may be used with a particular material are thus constrained.

The GeM model takes this further and explicitly includes the interaction of material with production, distribution and reception technologies. Thus, rather than talking of material by itself, the GeM model introduces the notion of the *virtual artifact*. The virtual artifact is the 'material' that is accessible to design decisions by virtue of both the actual physical properties of some material *and* available technologies and practices for using that material. Genres are then carried by the virtual artifact rather than the physical material directly. The reason for this extra level of indirection is that genres as social constructs may maintain themselves even in the face of changing physical properties. One good example of this process is offered by newspapers, where the virtual artifact of the print newspaper has developed over the past 200 years in a way that has its origins in physical properties of the paper used and the printing technology then available, but which is now reproduced for reasons of genre rather than technological or physical limitations. The virtual artifact of print newspapers still generally consists

of 6–8 narrow columns of text with limited use of headlines even though, in contrast to the situation in the early 19th century, there is nowadays no technological reason for this. Narrow columns then help newspapers to deploy the two-dimensional spatial extent of the page to express subtle distinctions of relative importance and newsworthiness that developed for the medium over its 200 year history (see: Bateman, 2008, pp. 17-18, p. 181). Design decisions for the medium are by no means free of this heritage and so proceed 'as if' the material was still imposing constraints. Maintaining the notion of virtual artifact independently of the actual physical 'canvas' is therefore useful to account for this. The actual physical canvas now includes few constraints concerning columns, font size, color of print, etc.; but the virtual artifact does.

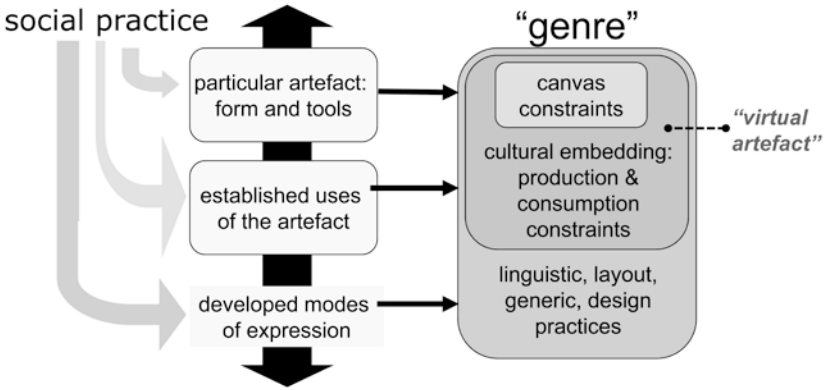

Figure 1: The basic *Genre and Multimodality* model (Delin et al. 2002; Bateman 2008, p. 16)

The notion of genre developed within the GeM model then originally included all of these constraints, involving most specifically use of, on the one hand, the concrete material of artifacts

in combination with the technological equipment necessary to manipulate it and, on the other, established practices for producing and consuming an artifact of a given type. The GeM model is therefore best described as an onion-like structure of embedded levels as suggested graphically in Figure 1. The 'innermost' levels, shown under the label 'genre' on the right-hand side of the figure, are made up of the virtual artifact, which itself combines the physical canvas employed for presentation and the conventionalized technological use of that canvas. Outside of this we find conventions of design and generic patterns of expression that change as genres and technological capabilities change. All of these viewpoints are situated within, and developed by, social practices, as indicated on the left-hand side of the figure. Historically-situated social practices develop particular artifacts with the help of particular technologies of production and distribution, supporting various ranges of communicative uses of those artifacts, which in turn lead to the emergence of particular modes of expression. It is these modes of expression that the GeM framework then describes with the help of its visually-based multi-layer annotation scheme. As set out in detail in Bateman (2008) and summarized in Delin et al. (2002), two levels from this model are the *layout layer* and the *rhetorical structure* layer. We will see below how both of these have natural correlates in non-static artifacts also.

3 Semiotic modes within the GeM framework

On the basis of the investigations of static documents that followed using the GeM model, several kinds of social semiotic practice were proposed that reoccurred across all of the documents studied. This gave rise to an extension of the notion of *semiotic mode* along lines very similar to those originally proposed by Kress et al.

(2000, p. 43). Under this view, a semiotic mode is a conventionalized way of using a material substrate for semiotic purposes. Moreover, and as argued at greater length in Bateman (2011), to describe such semiotic configurations it is useful to adopt a 'stratified' view in which each semiotic mode is characterized at three distinct semiotic levels.[1]

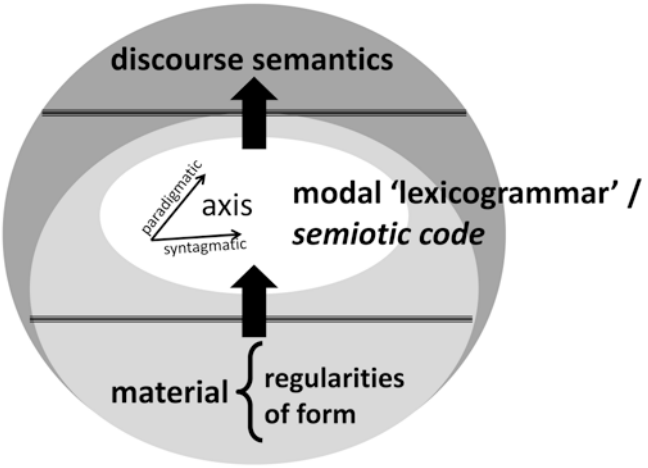

Figure 2: Semiotic modes as a combination of three semiotic 'strata': material substrate, 'grammar' and discourse semantics (cf. Bateman, 2011)

This semiotic stratification can again be summarized most succinctly in graphical form, for example as suggested in Figure 2. At the least abstract level, shown at the bottom of the figure, there is the material substrate that can be used for leaving traces of semiotic distinctions. At the next level, there is an organization analogous to lexicogrammar in verbal language, in which particular generalized patterns can be specified that hold over distinctions drawn in the material substrate. These patterns can vary in their complexity from simple 'lists' of different items (a 'lexical' organization) to complex structural configurations (a 'grammatical' organization). The role

of this level is to determine just which material distinctions are to be considered 'semiotically charged' and which not — there are consequently similarities to be drawn here with treatments of materiality pursued by, for example, Eco, Peirce and others (for some early discussion and references, cf., e.g., McCanles, 1977). Descriptions of this level can be built relying on traditional organizational dimensions, such as, for example, the Saussurean paradigmatic and syntagmatic axes. Finally, at the most abstract semiotic level, there is a semiotic *discourse semantics* stratum, which contains resources for linking configurations from the lower semiotic strata into connected and 'larger-scale' communicative unities. The particular function of the discourse semantics is to relate semiotic 'messages' or 'utterances' to their context of use. It is this component, well developed for verbal language, that marks the most significant extension of the model beyond previous accounts of semiotic codes. Traditionally, semiotic codes might be considered in terms of collections of signs; within the current model, those signs are themselves subject to 'orchestration' in order to construct more complex and richly textured semiotic acts. Specifying the discourse semantics of semiotic modes is then an attempt to make the functioning of this orchestration explicit and subject to investigation in its own right.

There is no pre-given list or closed set of semiotic modes. Semiotic modes can 'grow' *whenever* some community of users puts work into their use and the material the modes employ is sufficiently manipulable as to show the traces necessary for their recognition. An important consequence of this is that there are actually rather more semiotic modes to be discerned than generally discussed in the literature. The still widespread tendency to discuss semiotic modes in terms of sensory channels, rather than as diversely conventionalized ways of expressing meanings, typically serves to group together

semiotic modes that are more usefully distinguished — particularly when the analytic task is to determine how their respective meaning contributions combine.

Within the GeM work, for example, three visually-based semiotic modes were proposed to operate in many static documents: text-flow, image-flow and page-flow (Bateman, 2008, p. 175). Text-flow corresponds closely to Twyman's (2004) 'linear interrupted', where text is formatted in a fashion that comes as close as possible to a continuous unbroken stream of words — lines wrap at the end of columns, text is continued over pages, etc., but no additional meaning accrues from these segmentations. Image-flow is similar to text-flow but involves static pictorial elements instead of textual elements: simple comic strips would be an example. And page-flow involves the full two-dimensional extent of the virtual canvas (e.g., page, screen, window in a user interface, etc.) to express additional semantic relations of difference, similarity, relatedness, unrelatedness, and so on. These three semiotic modes are distinguished, as is the case for all semiotic modes, by virtue of their quite distinct discourse semantics (Bateman, 2011).

There are already several suggestive similarities to be drawn out here with semiotic artifacts that employ dynamic materials, such as films. Most straightforwardly, for example, we can investigate a potential connection between representations such as comics (or 'sequential art': Eisner, 1992) and film. Similarities and overlaps between these media are frequently discussed both from the side of film and from that of comics (cf. Lacassin, 1972; Groensteen, 2007; Ecke, 2010). Both comics and film typically share a reliance on iconic pictorial representations combined with a depicted unfolding over time employed for narrative purposes. The film semiotician, Christian Metz, accordingly characterized both representational

forms in terms of the single semiotic property of *multiplicity* (Metz 1974, pp. 227–232; Bateman & Schmidt 2012, p. 134): i.e., sequences of images (themselves either moving or static) are arranged successively over time. For this reason, we can now see the image-flow semiotic mode introduced above as occurring in both static and dynamic varieties. Their discourse semantics appear analogous both to each other and to that of conjunctive relations as defined for verbal language (Martin, 1983; van Leeuwen, 1991; Martin, 1992), although each exhibits interesting medium-specific differences requiring its own separate treatment.

The page-flow semiotic mode is less relevant for film in the form defined within the GeM work, although something similar does appear to be happening in segments of films that use a 'split-screen' effect (cf. Bordwell and Thompson, 2010, p. 187). This is analogous to having two images or other visual elements placed together on a page since, in contrast to the situation with image-flow, there is typically no necessary assumption of temporal succession involved. Split-screens most commonly involve simultaneity plus a strong sense of comparison or contrast — both typical semantic relations found within page-flow generally. Films are, however, now making increasing use of *dynamic* split-screen effects, where the sizes and shapes of elements within the screen change or overlap (cf. Bateman and Veloso, 2013). To capture the meanings being created here may well require a further dynamic variety of page-flow; it is at present too early to say how this might look. Considerable further empirical research is required

Finally, we need to combine the discussion of the previous section and the notion of semiotic modes introduced here. Whereas our earlier definition of semiotic mode discussed the manipulable material substrate necessary for any semiotic mode in terms of

physical properties, we must also open this up and use instead the GeM notion of virtual artifacts. This means that the material that is manipulated for the formation of semiotic modes is exactly as indicated for the 'virtual artifact' in Figure 1, i.e., a combination of physical material and technologies of production, dissemination and reception. This is naturally of particular importance for a medium such as film, which is crucially dependent on technology throughout.

4 The virtual artifact of film

The considerations of the previous two sections now provide sufficient basis to turn to film itself. We begin by characterizing in some detail just what we should consider the virtual artifact of 'film' to be; this will turn out to have rather more interesting properties than commonly assumed. One natural tendency previously has been to focus on the 'physical' side of the medium, that is, in this case, traditionally strips of celluloid. This does not, however, do justice to the virtual artifact that is manipulated in service of the semiotic modes operative in, and constitutive of, film and makes further meaningful cross-media statements more difficult than need be. Equally insufficient is a focus on what is 'shown' in the film — discussions of film in terms of the storyworld portrayed often take this path. Film-as-medium then becomes transparent and reduced to Peircian indexicality, in much the same way that photography is sometimes seen (cf. Lefebrve, 2007), only more so because of the increased immersive 'reality-effect' of film's synchronized sound and movement.

What we need to set alongside these perspectives is the 'raw material' available for *constructing* film: i.e., the manipulable material that provides the basis for the growth of semiotic modes within

appropriate communities of practice. This manipulable material consists of viewable film segments that may be joined together in various ways. Thus the manipulations carrying semiotic modes may be exercised both within segments, in terms of any of the particular properties that a moving audio-visual iconic image possesses, and across segments, in terms of which audio-visual segments are brought together and how. This is very much a perspective aligned with film production: typically described more loosely as what goes in the scene (*mise-en-scène*) and how scenes are combined (*montage*); extensive introductions to these terms are given in, for example, Bordwell and Thompson (2010). For current purposes, we will focus on the combination of film segments since this is one aspect of the use of the material that clearly distinguishes film from straightforward recordings such as video surveillance or medical imaging.

A basic design decision when filming some ongoing event is then between whether that event is presented as a single film segment — the equivalent of leaving the camera running and having the event play out in front of it — or whether the event is presented by means of several distinct segments that are placed together in sequence. This difference is simple, but fundamental and is analogous in many ways to discussion of the space between panels in sequential art (McCloud, 1994, p. 66): the elements on either side of this space, called the 'gutter', set a communicative challenge to the observer in order to see how the elements can be related. This is the key role of *multiplicity* as defined by Metz and is a decisive feature of all image-flow semiotic modes precisely because, once the 'unity' of the event is broken, space is opened up for considerable variation. What can happen *between* segments is almost endless: there may be omissions of uninteresting material, changes in camera angle for various

focusing and attention directing additions, or even the omission of *interesting* material in order to raise tension and suspense or introduce ambiguity. Moreover, the placing of distinct segments together in montage is by no means restricted to preserving the 'order' of succession of the original event or events and it is here that film really begins to become a fully fledged semiotic mode of its own.

This property was discovered very early on in the development of film and has continued in increasingly complex forms ever since. One of the earliest conventionalized examples is the parallel or alternative editing style promoted particularly in the films of David Wark Griffith in the early 1900s; as Mary Ann Doane describes this particular case:

> The yoking together of noncontiguous spaces through parallel editing forced a certain denaturalization of the filmic discourse. It required the spectator to accept enormous leaps in space and to allow the disfiguration of continuous time, its expansion or contraction. (Doane, 2002. p. 194)

From our perspective here, however, this 'denaturalization' of discourse is precisely what marks the birth of discourse proper, for it is only with this development that the resources of a potential semiotic discourse stratum are freed from the spatiotemporal contingencies of an iconic audio-visual representation and are able to begin developing their own styles of meaning.

The consequence of this is that we need to take *freely combinable sequences* of film-material as the 'raw' manipulable material substrate of film. Moreover, referring back to the onion-structure of Figure 1, this film-material is itself considered in terms of what an observer sees and hears when it is played back with appropriate technology. It

is crucial that a semiotic perspective is taken, therefore, rather than a physical, material one. It is not the fact that a film strip may be made up of images in frames on celluloid that is important: it is the audio-visual images that are perceptible when played that determines the manipulations that are relevant for carrying out on that material, manipulations which correspond to the middle semiotic stratum in the diagram of Figure 2. This position is relatively uncontroversial. What goes into individual shots and how they are edited together have long been seen as providing the basic dimensions for meaning-making within film. Our placement of this within a model of semiotic modes does, however, provide a stronger foundation for exploration than looser notions of film language or film grammar. As we shall see below, it is then possible to start characterizing the virtual artifact of the semiotic mode(s) of film in considerably more detail than hitherto the case.

The emphasis on semiotic modes and the fact that it is semiotic modes that define the distinctions that are to be seen as meaningful also moves our characterization away from particular medial realizations. For the semiotic modes that develop it is of little consequence whether the combinable film sequences are acetate, celluloid or data files. As long as these materials together with their supporting technology are subject to the same semiotic constraints, they can stand as equivalents for the semiotic modes that develop. Differences in modes come about only when perceptible differences are supported by the virtual artifacts involved. Thus, the move from silent to sound films certainly made possible a significant extension in the semiotic modes supportable; similarly, to the extent that extensions are used in ways that are indeed different to their predecessors, extension of the virtual artifact, on the one hand from single-track audio, to stereo, to 5.1 and 7.1 surround sound, etc. and,

on the other hand, from 2D to single-viewpoint 3D (e.g., current 3D films) to multiple-viewpoint 3D (e.g., holograms), etc. might well enable different semiotic modes to develop.

With this background on the nature of the 'filmic', we can now go on and say significantly more about the properties of the filmic material substrate. Since we are at all points dealing with 'semiotically-charged' material, there are many properties inherent to the material that contribute to its use. For example, if two film segments show images of what is recognizably the same location, the perceptual capabilities of the human visual system generally provide ready access to this fact. Alternatively, if the same object, or the same person, is shown, this referential information is also immediately accessible unless the film-maker takes pains to hide it. These *cohesive* qualities of the filmic-material (Tseng, 2013) provide richly organized linkages across segments as a film unfolds and are *manipulable* in precisely the manner required to support semiotic modes.

Film segments are therefore not unmarked, uninterpreted strips of potential. Due to the phenomenological situation that under normal circumstance we never perceive natural scenes as uninterpreted sense data, filmic material comes already 'labeled', both culturally and spatiotemporally. As Lena Jayyusi puts it, when viewing a scene where a policeman is arresting someone, that is what we 'see' and not "a man putting circular metal objects round another human being's limbs" (Jayyusi, 1988, p. 274). Thus cultural knowledge, to the extent that it is available, is always already present in the perception alongside attributions of spatial setting and temporal extent. We immediately categorize and classify what is being seen and heard and this equally forms part of the material substrate of film.

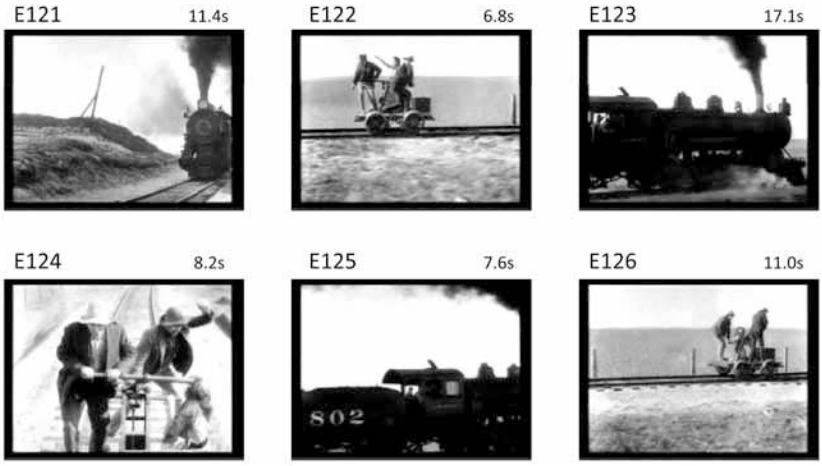

Figure 3: Shots 121–126 from Griffith's *The Girl and her Trust* (1912: 12:03–13:05)

We can see all these aspects at work in the simple film extract shown in Figure 3. This is an early example from Griffith taken from towards the end of his film *The Girl and her Trust* from 1912, in which a speeding locomotive is in hot pursuit of two vagabonds on a handcar who have stolen a chest containing a considerable sum of money. The segment shows the typical cross-cutting of pursuers and pursued that has long since become a staple of film of all kinds. This particular chase sequence has also been discussed at length from the perspective of film theory by, for example, Branigan (1992, pp. 20–25). Showing a train, some other scene, then the train again is almost necessarily (given sufficient supporting cohesive ties) taken as a view of a single train interrupted by a view of something else. It is then precisely the inherent labeling that comes along with each shot that shows that cross-cutting is occurring.

The virtual artifact of film can then be characterized very well by employing further constructs developed within formal modeling

approaches to documents. In particular, we will rely on the general distinction between the *logical* organization, typically related to what is being portrayed or depicted — i.e., the sociocultural, temporal and spatially labeled 'pro-filmic' material (that is, the material in front of the camera), and the *layout* organization, which characterizes how a logical organization is being presented on some display medium (cf. Schmidt 2008, Bateman & Schmidt 2012, pp. 48–58). Within the original GeM model, a particularly detailed view of 'visual' layout structure was set out for static documents. For film, we develop this further and consider layout in terms of the design decisions involved in combining and sequencing film segments. These possibilities are inherent to the virtual artifact as just described. The logical document structure for some filmic segment is then characterized as a collection of shots. These shots can be grouped according to their times of occurrence and the spatial regions that they depict. For the purposes of analyses, grouping of this kind should be carried out as conservatively as possible: that is, if it is not apparent from what can actually be seen and heard in the shot, then one can assume that those shots belong to distinct groupings and the task of relating them must be taken up by discourse considerations.

The current example segment readily decomposes into two labeled sets: one concerned with the vagabonds, the other with the pursuing train. Now, *each* of these could have been presented as a segment in its own right — this would then have corresponded to the simplest situation mentioned above in which a single 'event' is shown depicted in several shots. In terms of the logical organization, therefore, we do not need to distinguish these alternatives: both contribute to what can be termed a *scene*. For the current segment, we then have two scenes:[2] one consisting of the shots {E121, E123, E125}, and one consisting of the shots {E122, E124, E126}. Each has

its own defined 'space' of occurrence (which we will denote as S_{train} and $S_{handcar}$ respectively), and each shows subevents which occur (or which may be taken to occur) in the temporal order shown.

The one additional property of the artifact that turns the example segment from two scenes into something more interesting is the way in which the layout structure of the segment presents the scenes: rather than following one after the other, they are interleaved to give a classical 'alternation' structure. Interleaved structures of this kind are, as pointed out above, immediately recognizable and provide a further structural organization to which various discourse meanings can be attributed — that is, we have a manipulable property of the material substrate which can be semiotically 'charged' by use within a semiotic mode.

It is then the fact that a particular layout structure has been selected and is perceptible to viewers that triggers the *discourse* requirement that a semantic connection be found between the portrayed scenes. If a semantic bridge cannot be found, then the entire segment would be seen as unmotivated.

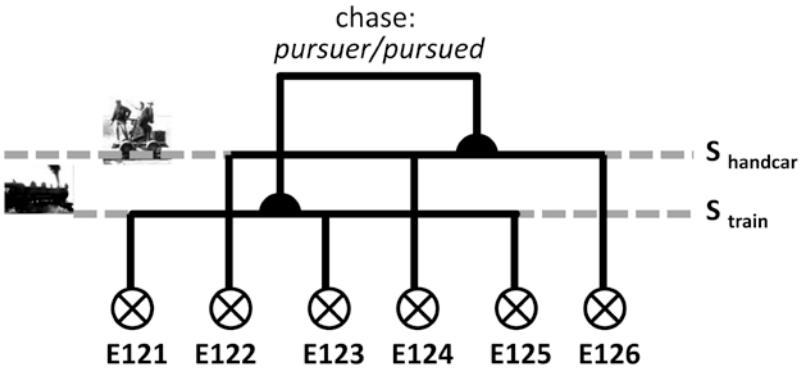

Figure 4: Layouted logical structure of shots 121–126 from Griffith's *The Girl and her Trust* (1912: 12:03–13:05)

This can be shown in the form of a structural diagram as suggested in Figure 4. In such diagrams, the shots and their depicted content are shown as crossed circles along the bottom of the structure. These are directly linked to their respective scenes, which are identified by their respective spatial labels. Here we have two such scenes as described above. Particular subsegments of these scenes may then be picked out to participate in 'higher-order' structures. The subsegments are indicated by semicircular connectors picking out the portions that are relevant. This is necessary because we may, for example, have been working with a film which showed the handcar (or the train) in many other configurations before or after the actual chase. It is not necessary, therefore, that an *entire* scene contributes to any particular higher-order structure. Alternation layout structures then demand that a higher-order organization be found to relate the contributing alternating scenes. This is possible by hypothesizing a symmetric semantic relation that holds for each pair of shots forming a transition from one scene to the other: i.e., in the present case, over the transitions {E121, E122}, {E122, E123}, {E123, E124}, etc. Since each of these can be fitted to the configuration 'chasing/being chased by', the given semantic connection can be taken to apply for the alternation as a whole as indicated in the diagram.

The properties that such logical and layout structures must exhibit in order to be considered an account of film can be articulated in considerable detail. This is undertaken at length in the corresponding chapters from Bateman and Schmidt (2012), on which the style of analysis presented here is based. We can take these levels of organization as determining the virtual artifact that is available for manipulation within the semiotic mode of dynamic image-flow. The particular structures that then result during the analysis of any film provide the necessary starting points for application of the discourse

semantics of that semiotic mode. Particular configurations within the artifact, such as the alternation layout illustrated here, call for specific kinds of discourse semantic hypotheses to be pursued. As with all discourse semantics, these hypotheses may then turn out to be incorrect or in need of correction — in other words, a discourse semantics always draws on *abductive* reasoning by definition (see Bateman 2011, again the relevant chapters from Bateman & Schmidt 2012, and Wildfeuer 2012).

Characterizing the detailed layout and logical structures instantiated in the virtual artifact of any film provides a solid basis for further empirical exploration of both the narrative organization of the film and any recipient's response to that film. The definitions that we have provided elsewhere concerning the properties of the layout and logical structures constrain possible segmentations so that a far higher degree of inter-analyst reliability can be sought than has traditionally been the case with semiotically-inflected approaches to film. The analysis only goes so far of course — there are no considerations of social import or aesthetic evaluation here; nevertheless, segmenting films in the ways suggested does offer a reliable place to start such further interpretative work whenever issues of discourse may be reasonably suspected to be at work.

5 Two more complex examples

To give more of a sense of the proposed analytic scheme in action, let us consider two somewhat more complex examples.

The first is a well-known episode from Alfred Hitchcock's *The Birds* from 1963, in which the main female character, Melanie Daniels (played by Tippi Hedren), is waiting outside the village school for classes to finish. The portion of the scene relevant for us here consists

of 15 shots and is shown in three rows of 5 shots in Figure 5. Since individual shots sometimes contain within them further useful detail, two images are shown for a single shot whenever necessary, positioned vertically within each of the three shot rows.

Figure 5: Segment from Alfred Hitchcock's *The Birds* (1963: 01:06:19–01:08:43)

This segment brings out again how crucial grouping by spatiotemporal labeling is for the virtual artifact of film. Applying

this criterion often directly provides a film segmentation strongly supportive of further levels of discursive analysis and interpretation. The introduction and maintenance of such regions during a film is accordingly one of the main tasks that design decisions made for the construction of its virtual artifact take on. It is then on the basis of these design decisions that narrative development and viewer response can be 'orchestrated'.

In terms of the storyline, the episode at issue here occurs as it is becoming all too apparent that the birds in the village present a threat. Groups of birds first flock together and then attack any humans they can find. Until they flock together, however, they remain harmless. At the beginning of the example segment, Melanie Daniels goes outside of the school building, walks along the outside of the school playground and sits down by the playground fence with a climbing frame in the playground behind her (shots 1–2). The link between shots 1 and 2 is a match-on-action as Melanie Daniels sits down, changing in the process from a long shot to a medium shot in front of her and slightly to her right. This effectively establishes the overall spatial region of the segment, which we will label S_0. What happens after this is then interesting for its use of space as a means to distribute knowledge. As Melanie Daniels sits on the bench, lights a cigarette and looks increasingly ill at ease, with frequent glances back towards the schoolhouse, birds gather in increasing numbers by perching on the climbing frame behind her. The main space of the action is, however, divided cleanly into two: one spatial region for Melanie Daniels and another for the climbing frame in the playground. Shots 1–2 inform the viewer what the relation between these spaces is, but shots 3–15 maintain them as separate.

This means that we are again dealing with an alternation-like structure, just as with the introductory example above. The two tracks

participating in the alternation here are that of the climbing frame, which we will label S_1, and that of Melanie Daniels, which we will label S_2. Each of these contributes to its own scene as before and, also as before, the layout structure for the segment interleaves the presentation of the shots of the scenes to form an alternation. However, in contrast to the previous situation where the semantic relations between the tracks were never in doubt, here the potential alternation remains unresolved for a full 70 seconds: that is, the shots alternate, but there is no connection drawn between them apart from the bare spatial positioning set up in shots 1–2. This, and what happens following shot 9, is brought out more clearly in the corresponding structure diagram as shown in Figure 6. To ease reference to the film material, indicative thumbnails of some of the scenes and subscenes involved are included in the diagram, although these play no role for the formal organization and its interpretation. Similarly, the duration of portions of the segment is also indicated below the shot numbers.

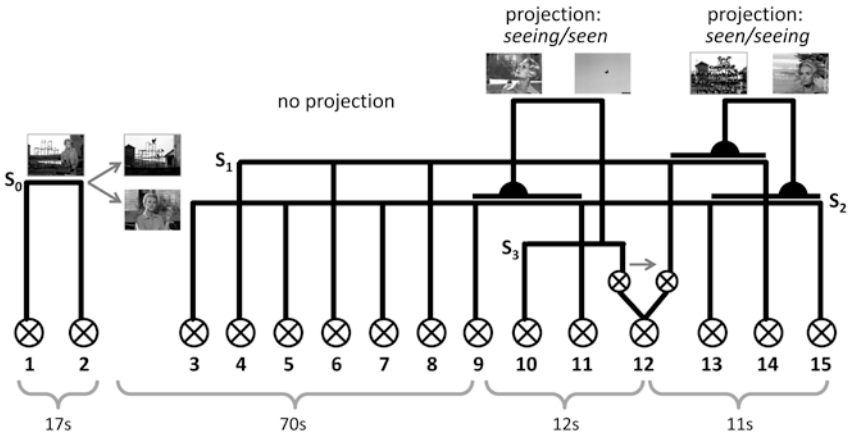

Figure 6: Layouted logical structure for the example segment from *The Birds* (1963: 01:06:19–01:08:43)

Particularly interesting about the segment is the manner in which tension is created. The two tracks of the alternation are only brought together narratively in the shots beginning with the last portion of shot 12. After this point there is a regular shot/reverse-shot structure alternating between the main character looking and what she is looking at (cf. Branigan, 1984). The semantic connection thus established is here one of, in systemic-functional linguistics terms, *projection* where a 'senser' (here Melanie Daniels) senses a 'phenomenon' (the birds on the climbing frame) (van Leeuwen, 1996). At this point the main character knows of her, and the schoolchildren's, danger and so heads back towards the schoolhouse.

In order to get to this, Hitchcock employs several delaying strategies which ensure that the viewer is always aware of the growing danger while also being equally aware that the main character does not know this. Moreover, rather than have the character suddenly discover the collected birds all at once, an intermediate episode is constructed whereby Melanie Daniels first notices a single bird flying high up towards the playground (shots 9 and 10). This introduces a further, previously unknown space, labeled S_3 in the figure. The Daniels character follows this bird on its trajectory for 12 seconds until she (and we) see that its final destination is indeed the climbing frame, which is now completely covered with birds. At this point, the previously disjoint spaces S_3 and S_1 merge (shot 12). The Melanie Daniels track S_2 and the climbing frame track S_1 therefore come to be related by projection only after 82 seconds of unrelated alternation.

Thus, in summary, the artifact's logical and layout organization in shots 3–9 is clearly indicative of an alternation and an alternation establishes a discourse requirement that a connection be found. But this explicit connection is uncomfortably withheld until the bird's trajectory beginning in shot 10 ends at the climbing frame.

Withholding this connection is an effective way of distinguishing the states of knowledge required for tension to result. The film viewer is well aware of what is going on behind Melanie Daniels but the possibility of filmic action to escape the danger is continually denied by the delay in establishment of a semantic connection. The film, and so the film viewer, knows more than the character (often treated in narratological approaches to film under 'focalization'; cf. Schlickers, 2009) and this is positively flaunted by the lack of explicit discourse connection.

Whereas this construction of the film could of course be described informally based on a careful viewing, one of the purposes of our more detailed formal analysis is that it focuses attention so as to more or less 'force' the pertinent details on the analyst. One of the most difficult issues for film analysis in general has long been to judge which of the myriad of technical details in film may be relevant in any particular case and in support of any particular analytic claim: our proposal here is that the structures uncovered in our description of the virtual artifact are extremely likely to be relevant because the appropriate construction of the virtual artifact is essential for any filmic artifact to be interpretable at all; more discussion of precisely this issue can be found in, for example, Bateman (2013). Obviously, these details do not exhaust what one needs to consider, but without them it becomes more difficult to characterize almost all other aspects of a film's organization.

For our second example, we move on another 40 years and consider a segment from Bryan Barber's *Idlewild* from 2006, a story that compares the paths taken by two friends, the quiet, piano playing Percival and somewhat wilder but essentially good-hearted 'Rooster', in a small southern town in the U.S. in the 1930s. This is a very different type of film to those we have just considered — part small-time gangster film, part musical. In addition, and as might be expected for a more recent film of this kind, there is considerable

camera movement and the film as a whole is cut far more rapidly than in the examples we have seen so far. Even relatively short extracts bring a considerable number of shots into play. The focus of our description in this case, therefore, will continue the line of development of the previous example but will also show how the kind of filmic structure we have defined supports a useful flexibility in the level of detail that needs to be considered. Particularly as we move to higher-level narrative constructions, it is by no means always necessary to track those constructions shot-by-shot, although this granularity is always available should it be needed.

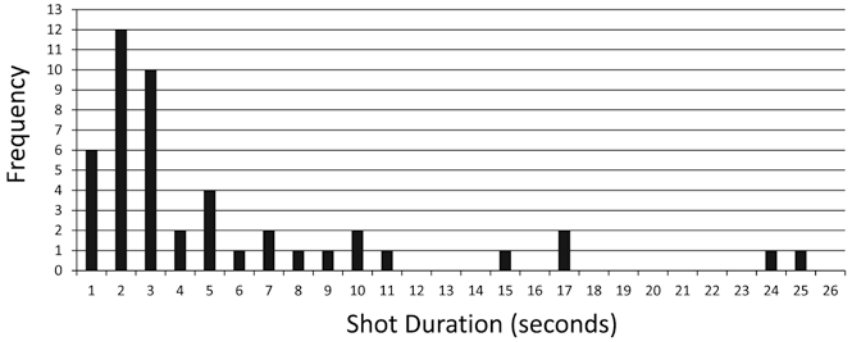

Figure 7: Frequency of occurrence of shot with specified lengths in the example segment from Bryan Barber's *Idlewild* (2006: 00:42:20–00:46:14)

The segment considered is made up of 46 shots and lasts just under 4 minutes. With an average shot length of 5s, this may not appear at first glance to be a particularly fast cutting rate; however this average is actually due to few rather long shots being balanced against an overwhelming majority of short shots. This is shown in the duration/frequency graph in Figure 7. In the example segment, two shots are over 24s in length whereas 18 shots last less than 2s.

As it happens, almost all of the very short shots occurring in

this segment are elements within shot/reverse-shot sequences in conversations. At these points, the film is showing sometimes quite heated interactions and so, focusing on first one conversational participant and then on the other, rapidly jumps from one angle to another. This behavior results in organizations of the virtual artifact that straightforwardly resemble the alternations we have seen above. The shot/reverse-shot structures each define two tracks interwoven in the layout structure. Therefore, as before, these alternations require semantic connections in order to be seen as motivated. This semantic connection is provided in all cases by the symmetrical relationship of 'talking to'/'being talked to'.[3] This means that, for present purposes, we can raise the level of abstraction of our account by treating all these fine-grained alternations as single units: their internal structure follows straightforwardly from the discussion above and so can be neglected in the following. The example segment then falls into the eight larger units shown in Figure 8.

Figure 8: Example segment from Bryan Barber's *Idlewild* (2006: 00:42:20–00:46:14)

Grouping the shots in this way builds naturally on what would have resulted had we carried out a fine-grained shot-by-shot analysis of the segment as above. Each group can be labeled as being associated with a particular spatial region: groups G1, G3, G5 and G7 take place at the home of Percival (played by André Benjamin), which is also a funeral parlour, while groups G2, G4, G6 and G8 take place at the home of Rooster's mistress Rose (Rooster played by Antwan A. 'Big Boi' Patton and Rose by Paula Jai Parker). In G1, the segment begins with Angel, a singer at the nightclub where both Percival and Rooster perform (played by Paula Patton), getting out of a taxi and going into the funeral parlour to talk to Percival. In G2, we see Rooster cautiously entering a house and drawing a gun, before being confronted by Rose wielding a frying pan. G3 then returns to the funeral parlour, where Angel surprises Percival by sitting up suddenly from an open coffin. In G4, G5 the two conversations are developed further, coming to a close in G6 and G8, with Rose leaving in a taxi, and in G7, with Angel leaving in a taxi. Both G7 and G8 also contain shots showing Percival and Rooster as the respective taxis leave, although in the case of Rooster the leave-taking is portrayed as more final while for Percival, the segment represents more the beginning of a relationship.

What is interesting about this segment for us here is the manner of its construction. As was the case with our original Griffith example, we have here what could equally have been presented as two separate scenes: one concerning Percival and Angel and another concerning Rooster and Rose. But, instead of this, the film interleaves them. Moreover, the segment employs a broad range of technical devices for tightly binding the scenes together. For example, at the end of group G1 Percival clearly hears and reacts to someone knocking at a door; at the beginning of G2, we see Rooster cautiously opening a door and

calling 'hallo' as he enters an apartment. It is therefore equally possible to link the knocking at the door with Rooster and with Angel. Then in the transition from G3 and G4, there is a link in the dialogue, with the subject matter directly taken over from Angel to Rooster, and again across G5 and G6 between Angel and Rose's landlady. G7 then ends with Angel getting into a yellow taxi and G8 begins with Rose getting into one. We could term many of these bridges *cohesive* (cf. Tseng, 2013) matches-on-action since there is only a similarity relationship involved — the common actions shown are carried out by different people in different places. There is also a common musical track running over the entire segment, punctuating particular high-points of action in a similar way regardless of whether that action is from the Percival-Angel interaction or the Rooster-Rose interaction.

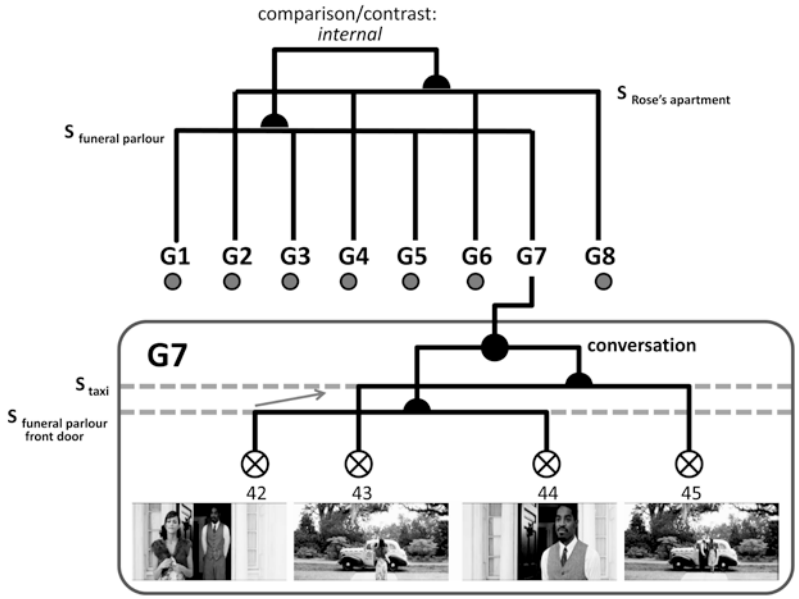

Figure 9: Virtual artifact structure of the segment from Bryan Barber's *Idlewild* (2006: 00:42:20–00:46:14)

The layouted logical structure of the segment is then as shown in Figure 9. To suggest the full analysis, one of the component units, G7, is shown in full. It would be possible to provide this structure for all of the units present in a similar fashion; this would bring out how the various spatial regions constructed by the film (suggested by the dashed gray lines in the rendition of unit G7 in the figure) run through the entire segment, participating along the way in alternations of the kind illustrated. The higher-order structure holding over all of these units is then *itself* an alternation of the kind we have seen operating in all of our examples so far. This alternation brings two broad spatial regions into a relationship of semantic connection by means of the interleaving of their respective units. And this, as we have argued above, signals a requirement that we find an appropriate discourse relationship to bind them together.

However, in the present case, there is no *content-based* connection that can reasonably be hypothesized. The events portrayed stand in no logical, or better, no *ideational* semantic relationship. As a consequence, we can state that the discourse relationship involved is, in the terms defined for verbal conjunctive relations in Martin (1992), an *internal* relationship (i.e., concerned with the textual organization) rather than an *external* relationship (i.e., concerned with the content that is being portrayed). The kinds of relationship most commonly occurring with this organization are temporal simultaneity and contrast/comparison; both are strongly suggested by interleaved layout structures. In particular, the very general sense of 'comparison' involved cannot be cancelled: showing shots with this layout organization *commits* to an assertion of relatedness and, even when this relatedness is not provided by the content, an internal comparison relation will remain.

This presents us with a particularly clear example, therefore, of how the forms inscribed within the virtual artifact can take on specific discourse interpretations that are not limited to causal, content-relations in the subject matter depicted. The structure of alternation itself serves the abstract function of realizing comparison and contrast between its tracks. In the present case, a viewer is strongly invited to make connections of comparison between the two scenes which would not otherwise have been foregrounded. These comparisons are strengthened by the deliberate parallels and cohesive matches designed into the scenes, but their explicit combination is already directly signaled by the (higher-level) cross-cutting between them.

In general, semantic relations between tracks and other elements of the filmic virtual artifact are provided by the particular discourse semantics of the dynamic image-flow semiotic mode. There appear to be a limited range of such relations, just as is the case for verbal language. This similarity with verbal conjunctive relations was first explored by van Leeuwen (1991); further proposals for this level of description are now motivated for film at length in Bateman (2007) and the corresponding chapters from Bateman and Schmidt (2012).

6 Conclusions, outlook and challenges

In this paper, we have seen how the layered model for investigating static documents developed within the GeM model can be naturally extended to consider dynamic documents such as film. The notion of layout employed for static artifacts corresponds to the filmic activity of putting segments of film-material into various structural arrangements. These arrangements are essentially

temporally organized rather than spatially organized as is the case for static documents. Nevertheless, whereas the use of a single temporal dimension here might have been thought to be more restricted than the two-dimensional organization available for static layout, in fact the richness of film's audio-visual iconic material substrate, and the ready recognition of 'content' that this supports, makes considerable complexity possible in the structural organization of the filmic virtual artifact. This complexity provides a strong foundation for a finely articulated discourse semantics of its own. We have suggested how this structural articulation can be described and by means of examples related this to discourse interpretation.

Of course this only scratches the surface of what needs to be taken into account when analyzing film. The structural configurations we have proposed for the virtual artifact correspond largely to what is allocated to the *syntagmatic* axis of filmic description in Bateman and Schmidt (2012); there we provide definitions and examples of the formal properties that this organization exhibits. The discourse semantic connections that we have mentioned at several places in the current discussion then correspond to what we elsewhere have investigated as part of the *paradigmatic* axis of filmic description. Both of these axes need to be combined in any comprehensive account. We have also simplified for the purposes of the present discussion the notion of filmic 'units' employed, essentially keeping these to shots. Again, for considerably more detail, the interested reader is referred to the relevant chapters of Bateman and Schmidt (2012).

Despite these simplifications, we believe that with this foundation in place it becomes possible to place many of the standard questions raised in film theory and film interpretation on a firmer empirical basis. Even complex filmic organization can be reliably segmented in order to provide an appropriate backbone for finer-grained analysis

of all kinds. What now remains to be done is to explore this line of development for a broader range of films, following through the predictions for structuring that follow from our model and showing how these support analyses at other levels of abstraction.

Notes

1. Kress and van Leeuwen (2001, p. 9) also talk of 'strata' involving multimodal artifacts — in particular, discourse and design (content plane), and production and distribution (expression plane). Their usage is, however, concerned with semiotic production as a social activity and so differs from the deliberately more restricted focus on semiotic modes that we pursue here.
2. In fact, we always make scenes maximal, so that they include all shots that could contribute spatially and temporally; we omit this for the present discussion since we are only going to discuss this short extract from the film in question.
3. There is rather more to discuss for such cases in general; often the spaces shown in such shot/reverse-shot configurations overlap and so it is also possible to consider them as providing differing views of a single event without diegetic alternation — a point also argued, for example, by Christian Metz. Further discussion here would take us too far afield, however.

References

Bateman, J. A. (2007). 'Towards a *grande paradigmatique* of film: Christian Metz reloaded', *Semiotica* 167(1/4), 13–64.

Bateman, J. A. (2008). *Multimodality and genre: a foundation for the systematic analysis of multimodal documents*. London: Palgrave Macmillan.

Bateman, J. A. (2011). The decomposability of semiotic modes, *in* K. L. O'Halloran, & B. A. Smith (Eds.), *Multimodal Studies: Multiple Approaches and Domains* (pp. 17–38), Routledge Studies in Multimodality, London: Routledge.

Bateman, J. A. (2013). Looking for what counts in film analysis: a programme of empirical research, *in* D. Machin, (Ed.), *Multimodal Communication*. Berlin: Mouton de Gruyter.

Bateman, J. A., Delin, J. L., & Henschel, R. (2007). Mapping the multimodal genres of traditional and electronic newspapers, *in* T. D. Royce, & W. L. Bowcher (Eds.), *New Directions in the Analysis of Multimodal Discourse*. (pp. 147–172). Mahwah, NJ: Lawrence Erlbaum Associates.

Bateman, J. A., Kamps, T., Kleinz, J., & Reichenberger, K. (2001). 'Constructive text, diagram and layout generation for information presentation: the DArt$_{bio}$ system', *Computational Linguistics* 27(3), 409–449.

Bateman, J. A., & Schmidt, K.-H. (2012). *Multimodal Film Analysis: How Films Mean*, Routledge Studies in Multimodality. London: Routledge.

Bateman, J. A., & Veloso, F. O. D. (2013). 'The Semiotic Resources of Comics in Movie Adaptation: Ang Lee's *Hulk* (2003) as a case study', *Studies in Comics* 4(1), 137–159.

Bordwell, D., & Thompson, K. (2010). *Film Art: An Introduction. Ninth Edition*, 9th edn. New York: The McGraw-Hill Inc.

Branigan, E. (1984). *Point of view in the cinema: a theory of narration and subjectivity in classical film*, number 66 *in* 'Approaches to semiotics', Berlin: Mouton.

Branigan, E. (1992). *Narrative comprehension and film*. London: Routledge.

Delin, J. L., Bateman, J. A., & Allen, P. (2002). A model of genre in document layout. *Information Design Journal* 11(1), 54–66.

Doane, M. A. (2002). *The emergence of cinematic time: modernity, contingency, the archive*. Cambridge, MA and London, England: Harvard University Press.

Ecke, J. (2010). Spatializing the movie screen: How mainstream cinema is catching up on the formal potentialities of the comic book page, *in* M. Berninger, J. Ecke, & G. Haberkorn (Eds.), Comics as a Nexus of Cultures: Essays on the interplay of media, disciplines and international perspectives (pp. 7–20). Jefferson, NC and London: McFarland & Company, Inc.

Eisner, W. (1992). *Comics and sequential art*. Princeton, WI: Kitchen Sink Press Inc.

Groensteen, T. (2007). *The system of comics*, Studies in popular culture, University Press of Mississippi, Jackson, Miss. translated by Bart Beaty and Nick Nguyen, from the original French *Système de la bande desinée* (1999).

Hiippala, T. (2011). The localisation of advertising print media as a multimodal process, *in* W. L. Bowcher (Ed.), *Multimodal Texts from Around the World : Linguistic and Cultural Insights* (pp. 97–122). Basingstoke: Palgrave Macmillan.

Jayyusi, L. (1988). Towards a socio-logic of the film text. *Semiotica* 68(3-4), 271–296.

Kress, G. (2010). *Multimodality: a social semiotic approach to contemporary communication*. London: Routledge.

Kress, G., Jewitt, C., Ogborn, J., & Tsatsarelis, C. (2000). *Multimodal teaching and learning*. London: Continuum.

Kress, G., & van Leeuwen, T. (2001). *Multimodal discourse: the modes and media of contemporary communication*. London: Arnold.

Lacassin, F. (1972). 'The comic strip and film language', *Film Quarterly* 26(1), 11–23. original articles translated by David Kunzle.

Lefebrve, M. (2007). The Art of Pointing. On Peirce, Indexicality, and Photographic Images, *in* J. Elkins (Ed.), *Photography*. (pp. 220–244). London: Routledge.

Martin, J. R. (1983). Conjunction: the logic of English text, *in* J. S. Petöfi, & E. Sözer (Eds.), Micro and macro connexity of discourse. number 45. *Papers in Textlinguistics* (pp. 1–72). Hamburg: Helmut Buske Verlag.

Martin, J. R. (1992). *English text: systems and structure*. Amsterdam: Benjamins.

McCanles, M. (1977). Conventions of the natural and the naturalness of conventions. *Diacritics* 7(3), 54–63.

McCloud, S. (1994). *Understanding comics: the invisible art*. New York: Harper Perennial.

Metz, C. (1974). *Language and Cinema*, number 26 *in* 'Approaches to Semiotics', Mouton, The Hague. Translated by Donna Jean Umiker-Sebeok.

Schlickers, S. (2009). Focalization, ocularization and auricularization in film and literature, *in* P. Hühn, W. Schmid, & J. Schönert (Eds.), *Point of View, Perspective, Focalization: Modeling Mediacy in Narrative* (pp. 243–248). Berlin: de Gruyter.

Schmidt, K.-H. (2008). 'Zur chronologischen Syntagmatik von Bewegtbilddaten (III): Deskriptive Syntagmen', *Kodikas/Code: Ars Semeiotica* 31(3–4), 217–270.

Thomas, M. (2009). Developing multimodal texture, *in* E. Ventola, & A. J. M. Guijarro (Eds.), *The world told and the world shown: multisemiotic issues*. Basingstoke: Palgrave Macmillan.

Tseng, C. (2013). *Cohesion in Film: Tracking Film Elements*. Basingstoke: Palgrave Macmillan.

Twyman, M. (2004). Further thoughts on a schema for describing graphic language. *Proceedings of the 1st International Conference on Typography and Visual Communication* (pp. 329–350). University of Macedonia Press.

van Leeuwen, T. (1991). Conjunctive structure in documentary film and television. *Continuum: journal of media and cultural studies* 5(1), 76–114.

van Leeuwen, T. (1996). Moving English: the visual language of film, *in* S. Goodman, & D. Graddol (Eds.), *Redesigning English: new texts, new identities* (pp. 81–105). London and New York: Routledge and the Open University.

Wildfeuer, J. (2012). Intersemiosis in film: Towards a new organization of semiotic resources in multimodal filmic text, *Multimodal Communication* 1(3), 233–304.

[Received in 24/09/2012. Approved in 08/04/2013]