

NB! Author's version: do not cite!

Use the following reference instead: Hiippala, T. (2012) *Reading paths and visual perception in multimodal research, psychology and brain sciences*. In Journal of Pragmatics. DOI: 10.1016/j.pragma.2011.12.008

Reading paths and visual perception in multimodal research, psychology and brain sciences

Tuomo Hiippala
Department of Modern Languages
University of Helsinki

Abstract

This paper argues that the concept of a reading path in multimodal research can be improved by previous research on visual perception in psychology and brain sciences, and particularly by the work done within eye-tracking studies. The paper argues that in its current state, the concept of a reading path is not sufficiently reliable due to the lack of empirical testing and therefore presents a methodological proposal to improve the current situation.

Thus, the paper identifies common areas of interest related to visual perception, where the research interests of the disciplines meet and enable reciprocal input. It is suggested that multimodal research is capable of describing the high-level factors that affect visual perception, whereas eye-tracking equipment can track the actual reader behaviour. Applicable state-of-the-art theories of multimodal analysis are then described, along with the technological requirements for the eye tracker and its software.

XML annotation, output and transformations are proposed for combining the results of multimodal analysis and the observer behaviour captured using an eye tracker. Finally, the paper presents a hypothesis on the relationship of visual perception and multimodal semiosis, which may be evaluated using the proposed method combining multimodal analysis and eye-tracking.

1 Introduction

The concept of a reading path has received increased attention in recent multimodal research, from the perspectives of both designer (Kress 2003) and observer (Lim 2004, White 2010). Reading paths constitute an important domain of multimodal research, as the concept seeks to shed light on how multimodal artefacts

are read and interpreted. At the same time, eye-tracking experiments by Holanovna & Holmqvist (2006) have shown limited support for otherwise influential hypotheses on reading paths in social semiotics (see e.g. Kress & van Leeuwen 1996, 2006). This discrepancy poses a significant problem, as multimodal research needs increased reliability, if it is to develop in an empirically responsible direction (Kaltenbacher 2004, p. 202).

Kappas & Olk (2008) have pointed out that any research into the visual domain is likely to benefit from the basic research and advances in psychology and brain sciences:

Whether we are watching a soap opera, browsing through a catalogue, admiring a sculpture at an exhibition or glancing at the face of a colleague for signs of approval, a complex set of processes in our brain related to vision is involved in making sense of the stream of information that our eyes provide. Vision is a highly complex interaction with our environment that relies on learned information and is shaped by biological constraints of our brain. (Kappas & Olk 2008, p. 162)

Based on this view, several focal points may be identified where the research interests of multimodality meet those of psychology and brain sciences. There is a growing body of work on film (see e.g. O'Halloran 2004, Bateman 2007, 2009*b*, Tan 2009, Tseng & Bateman 2010, Bateman & Schmidt 2011), print media (see e.g. Cheong 2004, Martinec & Salway 2005, Royce 2007), sculptures and exhibitions (O'Toole 1994, Hofinger & Ventola 2004, Stenglin 2009). However, a question remains to be answered: how can multimodality, psychology and brain sciences work together towards the description of visual perception and multimodal phenomena?

Multimodal research describes what is learned and how the **learned information** influences our interaction with multimodal artefacts, whereas psychology and brain sciences possess the necessary theories and methods to study the biological aspects of visual perception. Potential benefits are promising, as both fields have previously engaged in the study of similar data, such as biology textbooks (see e.g. Hannus & Hyönä 1999, Kress 2003, Guo 2004, Baldry & Thibault 2005). In such a context, this paper takes a step forward in bridging the gap between the two disciplines. Firstly, the paper draws on psychology and brain sciences for complementary perspectives to the study of reading paths and visual perception in multimodal research, and identifies areas of common interest between the disciplines. Secondly, the paper argues that multimodal analysis has to be methodologically reliable, if hypotheses are to be made about the relationship of visual perception and multimodal phenomena. Consequently, the paper presents a methodological proposal that increases the analytical reliability in studying multimodality and reading behaviour.

The paper begins with a brief introduction to multimodality. The subsequent sections continue with a review of the available work on reading paths in multimodal research, while also providing complementary perspectives from the fields of psychology and brain sciences. The review aims to identify key areas of interest that are contested by either field, in order to tease out the domains of research where the disciplines may benefit from reciprocal input. Finally, the paper concludes with a methodological proposal for developing integrated methods, in order to encourage cross-disciplinary work between the disciplines in both theoretical and applied research. Additionally, two hypotheses that may be tested using the proposed method are presented.

2 A brief introduction to multimodal research

In her introduction to the core theoretical concepts of multimodality, Jewitt (2009, p. 14) describes the general characteristics of the field as follows:

Multimodality describes approaches that understand communication and representation to be more than about language, and which attend to the full range of communicational forms people use — image, gesture, gaze, posture, and so on — and the relationships between them.

On the basis of our current understanding of multimodality, it is not an overstatement to put forward the idea that every communicative event is inherently multimodal. Spoken language is constantly combined with gestures, whereas many forms of written communication combine language with image in their representation, regardless of the medium used. Previously, aspects of multimodal communication have been studied independently in various disciplines, such as communication and media studies, anthropology, art history, design studies and semiotics (cf. Kaltenbacher 2004). Multimodal analysis, in turn, describes the various aspects of communication and semiosis in connection with each other, in order to tease out their internal structure, external relationships and functions in specific contexts. Most importantly, multimodal analysis is oriented towards the description of structure, whereas previous work in communication studies and semiotics has tended to focus on the description of content (see e.g. Barthes 1977, Williamson 1978).

However, despite over two decades of research, our understanding of multimodality is still relatively limited. As Bateman (2008, p. 11) points out, we need to acknowledge multimodality as a phenomenon of complex nature. He further notes that an understanding of the mechanics of multimodal meaning-making still requires considerable effort in research. Moreover, multimodal research needs to be increasingly reliable, evaluable and free of pre-structured conceptions about

the nature of multimodal phenomena. In future, digital tools and corpora will be prime candidates for enhancing multimodal research in terms of data analysis and annotation (cf. O'Halloran et al. 2010, 2011, Parodi 2010). As this paper shows, there is also a growing interest in how multimodal artefacts are perceived and interacted with. In this area, significant advances are likely to emerge from eye-tracking research (cf. Holsanova & Holmqvist 2006, Holsanova & Nord 2010). In order to provide a context for the discussion, the following section outlines the development of reading paths as a theoretical concept for describing the readers' interaction with multimodal artefacts and traces the emergence of the concept in multimodal research.

3 First steps: reading paths in early multimodal research

The interest in reading paths emerged at an early stage in multimodal research: van Leeuwen (1993) postulated that reading paths are construed by the spatial placement of verbal and visual elements, by their contrastive visual salience and by their configuration as a part of the layout. In his analysis, van Leeuwen (1993, p. 214-5) proposed that a reading path proceeds through visually salient images to visually salient text. At certain stages of a reading process, the verbal and visual semiotic resources combine to fulfil specific communicative functions, which van Leeuwen (1993, p. 215) termed "semiotic acts". An important contribution of van Leeuwen (1993, p. 214) came in the form of identifying three areas of further research, which he saw as necessary for developing a "semiotic theory" of reading paths:

1. Cultural patterns of reading (direction: left-right, right-left or top-bottom).
2. Perceptual salience, based on the psychology of perception (the hierarchies of contrast, colour hue and saturation, sharpness, etc.).
3. Semantic factors, which may override perceptual factors (such as the salience of the human figure).

Considering the early stage of multimodal research at the time, van Leeuwen's perspective was particularly insightful, especially in the light of future research.

Van Leeuwen's proposal may be complemented with insights from psychology and brain sciences, after a brief glance into psychological terminology. In psychological terms, the observed phenomena are referred to as stimuli for the visual sense. In describing the process of observation, eye-tracking studies use the term fixation to indicate the point of attention, whereas the jumps between fixations

are known as saccades. Saccades are high-velocity eye movements (up to 500° per second), during which no new information is obtained (Rayner 1998, p. 373). The time spent around a fixation point is called dwelling time. Following Kappas & Olk (2008), the term ‘observer’ is used in place of ‘reader’, in order to emphasise the multimodal characteristics of the visual stimuli.

As for point (1) above, recent psychological research suggests that cultural factors have both temporary and long-term effects on certain perceptual processes (for discussion, see Nisbett et al. 2001, Nisbett & Miyamoto 2005). On the basis of a growing body of research, Nisbett & Miyamoto (2005, p. 472) argue that:

People in Western cultures have been found to organize objects by emphasizing rules and categories and to focus on salient objects independently from the context, whereas people in East Asian cultures are more inclined to attend to the context and to the relationship between the objects and the context.

This suggests that Asian observers are oriented towards a holistic tendency in perception, whereas the Western observers pay more attention to individual objects. Cross-cultural differences have also been observed in multimodal research. In a study of English and Japanese procedural texts, Martinec (2003, p. 51) suggests that recipes in Japanese cookbooks engage the reader to a greater degree through the combined use of language and image, which results in a greater emphasis on detail than in their English counterparts.

	Description	Implications
Low-level factors	Contrast, colour, texture and luminance	Areas of higher contrast attract attention.
Intermediate factors	Shape and spatial relations	Shapes that differ from surrounding stimuli attract attention.
High-level factors	Short- and long-term memory	Previous spatial and semantic knowledge about similar stimuli guide perception.

Table 1: Low-, intermediate and high-level perceptual factors (based on Kappas & Olk 2008, p. 164-5)

In relation to point (2) presented earlier, we may draw on a tri-stratal division of the factors simultaneously affecting perception (Kappas & Olk 2008, p. 164-5). These are the low-, intermediate and high-level factors, which are described

in Table 1. The high-level factors in Table 1 constitute a particular point of interest for multimodal research, due to an emerging interest in the mechanisms of interpretation as a part of the research on semiotic modes (Bateman 2009a, 2011). A semiotic mode consists of three components: material substrate, semiotic resources and discourse semantics. The *material substrates* have emerged over time, as the substrates have established themselves as suitable carriers of meaning. The currently dominant substrates of page and screen allow the realisation of multiple *semiotic resources*, such as language and image. However, the multimodal combinations of language and image only become interpretable in context: this logic is provided by *discourse semantics* (Bateman 2011, p. 21-2). As a kind of learned information described by Kappas & Olk (2008), the discourse semantics, as a component of a semiotic mode, can be used to describe models of spatial (in terms of the material substrate and the space it affords) and semantic knowledge (configuration of the semiotic mode) that guide visual perception on a higher level. Thus, the concept of a semiotic mode is a particularly promising concept for cross-disciplinary research: we will return to the notion of a semiotic mode in Subsection 5.2.

In terms of the previously introduced point (3), there indeed are certain semantic factors that may override low- and intermediate level perceptual factors, such as the salience of the human face, and particularly that of eyes and lips, which are the most expressive elements of a face (Yarbus 1967, p. 191). Kappas & Olk (2008, p. 165-6) elaborate this point further by pointing out that for adults, a face is “a source of information about the identity of other human beings, their age and gender, but also their current intentions, attitudes and feelings”, while infants use faces to learn about themselves and their immediate social and physical surroundings. Therefore, it is not surprising that several multimodal frameworks have paid attention to the interpersonal function of gaze and included it in their models (see e.g. O’Toole 1994, Royce 1998, O’Halloran 1999, 2008).

Unfortunately, the development of a theory of reading paths that would integrate perceptual psychology as proposed by van Leeuwen (1993) has not been followed up in subsequent multimodal research. In fact, it may be suggested that the originally conceived concept of a reading path became swamped by semiotically-oriented, interpretative multimodal frameworks (for related criticism, see Bateman 2008, p. 13). The reliance on semiotic theories prevented the concept from evolving into a more sophisticated form through empirical research and its feedback into the theory. Therefore, the notion of a reading path needs to be reconsidered. With the broad principles of visual perception now established, the following section proceeds to deconstruct some of the semiotic notions related to reading paths and observer behaviour that emerged in later multimodal research.

3.1 Scanning

The notion of scanning was first introduced by van Leeuwen (1993) as an activity that precedes the observational process, whereby images are generally given a priority over language. This view was followed up in Kress & van Leeuwen (1998, p. 215), who described scanning as a process during which the observer sets up connections between different elements and their relations in terms of “relative importance”. The relative importance was thought to be determined by the contrastive visual salience of the elements. Lim (2004) further developed the notion of scanning by introducing the concept of “centre of visual impact” (CVI), which captures the observer’s attention and functions as an entry point to the text, thus marking the beginning of a reading path. At first sight, scanning appears as a plausible and natural process. Psychological research, however, contests this assumption: there are no separate processes of scanning and observation. In an extensive overview of eye-tracking research, Rayner (1998, p. 379) asserts that “because the nature of the search task influences eye movement behaviour, any statement about visual search and eye movements needs to be qualified by the characteristics of the search task”, which suggests that scanning and observation are two aspects of the same process, and therefore intertwined and inseparable.

The importance of the performed search task and its implications for visual perception lead us to an important observation in the seminal study of Yarbus (1967, p. 182-3). Yarbus concluded that the observer’s attention shifts towards the elements that are perceived as relevant to the performed task, regardless of their visual realisation in terms of level of detail and the use of colour. This early observation has far-reaching consequences for multimodal research, where multimodal semiosis as a part of the design process has been commonly regarded as the source of the reading paths (O’Halloran 1999, Lim 2004). For instance, O’Halloran (1999, p. 324) describes multiple semiotic resources as systems of choice, where “selections function within each system so that interactions between semiotics become the focal point at different stages”. According to O’Halloran, these focal points mark the reading path, bearing close resemblance to the concept of CVI (Lim 2004).

Indeed, semiotic resources tend to cluster, as the use of resolution reduction techniques has shown in a study of visual grouping in document layout (Reichenberger et al. 1996). The semiotic resources also form focal points in design, which exploit the interface between the rhetoric and the visual by emphasising rhetorical segments by typographic means (Delin & Bateman 2002). But unlike design, visual perception is largely task-driven and dependent on the sought information (cf. Rayner 1998). Furthermore, in their comparative study of multimodal theories and actual reading behaviour measured using an eye-tracker, Holsanova & Holmqvist (2006, p. 88) conclude that “readers do not scan the semiotic space

before taking a closer look at certain units”.

To conclude, it appears that at least some of the research in psychology and brain sciences contradicts the assumptions on the process of scanning in multimodal research. It should also be noted that multimodal perspectives on reading paths and scanning have evolved in parallel. Whereas van Leeuwen (1993, p. 214) proposed the priority of image over language in visual perception, Lim (2004, p. 228) advocates “perceptual equity” between these semiotic resources. Kress (2003, p. 159), in turn, suggests that scanning involves deciding whether one of the semiotic resources is dominant or whether they are equal, which has consequences for identifying their function in the multimodal artefact. These assumptions are subject to the same shortcomings as those related to scanning.

It may be argued that the main challenges in developing the concept of a reading path in multimodal research have resulted from a lack of attention to the roles of designer and observer. There has to be a relationship between the roles; otherwise there could be no agreement on the conventions of semiotic resources and their deployment, or their interpretation through the discourse semantics of the semiotic modes used. With this in mind, the following section will explore the relationship between designer and observer and its implications for the concept of a reading path.

3.2 The relationship between designer and observer

In his well-known study of multimodal literacy, Kress (2003, p. 4) puts forward an argument about various facets of reading paths and their construal:

Reading paths may exist in images, either because the maker of the image structured that into the image — and it is read as it is or it is transformed by the reader, or they may exist because they are constructed by the reader without prior construction by the maker of the image.

In short, Kress suggests that reading paths may be created in three ways: by the designer, by a process of transformation, or by the reader (see Figure 1). Kress hereby adopts a semiotic perspective on the construal of reading paths. However, from the perspectives of psychology and brain sciences, visual perception is a complex and multifaceted process shaped by both biological constraints, and our cultural and social knowledge. The designer can exploit some aspects of visual perception, but has little control over the perceptual processes in more general terms. Therefore, the real point of interest for multimodal research lies in what Kress calls the processes of transformation.

Transformation is obviously an abstract term to describe a group of processes between the designer and the observer. For instance, Kostelnick & Hassett (2003)

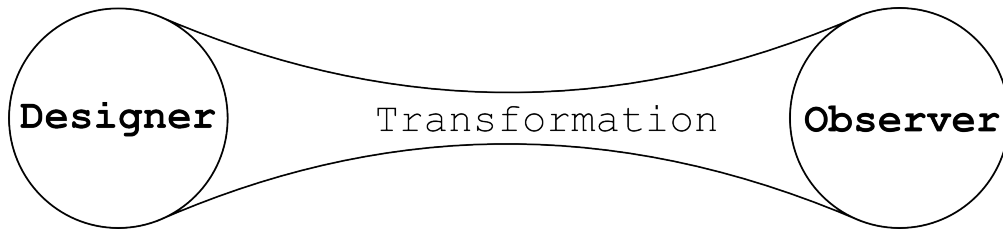


Figure 1: The transformation of reading paths (based on Kress 2003, p. 4)

approach the issue from the perspective of conventions, which they see as a set of constantly evolving social agreements between designers and observers. Naturally, the notion of convention is an equally abstract concept as that of transformation. Conventions, however, provide a connection to multimodality, as they are essentially configurations of semiotic modes in a given context. This observation on the semiotic modes may then be connected to the high-level factors in visual perception and to the processes of transformation: the observer requires previous knowledge concerning the configuration of semiotic resources in order to accomplish the search task. The previous knowledge of human semiosis and its conventions facilitates the search task by guiding visual perception towards the meanings that are considered relevant to the task at hand. These processes of inference may again be described using the discourse semantic component of a semiotic mode (cf. Bateman 2011), presuming we are able to capture the principles that govern the inferential processes.

A more practical example follows: as the previously discussed work of Yarbus (1967) showed, eye-movement is influenced by the search task. In his experiment, Yarbus (1967, p. 174) presented participants with *Unexpected Return* – a painting by the Russian realist painter Ilya Repin – and asked them to evaluate the material circumstances, age, relationships of the family portrayed in the painting, and to recall their spatial placement. Different tasks resulted in different patterns of fixation points and saccadic eye movements. From this result, it is possible to deduce that the observers had certain knowledge of where to direct their attention, that is, they were directed towards the meanings that were necessary to complete the search task. The construal of visual meanings in paintings and other works of art, explored in the seminal work of O’Toole (1994) could thus be studied in relation to the patterns of visual perception. It is exactly this kind of work that may be used to establish a contact point between multimodal research, psychology and brain sciences. The following section discusses the methodological requirements for such interdisciplinary research.

4 A step forward: cross-disciplinary benefits and challenges

So far we have established that multimodal research is likely to benefit from advances in psychology and brain sciences, at least when the study of visual perception is concerned. The question is whether and what multimodal research can contribute to the field of psychology and particularly to the description of the high-level factors that affect visual perception. In its current state, multimodal research resembles more of an analytical toolkit with various approaches and data than a fully developed theory of communication, although the ongoing work contributes to the continuous process of theory building (Jewitt 2009, p. 26). Multimodal research has also been scaled to accommodate both general and specific questions about the studied phenomena; we will now look at these approaches tie in with the research in psychology and brain sciences.

Beginning with the more abstract descriptions of multimodality in the work of van Leeuwen (2005) on multimodal genre, we again encounter the concept of a reading path, which is used to “reintroduce linearity in the case of spatially structured text” (van Leeuwen 2005, p. 81-2). In his analysis of a website for a home electronics manufacturer, van Leeuwen draws on systemic-functional models of genre, in which genres are described as staged, goal-oriented social processes (see e.g. Christie & Martin 1997, Martin & Rose 2008). In systemic-functional linguistics, genres are seen as recurrent configurations of meaning, which are used to enact the social processes within a culture (Martin & Rose 2008, p. 6). In terms of structure, as a staged process a genre has to unfold through multiple stages. Van Leeuwen uses the concept of a reading path to define the following stages of the website:

1. Welcome
2. Product choice
3. Product information
4. Price
5. Ordering

It is important to underline that van Leeuwen does not discuss the structuring of the multimodal artefact, but the structuring of the presumed reading process that completes the performed action, modelled using the concept of genre. Keeping in mind the perspective of genre and how genres are used to “accomplish things”, a multimodal artefact provides an environment for the staged process. As for

the specific stages identified by van Leeuwen, we have already established that visual perception is also goal-oriented: the genre stages may be thought of as consisting of specific tasks. Depending on what the observer wishes to accomplish, visual perception may be directed towards a particular stage: it is an active and dynamic process. What we need, then, are analytical tools to describe particular stages and their multimodal construal, from the perspectives of both structure and content (Holsanova & Nord 2010, p. 102).

For example, Martinec & Salway (2005) and Kong (2006) have presented frameworks for analysing the interaction between language and image. The phenomenon has also been studied using real-life data such as magazine advertisements and tourist brochures (see e.g. Royce 1998, 2007, Cheong 2004, Kvåle 2010). Despite the elaborate frameworks, the descriptive capability of these models remains restricted due to the limited availability of theories with sufficient empirical backing to contextualise these investigations. As Forceville (2007) has noted, the process of theory building necessitates not only detailed descriptions of phenomena, but also abstract, top-down descriptions to complement the detailed analyses. The benefits of a well-researched theory with multiple strata is evident in the description of visual perception: the low-, intermediate and high-level factors that affect perception complement each other and allow the analyst to make statements about them in relation to each other. A similar reach and capability is necessary for multimodal research, if it is to complement the research in psychology.

In the context of genre, van Leeuwen (2005, p. 85) also makes observations on future work in multimodality, which underline the previously mentioned key requirements for combining the perspectives of multimodal research and psychology. He identifies the following factors:

1. Visual analysis of the text, to study the environment of the staged, goal-oriented process, and the pathways it allows.
2. Observational, ‘ethnographic’ genre analysis of the user’s trajectory, to study actual staged, goal-oriented reading processes, and so access the (usually internalised) generic patterns that inform it.

Van Leeuwen’s observation provides an important line of demarcation that concerns both fields of study, as it outlines the areas of responsibility for each field.

Point (1) has to be reworked to explicitly include not only the visual, but the multimodal analysis of a text. Similarly, the analysis should not be approached only from the perspective of genre, but from multiple aspects and at various levels of detail. Most importantly, the analysis needs to provide a comprehensive picture of the principles of multimodal meaning-making across semiotic strata, providing the much needed theoretical backdrop. For this, we need more elaborate models of multimodal meaning-making, which are rigorous in their methods and based on

observation, as the structure of semiotic resources is known to be metaredundant (Martin 1997). This means that a semiotic resource forms patterns across strata, resulting in patterns of patterns of patterns and so on: capturing the patterns requires a theory that accommodates the strata required for their description.

Point (2) is much in line with the task-based view of visual perception. Eye-tracking methodology provides the tools to trace the observer's trajectory with precision in terms of fixation points, saccades and dwelling times. At the same time, van Leeuwen presents a significant challenge: how are we going to access the internalised generic patterns that guide our interaction with multimodal artefacts? In this case, we are obviously dealing with mainly high-level factors affecting visual perception, which can be complemented by multimodal research. To exemplify this issue, we may identify a particular stage that fulfils a certain task in a goal-oriented sequence. There are then at least two answers we need to provide: how a particular stage is construed multimodally and what are the circumstances that make the observer direct attention towards it? The following section proposes a method for undertaking such research.

5 Combining multimodal analysis and eye-tracking

This section presents a methodological proposal for combining multimodal analysis and eye-tracking research. At this point, it is necessary to emphasise that the method in question has not been tested in practice due to the lack of resources, but relies solely on the known capabilities of the described models and technologies. Finally, two hypotheses that may be tested using the introduced method are presented.

A double-page spread (shown in Figure 2) from Kara et al. (1986, p. 76-77) is discussed to illustrate the method and to highlight the critical issues related to multimodal research, eye-tracking research and their pedagogical applications. An identical double-page from a later pressing of Kara et al. (published in 1989) was used in an eye-tracking study by Hannus & Hyönä (1999). The double-page is a passage from a Finnish elementary school biology textbook, which describes the food consumption and reproduction of common flies. Although Hannus & Hyönä (1999) included other passages in their experiments and modified them for their purposes, we will not discuss the fly passage from the perspective of the experiment, but rather use it to illustrate the application of multimodal theories in its analysis. However, some background information is necessary: Hannus & Hyönä studied the utilisation of illustrations by elementary school pupils in biology textbooks using eye-tracking equipment. Their experiments showed that high-ability children performed better in integrating the relevant passages of text and illustrations, which were required to answer the more demanding comprehension

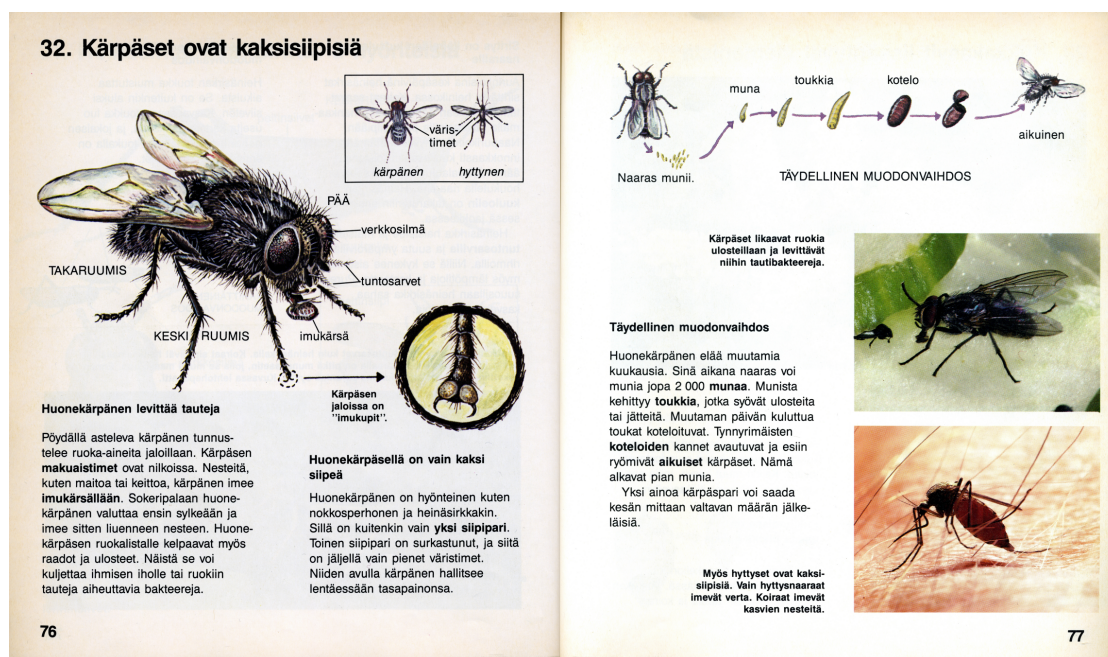


Figure 2: The fly passage in Kara et al. (1986, p. 76-7)

questions about the textbook passages (Hannus & Hyönä 1999, p. 107-108).

The performance of the high-ability students raises the question of multimodal literacies and the application of multimodal theories to the field of pedagogy. Both of these issues have been discussed in previous research (see e.g. Kress 1998, Cope & Kalantzis 2000, Kress 2003), but additional work is undoubtedly required, as many questions about multimodal meaning-making remain unanswered. As Livingstone (2004, p. 12) has observed, our limited knowledge of multimodality prevents us from making reliable statements about multimodal literacy:

Until we have a robust account of the media in which people might be judged literate, we can say little about the nature or uses of their literacy.

A sufficiently robust account of multimodality may only be produced by empirically-oriented research, which produces evaluable analyses that may be tested for support using, for example, corpora or eye-tracking equipment. With these methodological requirements in mind, the following subsections detail the methodological proposal presented in this paper.

5.1 Analysing the multimodal structure of a biology textbook

Previous multimodal research on biology textbooks has underlined their use of language, image, and their interrelations in the meaning-making process (Guo 2004, Baldry & Thibault 2005). In short, the biology textbooks rely on multimodality to perform their communicative tasks. This is also evident in Figure 2 and serves as our point of departure for a discussion of the passage’s multimodal characteristics. The passage consisting of the double-page spread is considered not only a part of a larger multimodal artefact, but a representative of a specific type of a multimodal artefact, that is, a textbook.

The structure of the textbook as a multimodal artefact has to be considered first. In this context, the relevant questions about the multimodal structure of the textbook are related to the artefact’s configuration of the semiotic modes in the two-dimensional space. What are the specific functions of language and image, are they organised into functional clusters, and do they use the two-dimensional space to communicate additional meanings? The passage in Figure 2 shows text paragraphs, callouts, captions and illustrations. How do we then move beyond these superficial labels to describe the fly passage?

Attempting to characterise the structure of a textbook as a multimodal artefact inevitably involves comparing it against other artefact types. The relations between artefacts and texts have been often approached from the perspective of genre (see e.g. Christie & Martin 1997, Lemke 1999, Baldry & Thibault 2005). In multimodal research, the state-of-art is represented by the *Genre and Multimodality* model (hereafter GeM; for a description, see Bateman 2008), which is described in Subsection 5.3. First, it is necessary to look at the concept of a semiotic mode in greater detail, in order to identify the specific semiotic modes at play in the fly passage.

5.2 Semiotic modes

The notion of a mode is a foundational concept in multimodal research and a focal point of research (for recent work, see e.g. Kress 2009, Elleström 2010). For the current purposes, we will draw on the work of Bateman (2011), whose tri-stratal model of a semiotic mode was introduced briefly in Section 3. To reiterate, the three strata are a material substrate, semiotic resources and discourse semantics. Together, they make up the notion of a semiotic mode. Bateman (2008) identifies three distinct semiotic modes, which are termed *text-flow*, *image-flow* and *page-flow*: their detailed descriptions can be found in Bateman (2009a, 2011).

The discussion begins with a description of text-flow. According to Bateman (2009a, p. 61), text-flow is a semiotic mode that organises text into linear-

illustration designated as (2b), where we move into the domain of page-flow.

The third and final semiotic mode is page-flow, whose defining feature is the use of the two-dimensional space to communicate additional meanings. As Bateman (2011, p. 26) writes:

[Page-flow] relies upon the complete two-dimensional space of the ‘canvas’ provided by the physical substrate and uses proximity, grouping of elements, framing and other visual perceptual resources in order to construct patterns of connections, similarity and difference.

However, the affordances of page-flow are not limited to those described above. It can also incorporate instances of text-flow and image-flow, as the double-page in Figure 3 shows in its entirety. As the fly passage indicates, page-flow is not subject to the principle of linear organisation. Instead, rhetorically organised units are formed based on the principles described in the quote above.

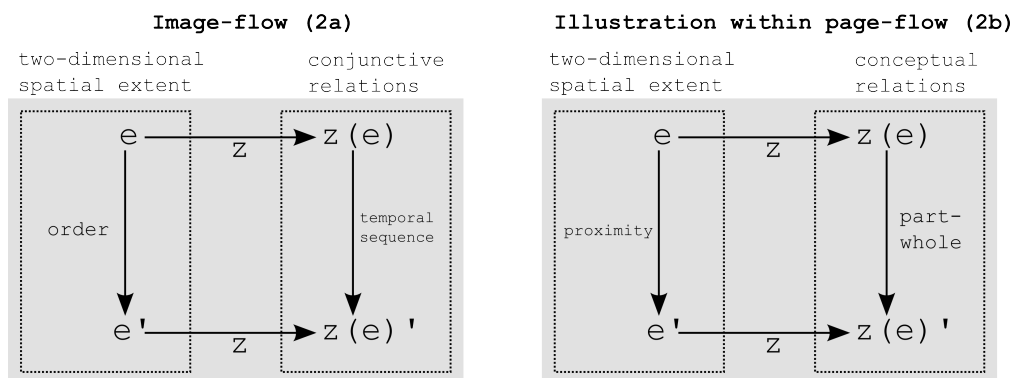


Figure 4: Formal back-and-forth mappings between two-dimensional space and discourse semantics in illustrations (2a) and (2b) in Figure 3 (after Bateman 2011)

Consider the hypothetical position of a student faced with the task of retrieving information about a common fly using the passage: the student has to possess an understanding of the discourse semantics of the semiotic modes at play, in order to arrive to the correct interpretations. Figure 4 shows a “back-and-forth” mapping of the discourse semantics in illustrations (2a) and (2b) and the logic behind the correct interpretations (for a further discussion of discourse semantics, see Bateman 2011): we will now take a closer look at the mappings.

For both illustrations (2a) and (2b), the left domain in Figure 4 represents the material substrate of the page. On the page, there is a relation between the entities e and e' . In the case of (2a), the two-dimensional spatial extent is mapped according to the principle of order, whereas in (2b) the principle is proximity. The domains on the right represent the discourse semantics: a mapping,

indicated by z , holds between the two domains. This means that if a relation holds between entities in one domain, a corresponding relation holds in the other domain as well. In the domain of discourse semantics, relationships of temporal sequence (2a) and part-whole (2b) hold respectively between the mapped entities $z(e)$ and $z(e)'$. In (2a), the two-dimensional space is mapped with time, whereas in (2b) the two-dimensional space is used to indicate a part-whole relationship. In order to successfully interpret the images, the student has to possess the necessary knowledge of the discourse semantics in the material substrate of the page.

Increasing our understanding of the processes of “defeasible inference” (cf. Bateman 2011, p. 22) that allow the student to arrive at the correct interpretation is a candidate area for improving multimodal literacy. Therefore, the formalisation of the principles of discourse semantics is a complex but necessary task, as it provides the means to discuss the meaning-making processes at an abstract level. This enables us to move the discussion beyond mere labels and to capture the principles behind them: we are not only capable of identifying the elements on the page and their interrelations, but we can also describe their internal logic and structure. As our knowledge of multimodality increases, we may be able to pinpoint configurations that typically result in erroneous inference, which has significant potential in pedagogical applications. With the theoretical concept of semiotic modes now established, the following subsection describes the analytical method.

5.3 The GeM model

The GeM model aims to provide the necessary analytical tools to describe the multimodal structure of an artefact: the analytical layers are described in Table 2. As the name of the model suggests, genre is a foundational notion within the model: it does not only provide a tool of comparison, but enables theorising about the relations between the genres and the historical and social factors that define them (Bateman 2008, p. 9-10).

The remainder of this subsection describes how the GeM model could be used to describe the fly passage in Figure 2: Table 2 provides information needed to follow the process. However, note that we are not performing an actual analysis here, but merely highlighting the distinct analytical aspects covered by the GeM model. Firstly, the *base layer* model allows the identification of a range of base units on the double-page. The base units are defined according to a list of recognised base units, which in the case of Figure 2 include orthographic sentences, illustrations, captions and arrows (Bateman 2008, p. 111). The task of the base layer is to identify each element on the page and assign them with a unique identifier, so that they may be picked up in the subsequent analytical layers: it also defines the analytical granularity, stating that an orthographic sentence is the minimal unit of linguistic analysis in the GeM model.

Layer name	Descriptive function	Analytical unit and examples
Base layer	Provides a list of base units that may be analysed as a part of other layers.	Base units: sentences, headings, drawings, figures, photos, captions, list items, etc.
Layout layer	Groups the base units together based on similar properties in the three domains below.	Layout units: paragraphs, headings, drawings, figures, photos, captions, list items, etc.
<i>Structure</i>	The hierarchical structure between layout units.	
<i>Area model</i>	The placement of layout units in a layout.	
<i>Realisation</i>	Typographical or visual features of layout units.	
Rhetorical layer	Describes the rhetorical relations holding between the identified rhetorical segments.	Rhetorical segments: base units with rhetorical functions.
Navigational layer	Describes the navigational structure by defining pointers, entries and indices.	Pointers, entries and indices: base units and layout units with navigation functions.

Table 2: The layers of the GeM model (Hiippala 2012, p. 108)

The base units are then grouped into layout units according to their realisation features and spatial positioning in the *layout layer*. In this case, realisation refers to the visual realisation of the base units, that is, their typographic and graphic features. For example, each of the paragraphs on the left-hand side in Figure 2 constitutes one layout unit due to their similar typographic realisation and spatial proximity. The layout units are also described in terms of their hierarchy. For instance, a paragraph and its header would be children of the same layout node, which indicates that they belong together. Finally, the placement of the elements in the two-dimensional space afforded by the layout is described. This is done using a grid to establish layout areas, where the layout units are positioned: Figure 3 shows the area model for the fly passage. Essentially, a representation of an area model may be deemed successful when exact placement information may be provided for each of the layout units.

So far, we have identified the elements that occupy the layout space, which we

have, in turn, described using the area model. The following step is to identify the elements in the *rhetorical layer*. The base units contributing to the rhetorical structure are referred to as rhetorical segments. For its description, the rhetorical layer uses an application of rhetorical structure theory (hereafter RST, for a description see Mann & Thompson 1988). RST is used to describe the rhetorical relations that hold between the elements on the page, that is, how they function together. The GeM model provides an extended set of relations to describe the interaction between verbal and visual rhetorical segments, also on the subnuclear level below the sentence (Bateman 2008, p. 162). The subnuclear relations provide access to the elements below the rank of a sentence, such as the labels in illustration (2b) in Figure 3. It should also be mentioned that RST has also recently gained interest (in connection with the GeM model) as a tool for describing multimodal designs in an eye-tracking study by Holsanova & Nord (2010, p. 90-1), which is duly acknowledged here.

Finally, we will briefly mention the *navigation layer*, which describes the structures that facilitate the access to the artefact. Examples of navigation devices include indices, page numbers, colour-coded elements and numbering. The fly passage contains certain navigation devices, such as chapter and page numbers, but we will not discuss their role any further in this paper. For a more detailed discussion of navigation structures and their contribution to structuring multimodal discourse, see Hiippala (2012, p. 118-9).

The analytical strength of the GeM model lies in its XML-based annotation scheme, which cross-links each of the layers described above. Each unit is cross-referenced across all layers by a unique identifier, enabling the analysis to pinpoint its position both in the layout and in the hierarchy, its realisational features and its function in the rhetorical structure. The GeM model is also scalable: additional layers of analysis may be defined and incorporated into the XML annotation as required. When deployed on a sufficient scale, the GeM model can be used to identify patterns, not only within the layers but also across them. For instance, the interface between the rhetorical structure and the use of the two-dimensional layout space is of high interest, especially as different multimodal artefacts use these structures in different ways (see e.g. Martinec 2003, Cheong 2004).

5.4 Eye-tracker configuration

In order to combine eye-tracking with the GeM model, two capabilities are required from the eye-tracker. The first requirement is the ability to designate focal areas on the screen and assign them with identifiers; the second requirement is the possibility of output in XML format (or a format that may be transformed into XML).

With the prerequisites now established, completing the proposed process has

to include at least the following steps:

1. The analysed page is annotated using the GeM model.
2. The focal areas in the eye-tracker are set up corresponding to the area model in the GeM annotation.
3. The identifiers for the area models in the eye-tracker and in the GeM annotation correspond to each other.
4. A set of pre-planned eye-tracking experiments are prepared.

Using the same layout area identifiers in the GeM model and the eye-tracker provide the link between the two data sets: this process is visualised in Figure 5. While the GeM model is used to provide the necessary tools for multimodal analysis, the configuration of the eye-tracker is expected to provide information about the fixation points, dwelling times and saccades, which show how the observer's attention is directed towards particular layout areas under specific conditions. If possible, another point of interest is the saccadic eye movements between layout areas, as indicated by transitions across layout areas.

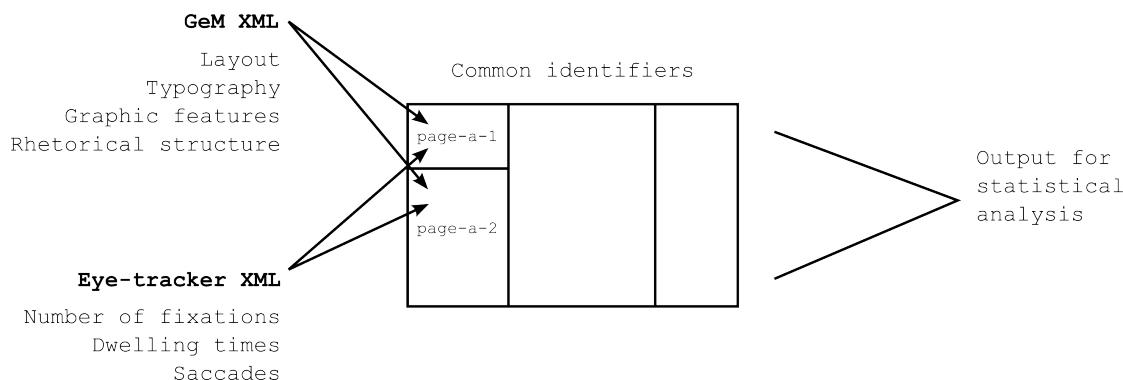


Figure 5: XML output and transformation

In theory, the proposed method should be able to provide specific information related to the multimodal structure of the artefact under observation, which may be accessed using the layout area identifiers. As the analytical layers are cross-referenced, the layout area identifier may be used to retrieve each base unit present in the particular area, their hierarchical structure, rhetorical organisation and visual properties. The semiotic modes and the notion of genre provide a backdrop for these observations, which may then be evaluated on the basis of the observer behaviour measured using the eye-tracker. The following subsection describes how the two data sets may be combined for statistical analysis.

5.5 The interface between the GeM model and the eye-tracker

The final task is to combine the multimodal analysis with the data collected from the eye-tracker, which may be done through a transformation, if both data sets are in XML format. There are several languages available for XML transformations, such as XSLT and XQuery. A precursory application combining the output of an external software with the GeM model is that of Thomas (2009, p. 243-4), who used a Perl script to enrich the output of optical character recognition software. The Perl script produced what Thomas referred to as “proto-GeM” annotation, which eased the annotation workload by producing a rough GeM annotation for human post-processing.

Fortunately, the transformation of the GeM annotation and the eye-tracker output has little or no need for post-processing, except for the process of importing the statistical data into the program that is used to analyse the data. Once the best practices for the analysis are established, the same experiment may be replicated with multiple participants. Finally, we will now move to present two hypotheses that may be investigated using the proposed method.

5.6 Example hypotheses

Holsanova et al. (2009) introduce the *dual scripting principle*, which suggests that the observer’s understanding of multimodal structure is enhanced by spatial and conceptual configuration of the verbal and visual elements. Their study found support for this hypothesis. In addition, the study also implied that spatial contiguity, that is, the spatial proximity increased the number of integrative saccades between the semantically related elements. Holsanova & Nord (2010), in turn, have shown how “redesigns” may be used to investigate the integration of verbal and visual elements in visual perception. With the GeM model, we are presented with the capability to capture the foci of these redesigns on the various strata of an artefact (cf. Delin & Bateman 2002), whereas the semiotic mode allows us to move beyond the artefact structure and investigate the discourse semantics at play. The remainder of this subsection describes how this potential may be transformed into hypotheses about the relationship between multimodality and visual perception.

The GeM model and the semiotic modes allow us to formulate hypotheses about visual perception in relation to different semiotic levels in the observed artefact. Therefore, to exemplify the range of phenomena that may be investigated, it is proposed that the following hypotheses may be tested using the method proposed in this paper:

1. Visual perception is sensitive to discourse semantics in the semiotic modes.

2. Visual perception is sensitive to genre structures in the multimodal artefact.

Hypothesis (1) states that in addition to spatial proximity, visual perception is sensitive to discourse semantics. The hypothesis predicts that breaking or altering the mapping between the two domains or the entities within them in Figure 4 increases the number of fixations and dwelling times directed at the redesigned element. A prerequisite for testing hypothesis (1) is the identification of semiotic modes deployed in the artefact and the formalisation of their discourse semantics, which is a considerable task. However, an investigation of the discourse semantics may yield significant benefits, as the formal back-and-forth mappings may be used to capture underlying principles behind the “temporal, spatial, semantic and logical arrangement of components”, which Holsanova et al. (2009, p. 1220) consider necessary for creating a coherent message.

Hypothesis (2) states that visual perception is also sensitive to genre structures. In short, hypothesis (2) suggests that the previous knowledge of a semiotic artefact is rooted in its configuration and deployment of the semiotic modes. The hypothesis may be tested by using the GeM model to analyse the structure of two functionally different multimodal artefacts and adapting the layout structure of one to another. In this case, for example, this could involve taking an instructive artefact, such as a manual (cf. Martinec 2003) and ‘pouring’ the content and the rhetorical structure of the textbook passage shown in Figure 2. The hypothesis predicts that the original artefact includes a higher number of integrative saccades, which indicates that the familiarity with the artefact structure facilitates its understanding.

This concludes the discussion on combining the methods of multimodal research and eye-tracking research. The paper has attempted to show that multimodal research may contribute to the understanding of the high-level factors affecting visual perception. However, to be effective, the methods applied in multimodal research have to be considered carefully, in order to restrict the interpretation on behalf of the analyst. In short, the capability to form hypotheses relies largely on the capability of multimodal research to provide statements about the nature of multimodal phenomena. The discussion now returns to more general matters in the form of a conclusion.

6 Discussion

To begin with, it appears that the concept of reading paths in multimodal research can gain significantly from the input of psychology and brain sciences. As eye-tracking research shows, multimodal analysis has largely worked with assumptions based on semiotic theories that have found only limited support in experimental psychology (cf. Holsanova & Holmqvist 2006). At the same time, it is crucial

that we continue to investigate the multimodal interaction that we engage in on a daily basis. As White (2010) has shown, new ways of interaction are constantly emerging, and they are, in most cases, both mediated and multimodal. This presents an enormous research challenge, as we are only taking the first steps in understanding the general principles of multimodality.

Furthermore, this paper has also attempted to argue that the fields of multimodal research, psychology and brain sciences may achieve significant synergies in the research of observer behaviour and reading paths. It may be said that the kind of work proposed here has become possible only recently, with the slow maturing of the multimodal theories and the availability and quality of the eye-tracking software, not to mention the digital tools (cf. O'Halloran et al. 2010, 2011). The potential to work together is there, but embarking on this task requires an interest from both disciplines.

Finally, it has to be noted that undertaking research projects such as the one proposed in this paper requires input from specialists in both fields. Multimodal analysts may do well in describing how the studied artefacts are construed, but the fields of psychology and brain sciences are undoubtedly more experienced in planning observer experiments, and in compiling and analysing statistical data. The positive results from this kind of work are finely illustrated by Holsanova & Holmqvist (2006), Holsanova et al. (2009) and Holsanova & Nord (2010): considerable efforts should be therefore directed towards fostering interdisciplinary work.

Acknowledgements

The author would like to thank — in no particular order — the following individuals who commented on earlier drafts of this paper: John Bateman, Ivan Berazhny, Eija Ventola and the two anonymous referees. In addition, the author thanks the WSOY publishing company for granting the permission to reproduce the textbook passage discussed in this article.

References

- Baldry, A. P. & Thibault, P. J. (2005), *Multimodal transcription and text analysis*, Equinox, London.
- Barthes, R. (1977), *Image-Music-Text: Essays*, Hill and Wang, New York. Translated by Stephen Heath.
- Bateman, J. A. (2007), 'Towards a *grande paradigmatique* of film: Christian Metz reloaded', *Semiotica* **167**(1), 13–64.
- Bateman, J. A. (2008), *Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents*, Palgrave Macmillan, London.

- Bateman, J. A. (2009a), Discourse across semiotic modes, *in* J. Renkema, ed., 'Discourse, of course: An overview of research in discourse studies', Benjamins, Amsterdam/Philadelphia, pp. 55–66.
- Bateman, J. A. (2009b), Film and representation: making filmic meaning, *in* W. Wildgen & B. van Heusden, eds, 'Metarepresentation, Self-Organization and Art', European Semiotics, Lang, Bern, pp. 137–162.
- 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 Studies in Multimodality, Routledge, London, pp. 17–38.
- Bateman, J. A. & Schmidt, K.-H. (2011), *Multimodal Film Analysis: How Films Mean*, Routledge Studies in Multimodality, Routledge, London.
- Cheong, Y. Y. (2004), The construal of ideational meaning in print advertisements, *in* K. L. O'Halloran, ed., 'Multimodal Discourse Analysis: Systemic Functional Perspectives', Continuum, London, pp. 163–195.
- Christie, F. & Martin, J. R., eds (1997), *Genre and Institutions: Social Processes in the Workplace and School*, Continuum, London.
- Cope, B. & Kalantzis, M., eds (2000), *Multiliteracies: literacy learning and the design of social futures*, Routledge, London.
- Delin, J. L. & Bateman, J. A. (2002), 'Describing and critiquing multimodal documents', *Document Design* **3**(2), 140–155.
- 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.
- Forceville, C. (2007), 'Book review: Multimodal Transcription and Text Analysis: A Multimedia Toolkit and Coursebook', *Journal of Pragmatics* **39**(6), 1235–1238.
- Guo, L. (2004), Multimodality in a biology textbook, *in* K. L. O'Halloran, ed., 'Multimodal Discourse Analysis: Systemic Functional Perspectives', Continuum, London, pp. 196–219.
- Hannus, M. & Hyönä, J. (1999), 'Utilization of illustrations during learning of science textbook passages among low- and high-ability children', *Contemporary Educational Psychology* **24**(2), 95–123.
- Hiippala, T. (2012), 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.
- Hofinger, A. & Ventola, E. (2004), Multimodality in operation: language and picture in a museum, *in* E. Ventola, C. Charles & M. Kaltenbacher, eds, 'Perspectives on Multimodality', Benjamins, Amsterdam, pp. 193–210.
- Holsanova, J., Holmberg, N. & Holmqvist, K. (2009), 'Reading information graphics: the role of spatial contiguity and dual attentional guidance', *Applied Cognitive Psychology* **23**, 1215–1226.
- Holsanova, J. & Holmqvist, K. (2006), 'Entry points and reading paths on newspaper spreads: comparing a semiotic analysis with eye-tracking measurements', *Visual Communication* **5**(1), 65–93.
- Holsanova, J. & Nord, A. (2010), Multimodal design: Media structures, media principles and users meaning-making in printed and digital media, *in* H.-J. Bucher, T. Gloning & K. Lehnen, eds, 'Neue Medie - neue Formate: Ausdifferenzierung und Konvergenz in der Medienkommunikation', Campus, Frankfurt/New York, pp. 81–103.
- Jewitt, C. (2009), An introduction to multimodality, *in* C. Jewitt, ed., 'The Routledge Handbook of Multimodal Analysis', Routledge, London, pp. 14–27.
- 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.
- Kappas, A. & Olk, B. (2008), 'The concept of visual competence as seen from the perspective of the psychological and brain sciences', *Visual Studies* **23**(2), 162–173.
- Kara, P., Koskenniemi, A., Kovakoski, H. & Nurmi, M. (1986), *Värikäs luonto 4*, WSOY, Porvoo.
- Kong, K. (2006), 'A taxonomy of the discourse relations between words and visuals', *Information Design Journal* **14**(3), 207–230.
- Kostelnick, C. & Hassett, M. (2003), *Shaping Information: The Rhetoric of Visual Conventions*, Southern Illinois University Press, Carbondale.
- Kress, G. (1998), Visual and verbal modes of representation in electronically mediated communication: the potentials of new forms of text, in I. Snyder, ed., 'Page to Screen: Taking Literacy into the Electronic Era', Routledge, London and New York, pp. 53–79.
- Kress, G. (2003), *Literacy in the New Media Age*, Routledge, London.
- Kress, G. (2009), What is mode?, in C. Jewitt, ed., 'The Routledge Handbook of Multimodal Analysis', Routledge, London, pp. 54–67.
- Kress, G. & van Leeuwen, T. (1996), *Reading Images: The Grammar of Visual Design*, Routledge, London.
- Kress, G. & van Leeuwen, T. (1998), Front pages: (the critical) analysis of newspaper layout, in A. Bell & P. Garrett, eds, 'Approaches to Media Discourse', Blackwell, Oxford, pp. 186–219.
- Kress, G. & van Leeuwen, T. (2006), *Reading Images: The Grammar of Visual Design*, second edn, Routledge, London.
- Kvåle, G. (2010), Invitasjon til destinasjon: Multimodal retorikk i turistkommunikasjon, in M. Engebretsen, ed., 'Skrift/bilde/lyd: Analyse av sammensatte tekster', Høyskoleforlaget, Kristiansand, pp. 39–55.
- Lemke, J. L. (1999), 'Typology, topology, topography: genre semantics', MS University of Michigan.
- Lim, F. V. (2004), Developing an integrative multi-semiotic model, in K. O'Halloran, ed., 'Multimodal Discourse Analysis: Systemic Functional Perspectives', Continuum, London, pp. 220–246.
- Livingstone, S. (2004), 'Media literacy and the challenge of new information and communication technologies', *The Communication Review* **7**, 3–14.
- Mann, W. C. & Thompson, S. A. (1988), 'Rhetorical Structure Theory: Toward a functional theory of text organization', *Text* **8**(3), 243–281.
- Martin, J. R. (1997), Analysing genre: functional parameters, in F. Christie & J. R. Martin, eds, 'Genre and Institutions: Social Processes in the Workplace and School', Continuum, London, pp. 3–39.
- Martin, J. R. & Rose, D. (2008), *Genre Relations: Mapping Culture*, Equinox, London.
- Martin, J. R. & Veel, R. (1998), *Reading Science: Critical and Functional Perspectives on Discourses of Science*, Routledge, London.
- Martinec, R. (2003), 'The social semiotics of text and image in [j]apanese and [e]nglish software manuals and other procedures', *Social Semiotics* **13**(1), 43–69.
- Martinec, R. & Salway, A. (2005), 'A system for image-text relations in new (and old) media', *Visual Communication* **4**(3), 337–371.
- Nisbett, R. E. & Miyamoto, Y. (2005), 'The influence of culture: Holistic versus analytic perception', *Trends in Cognitive Sciences* **9**(10), 467–473.
- Nisbett, R. E., Peng, K., Choi, I. & Norenzayan, A. (2001), 'Culture and systems of thought: Holistic versus analytic cognition', *Psychological Review* **108**(2), 291–310.
- O'Halloran, K. L. (1999), 'Interdependence, interaction and metaphor in multisemiotic texts', *Social Semiotics* **9**(3), 317–354.

- O'Halloran, K. L. (2004), Visual semiosis in film, in K. L. O'Halloran, ed., 'Multimodal Discourse Analysis: Systemic Functional Perspectives', Continuum, London, pp. 109–130.
- O'Halloran, K. L. (2008), 'Systemic functional-multimodal discourse analysis (SF-MDA): Constructing ideational meaning using language and visual imagery', *Visual Communication* 7(4), 443–475.
- O'Halloran, K. L., Tan, S., Smith, B. A. & Podlasov, A. (2010), 'Challenges in designing digital interfaces for the study of multimodal phenomena', *Information Design Journal* 18(1), 2–21.
- O'Halloran, K. L., Tan, S., Smith, B. A. & Podlasov, A. (2011), 'Multimodal analysis within an interactive software environment: critical discourse perspectives', *Critical Discourse Studies* 8(2), 109–125.
- O'Toole, M. (1994), *The Language of Displayed Art*, Leicester University Press, London.
- Parodi, G. (2010), 'Research challenges for corpus cross-linguistics and multimodal texts', *Information Design Journal* 18(1), 69–73.
- Rayner, K. (1998), 'Eye movements in reading and information processing: 20 years of research', *Psychological Bulletin* 124(3), 372–422.
- Reichenberger, K., Rondhuis, K., Klein, J. & Bateman, J. A. (1996), Effective presentation of information through page layout: a linguistically-based approach, Technical Report Arbeitspapiere der GMD 970, Institut für Integrierte Publikations- und Informationssysteme (IPSI), GMD, Darmstadt.
URL: <http://www.cs.uic.edu/ifc/mmwsproc/reichen/page-layout.html>
- Royce, T. D. (1998), 'Synergy on the page: Exploring intersemiotic complementarity in page-based multimodal text', *JASFL Occasional Papers* 1(1), 25–49.
- Royce, T. D. (2007), Intersemiotic complementarity: A framework for multimodal discourse analysis, in T. D. Royce & W. L. Bowcher, eds, 'New Directions in the Analysis of Multimodal Discourse', Lawrence Erlbaum, Mahwah, NJ, pp. 63–109.
- Stenglin, M. (2009), Space and communication in exhibitions: Unravelling the nexus, in C. Jewitt, ed., 'The Routledge Handbook of Multimodal Analysis', Routledge, London, pp. 272–283.
- Tan, S. (2009), A systemic functional framework for the analysis of corporate television advertisements, in E. Ventola & A. J. Guijarro, eds, 'The World Told and the World Shown: Multisemiotic Issues', Palgrave, London, pp. 157–182.
- Thomas, M. (2009), Localizing pack messages: A framework for corpus-based cross-cultural multimodal analysis, PhD thesis, University of Leeds.
- Tseng, C. & Bateman, J. A. (2010), Chain and choice in filmic narrative: an analysis of multimodal narrative construction in *The Fountain*, in C. R. Hoffmann, ed., 'Narrative Revisited: Telling a story in the age of new media', Vol. 199 of *Pragmatics and Beyond*, John Benjamins, Amsterdam, pp. 213–244.
- van Leeuwen, T. (1993), 'Genre and field in critical discourse analysis: a synopsis', *Discourse and Society* 4(2), 193–223.
- 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.
- White, P. (2010), 'Grabbing attention: the importance of modal density in advertising', *Visual Communication* 9(4), 371–397.
- Williamson, J. (1978), *Decoding Advertisements: Ideology and Meaning in Advertising*, Marion Boyars, London.
- Yarbus, A. L. (1967), *Eye Movements and Vision*, Plenum Press, New York.