9 Multimodal artefacts in digital media

The ninth chapter tests the framework developed in this book in a different setting by exploring multimodal artefacts in digital media. Obviously, a framework developed for the empirical study of multimodal artefacts cannot neglect digital media, given its growing influence in everyday life. Digital media, such as websites and social networking services, are consumed on a daily basis using a variety of devices brought about by technological development: desktop computers, laptops, and handheld devices, such as smartphones and tablets. In short, if the earlier technological advances revolutionised the production of multimodal artefacts, it may be argued that the mobile devices have done the same for their consumption.

These technological developments and their possible semiotic consequences have been naturally covered within fields of research too numerous and wide to be covered here comprehensively. It is, however, possible to sketch the relevant work in general terms. Whereas communication studies have described the technological developments and their impact from a broad perspective, particularly in terms of 'mediation' (see e.g. Crowley, 2013), multimodal research has focused on the more detailed phenomena, such as text-image relations and their patterns (Martinec and Salway, 2005; Knox, 2007, 2009), how digital media demands the user to cope with rapid transitions between different genres (Lemke, 2005), and what kinds of changes occur within genres as they cross over from print to digital (Bateman et al., 2007). These and many other phenomena explored within the field of multimodal research have been outlined in Djonov (2012) and O'Halloran and Smith (2012). To sum up, digital media can be highly dynamic, and therefore its analysis requires a robust framework to keep track of the multimodal goings-on. The previous chapters showed the benefits of adopting this kind of approach to static print media, and the same analytical principles may be proposed to apply to digital media.

Nevertheless, because digital artefacts related to tourism, such as tourism websites, have been described comprehensively elsewhere (see e.g. Hallett and Kaplan-Weinger, 2010; Nekic, 2014), the current chapter turns away from the tourist brochures for a moment. Instead, this chapter focuses on a very specific form of online communication: *digital longform journalism*, which takes advantage of the website medium and its dynamic capabilities, using written language, full-screen images, film, animation,

and most importantly, subtle transitions to organise the presentation of content. As the name suggests, longform journalism is concerned with in-depth storytelling, exposing topics that would not otherwise be covered by the news (Le Masurier, 2014). To do so, longform journalism combines multiple modes in a different way than traditional online news by taking advantage of the transitions, which may be suggested to be a borrowing from film. This chapter will show that these transitions play an important part in combining the various types of content into a seamless whole for the viewer.

The *New York Times*' Pulitzer Prize -winning publication, 'Snow Fall: The Avalanche at Tunnel Creek' by John Branch, has been heralded as a prime example of digital longform journalism. These kinds of lengthy multimodal narratives are mainly intended to be consumed using hand-held tablet computers (Dowling and Vogan, 2014). Since the publication of 'Snow Fall', the concept has been copied in various publications, leading Dowling and Vogan (2014: 2) to suggest that the digital longform also constitutes an emergent genre in journalism. The notion of genre, of course, brings us close to the interests of the current book: this chapter thus explores the structure of one artefact belonging to the proposed genre – a *New York Times* longform article entitled 'A Game of Shark and Minnow'. The chapter will exemplify how the framework developed in this book can bring out multimodally interesting aspects of state-of-the-art designs in digital media, while simultaneously underlining the benefits of having a carefully constructed theoretical framework at hand.

Now, to briefly introduce the artefact, 'A Game of Shark and Minnow' is a piece of investigative journalism, realised using the digital longform genre. The article attempts to capture the geopolitical dynamics of the Spratly Islands, a group of islands and reefs located in a remote corner of the South China Sea, whose proximity to major shipping lanes and natural resources lends strategic value to the otherwise barren islands. Consequently, the Philippines, China, Taiwan, and Vietnam all lay claim and occupy or maintain presence on some of the islands and reefs. To make this remote geopolitical struggle vivid to the Western consumer, the artefact combines text, static and looping photographs, short videos, and maps into a unified, multimodal narrative. Given the abundance of dynamic content in the longform artefact and the constraints set by the book medium for this research monograph, only limited attempts will be made to reproduce the content of the longform article under analysis in this chapter. The reader is therefore strongly encouraged to examine 'A Game of Shark and Minnow' while reading this chapter: a link to the longform article may be found in the bibliography under the citation for Himmelman, Gilbertson, Bostock, Cargill, Carter, Donaldson, Giratikanon, Xaguín, Maing and Watkins (2013).

9.1 Technology, production, and consumption

Before proceeding to examine a multimodal artefact intended to be consumed via mobile devices, it is useful to consider certain aspects of production and consumption arising from the new consumer technologies. The impact of digital and mobile technologies on multimodal communication has been well documented in previous research (for a concise overview of this work, see Djonov, 2012). For multimodal research, the major questions have been naturally related to if and how technological development has affected fundamental semiotic processes, such as how we make meaning and how we interact with other humans and the ubiquitous multimodal artefacts around us.

White (2010, 2012), for instance, has explored how mobile devices enable old and new technologies and media to converge, showing how a Bluetooth device attached to a traditional printed billboard advertisement may be used to engage passers-by. The Bluetooth-equipped advertisement sends a message to the passer-by's device, challenging the device owner to locate specific information in the billboard advertisement and to respond by submitting an answer via the mobile device. In this context, the mobile device shifts the pattern of engagement radically: a traditional poster – such as the famous 'Lord Kitchener Wants You' poster discussed in White (2010) – would attempt to capture the viewer's attention using a combination of semiotic resources. Here, however, the pattern is turned around: it is the user that engages voluntarily with the poster, prompted by receiving a message on the mobile device. The attention is thus distributed between two media: the mobile screen and the printed billboard.

To improve our understanding of how these kinds of multimodal interactions involving mobile devices unfold in different situations, Leander and Vasudevan (2014: 152) have proposed that multimodal research must be "unmoored" from texts and artefacts – the traditional objects of analysis within the field. Taking the goals of this book into account, this proposal naturally warrants further consideration. How mobile devices enable and facilitate multimodal interaction, for instance, using social networking services, is undoubtedly a timely topic for many fields ranging from economics to ethnology. For multimodal research in general and this book in particular, the important question is whether this kind of interaction may be described efficiently using approaches that attend primarily to the structure of multimodal artefacts. Bateman (forthcoming: 16–17), for instance, has proposed that interactions revolving around dynamic, user-generated content are likely to be described more accurately using approaches that take interaction and participation as their point of departure (Eisenlauer, 2013; Villi, forthcoming).

However, for studying multimodal artefacts produced for mass consumption using mobile devices, as opposed to facilitating interaction and participation, the theoretical concerns and methodological solutions of

the approaches focusing on artefact structure remain largely valid. Not surprisingly, from this perspective, the distinctions begin to emerge already during the fundamental process of production. Unlike most of the user-generated content, the artefact examined in this chapter – 'A Game of Shark and Minnow' – is a product of professional teamwork involving multiple roles, as the list of names in the citation of the longform article shows. Section 2.4 proposed that the production of a multimodal artefact, and particularly the roles involved in the process, can be examined efficiently using a value chain. For 'A Game of Shark and Minnow', however, the precise roles are obscured apart from the writer (Himmelman) and the photo- and videographer (Gilbertson). The rest are credited as producers, but other roles involved in creating the artefact are likely to include editors, information and graphic designers, and programmers. In short, the resulting artefact is a product of joint effort, which involves negotiating choices related to content, its realisation using different semiotic modes, and particularly, in planning the user interface and navigation (Tan, 2010; E., O'Halloran and Judd, 2011*b*).

As a team effort involving several specialised roles, longform journalism requires both time and resources. Compared to presenting "hard news" online, which often follows well-established patterns set out for the genre, for instance, in Knox (2014), the structure of longform journalism is far less predictable. Whereas hard news are often organised into a clearly compartmentalised and segmented layout, longform journalism benefits from having the entire layout space available for use. In this aspect, the proposal of Dowling and Vogan (2014: 4) that longform artefacts constitute "(cognitive) containers" is adept: unlike hard news, the longform does not have to compete for the user's attention with other content on the page, nor is it subject to the constraints arising from "news value", which determine the position of the content in the hard news genre (Bateman et al., 2007). In order to hold the user's attention, the longform is contained from other content on the website. This containment removes many of the constraints that determine what may be done with the layout space in terms of design, which represents a rare situation, because the layout space is a valuable commodity. The news outlets typically distribute their content over several sub-pages to maximise the number of advertisements exposed to the user.

However, as Dowling and Vogan (2014: 2) point out, longform journalism follows a different business model. For the publisher, longform constitutes a "signature product", which helps to drive to traffic to the publisher's main site, possibly leading to visits to other content on the site or to a subscription to premium content behind a paywall. To put it simply, the longform's goal is to generate traffic from social media as the artefact is interacted with, shared, and discussed in social networking services (Boyd and Ellison, 2007). The longform is, in short, a genre to be talked about. Not surprisingly, the genre makes frequent use of social web panels – a feature of the underlying website medium now increasingly common for

many other genres as well – which facilitate sharing the artefact (Bateman, forthcoming: 16). Considering that within a week of its publication, 'Snow Fall' had attracted 3.5 million page views that mainly originated from social networking services, combined with the average user viewing the article for 12 minutes, the business model and strategy appears to hold much promise (Romenesko, 2012), at least as long as the longform genre remains a novelty and retains the status of a signature product.

In terms of the users' attention span, 12 minutes is an eternity for online news. This also raises an interesting question for multimodal research, because the users are likely to be very familiar with written feature journalism, news photography, short videos, and animated maps. They are, in short, common features of ordinary online news. What, then, are the characteristics of digital longform journalism that captivate the users? Obviously, the longform genre is able to elicit discussion and sharing in the social media using the same means and resources for communication available to anyone working with the website medium. This observation stands in contrast with the views presented by Leander and Vasudevan (2014) above, who proposed that to truly appreciate mobile culture, multimodal research should divert its attention from the artefacts. Contrastingly, in the case of longform journalism, the mobile culture and the multimodal artefacts are so strongly intertwined that in order to understand the interaction between individuals, the artefact at the centre of attention cannot be neglected, because it is this artefact that generates the interaction. To explore the reasons why the longform is able to do so, the following discussion proceeds to dissect 'A Game of Shark and Minnow', contrasting its structure with that of traditional online news and other artefacts with dynamic content.

9.2 Semiotic modes in digital longform journalism

Like the tourist brochures studied in the previous chapters of this book, 'A Game of Shark and Minnow' is a multimodal artefact that works with a "page metaphor" (Bateman, 2008: 8). What this means is that the content is organised along the horizontal and vertical dimensions. For page-based media, these dimensions remain foundational, regardless of whether the underlying medium is a paper page or a tablet screen. The two dimensions provide the space on the page, in which multiple semiotic modes may be integrated. Due to their capability to integrate linguistic and visual contributions, Djonov and Knox (2014: 171) have proposed that webpages can be treated as "visual units."

To some extent, this reflects the view presented in Chapter 2, which suggested that the page warrants increased attention as a site of integration, in which multiple semiotic modes combine and work together towards a shared communicative goal. The analyses in the subsequent chapters, in

turn, showed that several overarching structures can indeed be identified on the page, which are geared towards supporting the kind of communicative work involving multiple semiotic modes. To draw on an example, the rhetorical and layout structures were considered responsible for joining the different pieces of content into larger units on the page (Hiippala, 2013). However, whether the epithet 'visual' is appropriate for the webpage as a unit of analysis is another question, as several scholars have pointed out the problems of using the verbal/visual dichotomy to build distinctions between multimodal artefacts or their parts (see e.g. Stöckl, 2004; Forceville, 2014).

Establishing the necessary distinctions, however, are precisely those required for making sense of 'A Game of Shark and Minnow' as a multimodal artefact. To clarify the terminology, the basic unit that works with a page metaphor is hereby termed a 'screen' in the following discussion. The longform genre relies on sequences of screens to establish a narrative structure, and the individual screens within this sequence combine multiple semiotic modes and resources. To establish how the individual pages contribute to the overall narrative realised in the longform artefact, the analytical tools need to be sharpened considerably, with the goal of advancing beyond the verbal/visual divide. In plain words, we need to be able to describe the screens in other terms than simply referring to language and images. The previous chapters of this book showed how the definition of a semiotic mode proposed in Bateman (2011), in combination with the *Genre and Multimodality* (GeM) model presented in Bateman (2008), could be used for a detailed description of the tourist brochures, while also relating the observations made in individual artefacts to a more abstract level, painting a picture of the tourist brochures as a genre. The same approach will be used in this chapter, albeit on a more general level and without performing an extensive corpus analysis. Moreover, for describing the dynamic content in 'A Game of Shark and Minnow', the framework will leverage the extension of the GeM model to examine cinematographic artefacts in the film medium (Bateman, 2013), together with related descriptions of films (Bateman and Schmidt, 2012).

Section 3.4 introduced three foundational semiotic modes: text-flow, image-flow, and page-flow, whose basic properties were set out concisely in Figure 3.9. Text-flow, in short, involves written language, which may be occasionally interrupted by graphic elements, such as photographs and maps. Image-flow, in turn, organises images into meaningful sequences. Finally, page-flow exploits the two-dimensional layout space to make additional meanings by establishing rhetorical relations between the pieces of content on the page. Moreover, because page-flow may combine the output of both text-flow, image-flow, and other semiotic modes, it may be characterised as a composite semiotic mode. All of the aforementioned semiotic modes – text-flow, image-flow, and page-flow – may be found in 'A Game of Shark and Minnow'. Their distribution is represented using

Figure 9.1: The content and its transitions in 'A Game of Shark and Minnow'

the boxes in Figure 9.1, which indicate the dominant semiotic mode in each screen: the grey boxes mark dynamic content, such as moving images or animated maps. These screens also correspond closely to the stages of the narrative realised using the longform article. Moreover, this description is complemented by the transitions connecting the different screens (wipe, dissolve, and zoom), whose contribution to the artefact will be discussed shortly below.

As Dowling and Vogan (2014) point out, the longform genre is primarily intended to be consumed using tablets. For a description of its multimodal characteristics, this constitutes an important observation, as allows the underlying material substrate to be defined clearly. This material substrate is the screen, which makes a set of semiotic modes available to the designers working with the website medium. Defining the website as a medium is useful at this stage for characterising certain features shared by both the digital longform journalism and hard news genres. In short, the semiotic modes made available by a medium can be put into use in realising a multimodal artefact. This involves taking the 'raw' material – the content produced and acquired in preparation for creating the artefact – and rendering the material for presentation in a multimodal artefact using the range of semiotic modes available for the medium.

In 'A Game of Shark and Minnow', the very first screen already takes advantage of the underlying medium and its dynamic capabilities. Unlike the printed page, the screen is able to realise dynamic image-flow, which is deployed here to show a person standing in the bow of a boat on the sea. The boat moves up and down with the waves, while the person's shirt flutters in the wind. This short video loops endlessly next to a header and a caption – an introductory sentence. The opening of 'Snow Fall' uses a similar instance of looping image-flow, which Dowling and Vogan (2014: 5) characterise as "a cinematic establishing shot." While the looping image-flow undoubtedly serves as a point of departure for the narrative also in 'A Game of Shark and Minnow', it is less clear to which extent the short loop warrants the epithet 'cinematic'. As such, the loop – a single shot – would qualify as a minimal unit of a cinematographic artefact, constituting a single content portion in a filmic narrative (Bateman and Schmidt, 2012: 58). Such single-shot loops are a common feature of many websites, from GIF animations to advertisements, but here the question is: how this instance of looping image-flow makes the entire artefact 'cinematic' (Dowling and Vogan, 2014: 5–6)?

In order to understand how the artefact achieves a cinematic effect, the instance of looping image-flow must be first related to the overall structure of the opening screen and the sequence that follows. As Section 3.2.2 set out in detail, the GeM model describes this kind of organisation using the layout structure, which captures the hierarchical relations holding between the content portions. Bateman (2013: 66) has argued that the layout structure is equally useful for explicating the organisation of content

portions in cinematographic artefacts, showing how the analysis of film may be pursued within the GeM framework, whose application has been so far limited to printed artefacts (Bateman, 2008). In the semiotic mode of film, the viewers are invited to draw discourse semantic interpretations about the relationships that hold between the different content portions. These interpretations are supported by a variety of mechanisms described in Bateman and Schmidt (2012), Tseng (2013*b*), and Wildfeuer (2013): an useful overview of this work is provided in Bateman (2014*a*). It must be understood, however, that in the case of the opening shot of 'A Game of Shark and Minnow', the kind of interpretative work required for making sense of a film is largely unnecessary, because the 'film' consists of a single shot. Despite the presence of diegetic sound, more elaborate structures and resources made available by the semiotic mode of film, such as discourse relations and cohesive devices, are not deployed. Therefore, describing the single shot as an instance of looping image-flow is likely to be a more appropriate choice.

For this reason, it may be argued that the absence of further shots prompts the viewer to seek out alternative discourse semantic interpretations, which *are not* appropriate for a cinematographic artefact, but are certainly relevant for making sense of the two-dimensional layout space on the page. In other words, the use of layout space in the longform artefact bears the hallmarks of page-flow, which is a composite semiotic mode capable of taking the output of other semiotic modes – such as image-flow and text-flow – and combining them in the two-dimensional layout space. Furthermore, it should be noted that the dominant multimodal structure on the page closely resembles the pattern described in Section 7.2.1 – an image-text-complex – a structure for scaffolding text-image relations on the page. As the following analyses will show, employing image-text-complexes is an effective strategy for introducing the narrative in a 'A Game of Shark and Minnow', which also lends certain cinematic qualities to the artefact. This strategy may be highlighted by examining the first three screens in the artefact, which are provided in Figure 9.2. The texts for the three captions in Figure 9.2 are given below.

> "In a remote corner of the South China Sea, 105 nautical miles from the Philippines, lies a submerged reef the Filipinos call Ayungin." (Caption 1)
> ...
> "In most ways it resembles the hundreds of other reefs, islands, rock clusters and cays that collectively are called the Spratly Islands." (Caption 2)
> ...
> "But Ayungin is different. In the reef's shallows there sits a forsaken ship, manned by eight Filipino troops whose job is to keep China in check." (Caption 3)

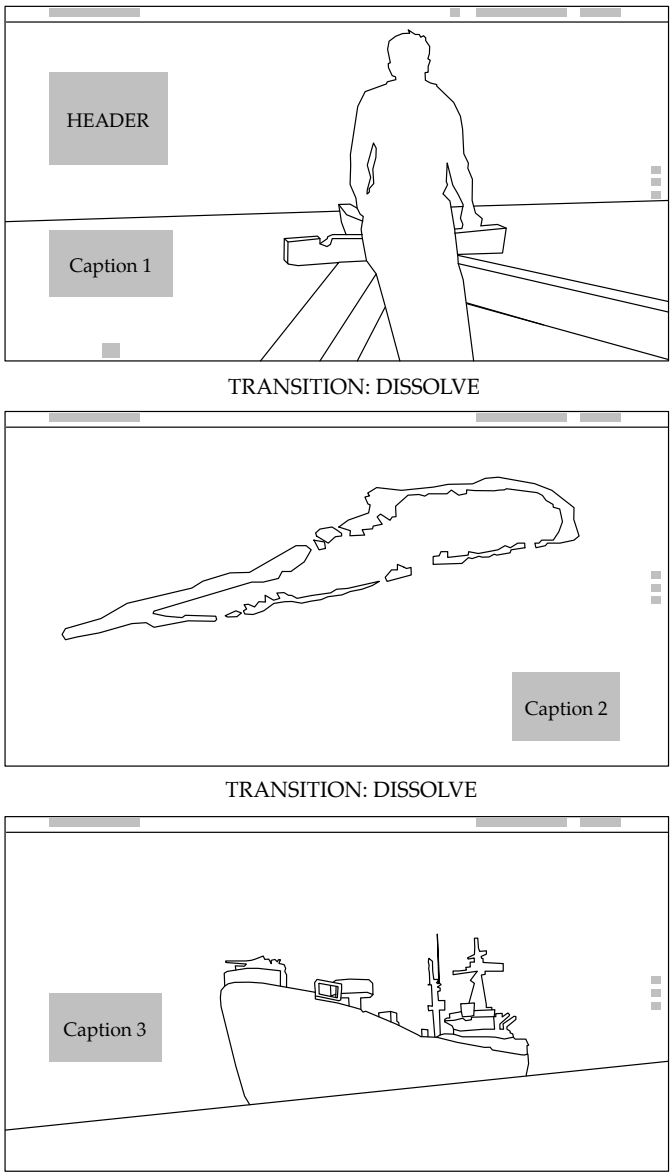

Figure 9.2: The opening sequence of 'A Game of Shark and Minnow'.

It may be argued that the sentences in the three captions participate in a unified rhetorical structure. Together, they form a coherent whole. As a property of a text, coherence is what enables the reader to relate these sentences to each other (for an overview of coherence research in multimodal contexts, see Bateman, 2014*b*). Aspects of coherence may be captured using Rhetorical Structure Theory (Mann and Thompson, 1988), which the GeM model applies to the description of relations holding within texts and between text and images (see Section 3.2.3). In this case, the introductory sentence in Caption 1 is described in greater detail in Captions 2 and 3, therefore establishing a relationship of ELABORATION between the captions. A further relation holds between Caption 2 and the first sentence in Caption 3, establishing a CONTRAST between Ayungin and other reefs in the Spratly Islands. This adds depth to the description: similar rhetorical configurations were used in the tourist brochures to introduce a destination and its locations, as described in Section 7.1. Moreover, the relations between the captions are signalled explicitly using demonstratives ('it') and conjunctions ('but') (Martin and Rose, 2003).

By the virtue of participating in an image-text-complex, each caption is also strongly contextualised by the accompanying graphic element. As pointed out above, the graphic element in the first screen is a single-shot video – in our terms, an instance of looping image-flow – followed in the second screen by a dynamic map, which zooms from an overview of South China Sea and the surrounding countries to a close-up of the Ayungin reef. The map is indispensable for achieving this kind of geographical contextualisation. Due to its capability to present geographical information dynamically, a single image-text-complex is sufficient to pinpoint the area of discussed by the longform article. Finally, the third screen presents a still photograph of the rusty ship tasked with guarding the reef, narrowing down the location even further, in preparation of the text-flow and the graphic elements that follow.

Together, these three screens set up the scene for the narrative by deploying the same structural configuration in every screen. This structure is the image-text-complex, which consists of a caption and an accompanying graphic element. In terms of the multimodal structure of the entire artefact, this is a simple solution, which is also reflected in the layout structure: each image-text-complex requires only two layout nodes, that is, the graphic element and its caption. Moreover, each screen achieves a sense of immediacy, because the image-text-complex constitutes the dominant unit on the screen, which may also contribute to the artefact being perceived as cinematic. When engaging in the process of interpretation, the user does not need to resolve a complex layout structure to make sense of the screen and its contents (see Section 8.2.1). This stands in stark contrast to the presentation of content in the hard news genre, which often requires a complex layout structure to organise the wealth of content on the screen (Bateman et al., 2007). In this connection, it should be noted that image-

text-complexes are also a common structure for presenting hard news, as Knox (2007: 26) has shown in his study of 'newsbites', which "highlight the stories valued by the institution of the newspaper as most important on a given day." In the longform genre, the image-text-complexes appear to serve a completely different purpose, that is, they move the narrative forward.

However, without the rhetorical structure, realised using written language, which joins the captions together, it would be difficult to establish how the three screens in Figure 9.2 relate to each other. To put it simply, it is the written language in the captions that binds these three screens together into an unfolding sequence. Because the graphic elements are structurally embedded into the image-text-complexes in each screen, they are also integrated into the unfolding sequence that the captions participate in. Naturally, language offers several mechanisms for supporting these kinds of unfolding sequences, such as coherence and other structures described above (Mann and Thompson, 1988; Martin, 1992; Taboada, 2004). From the user's perspective, integrating the accompanying graphic elements into the sequence is unlikely to present considerable challenges for interpretation, because these kinds of integrative structures commonly occur in far more complex configurations in other semiotic modes, such as that of comics (Bateman and Wildfeuer, 2014). Furthermore, cohesive ties may be identified in every screen between the captions and the associated graphic elements – they are a common feature of any multimodal artefact working with a page metaphor (see Section 2.2). A prime example of a cohesive tie is presented by the third screen, which portrays the rusty *Sierra Madre* in the photograph, whose caption includes the nominal group 'a forsaken ship'.

Nevertheless, the image-text-complexes are unlikely to suffice for realising the main story at the heart of the narrative – the product of investigative journalism – but carry aesthetic value, place an event in space and time, provide evidence for the event, capture its critical moments, and finally, tell a story (Knox, 2009: 157). As Figure 9.1 shows, the opening sequence is followed by an extensive instance of text-flow. This semiotic mode, which can draw on the entire meaning potential of written language, is naturally the most suitable option for realising the kind of narrative required for investigative journalism. In this aspect, the longform genre bears close resemblance to "slow journalism", which privileges storytelling over "the mechanistic expository style of hard news stories" (Le Masurier, 2014: 6). In terms of multimodality, it may be argued that the longform genre marks a further departure from the presentation of hard news, as the genre exploits the available semiotic modes and combines them into novel sequences, as shown in Figure 9.1.

What Le Masurier (2014: 6) terms as "mechanistic" in relation to traditional online news may also be considered from another perspective. Section 4.5.3 pointed out that many multimodal artefacts draw on patterns

for their design, because these patterns serve as cue structures that help the user to make sense of the artefact and its contents (Waller et al., 2012). They develop over time, and result in 'over-coded' expressions that characterise specific genres (Bateman, forthcoming). For presenting hard news in digital media, these patterns include access structures, such as headers and other typographically highlighted elements, and the positioning of content according to their news value, which are often realised using newsbites (Knox, 2007; Bateman et al., 2007). It should be noted that the patterns commonly used for hard news are largely absent in the longform genre, leading the user to an unknown territory. Consequently, the question is, how does the longform genre support user engagement and navigation?

Reading online news typically involves navigation around the website. Access structures, such as headers and newsbites, support the navigation by linking the home page to various subpages, where the actual news content resides. The longform genre, however, does not have these kinds of common access structures or links to other artefacts or services, apart from the social web panels. Due to absence of the familiar cue structures to support navigation, 'A Game of Shark and Minnow' facilitates navigation by introducing an alternative means of navigating the artefact, which does not involve clicking through links. The narrative proceeds with a simple scrolling movement, which is typically reserved for interacting with a single page. The scrolling movement also triggers a variety of transitions between the screens, whose contribution to the artefact will be evaluated below.

9.3 Filmic transitions in the website medium

As Dowling and Vogan (2014: 6) point out, the looping single-shot videos in 'Snow Fall' are not solely responsible for creating a cinematic effect: the "curtain effect" during the transitions also contributes to this impression. In film editing, this effect is described as a 'wipe' transition, in which the new shot gradually replaces the preceding one. This transition begins from a specific area of the screen. In the case of 'Snow Fall', the new 'shot' – a portion of content – begins to replace the preceding one from the bottom, spanning in the entire horizontal dimension of the screen. The new screen is revealed like a view when the blind covering the window is lifted: hence the curtain metaphor. The same transition is also used in 'A Game of Shark and Minnow', while simultaneously reducing the intensity of the pixels in the outgoing screen from light to dark.

In relation to the use of the "curtain effect" – or more appropriately, a wipe transition – Dowling and Vogan (2014: 6) observe that:

> Reading through graphics and interactive multimedia draws from the conventions of documentary film, particularly works saturated in primary source documents, interviews, and data vi-

> sualisation. Digital longform journalism's use of documentary film conventions, especially those that lay bare their sources, provides a visceral nearness to the subject.

Here, however, it is useful to draw a distinction between the actual content, such as primary source documents, and the filmic conventions used for transitions, which join together different pieces of content. As a filmic convention or device, these transitions are not dependent on the actual content, but become available with certain semiotic modes, which can be selected for presenting the content to the reader. The range of semiotic modes available for a given artefact is limited by materiality: to implement a wipe transition, the material substrate must be dynamic, just like film and screen are. As Bateman (2011: 23) points out, "if the physical substrate can 'do the job', then a community of users may put the effort into making it do so." To draw on an example, the printed page underlying this book cannot obviously realise the kinds of transitions described above effectively. Although digital reading applications may occasionally use a wipe to mimic the effect of turning a page, considering the actual process of turning a printed page as a transition is far-fetched.

In the film medium, these transitions have emerged within the semiotic mode of film as a means of establishing discourse semantic relations between multiple shots (for film as a semiotic mode, see Bateman and Schmidt, 2012). What is particularly interesting about the longform genre is that it takes advantage of the website medium and its dynamic capabilities to introduce these kinds of filmic transitions to the semiotic mode of page-flow. In plain words, the longform genre combines written language, dynamic maps, looping videos, and photographs into structures that are familiar for many websites, but at the same time, it begins to connect the screens that make up the artefact *using transitions commonly found in films*. Moreover, as Figure 9.1 indicates, these transitions are not limited to wipes, but also include dissolves.

A dissolve is another form of transition in film editing, which involves a gradual transition from one shot to another. Unlike a wipe transition, the dissolve does not begin from a specific area of the original shot, but affects the entire image simultaneously. In a digital environment, a dissolve is achieved by gradually replacing the pixels that make up the outgoing shot with the incoming shot, pixel by pixel. The number of pixels from the incoming shot grows with each frame, rendering the new image on the screen and removing the previous one. In the semiotic mode of film, a dissolve is sometimes used to indicate the passing of time, but as Tseng and Bateman (2012: 92) point out, the use of dissolve and other devices must be evaluated against their context of occurrence. This also applies to using dissolves in 'A Game of Shark and Minnow', which are obviously not meant to indicate the passing of time, but are certainly concerned with

moving the narrative forward – the specific functions of these transitions will be addressed shortly below.

Like the wipe transition, the dissolve may be considered an import from one semiotic mode (film) to another (page-flow). Such imports between semiotic modes are not rare: Bateman (2014*c*: 111) cites arrows as one example that originates in diagrams, but which has been later introduced to comics. As pointed out above, a prerequisite for these imports is that the underlying material substrate can support them, which constitutes one aspect of the common ground needed for the emergence of "semiotic blends" (Bateman, 2011: 29–30). In this case, another aspect of the common ground may be the general concern of organising content into narrative sequences, which characterises many instances of films and longform journalism. Bateman (2011: 29) argues that these blends may result in the creation of a hybridised semiotic mode, which is constructed by merging the discourse semantics of the participating semiotic modes. Film, for instance, is a prime example of a hybridised semiotic mode, which has since achieved maturity: the semiotic mode seamlessly combines image-flow, spoken language, diegetic sound, and other resources into structures, which may be used to do very different kinds of communicative work. However, in the case of 'A Game of Shark and Minnow', the combination of filmic transitions and page-flow may be more appropriately described as a *composite semiotic mode*.

In fact, the distinction between a hybridised and a composite semiotic modes is crucial for understanding the use of filmic transitions in the longform genre. If page-flow and film would constitute a hybridised mode that realises 'A Game of Shark and Minnow', this would likely involve a far more extensive integration of filmic devices and conventions into the artefact, which would also affect the organisation of content on each screen, as opposed to merely introducing transitions between sequences of multiple screens. Therefore, 'A Game of Shark and Minnow' is an artefact whose organisation is governed by a composite semiotic mode, page-flow, which has been enriched with certain features of the filmic semiotic mode: transitions in the form of wipes and dissolves. These may constitute a novelty in the particular genre – longform journalism – (and even the website medium) that lends the artefacts deploying these transitions with 'cinematic' properties. In order to understand how these transitions are used in 'A Game of Shark and Minnow', it is necessary to return to Figure 9.1 and examine what kinds of content they are used to connect, while keeping their properties in mind.

As stated above, a wipe transition proceeds from one direction to another to replace the content on the screen, whereas a dissolve involves gradually changing each pixel on the screen. Naturally, on the computer or tablet screen, both transitions involve the manipulation of pixels, but these pixels are manipulated in a different order. To understand the transitions in terms of multimodality, however, requires turning the analytical attention away

from the pixels and towards the artefact structure. As Figure 9.1 shows, the transitions in 'A Game of Shark and Minnow' alternate between wipes, dissolves, and zooms, which is another type of transition that occurs in connection with maps. The following discussion considers how these three transitions are used to connect multiple screens, which all use a different semiotic mode to realise their content.

The opening screens of the artefact, which consist of three instances, a looping image-flow, a zoom-in map, and a static photograph, are connected using two dissolve transitions, as shown in Figure 9.2. The looping image-flow, which consists of a single shot, dissolves into a map, whereas the map zooms in and then dissolves into a static photograph. What is worth noting here is that despite of realising both dynamic and static content, each screen makes extensive use of graphic elements. The image-text-complex is the dominant element on each screen. And as Figure 9.1 shows, this pattern is repeated throughout the artefact: a dissolve is reserved for transitions between screens in which the graphic elements are dominant. A wipe transition, in turn, is deployed when the artefact moves into text-flow. This becomes evident when observing each instance of text-flow in the artefact: each is surrounded by wipe transitions.

But what explains this division of work between dissolves and wipes? As said, dissolves are used for transitions between both dynamic and static graphic elements, whereas wipes are used for text-flow, which carries the main narrative of the longform article. One answer for the reason behind the division of work may be proposed to be the kinds of semiotic modes participating in the dissolve transition. Recall that the dissolve transition involves manipulating the pixels on screen, replacing the old screen with the new one, pixel by pixel. Unlike the wipe transition, in which the effect begins at a specific area of the screen – such as the entire horizontal span of the screen – the dissolve transition gradually interpolates the values of the pixels on the new and the old screen.

To put it simply, a dissolve is more appropriate for a transition between two full-screen graphic elements than between a graphic element and an instance of text-flow. This results from the distribution of pixels on screen and their intensities. Whereas a graphic element will have a broad range of values distributed into the pixels on the screen, an instance of black text-flow on a white background will not. If a dissolve transition would be performed between a full-screen graphic element and text-flow, the text would become rapidly illegible, as pixels would be removed from letters and introduced into the empty space in the margins. When occurring between two graphic elements, however, the effect is much more subtle, because the new screen is not introduced into what is perceived as an empty space by the viewer, but superimposed on another image until the transition is complete. This is precisely how the dissolve operates in the semiotic mode of film.

With this kind of filmic effect, 'A Game of Shark and Minnow' is able to move the narrative forward between the instances of text-flow, providing the means to rapidly contextualise the subject matter, as shown in Figure 9.2. The zoom transitions, in turn, are reserved for maps, as the second line of Figure 9.1 indicates. They are used to zoom in and out of the map, much in the style familiar from Google Earth and other digital maps, while using partial dissolves to add and remove various annotations, captions, and overlays explaining the geopolitical situation in the South China Sea. Although maps are an effective way of contextualising the narrative geographically, using them to present evidence or an argument requires additional support from written language and superimposed graphic elements. Despite the more complex interactions typically required for exploring a digital map, together with the increased cognitive load, it should be noted that all of the aforementioned transitions – dissolve, wipe, and zoom – are automatically activated by scrolling down the screen. Minimising the demands of navigation helps the reader to cope with occasional instances of more complex content.

To conclude, it may be argued that 'A Game of Shark and Minnow' creates its immersive experience by tightly controlling the user experience. The user interface, which involves only scrolling, prevents the user from digressing from the artefact. The transitions, in turn, help to drive the narrative forward and maintain the user's interest by unfolding the narrative in a controlled sequence. This is a conscious decision, because the underlying website medium would certainly allow rendering all of the content in the longform artefact on a single page. By drawing on the transitions that have evolved within semiotic mode of film, the composite semiotic mode of page-flow in 'A Game of Shark and Minnow' acquires the necessary mechanisms for narrative storytelling involving written language, photographs, maps, and videos. Moreover, as the introduction of dissolves and zooms alongside the wipe transitions used in 'Snow Fall', it appears that the affordances of filmic mode are being explored further by the designers. This sets certain requirements for further investigations, which are explicated in the concluding section below.

9.4 Emerging genres and media convergence

One of the main arguments of this chapter has been that understanding emerging genres, such as digital longform journalism, will require a toolkit equipped with an appropriate set of analytical tools. Moreover, these analytical tools need to be set in a clear relation to each other, in order to distinguish between the various contributions to a complex multimodal artefact such as 'A Game of Shark and Minnow'. It is important to acknowledge that deploying concepts such as medium, genre, and semiotic mode will not increase our understanding of multimodality, communication, and

human creativity, unless their definitions are carefully formulated and applied consistently to describe particular kinds of structures or activities. These concepts must not be considered as universal truths, but tools for answering research questions. Maintaining this kind of approach is necessary for tackling some of the more abstract concepts that have gained attention in recent years, such as the concept of media convergence.

Dowling and Vogan (2014: 6) suggest that 'Snow Fall' is a prime example of converging media, because the artefact complements the traditional tool of journalism – written language – with elements of theatre (the curtain effect), while acknowledging "the diverse forms of literacy that mark twenty-first-century audiences." Their observation captures several key aspects of media convergence, which were explicated above using linguistically-inflected methods derived from multimodal research. What must be understood is this: unless we possess a firm understanding of multimodality in what new media theorists often perceive as 'legacy' media – printed books, newspapers, brochures, and many more – there is little hope that we could describe digital media in a comprehensive manner, and even more importantly, explain how these media are able to converge in the first place. As the analysis of 'A Game of Shark and Minnow' showed, previous descriptions of film, for instance, can prove valuable when investigating multimodal artefacts in digital media.

Leveraging the previous work, however, requires that it to be a good fit for the framework at hand. In this case, the necessary compatibility could be achieved by drawing on work that follows similar principles in constructing the theoretical framework. Armed with these concepts, the analysis showed that page-flow, the semiotic mode responsible for organising the content in 'A Game of Shark and Minnow', borrowed certain transitions from the semiotic mode of film. However, the semiotic modes of page-flow and film were not integrated into a hybridised mode, but the transitions were simply added to the discourse semantic repertoire of page-flow. Like film, the longform genre used these transitions to pace the unfolding discourse. Their interpretation, however, remains strongly tied to the context. Unlike in film, where a dissolve may occasionally indicate the passing of time, their use in the longform genre does not encourage this kind of interpretation (Tseng and Bateman, 2012: 91–92). The longform genre uses the transitions to achieve rapid contextualisation of subject matter and to tightly control the user experience within an unfolding structure. All of this is enabled by the underlying website medium, which – together with film –shares the capability to realise dynamic content, making these imports possible.

Without the support from the previous work, these observations would not have been possible. Given the potential of digital media, however, this is unlikely to be always the case. As Bateman and Veloso (2013: 146) point out in their analysis of how Ang Lee uses the resources made available by the semiotic mode of comics in the film adaption of 'Hulk':

> ... the overlap and reuse of technical features and aesthetic choices across films, comics, videogames and computer interfaces is sufficiently prominent to have received attention from several perspectives but our understanding of such overlaps is still quite preliminary.

Sufficiently prominent features, such as the use of filmic transitions in the website medium, are likely to attract attention, before they become commonplace. For the analyst, in turn, these overlaps raise many questions: what enables us to rapidly integrate – with very little effort – content rendered using multiple semiotic modes, whose organising principles are radically different? Page-flow, for instance, is often spatially organised, whereas film is dominated by the principle of linearity. Such questions relate closely to what Dowling and Vogan (2014: 6) referred to as 21st century literacies and carry considerable pedagogical implications. This presents a significant opportunity to multimodal research, which can help to answer these questions by developing sufficiently sharp analytical tools to unpack the contemporary literacies and those required in the future, and the more general phenomenon of media convergence.