Author(s): Hiippala, Tuomo

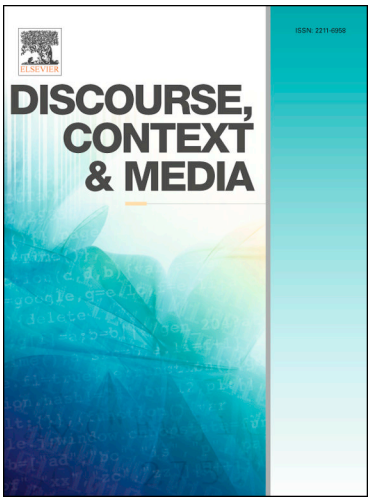
Title: An overview of research within the Genre and Multimodality framework

Year: 2017

Version:

Please cite the original version:

Hiippala, T. (2017). An overview of research within the Genre and Multimodality framework. Discourse, Context and Media, 20, 276-284. https://doi.org/10.1016/j.dcm.2017.05.004



An overview of research within the *Genre and Multimodality* framework

Tuomo Hiippala

PII: S2211-6958(16)30187-8
DOI: http://dx.doi.org/10.1016/j.dcm.2017.05.004
Reference: DCM 168

To appear in: *Discourse, Context & Media*

Accepted Date: 23 May 2017

Please cite this article as: T. Hiippala, An overview of research within the *Genre and Multimodality* framework, *Discourse, Context & Media* (2017), doi: http://dx.doi.org/10.1016/j.dcm.2017.05.004

An overview of research within the *Genre and Multimodality* framework

Tuomo Hiippala
Centre for Applied Language Studies
University of Jyväskylä

1 Introduction

Although research on multimodality, which studies how different means of expression interact and co-operate with each other, is blooming at the moment, concise overviews of the field and the issues under debate are rare or already somewhat dated (see e.g. Kaltenbacher 2004, Martinec 2005). This may be partly explained by the wealth of different approaches to multimodality, but also by the rapid take-up of the concept in recent years across various fields (Bateman et al. 2017). However, because multimodality as a field of study is now considered mature enough to be considered in relation to established fields such as ethnography and applied linguistics (Kress 2011, 2015), the time might be ripe for overviews of the research conducted so far.

Such overviews can benefit both old and new audiences: those who have worked on multimodality for some time already can reflect on the previous work, while newcomers to the field may find concise overviews useful for guiding their way. The breadth of the field, however, is likely to impose certain restrictions on conducting such overviews. For this reason, mapping the research on multimodality is likely to require a piecemeal approach, examining particular strands of research at a time before attempting to build a bigger picture.

This article sketches a small part of that picture by providing a systematic literature review of the research conducted within the *Genre and Multimodality* framework (hereafter abbreviated GeM), which has been used to describe a variety of page-based documents and other multimodal artefacts over the last 15 years. Although the notion of genre has also been invoked within other streams of research, such as the social semiotic approach to multimodality (cf. e.g. van Leeuwen

2005, van Leeuwen & Hestbæk Andersen 2017), no other framework apart from GeM has adopted genre as its point of departure for studying multimodal phenomena. Given that genre is a notoriously elusive concept (cf. Freadman 2012), it may be argued that with 15 years, the GeM framework has reached a stage where it warrants attention and critical evaluation.

The current overview does not, however, cover the work in a chronological order, but rather focuses on the central concepts of medium, mode and genre, while also outlining how the GeM framework has contributed to the research on multimodality. In this way, the article attempts to sketch how the aforementioned concepts have evolved over time, allowing the reader to trace the development of the GeM framework, while simultaneously serving as an accessible introduction to this stream of research.

The article begins by outlining the initial motivation for developing the GeM framework, before briefly presenting its methodological foundation for doing empirical research on multimodality. In the subsequent sections, the article discusses central concepts that have been given extensive consideration within the GeM framework: medium, mode and genre. The article then continues to outline domains of research where the GeM framework has been put into productive use. Finally, the article concludes by discussing the impact of this work and outlining several avenues of future research.

1.1 Inspiration and early beginnings

Recounting how the research in computational linguistics inspired the early development of the GeM framework, John Bateman (2014*b*, 25–27) – one of its lead developers – observes that the field of natural language generation had always attended to communicative 'goals' set for texts, and particularly, to the contextual constraints that had to be accounted for in order to meet the designated goals. Research on the organization of language, discourse and non-linguistic forms of communication – such as diagrams – for the purpose of generating their appropriate combinations then eventually led to consider the role of *layout*.

This presented a substantial challenge, because immense variation could be found in the use of layout space across different page-based documents. Although tools for describing layout – which were later incorporated into the GeM framework – emerged at the time, they offered few explanations for the variation commonly encountered in documents (cf. Reichenberger et al. 1995). This led to the question: why do specific documents adopt a particular kind of layout and organization (Bateman 2014*b*, 30)?

The *Genre and Multimodality* research project, which ran between 1999 and 2002 at the University of Stirling, Scotland, and Bremen University, Germany, deployed the notion of genre to account for variation in document structure. Drawing

on the linguistic notion of genre, which is traditionally understood as introducing context-dependent constraints to the selections made within language and discourse (for a comprehensive overview of various approaches, see Bawarshi & Reiff 2010), the GeM project departed from the hypothesis that the notion of multimodal genre

> "might similarly exercise constraints on selections within layout structures, on their typographical and spatial realisation, and on the transformation processes between layout structure and rhetorical organisation." (Bateman 2014*b*, 32)

Within the GeM framework, genre was conceptualized a space of possibilities, drawing on the proposal put forward by Lemke (1999). Moving around this space was assumed to be reflected in different configurations of multimodal structures, depending on what the document's communicative goals were, that is, whether the document was intended, for example, to instruct, to describe or to achieve both of these goals simultaneously.

To keep track of the multitude of contextual variables that could influence the selections made during the design process, which would then take a concrete form on the document surface, the GeM framework was designed to be corpus-driven from the outset. In order to bring the documents under analytical control, the project developed an annotation schema with multiple layers of description, which was intended "to function as a tool for isolating significant patterns against the mass of detail that multimodal documents naturally present" (Bateman 2014*b*, 33). This annotation schema was expected to enable a systematic empirical exploration of multimodality in documents, which would, in turn, provide a stronger basis for formulating theories about their principles of organization.

The findings of the GeM project are presented at length in *Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents* (Bateman 2008). This monograph is best contextualized by its preface, in which Bateman (2008, xix) identifies several established theories of multimodality (e.g. O'Toole 1994, Kress & van Leeuwen 1996, 2001) and raises the question: which direction should be taken in research after multimodality has become accepted as an inherent feature of meaning-making? For the study of page-based multimodal documents, Bateman's (2008, 16) proposed direction involves a systematic, empirically-motivated account of their production and consumption. These fundamental processes were intended to be examined using the corpus-based method proposed in the GeM framework, to which this article turns next, before considering its theoretical scaffolding.

1.2 Methodological orientation

As mentioned above, the GeM framework provides an annotation schema with multiple analytical layers to support empirical research on page-based documents. These layers include, but are not limited to, the following:

- **base** layer, which carries the content realized using different semiotic modes
- **layout** layer for the hierarchical organization of the content, its typographic and graphic characteristics, and spatial positioning in layout
- **rhetorical** layer for describing discourse relations that hold between content elements using Rhetorical Structure Theory (RST; see Taboada & Mann 2006)
- **navigation** layer describing structures that support the use of the document

To search for structural patterns that could potentially characterize the genre under investigation, each annotation layer is cross-referenced with other layers. The entire analytical process is visualized in Figure 1.

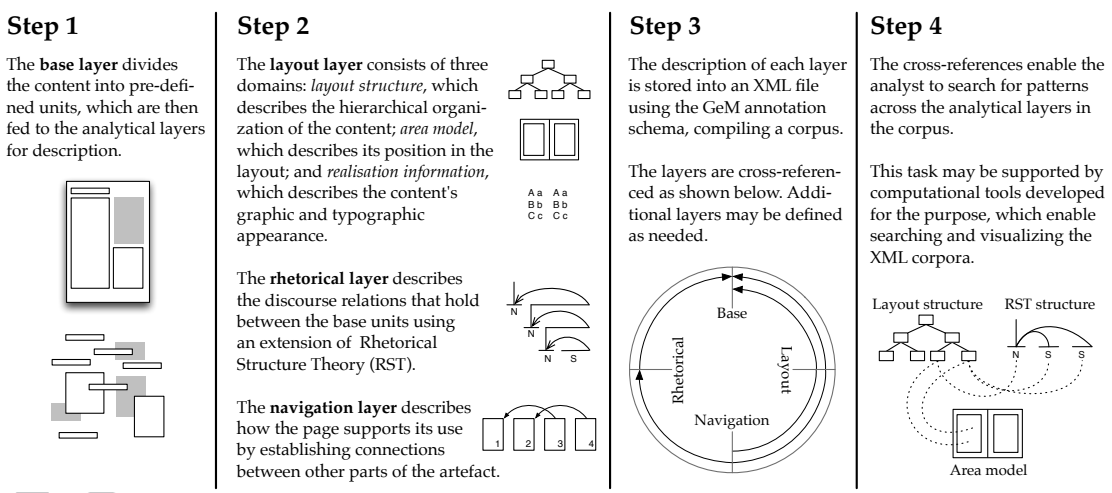


Figure 1: Methodological steps in applying the GeM framework

The annotation schema, which uses Extensible Markup Language (XML) for storing the annotation, is described extensively in Bateman et al. (2002, 2004), Bateman (2008, 254–264) and Hiippala (2015*c*, chapter 5). Among the early work, useful examples of applying the annotation schema to various artefacts may be found in Bateman et al. (2000) and Delin et al. (2003). Subsequent work and findings will be taken up for discussion later, as the following discussion will focus

on the advantages and limitations of adopting an XML-based annotation schema for multimodal corpora.

On the one hand, it may be argued that adopting XML, an industry standard markup language, allows the GeM framework to benefit from commercially-driven developments for handling XML data, which often outpace the tools developed for academic research (Bateman et al. 2002, 19). For a concrete benefit, XML has enabled the use of different programming languages for querying and visualizing GeM corpora, because they can work with XML data (Thomas 2007, 2009, Hiippala 2015*c*). Moreover, XML remains effective for representing structured data, which may be extended as necessary: Thomas (2014) shows how additional layers of description may be integrated into the GeM annotation schema to extend the model's capability to describe multimodal phenomena.

On the other hand, applying the GeM framework requires a considerable amount of time and resources. Although commercial XML editors facilitate the process, the annotation remains a tedious task involving a lot of manual work. To exemplify, annotating a corpus consisting of 58 double-pages in tourist brochures took roughly three years (Hiippala 2015*a*). In particular, the layout layer has been identified as a bottleneck in the annotation process, as numerous measurements are needed to represent the positioning of layout elements, their size and typographic features (Thomas 2009, 245).

To solve this challenge, Thomas (2009) explores the use of commercial optical character recognition software for automatically generating "proto-GeM" XML, but concludes that extensive post-processing required from the analyst rendered this approach impractical. Nevertheless, Thomas (2009, 243) also observes that computer vision techniques may contribute to automating parts of the annotation process in the future. Hiippala (2016*b*) discusses previous research in this area and proposes a solution for generating parts of GeM annotation automatically. Given the more general development within the humanities, future GeM corpora are likely to combine both manual and automatic annotation (Bateman et al. 2016).

2 Key concepts within the framework

An annotation schema alone, however, provides little explanatory power without support from theory. Thus the following sections discuss key concepts within the GeM framework – medium, mode and genre – which provide the theoretical scaffold necessary for separating different kinds of contributions to multimodal documents. Section 2.1 begins with the concept of medium, which describes the materiality, such as paper or screen. Section 2.2 then proceeds to discuss mode, a central concept to any account of multimodality, which describes the resources used for making and exchanging meanings. Finally, Section 2.3 concludes with the concept

of genre, which is used to characterize the patterns of semiotic modes commonly found across different documents.

2.1 Medium

Within the GeM framework, the concept of medium describes how some material – that is, the concrete physical substrate on which a document is realized – attains a stable form over time, as its production and consumption stabilizes. If the material is steadily used to fulfil some communicative purpose, it may establish itself into a full-blown medium. The newspaper, for instance, serves as a useful example of a medium that has evolved to support the fast-paced production and consumption of news by adopting a particular type of low-cost paper – newsprint. However, the material substrate of newsprint also sets certain limitations to the newspaper medium: although newsprint is durable enough to be run through a high-speed printing press, its capability to realize photographs is limited compared to, for instance, high-quality glossy paper reserved for monthly magazines.

To disentangle the effects of "technological imperative and cultural habit" that shape the form of multimodal documents (Bateman et al. 2007, 157), the GeM framework built on Waller's (1987) work on document design to define three constraints arising from materiality (Bateman 2008, 18):

1. canvas constraints, that is, constraints arising from the materiality, such as being able to realize colour photographs

2. production constraints, such as micro- and macro-economies of time and materials, as reflected in the use of cheaper newsprint for daily news and more expensive glossy paper for monthly magazines

3. consumption constraints, which determine the use and life cycle of the document – the content of a daily newspaper 'expires' faster than that of a monthly magazine

Defining these constraints helps to illuminate a key issue in describing multimodal documents: while these constraints have a concrete effect on the documents, they need to be treated separately from any semiotic contribution to the artefact. The *combination* of these constraints gives rise to "virtual artefacts" – established, recognizable ways of shaping some material for some communicative purpose (Bateman 2008, 18).

Again, a newspaper may be used to illustrate the concept of a virtual artefact: early printing technology could not produce columns of arbitrary width, therefore imposing a production constraint on newspaper layout. Similarly, the paper used at the time set a canvas constraint to font size, because small fonts were rendered

illegible by smudging ink, as a result of improper absorption. These constraints have largely disappeared due to developments in printing technology, but they influenced the form of the newspaper sufficiently to establish a virtual artefact, whose form has been upheld by the established use of the newspaper. As a consequence, newspapers continue to organize their content into compartments, whose size and positioning can be used to convey information about their news value.

As a theoretical concept, the notion of a virtual artefact bears close resemblance to what Bateman (2014*c*) expands into a full-blown concept of a medium. Previously, the GeM framework embedded the aforementioned technologically and culturally motivated constraints arising from the virtual artefact within the notion of genre. Conflating the notions of virtual artefact and genre, however, did not sufficiently differentiate between multimodal genres in different media, because some of the features ascribed to virtual artefacts were strongly rooted in the underlying materiality and therefore unlikely to be carried across to other media (Bateman 2014*c*, 255–256). To draw again on the example of newspapers, it is unlikely that page numbers – a key feature of the printed newspaper, which function as navigation devices to support its use – would be included in the electronic edition, in which navigation is handled using hyperlinks. In other words, these features of individual artefacts arise from the underlying medium.

Because the notion of medium was introduced to the GeM framework fairly recently, empirical research in this area has been limited. However, a theoretical perspective that aligns with the GeM framework can be found in Bateman et al. (2017), which discusses how different materialities are shaped for communicative purposes and consequently act as incubators for new semiotic modes *and* their combinations (Bateman 2014*c*, 257). Therefore, to move forward into this area, the following section examines how the fundamental notion of semiotic mode has been conceptualized in the GeM framework.

2.2 Mode

The concept of a mode is central to multimodal research, and streams of research in the field often differ in their definition (see e.g. Elleström 2010, Jewitt 2014). Within the GeM framework, the notion of a semiotic mode is conceptualized as an intermediate category between large- and small-scale semiotic phenomena, such as entire semiotic systems and individual signs (Bateman 2014*a*, 10). The motivation for situating mode in this intermediary space has been to enable the concept to be applied as an analytical tool without over-generalizing or getting lost in the detail. For this purpose, Bateman (2011) proposes a definition of semiotic mode with three distinct strata: an underlying *material substrate*, which is capable of carrying a number of *semiotic resources*, whose contextual interpretation is supported by *discourse semantics*. These three strata and the theoretical foundation

they provide will be covered before discussing the semiotic modes identified within the GeM framework.

As outlined above, the GeM framework considers a material substrate that may be manipulated to be a prerequisite for the emergence of a semiotic mode. Over time, this substrate may evolve into a full-blown medium, but the first step for a semiotic mode is to establish a material substrate capable of carrying traces of semiotic resources. Without such a steady material substrate at hand, the semiotic resources cannot stabilize and develop the kind of organization required for making complex meanings, that is, having fully developed semiotic resources and their discourse semantics in place (Bateman 2011, 21). These concepts will be explored in greater detail below.

Within the GeM framework, the semiotic resources have been treated as having an organization with paradigmatic and syntagmatic axes, which enable making paradigmatic choices and combining these choices into syntagmatic structures. These are precisely the structures that leave traces on the underlying materiality, for instance, in the form of written language and illustrations (Bateman 2011, 20). Basing the definition of a mode on the stratum of materiality has enabled the GeM framework to define the semiotic resources participating in a semiotic mode in a more precise manner, because broad definitions such as 'language' and 'image' are unlikely to suffice to bring out the finer distinctions that multimodal artefacts naturally present (cf. Bateman 2014*d*).

Finally, the semiotic resources realized on some material become interpretable by virtue of being embedded into unfolding discourse (Bateman 2011, 21). This capability arises from the stratum of discourse semantics, a property of any full-blown semiotic mode which ensures that the semiotic resources are organized in a way that supports their contextual interpretation. Without discourse semantics, the interpretation of semiotic resources would be limited to specific contexts. Contrastingly, a full-blown semiotic mode that has the stratum of discourse semantics in place extends the possible contexts of use for each semiotic resource by providing a mechanism that guides their contextual interpretation.

A central argument within the GeM framework has been that the first step in multimodal analysis should involve an effort to identify the semiotic modes, instead of assuming that these modes are known in advance (Bateman 2014*a*, 10). Moreover, this effort must be supported with empirical evidence based on the semiotic choices made in the artefact under analysis (Bateman 2011, 35). This argument is directly in line with the empirical approach advocated within the GeM framework, and underlines the need to exercise caution in defining a mode. Assuming that 'language' and 'image' constitute the semiotic modes active in page-based artefacts is often under-differentiating, as various kinds of 'images' ranging from illustrations to photographs can serve a number of different functions

in different genres (see e.g. Kong 2013, Hiippala 2015*c*).

As an alternative solution for identifying semiotic modes in page-based documents, Bateman (2011) proposes a set of abstractions termed *text-flow*, *image-flow* and *page-flow* (see also Bateman 2008, 175–176). These abstractions seek to capture the underlying logic that defines how different multimodal contributions on a page are organized. For instance, text-flow is built around the linearity of written language, whereas page-flow begins to make use of the entire two-dimensional layout space to organize its contents. Which kinds of expressive resources can be put into use within these semiotic modes depends on the underlying materiality: accessible introductions to these issues are given in Bateman (2009) and Hiippala (2014). The selections made within these modes, however, are dependent on the notion of genre, to which the article turns next.

2.3 Genre

Genre is often conceptualized as a "second-order phenomenon: a patterning of patterns" (Bateman 2014*c*, 241), that is, a phenomenon which cuts across different strata of organization. For language, the effect of genre is reflected, for instance, not only in lexical and grammatical choices, but also in their discourse organization: for multimodal documents, the same effect is manifested in selections made within semiotic modes *and* their combinations as a part of a larger whole such as a document.

The choices made across different strata are considered to be socially motivated. In other words, genres have designated communicative goals or social purposes: the particular selections made in the semiotic modes are seen to support the accomplishment of these goals (Bateman 2014*c*, 242). Such descriptions of genre as a goal-oriented process are often accompanied by statements about its organization being 'staged'. Bateman (2014*c*, 243) notes, however, that the property of staging is ontologically dependent on the property of *linearity*, which disappears quickly when documents began to take advantage of the two-dimensional layout space available on the page.

This does not mean that multimodal documents could not support linear organizations. They certainly do, because linear models of genre, such as the well-known proposal from Swales (1990), can be applied to instances of written language in any document drawing on this particular semiotic mode. The problem is that models built on the principle of linearity do not scale up to the entire page-based document, because a page inherently provides the potential for non-linear organizations built on the layout space. This has direct consequences to defining a multimodal notion of genre: with linearity gone, a multimodal definition must look elsewhere for the crucial support provided by structure.

As Bateman (2014*c*, 244) argues:

> The task of advancing multimodal application of genre cannot be separated, therefore, from that of providing a firm theoretical grasp of the properties of multimodal artefacts that are available for carrying generic patterning.

Put differently, detecting any patterns in documents that bear the hallmarks of some genre requires a sufficiently developed understanding of document structure. Achieving this kind of understanding must naturally involve pulling apart the contributions arising from the medium *and* the semiotic modes that are made available within this medium.

The annotation schema provided by the GeM framework is intended to make these structures available for systematic corpus-driven analyses. As pointed out above, however, genre patterns do not necessarily manifest themselves within a single analytical layer defined within the framework. Cross-layer analyses, such as those combining descriptions of layout and rhetorical structure, are often necessary for teasing out potentially genre-defining features in GeM-annotated corpora (Hiippala 2015*c*, 161). To draw on an example from the tourist brochures studied by Hiippala (2015*c*, chapters 6–8), choices made in the semiotic modes *and* their contributions to the rhetorical and layout structures differ depending on whether the brochures *guide* the tourists around or *describe* the locations to them.

The GeM framework posits that genres can actually be rather flexible in how they fulfil their communicative purposes, bearing dissimilarities when observed from one perspective and similarities when viewed from another angle. For this purpose, the notion of 'genre space' – that is, the topological perspective proposed in Lemke (1999) – can effectively capture this flexibility and its consequences to the genre identity of a document (Bateman 2008, 224). This can also explain how page-based documents can play around with their structure: borrowing the layout structure patterns from one genre provides a 'false identity', which is only revealed upon examining the content and rhetorical structures.

It is important to understand, however, that the genre space – as conceptualized within the GeM framework – is not intended to be populated by instances of actual document genres at this stage. This would require a large volume of data to separate a signal from the noise, which is caused by the variation inherent to all document genres. Instead, the notion of genre space is intended as a tool for contrasting how different genres are positioned along some dimension (cf. e.g. Hiippala 2015*c*, 37). Mapping the dimensions of genre space remains an open challenge, which can only be met when empirically well-founded descriptions covering a sufficiently broad range of document genres become available (Bateman 2014*b*, 33). To consider the contributions made towards this goal so far, the discussion now turns towards the application of the GeM framework.

3 Applications of the GeM framework

This section discusses how the GeM framework has been applied in multimodal research and beyond. Beginning with analyses of cross-cultural communication in Section 3.1, the discussion then proceeds to discuss analyses across different media in Section 3.2. Finally, Section 3.3 considers various points of contact to different disciplines that have emerged from interactions with the GeM framework, before proceeding to reflect on the past and future work.

3.1 Cross-cultural comparisons

To begin with, the GeM framework has been frequently used to study cross-cultural communication. As an example of the early work, Bateman & Delin (2003) seek to identify appropriate units of analysis for contrasting page-based multimodal documents produced within different cultures. Comparing a single pair of English and Japanese instruction manuals, Bateman & Delin suggest that cross-cultural analyses should target the entire page. They show that when applied to the entire page, the GeM framework is able to capture several cross-cultural differences, which are manifested in the staging, organization and appearance of document genres. To exemplify, whereas the English instruction manual relies mainly on typography to organize its contents and rhetorical structure, the Japanese counterpart integrates graphical elements such as icons and warning signs into its rhetorical structure. Bateman & Delin suggest that this allows the manual to build the kind of interpersonal relationship between the manufacturer and the consumer appropriate for Japanese culture.

Hiippala (2012*a*), in turn, conceptualizes the cross-cultural adaptation of documents as a form of multimodal localization – adapting documents from one 'locale', or the bundle of language and culture, to another. He applies the GeM framework to Finnish and English versions of the same tourist brochure with the goal of teasing out conflicts arising from multimodal localization. Working with four pairs of tourist brochures, Hiippala observes that rhetorical structures become easily mismatched with the intended layout structure. This arises from differing sentence and paragraph lengths across language pairs, which causes text to be misplaced, thus reducing usability by breaking rhetorical and layout structures on the page level. Hiippala suggests that occasional negligence towards multimodal localization can be traced back to lack of time and resources in the production of tourist brochures.

Kong (2013) uses the GeM framework to compare the multimodal structure of 'global' news items, that is, news items picked up in both English and Chinese tabloid newspapers. To do so, Kong compiled a corpus with 55 pairs of corresponding news items in both languages for an in-depth analysis of their structure.

Whereas the Chinese tabloids organized the semiotic modes into separate compartments in the layout space, the English tabloids organized their content into composite units consisting of layers partially overlapping each other. Kong identifies additional differences in the use of diagrammatic elements, such as arrows and icons, which the Chinese news items employ to guide the reader through the story. Like Bateman & Delin (2003), Kong relates this preference to building a culturally-appropriate relationship with the reader.

Yet another cross-cultural study is presented by Thomas (2014), who applies the GeM framework to study product packaging in the United Kingdom and Taiwan. Focusing on fast-moving consumer goods, such as toothpaste and shampoo packs, Thomas (2014) describes their genre structure and examines the differences between the two locales. His analysis builds on the more extensive work in Thomas (2009), which defines a set of common 'message types' in product packaging on the basis of a corpus of 24 packages. These messages range from expressing brand identity to providing consumers with instructions and contact information, whose multimodal structure Thomas interrogates using the GeM-annotated corpus, supported by tools developed for the purpose (Thomas 2007). In addition to extending the framework to the non-page-based medium of product packaging, Thomas also makes a significant methodological contribution by showing how additional layers may be introduced to the GeM framework to meet specific analytical needs.

Nekić (2015), in turn, draws on the GeM framework to analyse the landing pages of Scottish and Croatian tourism websites. While she mainly pursues a detailed linguistic analysis intended to uncover how locations are construed as tourist destinations, she also applies the GeM layout layer to a subset of four examples from her corpus of 48 website landing pages. Nekić shows how landing pages combine content elements into hierarchical organizations, which are then rendered for display on the layout. Although the landing pages exhibit considerable variation their organization, they also share several features: given the underlying digital medium and their generic purpose as an entry point to a site, all landing pages emphasize interactivity by investing heavily into navigation structures.

Finally, a recent dissertation by Zhang (2017) uses the GeM framework to compare public health posters in New York City and Hong Kong. Compiling a GeM-annotated corpus of 60 posters, Zhang extends the framework to yet another medium, while also complementing the multimodal description with a detailed linguistic analysis. In a corpus-driven study, Zhang explores how this genre is used to inform and educate audiences in the two cities. Her analysis reveals that posters in New York City prefer the use of imperatives to instruct the audience and often restate their message using written language and images, whereas their counterparts in Hong Kong prefers to communicate more indirectly, preferring nominal groups and symbolic images. Methodologically, Zhang's work also exemplifies how

the GeM framework may be interfaced with more detailed linguistic analyses.

3.2 Cross-media comparisons

In addition to comparing multimodality across cultures, the GeM framework has been used to explore contrasts between different media. Bateman et al. (2007), for instance, introduce the notion of "genre mapping" in a comparison of front pages in printed newspapers and digital news websites. According to Bateman et al., mapping differences between genres enables sketching dimensions of the genre space. As pointed out above, this task is crucial for accounting for variability, that is, what kinds of choices are available to page-based documents. Although the front page genre serves the same broad goal of delivering news at a glance in both media, printed newspapers use layout and typography to construe 'news values', whereas news websites do not necessarily do so. Instead, they invest in navigation structures to guide the reader through the site. This naturally bears close resemblance to Nekić's (2015) findings described above, suggesting that navigation structures constitutes a requirement arising from the underlying digital medium.

That being said, the GeM framework has also been applied to journalistic genres beyond landing pages. Hiippala (2017) presents a study of digital longform journalism, a genre that draws on traditional feature journalism, that is, journalism that focuses on topics outside the daily news cycle. As opposed to landing pages that compartmentalize their content and feature complex navigation structures, the 'longform' genre typically dedicates the entire screen to the story in question and avoids outgoing hyperlinks that might distract the reader. In addition, the genre features novel transitions borrowed from the semiotic mode of film, such as dissolves and wipes, which are used alongside the more traditional transitions of scrolling along a page and navigating using hyperlinks. Unlike in film, however, in which these transitions may have discursive meanings – a dissolve, for instance, may indicate passing time – such meanings are not invoked in the longform genre.

Considering that audiovisual media such as films are embedded into page-based artefacts in digital media such as those represented by the longform genre, their analysis has also been taken up for discussion within the GeM framework. To this end, Bateman (2013) treats film as a *dynamic* document composed of audiovisual portions, or content segments, which are set into various kinds of relations with each other. In order to make sense of such documents, the viewer must reconstruct their organisation during interpretation. Many of these organisations may be described using the analytical layers of the GeM framework: to draw on an example, films organize their content into similar hierarchies as static documents, which may be captured using the GeM layout layer (Bateman 2013, 71). This kind of perspective to 'film-as-document' has been discussed extensively in Bateman & Schmidt (2012) and applied in the aforementioned study of longform journalism

by Hiippala (2017).

3.3 Extensions and inspirations

Extensions of the GeM framework have not been limited to media, but also include theoretical and methodological openings as well. To exemplify, Waller et al. (2012) build on the GeM framework and the previous work in Waller (1987) to explore the notion of 'pattern language' in information design. The notion of pattern language, coined by the architect Christopher Alexander, advocates the search for functional, pattern-based solutions to specific design problems. Examining the relation between design patterns and genres, Waller et al. (2012) note that similar patterns may be found in different genres, which implies the kind of flexibility of movement across the genre space theorised within the GeM framework.

They observe that particular patterns may be more prototypical or peripheral within genres, with their strength of association depending on their prevalence. The GeM framework may provide one candidate for tracking these patterns, their diversity and frequency, but note that this remains a tedious task due to the time- and resource-intensive nature of building multimodal corpora (Waller et al. 2012, 24). Due to this constraint, the GeM framework will likely remain a candidate tool for critiquing individual designs, before larger multimodal corpora become available (for other examples, see Delin & Bateman 2002, Hiippala 2016*a*).

Another issue pertaining to multimodality and design is visual perception. Discussing the perception of multimodal documents, Hiippala (2012*b*) interrogates the notion of 'reading paths', which is often invoked when making assumptions about visual perception in multimodal research. As an alternative to making such assumptions, Hiippala advocates drawing on previous research in experimental psychology, which studies visual perception using eye-tracking in combination with other methods, such as verbal protocols, interviews and tests (see Holsanova 2014).

Hiippala adds to this toolkit by presenting a solution for interfacing a GeM-annotated corpus with the data provided by the eye-tracker. By using the same identifiers in the GeM corpus and the eye-tracker software, the analyst can retrieve the multimodal characteristics of what is being perceived and in which order. This does not only place the study of reception on a firmer ground within multimodal research, but opens up the possibility of contributing to experimental psychology as well, providing the means to construct more elaborate hypotheses about the perception and reception of multimodality.

Moving on to a different field, Seizov (2014) proposes a framework called *Imagery and Communication in Online Narratives* (ICON) for analysing political communication online. Describing the goals of ICON, Seizov (2014, 28) acknowledges the systematic approach and layered architecture of the GeM framework as a source of inspiration. Seizov is, however, keen to point out how their goals dif-

fer: whereas ICON invests in visual content analysis, the GeM framework restricts itself to considering how graphic elements relate to other elements present on the page.

To make up for the shortcomings of the GeM framework in visual analysis, which Seizov considers crucial for analysing political communication online, he combines political iconography with multimodal analyses and communication research to define up to 16 content-analytical categories in ICON (Seizov 2014, 54). As such, Seizov's work does not only underline the effectiveness of a layered approach to studying multimodal phenomena, but shows how the GeM framework can inform analyses conducted within other fields.

Another type of extension contributing to methodological development can be found in Hiippala (2015*c*), which presents a full-scale application of the GeM framework to 58 double-pages from tourist brochures published by the city of Helsinki, Finland, between 1967–2008. The motivation for compiling this corpus was to develop a data-driven approach to describing single genres and their change over time. To examine the corpus, Hiippala (2015*b*) provides a set of tools for querying and visualizing the data stored in the corpus. These tools are now publicly available and have been used to interrogate and visualize GeM corpora in Zhang (2017).

4 Discussion

This section proceeds to discuss the reception and impact of the GeM framework. Beginning with criticism presented towards the framework in Section 4.1, the discussion considers some of its proposed shortcomings. Section 4.2 then turns towards the impact of the framework on multimodal research, while also outlining areas crucial for future research.

4.1 Criticism of the GeM framework

Martinec (2013, 149–150) argues against the use of RST as a framework for describing discourse relations between contributions from different semiotic modes. In particular, Martinec points out the problem of pinpointing what exactly makes the analysts decide upon a specific relation. Although this may be examined statistically by measuring agreement between analysts, the process of interpretation remains a 'black box' (see, however, Taboada & Mann 2006, 444–445). However, this problem with RST emerges mainly when considering intersemiotic relations without relating them to their immediate context of occurrence, a problem which also extends to other schemas for describing similar relations (Bateman 2016). Within the GeM framework, however, RST is used to describe how documents

achieve coherence, that is, how their parts work together towards a common goal. Capturing coherence is one aspect of account for the constraining effect of genre to which RST contributes – therefore it should not be reduced to a mere schema for describing intersemiotic relations.

In her review of Bateman (2008), Santini (2010) raises concerns about the applicability of the GeM framework to corpus-driven multimodal research, noting the lack of annotated corpora. The time- and resource-intensive nature of manual annotation certainly presents a formidable 'logjam' that prevents the creation of larger multimodal corpora. As pointed out above, however, this area of research is likely to benefit from the advances in machine learning and computer vision in the future (Bateman et al. 2016, Hiippala 2016*b*). At the same time, it is important to understand that the automatic processing of pages must go far beyond extracting written language, which would suffice for linguistic research. Doermann & Tombre (2014) provide a comprehensive introduction to the state of the art in document analysis, which has long acknowledged the need to understand the entire document structure in addition to extracting and processing its contents.

Santini (2010) also points out the lack of a clear definition of genre and labels for specific genres in Bateman (2008). This monograph, however, should be mainly seen as a starting point – its annotation schema providing a bare minimum for describing document genres. In other words, the GeM framework is first and foremost intended to support the efforts to map the genre space. Adopting a predefined set of genre labels would work directly against the main arguments in Bateman (2008): any definition of genre should be built on a description of genre space and its dimensions of variation, which should be in turn based on empirical research. The annotation schema provided by the GeM framework provides the means for doing this work.

In another review of the same monograph, Scott (2010, 241) argues that the GeM framework's focus on the "choices that documents have taken up in their design" (Bateman 2008, 271) shifts the attention away from 'meaning' to 'design', which Scott considers, following Kress & van Leeuwen (2001, 121), a "largely technical process". For this reason, Scott proposes that the GeM framework separates the processes of production and design, thus focusing on the end product, while neglecting the individuals responsible for creating the document.

This represents a misunderstanding that arises from attempting to fit the GeM framework into that proposed by Kress & van Leeuwen (2001). Unlike Scott (2010, 241) claims, the GeM framework does not assign agency to documents: the choices made in documents are always traced back to the semiotic modes and the individuals making use of them. In fact, as Hiippala (2016*a*) shows by combining ethnographic methods with the GeM framework, this combination can reveal much about agencies involved in the 'technical' and more 'semiotic' design

processes, providing a more precise view of their contributions and the constraints affecting the agencies that participate in the production of multimodal artefacts.

4.2 Impact and future research

With the major points of criticism covered, it is time to turn towards the impact of the GeM framework. Measured in terms of citations, as of May 2017 the GeM monograph (Bateman 2008) has been cited 440 times according to Google Scholar. This number is, of course, not reflected by the breadth of previous work discussed above. In most cases, references to the GeM framework serve to point out that the concept of genre has also been applied in multimodal research. Actual applications of the framework remain relatively rare; corpus-driven analyses even more so. This obviously raises the point already made by Santini (2010) in her review: is the application of GeM framework feasible in practice?

It should be clear that pursuing analyses analogous to corpus analyses in linguistics is still far away: a challenge which is by no means limited to the GeM framework, but faces the entire field of multimodal research (Bateman 2014*e*). The intermediate goals of the GeM framework, however, are more modest. As Bateman (2008) explicates:

> "Our concern is to articulate a framework within which it is possible to frame precise questions concerning the mechanisms by which a multimodal document goes about creating the meanings that it does." (Bateman 2008, 13)

This itself represents a considerable task, given the variation inherent to all document genres. What previous studies show is that the GeM framework has considerable potential for systematic analysis of multimodal genres realized in different media, being capable of teasing out these mechanisms. As Thomas (2014) notes, this helps to combat 'circularity' in multimodal analysis, whereby data analysis begins to be guided by the researcher's a priori assumptions. This crystallizes the main contribution of the GeM framework: increased empiricism in multimodal research.

To encourage the wider adoption of the GeM framework, increased efforts must be directed towards making its application easier and more efficient. Given that digital services are increasingly web-based, one possible solution would be to develop an environment or interface for conducting analyses, which would provide a set of tools for both manual and automatic annotation through a web-based service (cf. e.g. Würsch et al. 2016). Such modern web-based technologies could also enable annotators to collaborate, allowing the distribution of work and annotation of artefacts in a piecemeal fashion. Existing tools such as Thomas (2007) and

Hiippala (2016*b*) could be converted into backends responsible for data processing in these systems.

Increasing the size of multimodal corpora, in turn, could involve using crowd-sourced annotations via services such as Amazon Mechanical Turk (cf. Shank 2016). These services, in which non-experts are paid to undertake annotation tasks requiring human input, are frequently used to annotate complex multimodal data for artificial intelligence research (for a recent example of complex dataset of 5000 diagrams, see Kembhavi et al. 2016). These annotation tasks are typically broken down into small steps such as marking constituents and categorizing them, in order to enable non-expert annotators to perform the assigned tasks and to maximize the level of agreement between them (Kembhavi et al. 2016, 9).

5 Conclusion

To conclude, this review article has attempted to provide an overview of the research conducted within the Genre and Multimodality (GeM) framework for describing the multimodality of page-based documents (Bateman 2008). By introducing the central theoretical concepts, their respective contributions and application in analysis, the article sought to demonstrate how the framework provides a systematic, corpus-driven approach to describing multimodal documents. The article also identified certain challenges in full-scale application of the framework, which mainly emerge from the time- and resource intensive nature of compiling multimodal corpora. These challenges may be partially met by automating parts of the annotation process, which would enable researchers to take advantage of the descriptive and empirical capabilities provided by the GeM framework.

References

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

Bateman, J. A. (2009), Discourse across semiotic modes, *in* J. Renkema, ed., 'Discourse, of course: An overview of research in discourse studies', Benjamins, Amsterdam, pp. 55–66.

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

Bateman, J. A. (2013), 'Multimodal analysis of film within the GeM framework', *Ilha do Desterro* **64**, 49–84.

Bateman, J. A. (2014*a*), The constitutive role of semiotic modes for the theory

and practice of multimodal analysis, *in* C. DeCoursey, ed., 'Language Arts in Asia 2: English and Chinese through Literature, Drama and Popular Culture', Cambridge Scholars Publishing, Newcastle, pp. 8–33.

Bateman, J. A. (2014*b*), Developing a GeM (Genre and Multimodality) model, *in* S. Norris & C. D. Maier, eds, 'Interactions, Images and Texts: A Reader in Multimodality', De Gruyter Mouton, Berlin and New York, pp. 25–36.

Bateman, J. A. (2014*c*), Genre in the age of multimodality: some conceptual refinements for practical analysis, *in* P. Evangelisti-Allori, V. K. Bhatia & J. A. Bateman, eds, 'Evolution in Genre: Emergence, Variation, Multimodality', Peter Lang, Frankfurt am Main, pp. 237–269.

Bateman, J. A. (2014*d*), *Text and Image: A Critical Introduction to the Visual/Verbal Divide*, Routledge, London and New York.

Bateman, J. A. (2014*e*), Using multimodal corpora for empirical research, *in* C. Jewitt, ed., 'The Routledge Handbook of Multimodal Analysis', second edn, Routledge, London and New York, pp. 238–252.

Bateman, J. A. (2016), Methodological and theoretical issues for the empirical investigation of multimodality, *in* N.-M. Klug & H. Stöckl, eds, 'Sprache im multimodalen Kontext / Language and Multimodality', De Gruyter Mouton, Berlin, pp. 36–74.

Bateman, J. A. & Delin, J. L. (2003), Genre and multimodality: expanding the context for comparison across languages, *in* D. Willems, B. Defrancq, T. Colleman & D. Nöel, eds, 'Contrastive Analysis in Language: Identifying Linguistic Units of Comparison', Palgrave Macmillan, Houndsmills, pp. 230–266.

Bateman, J. A., Delin, J. L. & Allen, P. (2000), Constraints on layout in multimodal document generation, *in* 'Proceedings of the First International Natural Language Generation Conference, Workshop on Coherence in Generated Multimedia', Association for Computational Linguistics, Mitzpe Ramon, Israel.

Bateman, J. A., Delin, J. L. & Henschel, R. (2002), XML and multimodal corpus design: experiences with multi-layered stand-off annotations in the GeM corpus, *in* 'Proceedings of the LREC'02 Workshop: Towards a Roadmap for Multimodal Language Resources and Evaluation', Las Palmas, Canary Islands, Spain, pp. 13–20.

Bateman, J. A., Delin, J. L. & Henschel, R. (2004), Multimodality and empiricism: preparing for a corpus-based approach to the study of multimodal meaning-making, *in* E. Ventola, C. Charles & M. Kaltenbacher, eds, 'Perspectives on Multimodality', Benjamins, Amsterdam, pp. 65–89.

Bateman, J. A., Delin, J. L. & Henschel, R. (2007), Mapping the multimodal genres of traditional and electronic newspapers, *in* T. D. Royce & W. L. Bowcher, eds, 'New Directions in the Analysis of Multimodal Discourse', Lawrence Erlbaum, Mahwah, NJ, pp. 147–172.

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

Bateman, J. A., Tseng, C., Seizov, O., Jacobs, A., Lüdtke, A., Müller, M. G. & Herzog, O. (2016), 'Towards next generation visual archives: image, film and discourse', *Visual Studies* **31**(2), 131–154.

Bateman, J. A., Wildfeuer, J. & Hiippala, T. (2017), *Multimodality: Foundations, Research and Analysis – A Problem-Oriented Introduction*, De Gruyter Mouton, Berlin.

Bawarshi, A. S. & Reiff, M. J. (2010), *Genre: An Introduction to History, Theory, Research, and Pedagogy*, Parlor, West Lafayette.

Delin, J. L. & Bateman, J. A. (2002), 'Describing and critiquing multimodal documents', *Document Design* **3**(2), 140–155.

Delin, J. L., Bateman, J. A. & Allen, P. (2003), 'A model of genre in document layout', *Information Design Journal* **11**(1), 54–66.

Doermann, D. & Tombre, K., eds (2014), *Handbook of Document Image Processing and Recognition*, Springer-Verlag, London.

Elleström, L. (2010), The modalities of media: A model for understanding intermedial relations, *in* L. Elleström, ed., 'Media Borders, Multimodality and Intermediality', Palgrave, London, pp. 11–48.

Freadman, A. (2012), 'The traps and trappings of genre theory', *Applied Linguistics* **33**(5), 544–563.

Hiippala, T. (2012*a*), The localisation of advertising print media as a multimodal process, *in* W. L. Bowcher, ed., 'Multimodal Texts from Around the World: Linguistic and Cultural Insights', Palgrave, London, pp. 97–122.

Hiippala, T. (2012*b*), 'Reading paths and visual perception in multimodal research, psychology and brain sciences', *Journal of Pragmatics* **44**(3), 315–327.

Hiippala, T. (2014), Multimodal genre analysis, *in* S. Norris & C. D. Maier, eds, 'Interactions, Images and Texts: A Reader in Multimodality', De Gruyter Mouton, Berlin and New York, pp. 111–123.

Hiippala, T. (2015*a*), 'GeM-HTB: A multimodal corpus of tourist brochures produced by the city of Helsinki, Finland (1967–2008)'.
URL: *http://urn.fi/urn:nbn:fi:lb-201411281*

Hiippala, T. (2015*b*), '*gem-tools*: Tools for working with multimodal corpora annotated using the Genre and Multimodality model'.

Hiippala, T. (2015*c*), *The Structure of Multimodal Documents: An Empirical Approach*, Routledge, New York and London.

Hiippala, T. (2016*a*), 'Individual and collaborative semiotic work in document design', *Hermes – Journal of Language and Communication in Business* **55**, 45–59.

Hiippala, T. (2016*b*), Semi-automated annotation of page-based documents within

the Genre and Multimodality framework, *in* 'Proceedings of the 10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities', Association for Computational Linguistics, Berlin, Germany, pp. 84–89.

Hiippala, T. (2017), 'The multimodality of digital longform journalism', *Digital Journalism* **5**(4), 420–442.

Holsanova, J. (2014), Reception of multimodality: applying eye-tracking methodology in multimodal research, *in* C. Jewitt, ed., 'The Routledge Handbook of Multimodal Analysis', second edn, Routledge, London, pp. 287–298.

Jewitt, C. (2014), Different approaches to multimodality, *in* C. Jewitt, ed., 'The Routledge Handbook of Multimodal Analysis', second edn, Routledge, London, pp. 31–43.

Kaltenbacher, M. (2004), 'Perspectives on multimodality: from the early beginnings to the state of the art', *Information Design Journal + Document Design* **12**(3), 190–207.

Kembhavi, A., Salvato, M., Kolve, E., Seo, M., Hajishirzi, H. & Farhadi, A. (2016), A diagram is worth a dozen images, *in* 'Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV', Springer, Cham, pp. 235–251.

Kong, K. C. C. (2013), 'A corpus-based study in comparing the multimodality of Chinese- and English-language newspapers', *Visual Communication* **12**(2), 173–196.

Kress, G. (2011), 'Partnerships in research: multimodality and ethnography', *Qualitative Research* **11**(3), 239–260.

Kress, G. (2015), 'Semiotic work: Applied Linguistics and a social semiotic account of multimodality', *AILA Review* **28**(1), 49–71.

Kress, G. & van Leeuwen, T. (1996), *Reading Images: The Grammar of Visual Design*, Routledge, London.

Kress, G. & van Leeuwen, T. (2001), *Multimodal Discourse: The Modes and Media of Contemporary Communication*, Arnold, London.

Lemke, J. L. (1999), 'Typology, topology, topography: genre semantics', MS University of Michigan.

Martinec, R. (2005), Topics in multimodality, *in* R. Hasan, C. M. Matthiessen & J. J. Webster, eds, 'Continuing Discourse on Language: A Functional Perspective', Vol. 1, Equinox, London, pp. 157–181.

Martinec, R. (2013), 'Nascent and mature uses of a semiotic system: the case of image-text relations', *Visual Communication* **12**(2), 147–172.

Nekić, M. (2015), *Tourist Activities in Multimodal Texts: An Analysis of Croatian and Scottish Tourism Websites*, Palgrave Macmillan, London.

O'Toole, M. (1994), *The Language of Displayed Art*, Leicester University Press,

London.

Reichenberger, K., Rondhuis, K., Kleinz, J. & Bateman, J. A. (1995), Effective presentation of information through page layout: a linguistically-based approach, Technical Report 970, Institut für Integrierte Publikations- und Informationssysteme (IPSI), GMD, Darmstadt.

Santini, M. (2010), 'Book review: Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents', *Linguist List* **21**(1606).
URL: *http://linguistlist.org/issues/21/21-1606.html*

Scott, M. (2010), 'Book review: Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents', *Visual Communication* **9**(2), 241–245.

Seizov, O. (2014), *Political Communication Online: Structures, Functions, and Challenges*, Routledge, New York and London.

Shank, D. B. (2016), 'Using crowdsourcing websites for sociological research: The case of Amazon Mechanical Turk', *The American Sociologist* **47**(1), 47–55.

Swales, J. M. (1990), *Genre analysis: English in Academic and Research Settings*, Cambridge University Press, Cambridge.

Taboada, M. & Mann, W. C. (2006), 'Rhetorical structure theory: looking back and moving ahead', *Discourse Studies* **8**(3), 423–459.

Thomas, M. (2007), Querying multimodal annotation: A concordancer for GeM, *in* 'Proceedings of the Linguistic Annotation Workshop', Association for Computational Linguistics, Prague, Czech Republic, pp. 57–60.
URL: *http://dl.acm.org/citation.cfm?id=1642059.1642069*

Thomas, M. (2009), Localizing pack messages: A framework for corpus-based cross-cultural multimodal analysis, PhD thesis, University of Leeds.

Thomas, M. (2014), 'Evidence and circularity in multimodal discourse analysis', *Visual Communication* **13**(2), 163–189.

van Leeuwen, T. (2005), Multimodality, genre and design, *in* S. Norris & R. H. Jones, eds, 'Discourse in Action: Introducing Mediated Discourse Analysis', Routledge, London, pp. 73–94.

van Leeuwen, T. & Hestbæk Andersen, T. (2017), 'Genre crash: The case of online shopping', *Discourse, Context & Media*.

Waller, R. H. W. (1987), The typographic contribution to language, PhD thesis, Department of Typography and Graphic Communication, University of Reading.

Waller, R. H. W., Delin, J. & Thomas, M. (2012), 'Towards a pattern language approach to document description', *Discours* **10**.

Würsch, M., Ingold, R. & Liwicki, M. (2016), 'DIVAServices – a RESTful web service for Document Image Analysis methods', *Digital Scholarship in the Humanities* DOI: 10.1093/llc/fqw051

Zhang, K. (2017), Public Health Education through Posters in Two World Cities:

A Multimodal Corpus-Based Analysis, PhD thesis, The Hong Kong Polytechnic University.