The multimodality of digital longform journalism

Tuomo Hiippala

Centre for Applied Language Studies

University of Jyväskylä

Abstract

Digital longform journalism has recently attracted increased attention among both academics and professionals. This study contributes to the growing body of research by dissecting the multimodal structure of digital longform journalism, that is, how the emerging genre combines written language, photography, short videos, maps and other graphical elements, and joins them together into a seamless narrative using subtle transitions. The data consists of 12 longform articles published in 2012–2013, which have been annotated for their visual and verbal content, their underlying principle of organization, and the transitions that hold between them. The annotation is stored into a digital corpus, which is then analysed to examine the multimodal structures that enable the longform genre to establish a narrative, and to explicate how the longform attempts to captivate its audience by creating a distraction-free environment.

1   Introduction

Digital longform journalism, which seamlessly integrates written language and multimedia, is increasingly recognized as a powerful way of journalistic storytelling. The longform, which Jacobson, Marino, and Gutsche (2015) describe as an emerging genre of digital journalism, seeks to captivate its audience by combining text, photographs, looping videos, dynamic maps and data visualizations into a unified whole. As a genre, the longform exhibits several novel features: simplified navigation and user interfaces, together with smooth transitions between multimedia content slow down the readers' interaction with the longform. Dowling and Vogan (2015) observe that longforms often attempt to create a "cognitive container", an environment that protects the readers from the distractions that plague the typical browsing experience, allowing them to become immersed in the narrative. This stands in stark contrast to the increasingly fast-paced consumption and fragmented structure of online news (Barnhurst, 2012).

Recent research on the longform has emphasized the role of multimedia features and content in containing the readers' attention (Hernandez and Rue, 2016). The 'parallax scroll', in particular, in which content moves across the screen at different paces, has been suggested to characterize the emerging genre (Jacobson, Marino, and Gutsche, 2015). Other prominent features associated with the longform include the use of looping videos consisting of a single shot, and subtle transitions between verbal and visual content (Hiippala, 2015, chapter 9). Combining a minimal user interface with the aforementioned visual features results in what Dowling and Vogan (2015, 214) describe as a 'cinematic' effect. Jacobson, Marino, and Gutsche (2015, 14), who explore how the longform genre uses multimedia, conclude that integrating technology in the unfolding narrative has enabled the longform to recontextualize traditional techniques of literary journalism. For this reason, it may be argued that the longform is more than a simple display of technological prowess: multimedia is at the heart of its narrative structure.

What is traditionally considered 'multimedia' has received much attention in recent years, particularly within the field of multimodal research. This field of research studies how language, images, typography, layout and other modes of communication interact and co-operate in different contexts (for recent overviews of the field, see e.g. Jewitt, 2014; Norris and Maier, 2014). In this article, I argue that theories of multimodality can provide a more accurate view of multimedia in digital longform journalism, circumscribing the features that distinguish the longform from other journalistic genres. Understanding this distinction is, I propose, crucial for explaining the impact of digital longform journalism, as both traditional online news and longforms often draw on the same modes of communication, yet the genres are able to configure these modes in a manner that elicits a completely different reaction from the audience (for a discussion of user engagement with the longform, see Dowling and Vogan, 2015, 210–211). In short, multimodal theories may help to explicate how the longform is able to establish a "cognitive container" and to captivate its audience.

At this point, it is also useful to situate the current study in relation to the previous work on digital longform journalism (Dowling and Vogan, 2015; Jacobson, Marino, and Gutsche, 2015; Hernandez and Rue, 2016). The goal of this study is to complement the previous work by sketching a big picture of the different modes of communication deployed in the longform genre. Providing this kind of a broad scaffolding, I argue, may prove useful for contextualizing future work that dives deeper into the journalistic content and narrative structure of the longforms. Consequently, this study does not place in-depth analyses of content, narrative or aesthetics in the spotlight, but rather seeks to illuminate (1) which modes of communication are brought into play in the longform genre and (2) how they are joined together.

The article itself is structured as follows. To establish a point of departure, I begin by reviewing the relevant work on the multimodality of digital journalism. I then illustrate how multimodal theories may be applied to circumscribe certain features of the longform genre, before proceeding to present the methods and data. The following analyses focus on how multiple modes of communication co-operate in the longform genre, while also accounting for the transitions that join their contributions together. Finally, the conclusion attempts to place the findings at the intersection of journalism studies and multimodal research.

2   Multimodal perspectives on digital journalism

Early work on the multimodality of digital journalism focused mainly on the presentation of daily online news. Pioneering studies in this area were conducted by Knox (2007), who studied "newsbites" on the landing pages of news websites. He defined newsbites as small-sized, concise items that included a header, a brief verbal news item and optionally, a visual illustration. Knox argued that the multimodal structure of newsbites was specifically geared towards interacting with the user on very short timescales, that is, during the initial observation of the landing page. Knox (2009) later expanded the discussion to the visual illustrations in the newsbites, which he termed "thumbnails" due to their small size. In most cases, these thumbnails were used to portray participants mentioned in the newsbites. Yet Knox observed that the size of thumbnails constrained their capability to express meanings, which he found limited in contrast to those available to photojournalism proper, whose potential for conveying complex meanings has been recently explored extensively from a multimodal perspective in Caple (2013) and Caple and Bednarek (2015). For this reason, Knox argued that the thumbnails represented an expansion of the typography in online news, which had evolved to support navigation rather than doing actual journalistic work.

Similar propositions were put forward by Bateman, Delin, and Henschel (2007), who explored the generic structure of landing pages, contrasting their multimodal structure with the front pages of printed newspapers. They observed that the landing pages devoted considerably more layout space to navigation structures than their printed counterparts. Considering the interconnected nature of an entire news website, the move from print to digital has likely driven the emergence of structures such as the newsbites. At the same time, however, news websites retained aspects of traditional newspaper design, whose column-based layout originally emerged from the limitations of early printing technology, which could not produce columns of variable width (Bateman, 2008, 17). Over time, this kind of conventionalized layout – which emerged as a consequence of a technological constraint – evolved into a more abstract organization to support the presentation

4

of news, which was subsequently carried over to digital media. Given the kind of trajectory described above, it is reasonable to expect that certain predecessors may be identified for digital longform journalism as well, not only in terms of journalism practice, but also in terms of multimodal structure.

Generally, the previous research shows that even the most common genres of digital journalism appear to thrive on multimodality, which allows them to develop new means of expression, such as the newsbites and to recontextualize traditional forms of journalism (Jacobson, Marino, and Gutsche, 2015). This capability is founded on being able to coordinate the use of multiple semiotic systems, such as language and image (Knox, 2009, 147). In multimodal research, these systems are often conceptualized as *semiotic modes*, which are communicative resources that are socially shaped among the members of a culture (Kress, 2010, 2014). The modes are, in short, shared resources for making and conveying meanings. Due to its central role within theories of multimodality, the definition of a semiotic mode remains debated in the field (for discussion and various proposals see e.g. Constantinou, 2005; Elleström, 2010; Bateman, 2011). Given the target of the current study, I argue that Bateman's (2011) definition presents an appropriate choice for several reasons.

To begin with, Bateman (2011) postulates that any full-blown semiotic mode must have three strata. Firstly, a mode needs to have a *material substrate*, which may be manipulated to carry physical traces of semiotic activity. In the case of digital longform journalism, the underlying materiality is that of a screen. Secondly, this material substrate determines the available *semiotic resources*, which in this case include – among others – written language, illustrations, photographs, diagrams and audiovisual moving images. These semiotic resources are inherently structured in a way that provides their user with a range of choices, and the selections made within these choices may be combined into more complex structures. Thirdly, all semiotic modes have a specific component – *discourse semantics* – to support the contextual interpretation of their semiotic resources. In other words, discourse semantics provide candidate interpretations for semiotic resources and their combinations across different contexts of use.

The theoretical point of departure may be reiterated as follows: when brought together, the three strata described above – material substrate, semiotic resources and discourse semantics – help to characterize *entire semiotic modes*, which serve as abstractions for bringing multimodal phenomena under analytical control. These abstractions seek to transcend the language/image dichotomy, which is increasingly questioned across different fields concerned with multimodality (Bateman, 2014b). In practice, however, the situation is far more complex, as multimodal genres often deploy multiple semiotic modes simultaneously (van Leeuwen, 2005; Bateman, 2008, 2014a).

To exemplify, the landing page of a news website may combine written language, photographs and pictograms into a spatial organization that I will characterize in terms of a semiotic mode shortly below, while simultaneously embedding another type of multimodal organization in the form of a video. Whereas a video organizes audiovisual moving images into a time-based organization composed of shots and scenes, the landing page combines written language, photographs and other graphic elements in two-dimensional layout space. These combinations, such as the newsbites described above, require the viewer to resolve discourse relations that are primarily spatial, that is, determining which pieces of content belong together. Engaging with the video, however, demands a shift to a different mode of interpretation, which is appropriate for an organization that unfolds in time, that is, resolving relations between shots and their groupings. Tracking these kinds of hybrid organizations represents a major challenge to multimodal analysis, which applies equally to the study of digital longform journalism.

To illustrate how this rather theory-heavy discussion above can provide valuable insights into the multimodality of digital journalism, I will begin by discussing the landing page of the international edition of *The New York Times* (hereafter *NYT*), which also serves as point of contrast for the subsequent analyses of the longform genre. Figure 1 represents the layout of the landing page, which has been marked for the semiotic resources present: written language, pictograms, illustrations and photographs. In addition, I have designated several points of interest, which are numbered from one to seven. These include combinations of multiple semiotic resources commonly found in the news website genre, such as (1) user interfaces combining pictograms and language; newsbites consisting of alternative configurations of language, photographs and illustrations – provided by both the *NYT* (2, 3, 4, 6) and other news outlets (5) – and combinations of pictograms, written language and images (7).

By populating the layout space with these kinds of familiar multimodal structures, the news website is able to fulfil the many functions of "online journalisms" originally identified by Deuze (2003), which are now brought together on the landing page: the *NYT* website provides original content, aggregates content from other sources, allows the users to comment and interact with news items, and encourages sharing them in social media. This does not only show how the functions of the news websites have evolved: I argue that is equally important to acknowledge that performing all of these functions simultaneously is enabled by appropriate combinations of semiotic resources. The task of a genre, then, is to conventionalize these combinations, allowing emergent structures to turn into established patterns over time (Waller, Delin, and Thomas, 2012). This is precisely what Knox (2007, 2009) observed in the case of newsbites.
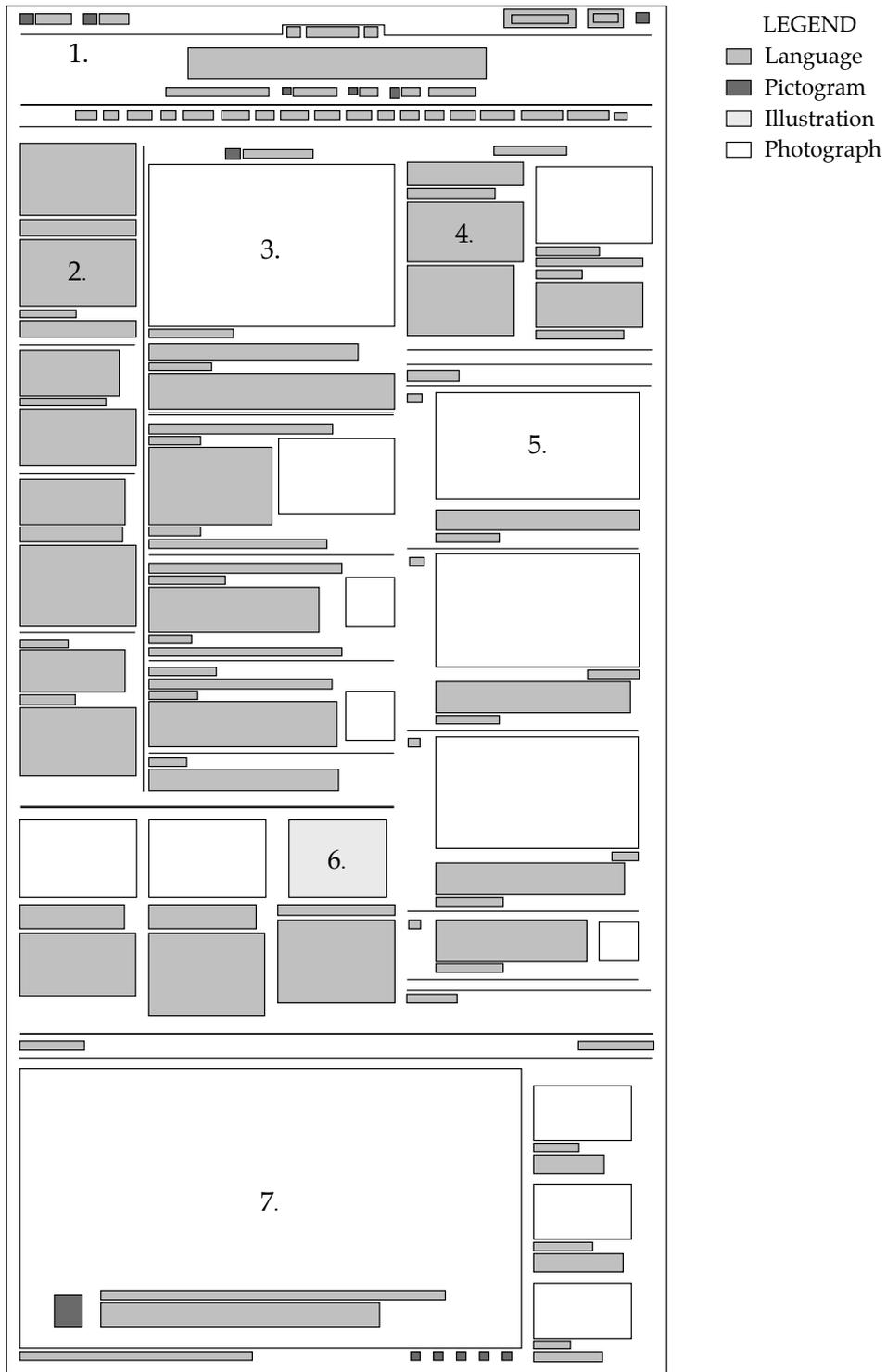
Figure 1: A part of the landing page for the international edition of *The New York Times*, 23 December 2015.

Point (7) exemplifies one such development by superimposing a pictogram – a play button – on a photograph, together with a header and author attribution. Additional pictograms, which consist of social media panels, are positioned below the photograph. This combination resembles a newsbite (Knox, 2007, 26), but the pictogram superimposed on the photograph allows the user to infer that this combination is not only a newsbite, but also a user interface for a video. In terms of multimodal structure, the user interface enables the spatial organization of the landing page to impose control on the time-based organization of the video. While this observation may appear trivial, formulating hypotheses about combinations of semiotic resources is a common task of interpretation demanded from every user interacting with multimodal phenomena (see e.g. Bateman and Wildfeuer, 2014; Wildfeuer, 2014; Hiippala, 2015). Providing these candidate hypotheses is the task of *discourse semantics*, which guide the contextual interpretation of semiotic resources across the landing page. Now, in the light of the stratified model of a semiotic mode proposed by Bateman (2011), the next step is to characterize the semiotic mode at play on the landing page.

Because of the compartmentalized structure of the *NYT* landing page, I argue that the dominant semiotic mode is that of *page-flow*. Page-flow is a *composite* semiotic mode, which means it can incorporate inputs from other semiotic modes and organize them in the layout space. Bateman (2008, 176) accordingly characterizes page-flow as exploiting the "full two-dimensional spatial extent of the page for expressing rhetorical and other functional organizations". That being said, the functional organization of the *NYT* landing page appears to be based on both newsworthiness and subject matter. The news items seem to be organized according to newsworthiness, with a report on upcoming elections occupying the central position (3), accompanied by current events (2). This is a common position of the "news frame", to which the readers attend first on the landing page (Leckner, 2012, 170–171). Further down the news frame, a newsbite (6) presents a recap of the past year's top stories, whose position indicates lower newsworthiness. Editorials and opinion pieces (4), news sourced from other outlets (5) and videos (7) exist outside the news frame, and therefore their position in the layout does not convey newsworthiness. In short, their functional organization is motivated subject matter.

It is important to understand that on the *NYT* landing page, the semiotic mode of page-flow is exploited for the efficient presentation of content. This does not, however, reflect the multimodal structure of the entire news website. In addition to news stories, many other genres may be found beyond the landing page, which include, among others, photo galleries and visual essays, which can present narratives that extend across pages (Caple and Knox, 2012), and more traditional language-driven feature stories (Steensen, 2009). I argue that many of the distin-

guishing features between these genres may be traced back to the selections made in the semiotic modes and their organization in the two-dimensional layout space. This also applies to the longforms, whose structure I now begin to unpack using the concept of "layouting", which allows dismantling their structure.

Discussing the multimodal structure of film, Bateman and Schmidt (2012, 51) define layouting as

> "an abstract mechanism that applies equally to both static and dynamic media and which offers a powerful conceptual tool for setting out quite explicitly the kind of segmentations that film, and other media, regularly employ."

Two key concepts in this quote are media and segmentations. Within this stream of multimodal research, the concept of media (or its singular form medium) refers to a "historically stabilized site for the deployment of some semiotic modes for the achievement of varied communicative purposes" (Bateman, 2014a, 256). Media may be either time- (e.g. film) or space-based (e.g. traditional newspapers, news websites). That being said, I consider the news website a medium, which may carry multiple genres, such as landing pages, news stories, photo galleries and longforms, to name a few. The second concept, segmentation, refers to the configuration of semiotic modes within a medium, which emerges as a result of the layouting process (Bateman and Schmidt, 2012, 53). For modes unfolding in time, such as film, the segmentation may involve organizing shots into scenes and further into more abstract units of narrative, whereas spatial media exploit the layout space organize their content according to some underlying principle, as exemplified by the landing page in Figure 1. As Jacobson, Marino, and Gutsche (2015) observe, the longforms frequently combine aspects of both, which sets a clear requirement: any analytical tool used to interrogate the multimodal structure of the longform needs to be able to handle both time- and page-based organizations.

On these grounds, I argue that the notion of layouting may be put into productive use in the study of digital longform journalism, particularly in tracking how the longform organizes contributions from multiple semiotic modes. This may be exemplified using the opening sequence of 'A Game of Shark and Minnow' (*The New York Times*, 2013), whose first four screens are visualized in Figure 2. 'A Game of Shark and Minnow' reports on the geopolitical tensions in the South China Sea, telling the story of Philippine marines guarding a desolate reef, stationed on a permanently grounded, rusty WWII-era ship. The longform's opening sequence consists of four screens:

1. a looping single-shot video illustrating the journey to the reef,

2. a dynamic, zoom-in map that situates the reef geographically,

3. a still photograph showing the grounded ship,

4. paragraphs of written language comprising an introduction.
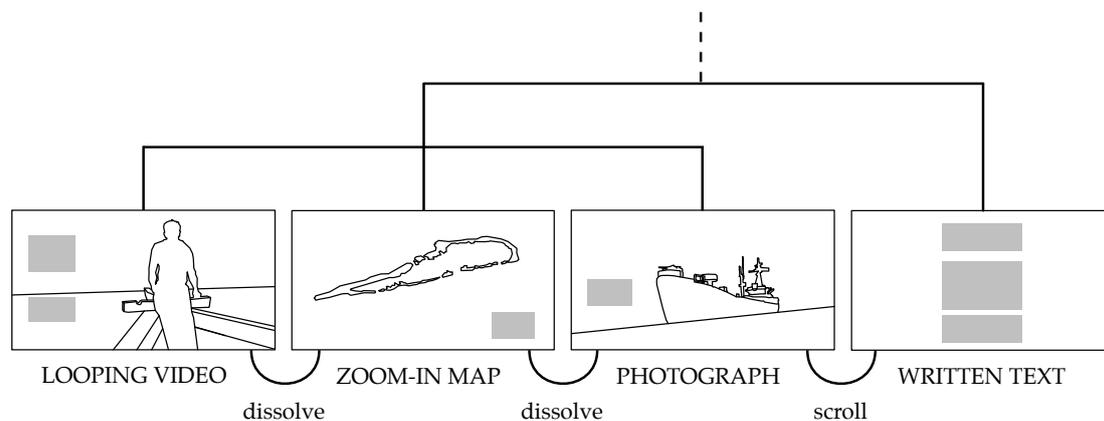

Figure 2: The opening sequence of *A Game of Shark and Minnow*

All four screens feature written language functioning as headers, captions or body text, which are marked using grey boxes in Figure 2. The looping video, the dynamic map and the static photograph (1–3) that precede the longer body of written text (4) are organized into a hierarchy during the layouting process, which is visualized using the tree diagram above the screens (for a more detailed analysis of this sequence, see Hiippala, 2015, 199–201). The hierarchy is established using two distinct transitions, *dissolve* and *scroll*, whose functions I will describe shortly below.

Two aspects of this sequence warrant attention as key features of the longform genre: (1) the use of multiple semiotic resources in each screen and (2) the use of transitions to join them. The screens in Figure 2 include static and dynamic images, written language, diegetic sound and cartography – all of which are carefully orchestrated in order to carry the narrative forward. Making sense of their *combinations* requires the viewer to select an appropriate strategy of interpretation for each screen: a prerequisite for doing so is identifying the active semiotic mode, which provides the discourse semantics for making inferences about the content and its organization (Bateman, 2011, 32). To exemplify, attempting to interpret the opening sequence in Figure 2 by invoking the strategies for making sense of news value in Figure 1 serves to illustrate how different kinds of structures invite different kinds of interpretations. This is an essential component of contemporary media literacy, which is particularly relevant for the longform, in which contributions from distinct semiotic modes and media, to quote Lemke (2005, 51),

10

"synthesize and integrate the older conventions of the printed text-and-image page and the dynamic image-and-sound streams of film, video, and animation."

One such contribution are the dynamic transitions between the screens. As Dowling and Vogan (2015, 214) note, the longform "converges the powerful elements of old media such as print and theater and honors the diverse forms of literacy that mark twenty-first-century audiences". These transitions may be described as an import from the semiotic mode of film, which the longform genre is also able to incorporate, as the underlying material substrate – the screen – allows rendering content instantly for display. In recent years, multimodal researchers have sought to contribute to the already extensive body of research on film by investigating how the filmic mode exploits the available semiotic resources to signal discourse relations between shots (Tseng, 2012; Wildfeuer, 2014). Tseng and Bateman (2012, 92), in particular, have emphasized that the meanings associated with specific transitions are not predefined, but recovered from their immediate context of occurrence, that is, from the preceding and following shots. A dissolve, for instance, may be used to indicate the passing of narrative time in film, yet this is not always the case (cf. Branigan, 1984, 29). Because transitions may be embedded in different contexts, they can also be extended to other media, such as news websites, in which they can have a completely different function.

In digital longform journalism, the transitions seem to mark a switch between different semiotic modes. The dissolves in Figure 2, whereby the pixels of the preceding screen are gradually replaced by those in the following one, appear suitable for transitions between visual content (1–3). For moving into textual content (4), the appropriate choice is a scroll; a traditional transition in the website medium (for transitions and their descriptions, see Table 2). Arguably, the viewers may recognize the dissolve as an import originating in the filmic mode, yet they are unlikely to carry over interpretations commonly associated with this transition in film to the longform genre. This is precisely how the longform can acquire seemingly cinematic qualities (Dowling and Vogan, 2015, 214), which also shows that media convergence is not restricted to content, but also extends to multimodal structures (cf. Jenkins, 2006).

Next, building on the review of previous research on the multimodality of digital journalism and the example analyses presented above, I proceed to present the methods and data of the current study.

3   Methodology

To identify preliminary characteristics of the longform as a multimodal genre, I describe their structure using several abstractions. These include the semiotic *modes* of text-flow, image-flow and page-flow, which describe how the longforms use the

Table 1: The semiotic modes and their definitions

Mode	Description
Text-flow	**Written narrative** consisting of paragraphed written text, which unfolds linearly. Text-flow may be interrupted or accompanied by photographs, diagrams, illustrations, videos and other semiotic resources (see e.g. Hiippala, 2016).
Photography	**Photographs** that take up an entire screen and adhere to the conventions of photojournalism (see e.g. Caple, 2013). They may have other semiotic resources, such as written language, superimposed on them.
Dynamic image-flow	**Video**, consisting of audiovisual moving images organized into one or more shots, which make up an unfolding sequence (see e.g. Bateman and Schmidt, 2012).
Map	**Maps** and other representations of geographic information that occupy the entire screen.
Page-flow	**Layouts** that use the two-dimensional layout space to organize the content into units that work together towards a common communicative goal (see e.g. Bateman, 2008, 176).
Animation	**Animations** consisting of computer-generated images organized into shots, which make up an unfolding sequence.
Dynamic illustration	**Animated illustrations** that occupy the entire screen; either hand-drawn or computer-generated.
Static illustration	**Static illustrations** that occupy the entire screen; either hand-drawn or computer-generated.
Static image-flow	**Comics** and other static images organized into meaningful sequences in the two-dimensional layout space (see e.g. Bateman and Wildfeuer, 2014).

layout space to organize multiple semiotic *resources*, such as written language, pictograms and illustrations (Bateman 2008, 175–176; Hiippala 2015, 53–59). In addition, I treat videos, photographs, maps and animations that occupy the entire screen as individual semiotic modes, in order to achieve a sufficiently fine-grained description of their use in the longform genre. The identified modes and their definitions are provided in Table 1.

Descriptions of semiotic modes present in each longform were stored into a corpus – a digital database – by applying the annotation schema exemplified below, which uses XML mark-up language:

```
<screen identifier="009" mode="text-flow" type="static"
 transition_in="scroll" transition_out="scroll"></screen>
```

The *screen* element identifies each screen in the longform, defining its dominant semiotic **mode**, **type** (static or dynamic) and the transitions that lead **in** and **out** of the screen using the attributes contained within the element. The boundary of a screen is defined as a transition to another semiotic mode, for instance, from paragraphed text (e.g. text-flow) to a full-screen looping video (e.g. dynamic image-flow). These transitions are described in Table 2.

Table 2: Transitions between the semiotic modes

Transition	Description
Click	Transition between screens requires clicking a hyperlink.
Dissolve	The pixels on the previous screen are gradually replaced by the pixels in the following screen.
None	Indicates the beginning or the end of a longform.
Scroll	The previous and following screens exit and enter the display simultaneously. This transition is identical to exploring a layout that does not fit on the screen.
Wipe	The previous screen remains static, as the following screen is drawn on top of it: "the curtain effect" or parallax scroll.
Zoom	A zooming movement either in or out of a screen.

Storing the annotation into an XML-based corpus allowed the description of each longform to be searched for particular combinations of semiotic modes and the transitions that occur between them. The data were collected from various sources listed in Table 3, containing a total of 206 screens from longform articles published between 2012–2013. The URL for each longform is given in Appendix A.

Table 3: The corpus of longform articles

Name	Publisher	Year
A Game of Shark and Minnow	The New York Times	2013
Kowloon Walled City	The Wall Street Journal	2013
Girl in the Shadows: Dasani's Homeless Life	The New York Times	2013
Greenland Melting	Rolling Stone	2013
Snow Fall: The Avalanche at Tunnel Creek	The New York Times	2012
The Dream Boat	The New York Times	2013
The Geeks on the Front Lines	Rolling Stone	2013
The Jockey	The New York Times	2013
The Lobotomy Files: Forgotten Soldiers	The Wall Street Journal	2013
Trials	The Wall Street Journal	2013
Tomato Can Blues	The New York Times	2013
Wrappers Delight	The Telegraph	2013

The corpus was intended to represent how the longform genre was shaped immediately following the publication of 'Snow Fall', which Dowling and Vogan (2015, 212) describe as the 'urtext' of the longform genre, as competing media outlets sought to enter the longform market. This is reflected in the corpus, which consists of longforms produced by news organisations, as opposed to independent publishers such as *The Atavist* or *Long Play* (Le Masurier, 2015, 144). Note also that the corpus describes the longforms as viewed on a laptop or desktop screen: mobile devices may render the content differently.

The annotated corpus is available at the following URL:

http://dx.doi.org/10.5281/zenodo.47729

The tools developed for interrogating the corpus, written in the Python programming language and distributed as interactive Jupyter notebooks (Pérez and Granger, 2007), are provided at the same location. These tools were used for the corpus analysis, whose results are presented in the following section.

4 Analysis

In this section, I describe the multimodal structure of the longform genre using the corpus introduced above. Table 4, which presents the distribution of the modes identified in the corpus, serves as a point of departure for the discussion. I begin by examining the most prominent modes in greater detail, in order to evaluate

their structure and functions in the longform genre, while also considering how the modes are combined.

Table 4: Semiotic modes in the corpus ($n = 206$)

Mode	n	$\%$
Text-flow	86	42
Photography	60	29
Dynamic image-flow	36	18
Map	9	4
Page-flow	7	3
Animation	2	1
Dynamic illustration	2	1
Static illustration	2	1
Static image-flow	2	1

As Table 4 shows, the most common mode in the corpus is text-flow, a foundational semiotic mode organized around continuous written text (Bateman 2008, 175; Hiippala 2016). The organizing principle underlying text-flow is linearity: although photographs, videos, maps and pictograms – among other semiotic resources and modes – may occur within text-flow, this semiotic mode does not use the layout space to make additional meanings, for instance, by using space to convey news value, as demonstrated by the *The New York Times* landing page in Figure 1. Curiously, the semiotic mode active on the landing page, page-flow, appears to be a rare feature of the longform genre (n = 7). This raises the question of why the longform prefers a linear organization over a spatial one: I address this issue in the following discussion.

First of all, a deceptively simple linear organization does not rule out complex interactions between text-flow and other semiotic modes deployed in the longform. To exemplify, the co-deployment of language and image has been theorized using concepts such as 'resemiotization', which traces "how meaning making shifts from context to context, from practice to practice, or from one stage of a practice to the next" (Iedema, 2003, 41). The epitome of the digital longform, *The New York Times*' 'Snow Fall: The Avalanche at Tunnel Creek' (published 2012) aptly exemplifies how other semiotic modes are brought into play alongside text-flow to advance the narrative. In 'Snow Fall', the author John Branch describes how Elyse Saugstad, a professional skier, experienced being caught in an avalanche: "She had no control of her body as she tumbled downhill. She did not know up from down." Branch's written narrative in the third person is accompanied by a

brief video that alternates between medium and close shots, in which Saugstad recounts her experience in the first person:

> I had no ability to control what was happening to me. I was being tossed over and over and over ... it was like being in a washing machine.
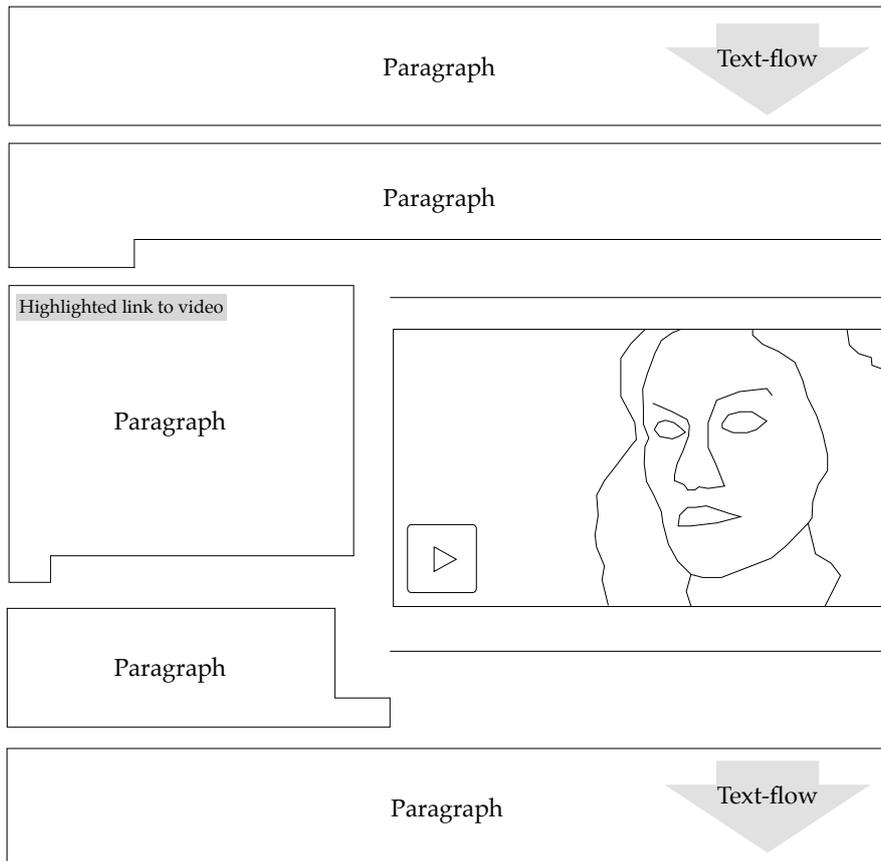

Figure 3: Text-flow in 'Snow Fall'. Paragraphs of written text move the narrative forward: they are accompanied by a short 45-second video in which Saugstad recounts the event.

Essentially, 'Snow Fall' narrates the same event from two different perspectives using two different semiotic modes, as visualized in Figure 3. Branch's description in third person is followed by a resemiotized recount of the event on video, giving a first person voice to Saugstad, as the underlying medium – a news website – allows the longform to integrate dynamic image-flow within text-flow (Bateman, 2014a). Taking advantage of this capability, the narrative resemiotizes a direct quotation typically found in traditional printed newspapers in an audiovisual form. To do so,

16

'Snow Fall' exploits the semiotic resources available within dynamic image-flow for cross-modal visual and verbal identification, as exemplified by continuous shots of Saugstad and the accompanying caption (Tseng, 2013, 43). These resources allow specifying Saugstad's identity and reinforce the interpersonal meanings associated with a personal testimony (Feng and Qi, 2014). Moreover, in yet another part of 'Snow Fall', Branch resemiotizes Saugstad's personal narrative by using a direct quote from the video interview again within text-flow. This shows how the longform can weave together contributions from different semiotic *modes* – text-flow and dynamic image-flow – without drawing explicitly on the additional meaning potential offered by the layout space. This stands in stark contrast to the compartmentalized structure of of the *NYT* landing page in Figure 1, which relies on the layout space to organize the content (Knox, 2014).

Now, considering Dowling and Vogan's (2015) description of the longform as a cognitive container, which seeks to minimize distractions to the reading process, I argue that text-flow represents an appropriate choice for the longform genre for several reasons. Firstly, the semiotic mode of text-flow can incorporate common semiotic modes typically encountered in multimodal news discourse into its linear structure as required (e.g. video, news graphics, photojournalism; cf. Bednarek and Caple, 2012). Secondly, the linear structure of text-flow does not require resolving discourse relations *across* the layout space, allowing the reader to remain focused on the unfolding narrative. It is therefore not surprising that with 86 screens (42%), text-flow is the most common semiotic mode in the corpus. Moreover, text-flow appears to be the hook on which other semiotic modes are hung in the longform genre, as indicated by Table 5, which shows the semiotic modes preceding and following each instance of text-flow in the corpus.

Table 5: Semiotic modes preceding ($n = 86$) and following text-flow ($n = 85$)

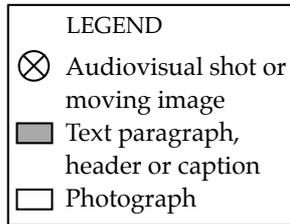| Preceding mode	n	%	Following mode	n	%
Full-screen photograph	45	52,3	Full-screen photograph	42	49,4
Dynamic image-flow	26	30,2	Dynamic image-flow	25	29,4
Map	3	3,4	Map	4	4,7
Text-flow	3	3,4	Page-flow	3	3,5
Animation	2	2,3	Text-flow	3	3,5
Page-flow	2	2,3	Dynamic illustration	2	2,3
Static image-flow	2	2,3	Static illustration	2	2,3
Dynamic illustration	2	2,3	Static image-flow	2	2,3
Static illustration	1	1,1	Animation	2	2,3

Inserting full-screen photographs and videos between instances of text-flow appears to be a choice preferred for the longform genre. Hernandez and Rue (2016, 154), who interview the writers and designers responsible for Snow Fall, provide valuable insights into the reasons behind this preference. Both the head of graphics department and the editor of digital news design agreed that written language is still the dominant form of expression for carrying the narrative forward, but occasionally, the screen is given over to other semiotic modes to deliver visually pleasing and surprising moments. Together with a carefully construed written narrative, these multimedia features are assumed to uphold the reader's interest.

Next, in order to provide a more detailed view of how the most prominent semiotic modes – text-flow, static photographs and dynamic image-flow – are integrated in the longform genre, I apply the analytical tool of layouting to two examples from the corpus (Bateman and Schmidt, 2012, 51). Whereas Figure 2 illustrated the general operation of layouting, that is, how the longform takes contributions from multiple semiotic modes and joins them together using transitions, Figure 4 shows a more fine-grained description of the organization that dissects the introductory phases of 'The Jockey', a portrait of the jockey Russell Baze, and 'Trials', an article about a rare genetic disease and the search for its treatment.
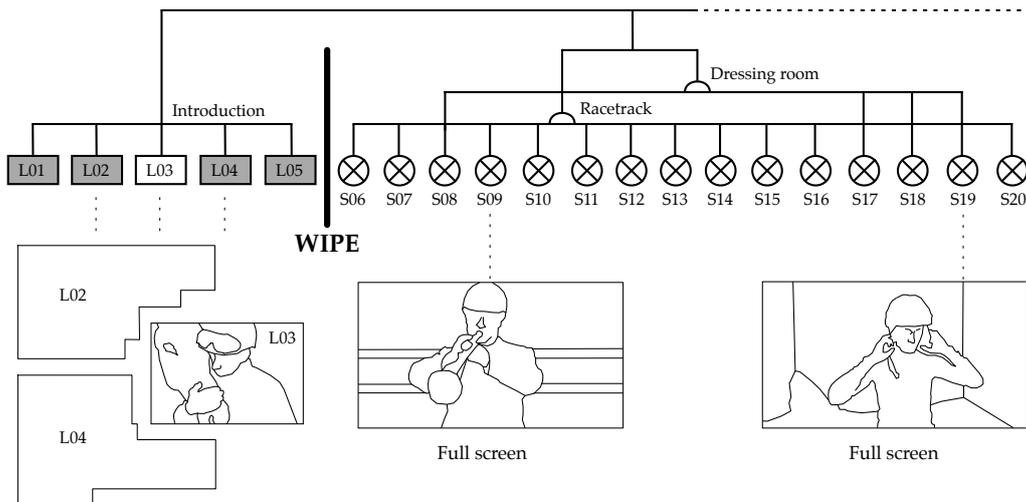
The grey boxes in Figure 4 stand for headers, paragraphs and other elements typically identified in page-based artefacts (Hiippala, 2015, 40–42). The white box indicates a static photograph, whereas the circles with diagonal lines mark individual shots of dynamic image-flow, a basic unit in filmic artefacts (Bateman and Schmidt, 2012, 149). The connecting lines between these elements indicate structural relationships, which form hierarchies of varying depth.

Bateman (2013) proposes that hierarchies are an efficient way of representing the structure of both page-based and filmic artefacts: I will refer to them as layout structures in the subsequent discussion (for layout structures in page-based artefacts, see Bateman, 2008, 116–117). Applying the same principle to represent the structure of all modes involved is a prerequisite for understanding how the longforms exploit the underlying dynamic medium. From a purely theoretical perspective, it is clear that the website medium is fully capable of realizing the semiotic mode of film, as exemplified by the success of various on-demand streaming services. Despite this capability, the longform appears to avoid weaving complex narratives using dynamic image-flow, preferring contributions from the filmic mode that consist of a limited number of shots.

One explanation for avoiding complex narratives may be put forward by examining the layouting process in greater detail. The layout structure for 'The Jockey' in Figure 4 shows a 15-shot sequence of dynamic image-flow (S06–S20), which follows a brief introduction to the article realized using text-flow (L01–L05). This video, which features the jockey Russell Baze and preparations for a horse

LEGEND
⊗ Audiovisual shot or moving image
▢ Text paragraph, header or caption
□ Photograph

THE JOCKEY

Introduction

L01 L02 L03 L04 L05

WIPE

Dressing room
Racetrack

S06 S07 S08 S09 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20

L02
L03
L04

Full screen

Full screen

TRIALS

S01 L02 L03 L04 L05 L06 L07 L08 .... L13 L14 L15 S16 L17 L18 L19 L20

SCROLL

L02

Full screen
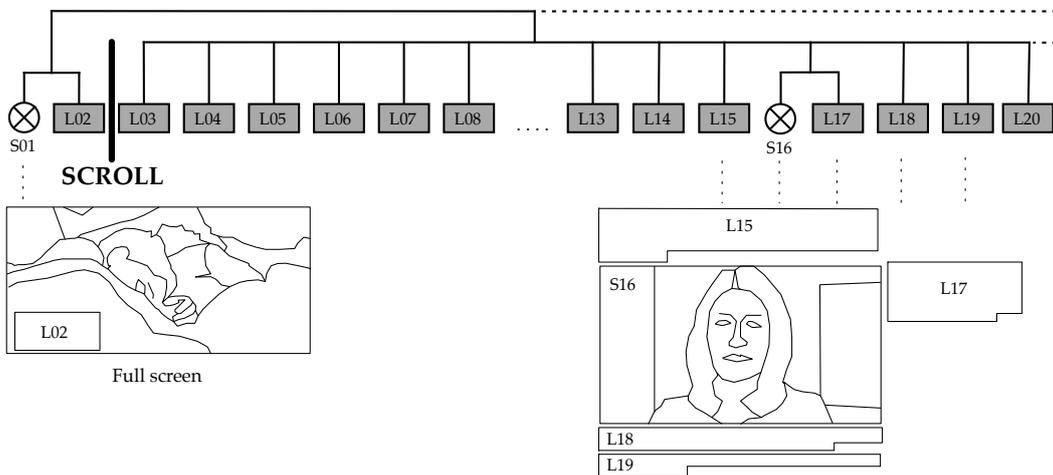
L15
S16   L17
L18
L19

Figure 4: Partial layout structures for 'The Jockey' (The New York Times, 2013) and 'Trials' (The Wall Street Journal, 2013)

19

race, begins to draw increasingly on the semiotic mode of film, which results in additional levels of hierarchy to organize the shots. The shots are organized into two scenes taking place at different locations (labelled 'racetrack' and 'dressing room').

Previous research on film as a semiotic mode has shown that 'proper' filmic narratives can involve very complex layout structures (see e.g. the example analyses in Bateman, 2013, 77–78). Some of the mechanisms involved in the formation of these structures have been outlined for shot/scene organization in Bateman (2007), for cohesion in Tseng (2013) and for discourse interpretation in Wildfeuer (2014). For the example at hand, however, it is sufficient to note that when filmic narratives become increasingly complex, their hierarchical organization becomes deeper as additional shots and scenes are introduced. However, as I will show shortly below, the longform genre does not prefer this kind of deep layout structure. This preference, I argue, reflects a constraint that affects the longform genre, which requires maintaining a shallow layout structure to ensure linearity.

Moving on the next example, it may be observed that the filmic structure of 'The Jockey' stands in stark contrast to how individual shots of dynamic image-flow are embedded into the structure of 'Trials', as shown in the lower part of Figure 4. This example features two single-shot instances of dynamic image-flow: the first instance consists of a full-screen looping video (S01) accompanied by the article header (L02), whereas the second instance is a small single-shot loop (S16), placed within text-flow (L15, L18–19) and accompanied by a caption (L17). What is worth noting here is the generally shallow layout structure, which is a common feature of page-based artefacts, in which text-flow is the dominant semiotic mode, such as an academic journal or a monograph (Hiippala, 2016).

Artefacts with shallow layout structures embed contributions from other semiotic resources by adding an additional level of hierarchy, as indicated by the combinations of image-flow and written language in S01–L02 and S16–L17, whose structure closely resembles the kind of image-text-complexes found in other page-based artefacts (Hiippala, 2015, 65–66). Because 'Trials' does not exploit the semiotic mode of film to the extent of 'The Jockey', the longform is able to retain a shallower layout structure. This represents the simplest structural configuration available to the longform genre, which places special emphasis on maintaining linearity. To put it simply, when shots of image-flow begin to be organized into increasingly complex narratives that feature multiple characters and locations, the hierarchical layout structure becomes deeper.

What, then, is the effect of the shallow layout structure on user experience? I argue that it is the shallow layout structure that establishes the 'cognitive container' described by Dowling and Vogan (2015), which is responsible for "forging a space for linear narrative on the web" (Jacobson, Marino, and Gutsche, 2015,

13). This argument is best explained by contrasting the longform with the landing page shown in Figure 1. As pointed out above, page-flow – the semiotic mode preferred by the landing page genre – takes advantage of the layout space the compartmentalize the content. Establishing these compartments, however, requires a deep hierarchy in the layout structure (cf. Hiippala, 2015, 180). For the landing page, this involves clearly segmenting headers, paragraphs, photographs and other components of 'newsbites' for each news item (Knox, 2007, 2009), before considering how these items should be grouped on the page according to their news value or subject matter (Bateman, Delin, and Henschel, 2007).

The longform genre, in turn, attempts to make sure that the reader does not have to resolve complex layout structures commonly featured in page-flow, as engaging in this kind of interpretative process would likely distract the reader from the narrative. This would work directly against the kind of close reading warranted by the longform, which is precisely how the genre is able to keep the reader focused (Waller, 2012). At the same time, it is important to understand that unlike older media such as books, the news website medium does not have a tradition of delivering long narratives, although a trend of 'slow journalism' appears to be emerging at the moment (Le Masurier, 2015; Belt and South, 2015). Such developments, however, do not take place in a vacuum, and 'cross-pollination' between media occur frequently.
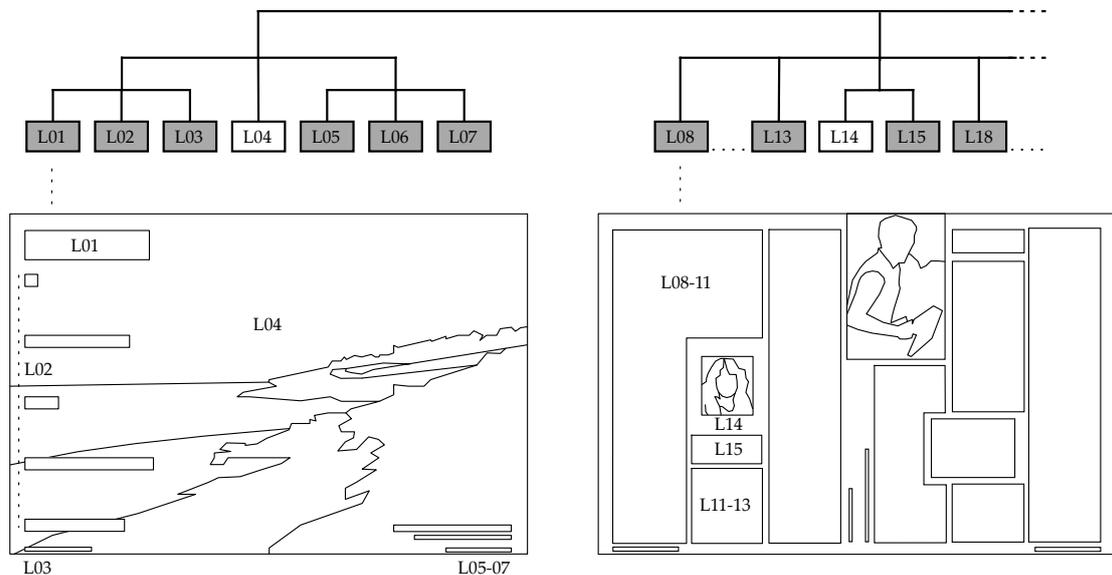

Figure 5: The layout structures of two pages in a feature story published in *The Atlantic*, October 2015

21

As said, genres that populate new media rarely emerge out of nowhere: in terms of multimodal structure, I argue that certain predecessors to the longform genre may be identified in print media. Monthly magazines such as *The Atlantic* frequently use photographs that fill an entire double-page spread to introduce longer feature articles, whose main narrative is delivered using text-flow (see Figure 5 for an example). In addition to a shallow layout structure, what connects the printed feature article and the digital longform is that both work with a 'page metaphor' (Bateman, 2008, 9), that is, they must operate with a two-dimensional space for deploying and configuring semiotic modes. Both also dedicate most of the layout space to presenting the content, as opposed to lending the real estate to navigation or advertising (cf. Bateman, Delin, and Henschel, 2007). This is something the traditional news story can rarely afford in print or digital media.

In the end, however, the distinguishing features of the longform genre appear to arise from the materiality of screen and its capability to render dynamic content. In addition to replacing static photographs with dynamic image-flow, the screen sets the longform free from most restrictions affecting the length of the article. Whereas the pages available to a feature article in a magazine are nearly always limited, the digital longform can repeat sequences of visual semiotic modes and text-flow as long as the longform is able to hold the reader's attention and interest. To do so, the designers introduce transitions borrowed from the filmic medium, which they assume to create a sense of wonder and encourage the reader to stay with the narrative (cf. Hernandez and Rue, 2016, 154)

Therefore, I now continue by examining the transitions in the longform genre from two different perspectives. First, I consider the transitions leading in and out of a screen, placing the semiotic mode at the centre of attention. I then shift the perspective by zooming in on the transitions, while focusing on the semiotic modes surrounding them. The overall distribution of transitions in the corpus is given in Table 6. The transitions are listed in pairs, providing the transitions leading in and out of a screen. The transitions at the beginning and the end of a longform have been marked as 'none'.

Notably, the most common transition pattern involves entering and leaving the screen by scrolling ($n = 43$), followed by the wipe–wipe transition ($n = 33$), which indicates the use of parallax scroll. Jacobson, Marino, and Gutsche (2015, 14) also pay attention to these two transitions, suggesting that the longforms use the traditional scroll to mimic a parallax scroll "by placing large video loops or photographs between sections of the story rather than creating a parallax transition." However, the corpus does not support the proposal that the parallax scroll constitutes a standard transition in the longform genre. Moreover, the inclusion of full-screen photographs and dynamic image-flow appears to be motivated by the needs of the narrative, not the creation of a pseudo parallax scroll. As Table 6

Table 6: The distribution of transition patterns in the corpus ($n = 206$)

Pattern	n	Pattern	n	Pattern	n
Scroll–scroll	43	Click–wipe	5	Scroll–wipe	2
Wipe–wipe	33	None–wipe	4	None–click	1
Dissolve–scroll	21	None–scroll	4	Wipe–scroll	1
Scroll–click	16	Zoom–zoom	3	Dissolve–none	1
Click–scroll	13	Click–click	3	Dissolve–wipe	1
Dissolve–dissolve	13	Zoom–scroll	2	Dissolve–click	1
Scroll–dissolve	13	Wipe–none	2	None–zoom	1
Wipe–click	10	Click–dissolve	2	Click–none	1
Scroll–none	7	None–dissolve	2	Scroll–zoom	1

shows, the longform can choose between many alternatives, of which the parallax scroll is just one.

It is also worth noting that in 46 percent of the cases ($n = 95$), the transition patterns are symmetric, that is, the same transition is used to enter and exit the screen. As the annotation scheme would lead to expect, there are five unique symmetric transitions: scroll, wipe, dissolve, click and zoom. However, 54 percent ($n = 111$) of the transitions are asymmetric, which form 22 unique combinations. The prominent role of scroll and wipe transitions in Table 6 may result from the fact that they are highly compatible with many different semiotic modes, which allows them to be used regardless of the modes participating in the transition. This claim is supported by Table 7, which lists the ten most common transitions between specific semiotic modes in the corpus: I will refer to them as mode-transition patterns. Each row in Table 7 lists the transition from one semiotic mode to another, while also providing the modes and the transition that holds between them.

In total, the corpus includes 194 mode-transition patterns, which feature 44 unique instances, the most common being a scroll transition from a full-screen photograph to text-flow with 29 instances (for an example, see the sequence presented in Figure 2). Scroll is also the most prominent transition when moving from text-flow to full-screen photography ($n = 19$) and from dynamic image-flow to text flow ($n = 15$). Given that these three semiotic modes – text-flow, full-screen photography and dynamic image-flow – are the most common semiotic modes in the corpus (see Table 4), it is worth noting that these semiotic modes prefer a scroll transition, which is arguably one of the most ordinary ways of traversing the genres carried by the website medium.

23

Table 7: Ten most common mode-transition patterns ($n = 122$; % of the entire corpus)

From	To	Via	n	%
Full-screen photograph	Text-flow	Scroll	29	14,9
Text-flow	Full-screen photograph	Scroll	19	9,7
Dynamic image-flow	Text-flow	Scroll	15	7,7
Full-screen photograph	Text-flow	Wipe	10	5,2
Text-flow	Dynamic image-flow	Click	10	5,2
Dynamic image-flow	Text-flow	Wipe	9	4,6
Full-screen photograph	Full-screen photograph	Dissolve	8	4,1
Text-flow	Full-screen photograph	Click	8	4,1
Text-flow	Full-screen photograph	Wipe	7	3,6
Text-flow	Dynamic image-flow	Scroll	7	3,6

Other prominent transitions in Table 7 include a wipe – the so-called parallax scroll – and a click. In the light of the corpus, it appears that a wipe transition with a parallax scroll is not the most common choice for transitions between semiotic modes in the longform genre, but simply constitutes one alternative among traditional scroll and click transitions. Not surprisingly, click transitions can be found in the longforms divided clearly into chapters, which use clicks to signpost the narrative. Although some click transitions may be found in the narratives centered around a location or an event (e.g. 'Snow Fall', 'The Jockey'), they are particularly prominent in human interest stories, such as 'Girl in the Shadows' and 'Trials'.

Of particular interest is the dissolve transition, in which the preceding screen is gradually replaced by the following one, pixel by pixel. I have suggested elsewhere that this transition is an import from the semiotic mode of film, which can be deployed by any artefact realized in a dynamic medium, such as the news website (Hiippala, 2015, 205). Table 7 shows that among the most prominent transitions, dissolves occur only between photographs. This reflects another genre constraint, which is revealed by examining the number of dissolve transitions leading in or out of photographs, image-flow and maps, whose total number adds up to 33. At the same time, a dissolve transition is used to enter a screen with text-flow in only 8 instances. For this reason, I argue that the use of a dissolve is constrained to some extent by the surrounding semiotic modes, because a pixel by pixel transition does not render written language fully legible until the transition is nearly complete (Hiippala, 2015, 206). Visual semiotic modes, which are perceived differently than linear written language, appear to be exempted from this constraint.

Finally, I wish to draw attention to the overall distribution of unique mode-transition patterns ($n = 44$) in the corpus. Some longforms, such as 'A Game of Shark and Minnow' ($n = 13$) and 'Snow Fall' ($n = 11$), feature a high number of unique mode-transition patterns ($M = 5.6$, $SD = 3.4$). This suggests that the transitions can be deployed flexibly, and given the success of these articles, introducing this kind of variation does not obviously pose a challenge for the reader. However, determining whether the transitions are used to convey or reinforce a specific discourse relation between the content in the preceding and following screens can only be achieved through a more detailed analysis (cf. e.g. Hiippala, 2015, 199–201). Although this falls outside the scope of the current study, pursuing this line of investigation could help to establish whether the transitions are motivated by the semiotic modes selected for presenting the content or the narrative structure of the story.

This section concludes the analysis, which has attempted to describe some of the multimodal features that characterize the longform genre. In the following section, I proceed to consider the implications of the study, and circumscribe potential areas of further research.

5   Conclusion

In this article, I described the multimodal structure of 12 longform articles published in 2012–2013, in order to characterize the emerging genre of digital longform journalism. My aim was to explicate how the longforms combine written language, photojournalism, short videos and other semiotic modes, enabling them to realize a seamless narrative and consequently, to provide a user experience that stands in stark contrast to that delivered by other journalistic genres, such as landing pages, photo galleries and news stories.

Working with abstractions such as text-flow, image-flow and page-flow to capture how written language, photographs, videos and other types of content are combined, the analysis revealed that the longform genre prefers to organize the content into a linear structure. This means that unlike landing pages, which exploit the layout space to simultaneously present different types of content to the reader, the longforms dedicate the entire screen to a single semiotic mode at a time. The longforms join this content together using a number of novel transitions, of which two in particular – wipe and dissolve – appear to have been borrowed from film.

In terms of limitations, it is important to acknowledge that this article provided a rather broad description of the longform genre, but at the same time, this description can be used to contextualize future investigations. In particular, the scope of the corpus compiled for this study could be expanded as necessary,

because XML mark-up language is inherently extensible. The annotation could be extended, for instance, to cover more detailed multimodal analyses of content, which could also help to contextualize the use of different modes and transitions. Extending the corpus would also help to consolidate the broad description provided by this study with the previous analyses presented in Dowling and Vogan (2015) and Jacobson, Marino, and Gutsche (2015).

More generally, theories of multimodality appear to complement those currently used for studying digital journalism well (cf. Steensen and Ahva, 2015). As this article has attempted to show, adopting a multimodal perspective may provide additional insights into phenomena that have been previously taken for granted in discussions of multimedia narratives. Moreover, given the ongoing processes of digitalization, theories and methods geared towards describing the inherently multimodal nature of communication may come in handy, as media outlets continue to develop their journalistic products. Such analyses, however, cannot be pursued without a knowledge of journalistic genres in old media, which may end up being recontextualized in new media. This clearly defines a domain which research on digital journalism and multimodality should explore together.

References

Barnhurst, Kevin G., 2012. "The form of online news in the mainstream US press, 2001–2010." *Journalism Studies* 13 (5–6): 791–800. doi: 10.1080/1461670X.2012.664346.

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

Bateman, John A., 2008. *Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents.* London: Palgrave Macmillan.

Bateman, John A., 2011. "The Decomposability of Semiotic Modes." In "Multimodal Studies: Multiple Approaches and Domains," , edited by Kay L. O'Halloran, and Bradley A. Smith. London: Routledge, 17–38.

Bateman, John A., 2013. "Multimodal analysis of film within the GeM framework." *Ilha do Desterro* 64: 49–84. doi:10.5007/2175-8026.2013n64p49.

Bateman, John A., 2014a. "Genre in the age of multimodality: some conceptual refinements for practical analysis." In "Evolution in Genre: Emergence, Variation, Multimodality," , edited by Paola Evangelisti-Allori, Vijay K. Bhatia, and John A. Bateman. Frankfurt am Main: Peter Lang, 237–269.

Bateman, John A., 2014b. *Text and Image: A Critical Introduction to the Visual/Verbal Divide.* London and New York: Routledge.

Bateman, John A., Delin, Judy L., and Henschel, Renate, 2007. "Mapping the multimodal genres of traditional and electronic newspapers." In "New Directions

in the Analysis of Multimodal Discourse," , edited by Terry D. Royce, and Wendy L. Bowcher. Mahwah, NJ: Lawrence Erlbaum, 147–172.

Bateman, John A., and Schmidt, Karl-Heinrich, 2012. *Multimodal Film Analysis: How Films Mean.* London: Routledge.

Bateman, John A., and Wildfeuer, Janina, 2014. "A multimodal discourse theory of visual narrative." *Journal of Pragmatics* 74: 180–208. doi: 10.1016/j.pragma.2014.10.001.

Bednarek, Monika, and Caple, Helen, 2012. *News Discourse.* London: Continuum.

Belt, Don, and South, Jeff, 2015. "Slow Journalism and the Out of Eden Walk." *Digital Journalism* doi:10.1080/21670811.2015.1111768.

Branigan, Edward, 1984. *Point of View in the Cinema: A Theory of Narration and Subjectivity in Classical Film.* Berlin: Mouton de Gruyter.

Caple, Helen, 2013. *Photojournalism: A Social Semiotic Approach.* London: Palgrave Macmillan.

Caple, Helen, and Bednarek, Monika, 2015. "Rethinking news values: What a discursive approach can tell us about the construction of news discourse and news photography." *Journalism.* doi:10.1177/1464884914568078.

Caple, Helen, and Knox, John S., 2012. "Online news galleries, photojournalism and the photo essay." *Visual Communication* 11 (2): 207–236.

Constantinou, Odysseas, 2005. "Multimodal Discourse Analysis: Media, modes and technologies." *Journal of Sociolinguistics* 9 (4): 602–618.

Deuze, Mark, 2003. "The web and its journalisms: considering the consequences of different types of newsmedia online." *New Media & Society* 5 (2): 203–230.

Dowling, David, and Vogan, Travis, 2015. "Can We "Snowfall" This? Digital longform and the race for the tablet market." *Digital Journalism* 3 (2): 209–224. doi:10.1080/21670811.2014.930250.

Elleström, Lars, 2010. "The Modalities of Media: A Model for Understanding Intermedial Relations." In "Media Borders, Multimodality and Intermediality," , edited by Lars Elleström. London: Palgrave, 11–48.

Feng, Dezheng, and Qi, Yujie, 2014. "Emotion prosody and viewer engagement in film narrative: A social semiotic approach." *Narrative Inquiry* 24 (2): 347–367. doi:10.1075/ni.24.2.09fen.

Hernandez, Richard, and Rue, Jeremy, 2016. *The Principles of Multimedia Journalism: Packaging Digital News.* New York and London: Routledge.

Hiippala, Tuomo, 2015. *The Structure of Multimodal Documents: An Empirical Approach.* New York and London: Routledge.

Hiippala, Tuomo, 2016. "Aspects of multimodality in higher education monographs." In "Multimodality in Higher Education," , edited by Arlene Archer, and Esther Breuer. Leiden: Brill, 53–78.

Iedema, Rick, 2003. "Multimodality, resemiotization: extending the analysis of discourse as multi-semiotic practice." *Visual Communication* 2 (1): 29–57.

Jacobson, Susan, Marino, Jacqueline, and Gutsche, Robert E., 2015. "The digital animation of literary journalism." *Journalism.* doi:10.1177/1464884914568079.

Jenkins, Henry, 2006. *Convergence Culture: Where Old and New Media Collide.* New York and London: New York University Press.

Jewitt, Carey, ed., 2014. *The Routledge Handbook of Multimodal Analysis.* London: Routledge, 2nd ed.

Knox, John S., 2007. "Visual-verbal communication on online newspaper home pages." *Visual Communication* 6 (1): 19–53.

Knox, John S., 2009. "Punctuating the home page: image as language in an online newspaper." *Discourse and Communication* 3 (2): 145–172.

Knox, John S., 2014. "Online newspapers: Structure and layout." In "The Routledge Handbook of Multimodal Analysis," , edited by Carey Jewitt. London and New York: Routledge, 2nd ed., 440–449.

Kress, Gunther, 2010. *Multimodality: A Social Semiotic Approach to Contemporary Communication.* London: Routledge.

Kress, Gunther, 2014. "What is mode?" In "The Routledge Handbook of Multimodal Analysis," , edited by Carey Jewitt. London: Routledge, 2nd ed., 60–75.

Le Masurier, Megan, 2015. "What is slow journalism?" *Journalism Practice* 9 (2): 138–152. doi:10.1080/17512786.2014.916471.

Leckner, Sara, 2012. "Presentation factors affecting reading behaviour in readers of newspaper media: an eye-tracking perspective." *Visual Communication* 11 (2): 163–184.

Lemke, Jay L., 2005. "Multimedia Genres and Traversals." *Folia Linguistica* XXXIX (1-2): 45–56.

Norris, Sigrid, and Maier, Carmen D., eds., 2014. *Interactions, Images and Texts: A Reader in Multimodality.* Berlin: De Gruyter Mouton.

Pérez, Fernando, and Granger, Brian E., 2007. "IPython: A system for interactive scientific computing." *Computing in Science and Engineering* 9 (3): 21–29. doi:10.1109/MCSE.2007.53.

Steensen, Steen, 2009. "Online feature journalism." *Journalism Practice* 3 (1): 13–29. doi:10.1080/17512780802560716.

Steensen, Steen, and Ahva, Laura, 2015. "Theories of Journalism in a Digital age: An exploration and introduction." *Digital Journalism* 3 (1): 1–18.

Tseng, Chiao-I, 2012. "Audiovisual texture in scene transition." *Semiotica* 192: 123–160. doi:10.1515/sem-2012-0091.

Tseng, Chiao-I, 2013. *Cohesion in Film: Tracking Film Elements.* London: Palgrave Macmillan.

Tseng, Chiao-I, and Bateman, John A., 2012. "Multimodal narrative construction in Christopher Nolan's *Memento*: a description of analytic method." *Visual Communication* 11 (1): 91–119.

van Leeuwen, Theo, 2005. "Multimodality, genre and design." In "Discourse in Action: Introducing Mediated Discourse Analysis," , edited by Sigrid Norris, and Rodney H. Jones. London: Routledge, 73–94.

Waller, Robert H. W., 2012. "Graphic literacies for a digital age: The survival of layout." *The Information Society* 28 (4): 236–252.

Waller, Robert H. W., Delin, Judy, and Thomas, Martin, 2012. "Towards a pattern language approach to document description." *Discours* 10. doi: 10.4000/discours.8673.

Wildfeuer, Janina, 2014. *Film Discourse Interpretation: Towards a New Paradigm for Multimodal Film Analysis.* London: Routledge.

Appendix A

Name	URL
A Game of Shark and Minnow	http://www.nytimes.com/newsgraphics/2013/10/27/south-china-sea/
Kowloon Walled City	http://projects.wsj.com/kwc/
Girl in the Shadows: Dasani's Homeless Life	http://www.nytimes.com/projects/2013/invisible-child/
Greenland Melting	http://www.rollingstone.com/feature/greenland-melting
Snow Fall: The Avalanche at Tunnel Creek	http://www.nytimes.com/projects/2012/snow-fall
The Dream Boat	http://www.nytimes.com/2013/11/17/magazine/the-impossible-refugee-boat-lift-to-christmas-island.html
The Geeks on the Front Lines	http://www.rollingstone.com/feature/the-geeks-on-the-frontlines
The Jockey	http://www.nytimes.com/projects/2013/the-jockey
The Lobotomy Files: Forgotten Soldiers	http://projects.wsj.com/lobotomyfiles/
Trials	http://projects.wsj.com/trials/
Tomato Can Blues	http://www.nytimes.com/projects/2013/tomato-can-blues/
Wrappers Delight	http://s.telegraph.co.uk/graphics/tunnocks