http://dx.doi.org/10.21622/ILCC.2025.05.1.1087

Vol. 5, Iss. 1, June 2025

Multimodal Analysis of Film and Video Production: Reviewing the Field

Marwa Mohamed Abd Allah[1], Marwa M. Khamis El-Zouka[2], Abeer M. Refky M. Seddeek[3]

[1&3]College of Language and Communication (CLC), Arab Academy for Science, technology, and Maritime Transport (AASTMT).
[2]Faculty of Arts, Alexandria University.

E-Mails: marwaabdallah@aast.edu, m.khamis@alexu.edu.eg, dr.abeer.refky@aast.edu

Received on: 11 November 2024 Accepted on: 19 December 2024 Published on: 06 April 2025

ABSTRACT

It has been found that the literature on film analysis and video production primarily contains four proposed attempts at applying multimodal terms, techniques, and procedures to film and video production: O'Halloran (2004), Tan (2009), Baldry and Thibault (2006), and Bateman (2008, 2012). The core insight and principle these four models share is how semiotic resources or choices are combined and interact to produce meaning. They all emphasize, each to their own, that semiotic resources or modes are organized into a hierarchy of systems, planes, strata, or taxonomies where semiotic features can be identified, classified, and analyzed to form patterns and connections that ultimately lead to a better understanding and interpretation of multimodal phenomena. In addition, they highlight the importance of global coherence and how it is achieved through the repeated co-deployment of semiotic modes to form patterns in dynamic texts. The four frameworks touch upon the notion of genre and how patterns of intersemiotic relations can be instrumental in identifying genres. Finally, they point out that the construction of meaning in dynamic texts is impacted by how the text unfolds in real time.

Keywords: *Multimodal analysis, Film analysis, Intersemiotic analysis, Meaning-making, Visual resources.*

1. INTRODUCTION

Multimodal analysis has gained prominence and imminence in contemporary research studies due to the constant rise, influence, and consumption of visual and digital media, offering an array of tools and strategies that shed light on how visual modes, means, and resources are utilized in the construction and production of meaning-making. The article at hand reviews four major analytical models proposed by influential scholars, namely O'Halloran, Tan, Baldry and Thibault, and Bateman. The article is divided into five sections; the first four sections provide the general underpinnings, tools, and strategies put forward by each scholar. The last section pinpoints research gaps and critiques of the models presented. Through a comparative lens, the article attempts to highlight the contributions and potential limitations of the analytical approaches proposed by these eminent scholars.

2. O'HALLORAN'S FRAMEWORK (2004)

O'Halloran (2004) believes that the spatio-temporal unfolding of semiotic choices, along with their interaction with other resources, contribute to meaning-making. She uses a web-based instrument called (MCA) designed to analyze dynamic texts that display varying configurations of sound, image, gesture, text, and language as they unfold in time. The MCA segments a video clip into sections according to frame numbers or time intervals. Moreover, the software allows the user to manipulate visual footage in many ways; for example, the image may be adjusted for brightness, contrast, and color. Special effects, such as blurring, distortion, perspective, edge definition, and shadowing, can be applied as well. The software also allows for the insertion of text, lines, vectors, figures, outlines, and shadings. In addition,

visual transitions between parts of the footage can be marked in several ways. The linguistic text can be tagged so that the visual images can be analyzed through direct textual engagement. According to O'Halloran (2004), video-editing tools allow the user to highlight the different semiotic choices visually and view their impact when they combine with the text in real time (113).

As shown in Figure 1, the proposed systemic-functional framework classifies the film according to type, form, and genre. Then, O'Halloran goes on to suggest that the semiotic analysis is based on a metafunctionally organized rank structure that consists of the film plot, sequences, scene, mise-en-scene, and frame (2004,114). Central to this framework is the idea that a film plot is constructed from a series of sequences motivated by similarity and repetition, difference, and variation. Mise-en-scene is concerned with everything seen within the frame as it unfolds in time; a change in mise-en-scene is motivated by a change in camera shot. Series of mise-en-scene from the scene, and scenes form sequences. "Sequence" is the term used to divide the film into segments. In O'Halloran's framework, mise-en-scene is analyzed according to visual imagery, speech, music, sound effects, and how visual imagery is interwoven with the soundtrack.

Film type:	fiction, documentary, experimental and animated
Film form:	narrative, categorical, rhetorical, abstract and association
Genre:	multiple types; for example, narrative films include science fiction, western, musical, comedy, suspense, and action thrillers with sub-genres horror, detective, hostage and gangster
Ranks:	Film Plot Sequences Scenes Mise-en-Scène (the shot) Frame

Figure 1: O Halloran's Film Classification and Ranks (O'Halloran 2004, 115)

Visual Imagery contains the ranks of Movement-Action-Event in a shot, temporal episode, temporal figure, and temporal member. Mise-en-scene includes systems for: (a) Interpersonal meaning, such as Patterns (Kinesic, Proxemic, Rhythm, Gaze, and Shape), Duration of the Image, Speed of Motion, and Point of View; (b) Representational meaning, for example, Movement-Action Sequence; (c) Logical meaning, for example, Narrative Cause-Effect Relations; and (d) Compositional meaning, for example, Changes in Gestalt, On-Screen/Off-Screen Space, Camera Angle, Camera Level, Camera Distance, and Mobile Frame. The Mobile Frame allows a change in camera position in the mise-en-scene. The Mobile Frame interpersonally orients the viewer toward the image and contributes to the representational meaning in the form of the Point of View constructed within the film (O'Halloran 2004, 118). Figures 2, 3, and 4 delineate and break down the aforementioned ranks and constituents and exhibit how they are realized via representational, logical, and compositional meanings.

Semiotic Resources/Rank	Modal	Representational	Logical	Compositional
MISE-EN-SCÈNE COMPLEX (the edited scene)	Contrasts	Narrative continuity and discontinuity	Cause-effect relations	Continuity and discontinuity
MISE-EN-SCÈNE The Temporal-Spatial Frame Complex Relation: The Shot				
Visual Imagery				
Movement-Action-Event in a Shot	Patterns: Kinesic Proxemic Rhythm Gaze Shape Colours and Contrast Lighting Quality Light Intensity Lighting Direction Lighting Source Clarity Focus Film Tonality Special Effects Duration of Image Speed of Motion Point of View (Viewer)	Movement-Action-Event Sequence Figures/Objects Nature of Scene Props Lighting Colour Narrative as Cause Effect Relations Point of View Visual Motifs	Narrative Cause-Effect Relations	Frame Dimension Frame Shape Changes in Gestalt: Framing Horizontal Vertical Diagonal Colour Cohesion/ Contrast Perspective Relations On-Screen/Off-Screen Space Camera Angle Camera Level Camera Distance Mobile Frame Film Editing

Figure 2: Functions and Systems in Mise-en-scene, (O'Halloran 2004, 120)

Semiotic Resources/Rank	Modal	Representational	Logical	Compositional
Temporal Episode	Relation to Movement-Action-Event: Scale Depth Centrality Relative Prominence Duration Clarity Focus Light	Sequence of Sub-Actions, Side Sequences and Events Interplay of Actions	Contribution to Narrative Cause-Effect Relations	Relative Relation of Action in Changing Gestalt Subframing Parallelism and Opposition Relative On-Screen/Off-Screen Space Camera Angle Camera Level Camera Distance
Temporal Figure	Colour Coordination/ Contrast Colour Intensity Costume Style Frontal View Change in Size Change in Prominence Gaze Pattern Focus Depth Light	Character of Figure Costume Body Behaviour/Gesture Props	Contribution to Cause-Effect Relations through Intertextual Motif	Relative Position in Changing Gestalt Subframing Parallelism and Opposition Relative On-Screen/Off-Screen Space Camera Angle-Camera Level Camera Distance
Temporal Member	Colour Colour Intensity Style of Costume Part Makeup Facial Expression	Body Part Makeup Facial Expression Gesture Role in action	Contribution to Cause-Effect Relations through Intertextual Motif	Relative Position in Changing Gestalt Subframing

Figure 3: Ranks and Systems of Temporal Episode, Temporal Figure, and Temporal Member (O'Halloran 2004, 121)

Semiotic Resources/Rank	Modal	Representational	Logical	Compositional
	Gesture Light Change in Size Change in Prominence Focus Depth			Parallelism and Opposition Relative on-screen/off-Screen space Camera level and angle Camera Distance
Soundtrack				
Speech	Negotiation Speech Function Mood Modality Polarity Attitude Comment Appraisal Lexical 'Register' Tone Pitch Volume	Ideation Transitivity Tense Lexical Content Ergativity Verbal Motifs	Conjunction and Continuity Logico-Semantic Relations	Identification Theme Cohesion Information
Music	Volume Pitch Timbre Rhythm Fidelity Beat	Genre: Experiential Context Intertextuality Musical Motifs	Narrative Cause-Effect Relations	Sound Perspective (Diegetic, Non-Diegetic)

Figure 4: Ranks and Systems of Soundtrack (O'Halloran 2004, 122)

O'Halloran (2004) admits that the proposed framework is not without fault, as it presents a range of difficulties. According to her, it was nearly impossible to simultaneously and dynamically record the metafunctional choices across the different semiotic systems due to the complexity and range of systems from which options are chosen and the temporal unfolding of these choices in real time. For instance, recording on-screen space for compositional meaning precluded including choices for color cohesion and contrast because the resulting footage became too dense and confusing. In addition, choices from interpersonal systems such as lighting and color could not be combined with the analysis of gaze and proxemics. Not to mention that the temporal unfolding of metafunctional analysis impacted the resultant footage, which was too fast for the viewer to grasp.

3. TAN'S MODEL (2009)

Tan (2009) employs a horizontal format as it supports a continuous presentation of visual frames based on shot length or duration. She contends that such a format aids intersemiotic analysis captures the ways in which the different resources are co-deployed across modes, and allows for the analytical categories to be expanded (160). As shown in Figure 5, the proposed transcription template consists of four analytic categories or blocks divided into sub-categories. The first category involves the sequences of frames, shots, scenes, phases, and sub-phases. The second category is concerned with aspects of the soundtrack. The third category captures the manifestations of experiential/ representational, interpersonal, and textual/compositional meaning potentials conveyed via elements of the visual message. Tan's (2009) framework examines intersemiotic meaning potential on both micro and macro levels and makes use of O'Halloran's (2004) notions of film form, type, genre, and mise-en-scene.

Block 1	Phase/Sub-phase	1q – *Everything's OK*							
	Visual Frame	SHOT 54					SHOT 55		
									
		Frame 292	Frame 293	Frame 294	Frame 295	Frame 296	Frame 297	Frame 298	Frame 299	Frame 300
Block 2	Sound: Music	♫ background rock music continues → Volume (p); Tempo: F					♫ background rock music continues → Volume (p); Tempo: F		
	Song	♪ song continues, but inaudible →					♪ song continues, but inaudible →		
	Speech	thing's every					O K		
Block 3	Verbal Description	[Two figures on snow-capped mountaintop; figure on the right waves it's arm. Argentine flag flaps in foreground.]					Butcher, clad in green T-shirt, surrounded by flanks of brightly colored red meat, holds up hand in *"Everything's OK"* gesture.		
Experiential/ Representational Meaning	Narrative Representations	P:2 +; Vector: Y:gaze:off-screen:viewer + Process: Existential/Circumstance of Location + Movement (flag): Y, Process: non-transactional, intransitive, material process of action					P:1; Vector: Y:gaze:off-screen:viewer + Circumstance of Means		
	Conceptual Representations	Relational Process: Symbolic Attributive: circumstantial:attributive ▬ *participants are mountaineers from (OR in) Argentina* Semiotic Process: Denotation:Categorization/ Typification:Setting, props, Visual Collocation/Iconographical Symbolism ▬ *Argentine flag;* Visual Metaphor + Visual Theme/Motif ▬ *impaired vision* + (implied) Conceptual/Narrative Theme ▬ *Everything's OK*					Relational Process: Symbolic Attributive: intensive:attributive ▬ *participant is a butcher* Semiotic Process: Denotation:Categorization/ Typification: props, Visual Collocation ▬ *meat* + Conceptual/Narrative Theme ▬ *Everything's OK*		

Figure 5: Tan's Transcription Template (Eija Ventola & Arsenio Jesus Moya Guijarro 2009, 172-17)

Interpersonal Meaning	Mood	Direct Address: Y:demand; Size of Frame: extreme long to long shot; Social Distance: public to far social; Angle/Power: HP:slightly oblique/detached, VP:low; CM:stat	Direct Address: Y:demand; Size of Frame: medium shot; Social Distance: far personal/close social; Angle/Power: HP:frontal:involved, VP:median; CM:stat
	Modality	Color: less than naturalistic S/D; CX: median-low; Depth: medium-shallow:angled; CD: Exposure:under	Color: naturalistic S/D; CX: low; Depth: shallow:central
Textual/ Compositional Meaning	Composition	Salience: Figure:placement	Salience: Figure:Meat:perspective+contrast:color
	Graphic/ Rhythmic/ Spatio-Temporal Relations	↔Graphic Conflict: Setting + color + lighting↔ ↔Rhythmic/Dynamic Match: CM↔ + Conceptual/Narrative Relation to SHOT 12,22,26,38,39,45,52,53,55,56,(60)	↔Graphic Conflict: Setting + color + lighting↔ + Graphic Relation to SCENE 3 + SEQUENCE 6; ↔Rhythmic/Dynamic Match: CM↔ + Conceptual/Narrative Relation to SHOT 12,22,26,38,39,45,52,53,54,56,(60)
Block 4	Intersemiotic Relations	Intersemiotic Complementarity: Intersemiotic Repetition Iconographical Symbolism	

Figure 6: Cont. Tan's Transcription Template, (Eija Ventola & Arsenio Jesus Moya Guijarro 2009, 172-17)

On the micro-level, Tan (2009) examines the impact of editing devices, such as straight cuts, dissolves, fades, flash, or swoosh. Flash is a burst of psychedelic lights and colors. A swoosh is characterized by a rapid diminishing of sharpness and focus or blurring of the image. These are used to segregate shots and create shot boundaries (164). Next, Tan (2009) explores the impact of *conjunctive relations*. She (2009) maintains that the viewers' understanding of how filmic events unfold depends on the *Logical Metafunction*; in other words, the ways in which one event is related to another in the overall structure of the film text (164). In dynamic texts, actions and events are linked based on *Temporal Sequences*. The logic of these sequences is presented through *continuity editing* via "match on action," where a person's action is shown at the beginning in one shot, then continued in the following shot, or mobile framing, where the action is shown from one camera angle, then captured from a different angle in the following shot. Another aspect of temporal conjunction is simultaneity, where the first shot shows one action or event, then another event or action

happening at the same time is shown in the following shot. The impact of *graphic relations* is also examined on the micro-level. These are similarities in shapes, colors, lighting conditions, or camera orientations that bind the logical continuity of scenes and sequences (Tan 2009, 165).

On the macro level, Tan (2009) moves to the wider organizational ranks of *Phase* and *Work as a Whole*. Television advertisements routinely unfold in wave-like, rhythmical patterns or *Phases*, which arise out of the constant shift in choices selected from one or more semiotic modes or resources. A phase is a set of co-patterned semiotic selections that are co-deployed consistently over a given stretch of text (166). Phases do not necessarily correlate with the narrative stages of thematic development: *Orientation*, *Complication*, and *Conflict*. Rather, they coincide with the *Given and New* information structures of the text. The transition between phases is often motivated by a change in the elements of mise-en-scene, like camera movement, for example, or graphic relations.

4. BALDRY AND THIBAULT'S FRAMEWORK (2006)

Baldry and Thibault (2006) adopt a scalar approach to the analysis of multimodal meaning-making by exploring the organization of multimodal texts in terms of different levels. Throughout their book, they put forward numerous *Insets*, basically tenets or principles, which provide bases for their multilayered framework. They also examine how semiotic modes interact and function in relation to one another, on the one hand, and how contexts of situation and culture impact their meaning-making potential, on the other. They posit that context is not something extrinsic to the text; rather, it is created when text users' knowledge of culture and society interacts with the internal features of the text's organization while analyzing and interpreting the text (3). Transcription also helps recognize typical patterns of resource integration as well as the variations within these patterns. Baldry and Thibault (2006) assume that transcription helps better understand the relationship between a certain genre, in their view, a text- and its typical features because transcription techniques can be used to compare different texts from the same genre to highlight their functions within the genre. They seek to establish a systemic way to analyze and interpret multimodal texts.

The resource integration principle is one of the main *insights* proposed by Baldry and Thibault (2006). Basically, meaning making depends on the combinations of semiotic resources, and semiotic resources construct meaning through their mutual interdependence. Baldry and Thibault (2006) move on to define clusters as groupings of resources that form recognizable textual subunits that carry out specific functions within a specific text. To them, multimodal transcription aims at identifying the components of each cluster and the function that each cluster plays within a text (11). Baldry and Thibault (2006) note that the complexity of how resources are co-deployed in any cluster is contingent upon social and technological developments. Another principle they suggested is the *meaning compression principle*. They define it as "a principle of economy whereby multimodal patterned visual and verbal resources are used to identify and provide a model for a larger complex reality that individuals engage with" (19).

In addition, Baldry and Thibault (2006) postulate that a multimodal text should be examined in terms of four types of meaning: *Logical*, *Textual*, *Experiential*, and *Interpersonal*. Logical meaning involves relations of cause, time, continuity, comparisons between events in a given sequence, and why certain changes occurred. Textual meaning constructs the ties between the participants in each sequence. Experiential meaning is concerned with expectations associated with participants' roles and behavior in each situation. Interpersonal meaning entails how the reader is positioned to take a certain evaluative stance towards the world depicted, the participants involved, and the experiences they undergo.

Baldry and Thibault (2006) propose three basic meaning-making units for analyzing film or dynamic texts in general: *phase*, *transition*, and *transitivity frames*. A phase is a set of co-patterned semiotic selections consistently co-deployed over a given stretch of text. Phases are salient local moments in the global development of the text as it unfolds in real time (47). Transcription allows for the revelation of the patterned choices from different systems while the text unfolds in real time. Transition points or boundaries between phases play an important role in how viewers recognize the shift from one phase to another, as well as how a particular phase relates to the overall meaning and organization of the text. They can be signaled via a change in music, camera movement, or body movement, to name a few. Visual transitivity is basically the visual configuration of a process, the participants involved, and the circumstances associated with that process. The meaning of visual transitivity frames is derived from the experiential dimension of meaning in visual texts (122). A transitivity frame can occupy a single shot or can be distributed over several shots. The former is called intra-shot transitivity frames; the

latter is called inter-shot transitivity frames. Baldry and Thibault (2006) consider transitivity frames very important parts of narrative development, for they show actions and how they bring about change or relate to other actions.

Baldry and Thibault's (2006) transcription model comprises six vertical columns: *Time*, *Visual Frame*, *Visual Image*, *Kinesic Action*, *Soundtrack*, and *Metafunctional Interpretation*'s *phases and sub-phases*. The first column specifies the time in seconds determined by the time indicator in the *Windows Media Player*. The second column refers to the visual frame corresponding to the time indicated in the first column. It presents the segmentation of the video track into shots and specifies the transition between shots. The third column presents notational glosses on the reproduced frame. It involves the visual options that orient the viewer to the depicted world in the text, such as camera movements, camera position, camera angles, salience, color, and participants' gaze. The fourth column is concerned with body movements and facial gestures initiated or performed by a certain participant or directed toward another participant. The fifth column includes all aspects of the soundtrack: speech, music, and other sounds. It encompasses the degree of loudness, continuity and pausing, duration, tempo, and relations among auditory voices, such as sequentiality, overlap, and turn-taking. The sixth column specifies the metafunctional bases of all acts of semiosis.

Baldry and Thibault (2006) posit that the basic reality of the visual image projected onto the video screen revolves around what they call a *delimited optic array*. According to them, the optic array is divided into *ambient* and *delimited*. *An ambient optic array* allows the viewer to pick up information about events in his environment, unlike a delimited optic array, which limits the viewer's perception only to what goes on the screen. The surface of the screen displays visual invariants and their transformations in time. In other words, the structure of the array undergoes change and transformation in time, and this change or transformation creates the effect of movement (224). Such change provides information about the movement of participants and objects in the depicted world of the film and information about the viewer's movement in relation to that depicted world – what Baldry and Thibault (2006) refer to as *visual event perception* and *visual kinaesthesis*, respectively.

Visual resources, such as lines, dots, light, shade, and color, comprise what Baldry and Thibault (2006) call the *expression stratum*. Information about visual invariants is manifested in the ways in which these lines, dots, shades, and colors are connected to provide information about shapes, surfaces, and textures, among others. Different visual forms and categories of information in the delimited optic array viewers pick up with their perceptual systems are equated to what Baldry and Thibault (2006) call *expression form*. The delimited optic array specifies information about the operations of transformations, substitutions, deletions, and additions of features employed in the structure of the optical array and the visual kinaesthesis of the observer. Baldry and Thibault (2006) differentiate between *expression* and *content* strata; the former is based on the display of visual invariants and their transformation on a video screen; the latter is based on the depiction of a visual scene consisting of actions, events, persons, and objects in the depicted world (225). The *discourse stratum* is the global level of the text as a meaning-making event in a given social or cultural context.

Furthermore, Baldry and Thibault (2006) maintain that, in visual depiction, a visual image represents a certain phenomenon spatially and temporally grounded in a real or imaginary situation. Accordingly, a visual image can be analyzed into two components: vectors signifying processes and volumes signifying participants in the process. Participants are also linked via *Identity chains*, which show the repeated patterns of interaction between participants on a shot-by-shot basis (233). Chains are linked to each other by visual processes in different kinds of *transitivity frames*. Shots are connected in terms of dependency relations; they are of three kinds: *elaboration* (represented by the equals sign =), extension (represented by the addition sign +), and *enhancement* (represented by the X sign) (235). Narrative dependency relations between temporal and spatial sequences in film or video texts are of two kinds: complication and resolution. Raising questions and providing answers are typical characteristics of narrative discourse organization (238).

Besides, Baldry and Thibault (2006) attach great importance to the identification and determination of perceptually salient units and how these contribute to meaning making. Like O'Halloran (2004) and Tan (2009), they also stress that meaning is always relative to an observer or participant and that meaning-making patterns can be perceived in different ways by different observers.

Baldry and Thibault's (2006) attempt to formulate better multimodal transcription techniques and procedures for analysing multimodal texts as well as constructing multimodal corpora that are inter-semiotic in nature. They believe that multimodal data should be accumulated and referenced to specific

transcriptions and electronically stored databases. The systematic relations between language and other semiotic modes will be quantified on a large scale, thereby explaining how meanings are constructed and manifested in certain genres.

5. THE GEM FRAMEWORK (2008, 2012)

GeM refers to *Genre and Multimodality*. Bateman (2013) views multimodal documents as visually realized artifacts. He believes that the term "document" is justifiable and beneficial for dynamic artifacts, such as film, as it paves the way for developing constrained analyses to interpret film. Bateman (2013) subscribes to the idea that models relying on communicating goals and intentions leave many design decisions open (50). He adds that these models do not take into consideration the constraints a given genre imposes. Bateman's (2008) model aims to devise a scheme that allows for the multimodal exploration of the genre as well as empirical identification and investigation of the design constraints of different classes of documents. The GeM framework offers a multilayered analysis and annotation scheme that can be used to decompose any multimodal document at several levels. Recurrent patterns at different levels are described in terms of constraints, which, in turn, bring about or put forward proposals regarding the definition of multimodal genres. Bateman (2013) states that the GeM model was first applied to static multimodal documents, yet he believes that applying the model to narrative film is an opportunity to evaluate the framework, on the one hand, and solve issues of reliable segmentation common in film studies, on the other (51). Bateman (2008) claims that the GeM framework provides a strong foundation for formulating hypotheses and conducting analysis since it relies on constructing multimodal corpora (15).

A central concept of the GeM framework is that of *materiality*. Bateman (2013) contends that materiality has such a significant impact on meaning-making. As shown in Figure 7, multimodal analysis should encompass the physical properties of the artifact under investigation and how these, in turn, contribute to meaning making as well as impose constraints on the design decisions. Bateman (2013) equates material that influences design decisions due to its physical properties and the technological practices allowing for the use of such material with the term *virtual artifact*. Virtual artifacts, not physical properties, carry genres, yet they are physical properties that impose design constraints. Genres, as social constructs, maintain themselves in the face of changing physical properties.

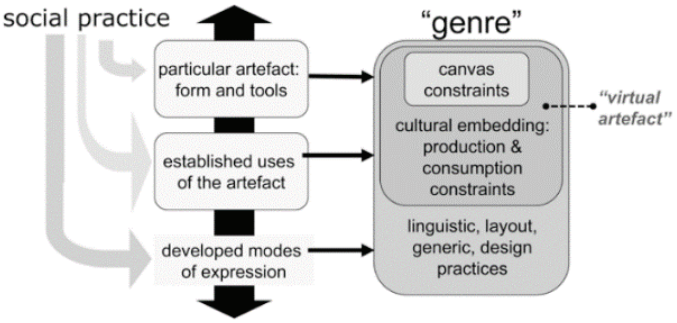

Figure 7: Genre and Multimodality Model (Bateman 2008, 16)

Bateman (2013) asserts that adopting a stratified view to describe semiotic configurations is useful. Figure 8 shows how material substrates give rise to semiotic distinctions; lexicogrammar organization contains generalized patterns, and these patterns can vary in their complexity from simple lists of different items to complex structural configurations. This level is also concerned with determining what material distinctions can be described as semiotically charged and what are not; descriptions can be attributed based on traditional organizational dimensions, such as Saussurean paradigmatic and syntagmatic axes. The semiotic *discourse semantics* stratum contains resources for linking configurations from the lower semiotic strata into connected, larger-scale communicative units and is concerned with relating semiotic messages to the context of use. Semiotic codes are collections of signs, and signs are orchestrated to construct complex and textured semiotic acts; orchestration should be made explicit and should be subjected to investigation.

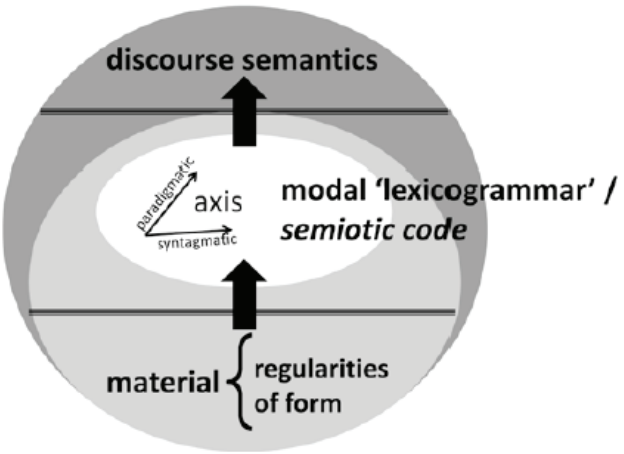

Figure 8: Semiotic Modes as a Combination of Three Semantic Strata: Material Substrate, Grammar, and Discourse Semantics (Bateman 2011, 30)

The predominant semiotic modes employed in dynamic artifacts belong to the image-flow category, as proposed by Bateman (2008). Films rely on combinations of iconic pictorial representations unfolded over time for narrative purposes. However, some films have a split-screen effect, and no succession of time is included. These indicate simultaneity and a sense of comparison or contrast. The semiotic modes employed in this case belong to the category Bateman (2013) calls *page-flow*.

Bateman (2013) views the film as a virtual artifact, a combination of physical material and technologies of production, dissemination, and reception (59). He maintains that film consists of viewable manipulable material that allows for the growth of semiotic modes within communities of practice. This material consists of segments that can be joined in various ways. Manipulations of semiotic modes, in terms of which and how they are brought together, can be achieved within one segment or across segments. In the film production, this is called *mise-en-scene* and *montage*, respectively.

Another key concept that Bateman (2008, 2012, 2013) relies on is *multiplicity*. In simple terms, *Multiplicity* highlights how sequences of moving images are presented on the screen. It is pertinent to what is shown to the viewer, what is omitted, and how these choices raise tension or ambiguity. A montage plays a vital role here because film sequences or shots can be manipulated to be either successive or non-linear. How shots are edited together is essential for meaning making. *The logical* organization involves the sociocultural, temporal, and spatial pro-filmic material, i.e., the material in front of the camera and the collection of shots grouped according to their spatiotemporal occurrence. *Layout* organization is related to how logical organization is presented, particularly the design decisions involved in combining and sequencing film segments (Bateman 2013, 65).

Bateman (2013, 2014) also stresses the importance of empirical examination of narrative organization and recipients' responses. He believes that his proposed framework ensures a highly systemic degree of reliability as far as analysis and interpretation are concerned. By empirical Bateman (2014) explains that constructing hypotheses concerning what a film means should be checked against a larger sample of data, in this case, other films belonging to similar genres, produced in a certain epoch, or directed by a certain director, to see whether these hypotheses can be supported or refuted, and then generalized (368). Therefore, he adopts a corpus-based approach to film analysis, for this, from his point of view, could yield a tighter relationship between filmic material, reliable as well as applicable analytic categories, and recurrent patterns that ultimately contribute to a better understanding of film mechanisms and audio-visual media in general. Bateman (2014) also believes that a corpus-based approach to film analysis results in the provision of databases that are accessible to researchers. However, Bateman (2013, 2014) admits that film as an audio-visual medium is very complex, thereby presenting challenges to developing a reliable scheme of technical descriptions necessary for the application of this methodology.

6. CRITIQUE OF THE FRAMEWORKS AND RESEARCH GAP

As for Bateman's (2008, 2012) model for film analysis, the researcher notes that he pays attention to the structure of semiotic modes and patterns without relating that to the underlying socio-political meaning behind the co-deployment of semiotic resources. He claims that his model stresses the importance of reliability of interpretation and analysis; he relies on quantitative methods to identify common properties of film segments while being compared to a large amount of data, yet the sample analyses he provides are dedicated to brief scenes of films that are not compared to their counterparts of similar genres, directors, or eras. O'Halloran (2004), on the other hand, admits that software analysis presents several challenges yet puts forward a very detailed structure of semiotic options that is impossible for any software or researcher to contain or process. Interestingly, O'Halloran (2004) resorted to Tan's model (2009) when she analyzed a live TV debate (2011), which is considered a dynamic text, instead of using her own model. Another drawback the researcher observes is that the four frameworks highlight the importance of editing tools in the analysis despite the technical challenges and difficulties such tools pose for researchers.

O'Halloran (2004), Baldry and Thibault (2006), and Bateman (2008, 2012) highlight the importance of connections of similarity between shots or sequences, yet they briefly touch upon how contrasting semiotic modes or options are revealed or how they impact meaning making. Tan (2009) and Baldry and Thibault (2006) dedicate their frameworks to analyzing TV advertisements, not feature films, and despite the fact that advertising nowadays utilizes narrative to attract the viewers' attention and communicate socio-political messages, one cannot be a substitute to exemplify how the other is analyzed or interpreted, at least for considerations of the complexity of audio-

visual semiotic combinations and temporal duration. Bateman's (2008, 2012) framework pays a great deal of attention to the design and look of the artifact or document, in his own terms, and insists on empirical research, yet the significant impact of sociocultural factors on meaning making while conducting film analysis must be considered and cannot be relegated in importance.

Baldry and Thibault's (2006) framework could be problematic for several reasons. Forceville (2007) reviews *Multimodal Transcription and Text Analysis: A Multimodal Toolkit and Coursebook* and presents several significant points. He calls their work "long and laborious" and "a real chore to read" (2). On the other hand, Baldry and Thibault keep on introducing so many concepts and definitions, which makes reading their book harder to grasp and their framework denser, more confusing, and harder to apply. Therefore, Forceville (2007) believes that Baldry and Thibault's (2006) book is "not the best book to dispatch students onto the vast ocean of multimodal discourse" (3). Furthermore, to the best of the researchers' knowledge and understanding, they do not dwell much on the notion of genre, and they sometimes confuse it with the notion of text. This is also echoed by Forceville's (2007) review. Finally, he draws upon the fact that they do not use terminology common in film studies and sheds light on how their description, numerous technical terms, and "Inserts" seldom convey what the texts are supposed to convey (Forceville 2007, 2). The researcher observes that no socio-political angle is adopted while Bauldry and Thibault analyze and interpret texts; they just provide descriptions of what takes place visually.

7. CONCLUSION

This article attempted to present a comprehensive review of four analytical frameworks put forward by eminent scholars in the field of multimodal discourse analysis: O'Halloran, Tan, Baldry and Thibault, and Bateman, respectively. These frameworks provide a variety of perspectives on how meaning is constructed via multiple visual resources, means, and modes. These analytical models offer diverse techniques, strategies, and tools that could aid researchers in the analysis of dynamic texts, such as film and video production of different formats. While each framework has its own unique strengths and potential limitations, they undeniably contribute to unraveling the complexities of meaning-making, not to mention paving the way for future research in film and media studies.

REFERENCES

Baldry, Anthony, and Paul J Thibault. *Multimodal Transcription and Text Analysis A Multimodal Toolkit and Coursebook*. David Brown Book Company, 2006.

Bateman, John. "Critical Discourse Analysis and Film." In *The Routledge Handbook of Critical Discourse Studies*, edited by John Flowerdew and John E. Richardson, 612–25. Routledge, 2017. https://doi.org/10.4324/9781315739342.

Bateman, John. "Looking for What Counts in Film Analysis: A Programme of Empirical Research." In *Visual Communication*, edited by David Machin, 365–69. DE GRUYTER, 2014. https://doi.org/10.1515/9783110255492.

Bateman, John. "The Decomposability of Semiotic Modes." In *Multimodal Studies: Multiple Approaches and Domains*, edited by Kay O'Halloran and Bradley Smith, 17–38. Routledge, 2012. https://doi.org/10.4324/9780203828847.

Bateman, John A. "Multimodal Analysis of Film within the Gem Framework." *Ilha Do Desterro A Journal of English Language, Literatures in English and Cultural Studies* 0, no. 64 (July 25, 2013). https://doi.org/10.5007/2175-8026.2013n64p49.

Bateman, John. "Multimodal Cohesion and Text-Image Relations." In *Text and Image: A Critical Introduction Visual/Verbal Divide*, John Bateman., 161–74. Routledge, 2014. https://doi.org/10.4324/9781315773971.

Bateman, John. *Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents*. *Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents*, 2008. https://doi.org/10.1057/9780230582323.

Bateman, John. "Text-Image Relations and Empirical MethodsText and Image." In *Text and Image: A Critical Introduction Visual/Verbal Divide*, 239–50. Routledge, 2014. https://doi.org/10.4324/9781315773971.

Bateman, John, and Karl-Heinrich Schmidt. *Multimodal Film Analysis*. Routledge, 2013. https://doi.org/10.4324/9780203128220.

Bateman, John, Janina Wildfeuer, and Tuomo Hiippala. "Film and the Moving Audio-Visual Image." In *Multimodality*, edited by John Bateman, Janina Wildfeuer, and Tuomo Hiippala, 327–39. De Gruyter, 2017. https://doi.org/10.1515/9783110479898.

Forceville, Ch.J. "Multimodal Transcription and Text Analysis: A Multimedia Toolkit and Coursebook." *Journal of Pragmatics* 39, no. 6 (June 2007): 1235–38. https://doi.org/10.1016/j.pragma.2007.02.007.

O'Halloran, Kay L. "Multimodal Discourse Analysis." In *In Continuum Companion to Discourse Analysis*, edited by Ken Hayland and Brian Paltridge, 120–37. Continuum, 2011.

O'Halloran, Kay L. *Multimodal Discourse Analysis: Systemic-Functional Perspectives*. *Multimodal Discourse Analysis: Systemic-Functional Perspectives*, 2004.

Smith, Bradley A., Sabine Tan, Alexey Podlasov, and Kay L. O'Halloran. "Analysing Multimodality in an Interactive Digital Environment: Software as a Meta-Semiotic Tool." *Social Semiotics* 21, no. 3 (June 2011): 359–80. https://doi.org/10.1080/10350330.2011.564386.

Tan, Sabin. "A Systemic Functional Framework for the Analysis of Corporate Television Advertisements." In *In The World Told and the World Shown: Multisemiotic Issues*, edited by Eija Ventola and Arsenio Jesús Moya Guijarro, 157–82. Basingstoke: Palgrave Macmillan, 2009. https://doi.org/10.1057/9780230245341.