Making sense with multimedia. A text theoretical study of a digital format integrating writing and video

Martin Engebretsen

Associate Professor, PhD
Agder University College, Norway
Email: martin.engebretsen@hia.no

Abstract

Digital text formats that allow a close interaction between writing and video represent new possibilities and challenges for the communication of educational content. What are the premises for functional and appropriate communication through web-based, multimedial text formats?

This article explores the digital writing-video format from a structural, theoretical perspective. To begin with, the two media's respective characteristics are discussed and compared as carriers of complex signs. Thereafter, the focus is upon how writing and video elements can be accommodated to web media. Finally, the article discusses the conditions for optimal co-ordination and interaction between the two media types within the framework of an integrated design. A design example is presented.

Keywords

Educational screen texts, multimedia, multimodality, text theory, web-design

Introduction

Digital media offer us the opportunity to describe the world and communicate with our surroundings in ways that have not previously existed. In particular have technologies connected with hyper linking and the mixing of media transformed the conditions for the creation of meaning. As broadband connection to the Internet becomes standard in homes, schools and offices, new opportunities for dynamic and "rich" forms of expression will seriously influence the different text genres of our culture. A common denominator with many of these forms of expression is that media forms that earlier appeared in separate media channels, now meet, mix and at times merge into new, distinctive media forms. They converge.

When written text, speech, photography, music, video and graphics are combined and integrated in digital texts, we are dealing not only with the convergence of media forms. On a more fundamental level it involves a convergence of semiotic systems, reading conventions and rhetorical patterns. A considerable challenge faced by contemporary media producers, is to develop text and genre forms that fulfil their rhetorical tasks within these frames. How will multimedia-readers be informed, touched, persuaded and

activated?[1] At the same time a new form of literacy will be required by recipients. And not the least, researchers will have to test and challenge both new and existing theories and methods in their exploration of new genres. This article is intended as a contribution from the field of text research.[2]

New rhetoric and educational potentials

In the light of current broadband developments certain media formats are of particular interest, not the least the formats that combine writing and video.[3] Such formats have for a long time been explored experimentally, e.g. within the production of educational CD-roms. But general usage has only arrived in the wake of broadband, so that is has become possible to distribute sound and video over the Internet, almost as swiftly as written text and images.

It is, on the other hand, not merely the possibilities of distribution that make text-video formats interesting. From a text theoretical perspective it is interesting because it entails a new phase in the development of *multimodal* types of text, i.e. types of text where different semiotic resources (writing, images, graphics, speech...etc) are combined and integrated.[4] From the perspective of rhetoric, text-video formats are of interest because they make it possible to undertake rhetorical tasks that are not possible with traditional media formats. They represent in other words a new *rhetorical* potential. It should be obvious that these   formats also are of significant *educational* interest, as they imply that educational material can be presented and treated in new ways.

When reading newspapers and magazines, we are rarely conscious of how interaction between words, images and graphical design influences our understanding of the issues presented. This form of *multimodality* is to a large extent conventionalised, and our ability to read such complex, but static media texts is so well developed that we read such texts more or less intuitively.[5] These kinds of text have already been comprehensively discussed in the research literature.

It is different with texts that combine writing with video.[6] We lack the grammar and aesthetic tools to describe the rules and effects concerning the combination of their meaning-carrying units, such that the creators of texts can predict how the they will be read, interpreted and experienced.

The goal of this article is to give a text theoretical presentation of the premises for the creation of meaning, when media types such as writing and video are applied to the web-medium. The discussion is based upon an analysis of the characteristics of writing and video as carriers of complex sign systems; i.e. their semiotic and user oriented *affordances,* and a description of the web-medium as a *hosting medium*.[7]

Design of content in a digital perspective

Günther Kress and Theo van Leeuwen's book *Multimodal Discourse* offers a theoretical framework for the study of multimodal texts.[8] Following the ambitious goal of social semiotics to reveal the premises for the formation of meaning through media-based communication, Kress and van Leuween identify four dimensions, so-called "strata", in communicative acts: *discourse, design, production* and *distribution*.

In the discussion about *how* and *what* one can represent and communicate in a text-video format, the concept of *design* proposed by Kress & van Leeuwen is particularly useful. *Design* is presented in a broad manner as the *planning* – the sketching – of how the mediated message is to be constructed and given a

shape, much in the same way as the architect plans the choice of materials, supporting structures and details of a building. An important component in the design of communication acts is to decide which semiotic resources, or *modalities* are to be used and how they are to be co-ordinated.

In the design, discursive content, semiotic resources and communicative intentions are joined together. But the interaction between these parts cannot be understood as a successive process, where one first defines content and thereafter selects a suitable form. Different semiotic modalities point towards different types of discursive content; different types of ″knowledge″. This means that the semiotic resources possessed by a particular media technology (and mastered by the designer of the text) decide the kind of content one might potentially deal with and thus the kinds of rhetorical and educational tasks that can be undertaken.

To take an example; curriculum material about the national political system can be communicated through writing or by the spoken word alone, or through an integrated mixture of words, graphic, photography, sound, video and animation – all dependent on the possibilities offered by the production and distribution technologies applied. And within the framework of digital technology, the different media components can be co-ordinated through a complete, narrative overarching structure, or they can be presented as a more or less ordered group of links leading to short autonomous pieces of information.

Such variations in media and format will naturally also exert an influence on the discursive content of the genre, on its *field of knowledge*. What can be expressed about a topic, changes when new modalities and structural principles are applied. And thereby the social expectations and hermeneutic frameworks supporting the genre will also be transformed.

Media typological traits of writing and video

How can media types such as writing and video complement each other in an integrated design? As a point of departure in attempting to answer this question, it is necessary to explore the different structural traits of the two types of media. In the presentation below, the following aspects of the two media will be explored: *form of representation, basic unit of syntax, grammar, structuring principles* and *reception.*

Form of representation

Writing and video represent the world in different ways. While writing *refers to* a (real or fictive) world, video seeks to *show* (a segment of) the world. With a re-casting of Platonic concepts, it is possible to assert that writing is in principle a *diegetic* (narrating) medium, while video is a *mimetic* (imitating) medium.[9] While writing requires considerable *interpretative* work, based among other things on advanced code and genre competence, the video clip is immediately experienced as "a piece of reality".

Also the concept of *transparency* can be used to illustrate basic differences in the manner of representation between the two media. Video is a more transparent medium, in the sense that the interpreter′s attention is to a greater degree directed towards *the mediated* and to a lesser degree towards the medium.[10] This means that video has the ability to bring forth an
 authentic emotional response.[11]

The proximity to the reality represented, means that the viewer of video is emotionally touched in a different and more immediate way than the reader of

written text, who must carry out a quite different interpretational work before a subjectively relevant meaning can be established and an accompanying emotional response can be released.

The emotional response includes among other things the experience of identification and sympathy – or antipathy – regarding the people and surroundings represented.[12] This experience is connected to the medium's ability to communicate *eye* and *face;* important elements in human communication.[13]

The video medium is, with its complex multimodality, the carrier of a "rich" and intricate – also in the sense of manifold – semiotic content.[14] The medium's semiotic richness and dynamic progression also have a "pasting" effect upon the viewer. The media expression is experienced as sense stimulating in a way that makes it more difficult to "hop off" a video sequence than a written sequence.[15]

These differences are primarily connected with the fact that these two types of media communicate different types of signs, respectively signs that possess significance through *likeness* – as moving images and accompanying sound recordings – and signs that have significance through *convention* – as words and concepts. This means that the written medium can communicate a relatively *precise* content through a clearly defined expression. It can also mediate thoughts and emotions that are not reflected in the physical scenario – though only through the writer's filtering formulation.

While writing is a medium for *semantic information*, video is to a higher degree a medium for *aesthetic information.*[16] Semantic information is based upon a conventional relationship between expression and content, and can thus be translated from language to language and partly from medium to medium. It is different with aesthetic information, which lacks a denotative level, and which is very difficult to trans-mediate. Aesthetic information deals more with atmospheres than with conceptual content, more with feelings than with ideas. How is it possible to translate atmospheres created by colours into a medium without colours? Or the feelings connected with a piece of music into a medium without sound?

The realism of video is understandably the result of its technological process of creation, it's making through (digital) storage of optical and acoustic recordings. Even though all forms of digital signal can be manipulated, it is more difficult to lie through video than through writing. This means that video is normally considered a stronger source of proof of an empirical case than writing (cf. the use of video recordings as evidence in criminal cases).

Basic unit of syntax

In both writing and speech it is the word, or the *concept*, that is the basic building block – acting as the connecting link between a particular linguistic expression and a particular *idea*. The content of the idea can vary strongly on the level of abstraction, from the completely concrete, such as the idea of a sun-ripe orange, to the extremely abstract, like the idea of Marxism as an ideology and way of life. This means that written language can be used for the communication of information both on a concrete level of detail and on an abstract and general level.

With video it is different. It is not the *concepts* or the ideas that form the starting point for the syntax of videos, i.e. for its combinations of meaningful elements. Rather, it is *time* and *place*, in other words the dimensions that help orientate us in physical time and space.[17] The basic unit of syntax can be described as a limited film sequence with organic time in one spatial context.

It is open to discussion how reasonable it is to regard each clipped sequence as a basic unit, irrespective of the use of zoom, panning etc., or if one should understand changes in visual focus and perspective as constituting new building blocks in the syntactic structure of the video film.

Differences between the two medias' principles of representation and syntax mean that they are suited to mediate different aspects of the world. If one in an educational context wishes to communicate a political event, for example an important public meeting, video in a way different to writing will be able to reveal details in the actual scenario – physical aspects connected to *who* and *where*: What kinds of impressions are made by the participants? What does the place look like? Writing (and speech) will be better able to provide a summary (*what* happened?) and describe the abstract aspects of the event: *when* and *why* did it happen?

This means that the two media types will provide different kinds of context to the events and actions that are mediated. The strength of video is the way it can provide situational context, while writing can better provide historical and socio-cultural context.

Grammar

Writing has verbal language as its primary modality – a sign system with a highly developed and strongly conventional grammar. This means that the rules for how the basic units can be established and put together into larger and more complex units, is well-known and clear, and they can be discussed on an abstract, theoretical level. One can for example discuss the communicative effects of placing an adverb at the beginning of a sentence, or summarising a sentence with a so-called "key sentence".

Video has moving images and accompanying sound recording as its primary modalities; spoken language, as well as writing and music are secondary modalities. This means that video is based upon primary modalities with a weakly developed grammar. It is not correct to say that it totally lacks a grammar. Both the language of images and that of graphical form have been described by Kress & van Leeuwen and Jacques Bertin, among others.[18] The categories and syntagms of film have been described e.g. by Christian Metz and James Monaco.[19] But the descriptions are not, as with the grammar of verbal language, refined and elaborated through years of discussion and criticism. They have not therefore attained the same level of complexity and functionality. This means that it is difficult to discuss the appropriateness of specific structures, or their rhetoric effects, without using actual examples. Collections of rhetoric examples – of metaphors, figures and forms – can hardly replace an abstract grammar as a basis for meta-discussions about the functionality of language and its development.[20]

The principle of signification though social convention, the rich lexicon and the well-developed grammar together make the verbal language the most precise and clear of all semiotic systems. And according to some, the most powerful.[21] It can be discussed whether the precision of writing or the fascination of images occupy the most prominent position in contemporary media culture.

Structuring principles

By *structure* is meant the planned combination of parts (components) into a whole (gestalt, form).[22] The complexity of structure deals with the *number of parts* constituting the totality, the degree of *likeness/contrast* between the parts, as well as the *principle of co-ordination* (for example sequence,

hierarchy or network). The structuring principles of writing and video are based upon semiotic and media technological premises.

Even though the content of the written text is *presented* in the form of a more or less linear chain of words, it is *organised* in a hierarchy of titles, paragraphs, sentences and phrases. This means that a competent reader will be able to distinguish central and over-ordinate content from more detailed content. A writer can quickly sum up a comprehensive text with the help of a shorter text, a resumé, such that only the ideas and assertions at the top of the content hierarchy are included. [23]

In the video medium *place* and *time* constitute, as mentioned, the basal content dimensions of the syntactic foundation, and the over-ordinate structure deals with organising the spatial as much as the temporal sub-structures. The progression and development of the video film are based on planned changes in both dimensions.[24] But for the film to be experienced as coherent, in the sense of its parts being logically inter-connected, there must also exist *continuity* in both dimensions. The temporal and spatial dimensions in a clip must have a clear relation to the temporal and spatial dimensions in the previous clip if one is to be able to guide the viewer from one point in time and a place in the total structure to another point in time and space. This tempo-spatial structure is readily called the *image rhythm.*[25]

The image rhythm in a video film has to a great extent the same ordering function as the paragraph structure has in written text. It assists in the integration of parts to a whole. In addition, the video film often has a verbal linguistic structure, which must be co-ordinated with the tempo-spatial structure if the videotext as a whole is to be experienced as coherent. In some cases the verbal linguistic structure will be over-ordinate. This happens for example when the video film shows a conversation or a monologue which is very little edited and which is filmed with small variations in visual perspective and focus. This also happens when the verbal text – for example in the form of a voice commentary, a so-called *speak* – joins together clips, which would otherwise have a broken and unconnected tempo-spatial structure.[26]

Reception

In media theory, as already indicated, it is possible to distinguish between *static* and *dynamic* media types.[27] Writing and photographs are examples of static media types, where the reader is invited to explore a fixed sign structure at their own tempo and (more or less) in the direction they choose. Sound recordings and video sequences are examples of dynamic media types. The procession of signs, and thereby both the tempo and direction of the reception process, is primarily decided in advance by the producer.

The fact that the reader of written text has to explore and interpret the sign structure on the basis of specialised code and genre competence, means that a certain mental effort is required in order for it to be experienced as meaningful. The *imitation of reality* experienced when a video sequence is viewed, does not place the same demands on semiotic competence. The mental effort which writing at the outset demands, can also mean that the process of interpretation taking place in the reading, reaches more deeply than is the case with video reception. This is presumably because one invests more effort in relating individual sequences to the over-arching topic and to one's own life.[28]

Both writing and video can be read in time as well as in space, but in different ways. The spatial dimension of writing, as it is realised through the depositing of coloured material on a flat surface, means that certain content factors can be realised through visual, non-verbal codes. Typography and page lay-out signalise certain meta-linguistic relationships: "this assertion is a title", "here

begins a new paragraph", etc. However, the most important content will be readable even if the visual formatting breaks down, such as sometimes is experienced when text is transported between different computer systems.

During video reception the dimensions of time and space are more equal. Content is connected to the dynamic procession of signs: through the steady changes in visual and auditory sign structures, attention is continually demanded for a particular period of time. At the same time, content in the video's two-dimensional picture frame is perceived according to the same principles as in photography. For example, the objects in the foreground and the ones in the background are represented *simultaneously,* they are perceived in *parallel* (even though they are usually not offered the same amount of attention).

Writing is thus a medium that is read in time more than in space, but where the reader, because of the media's static expression, gains a significant influence over sequence and pace. Video, on the other hand, is a medium that is read both in a linear and a spatial manner. However, the reader has little influence over sequence and pace because the signs are realised successively in a pre-determined manner.

Summary

Writing and video are media types organised according to different premises for both representation, organisation and communication, and they are therefore suited to different kinds of communicative and rhetorical tasks.

Through written text one can express relatively *unambiguous* assertions about the world and one can *summarise* a complex and detailed content with a sample of assertions at a high level of abstraction. Writing is therefore particularly suited to rhetorical tasks which require precision and economy. Writing is also suited to the expression of abstract, conceptually dependent aspects of actual events, for example connected to questions such as *why, how* and *with what legitimacy*?

Through video one can show *details* of a scenario which are difficult to communicate through writing, for example connected to the appearance of people and their personality. A discourse can be added *aesthetic information* of a kind that can hardly be trans-mediated: atmospheric sound, colour variations, body movements, etc. Video also carries with it a sense of credibility, based on video's realism and the fact that it is difficult to make video without having anything to film. The medium is therefore especially suitable for rhetoric tasks requiring evidence of proof, emotional response and sense-based fascination. The video is capable of accommodating all these things because of its semiotic richness and its reliance upon a dynamic procession of signs.

Adaptation to the web

In an integrated, multimodal design, the use of the individual modalities must be adjusted to the medium that is to support the production and distribution of the total medial expression. If we talk of the web medium, design must be adjusted to the possibilities and limitations of the digital network's technology, as well as the formats and user conventions associated with this medium.

The most fundamental premises that can be connected to web-based communication and publishing, are most likely that it is based on *fragmented* and *digitally stored* units of content, and thus *individualised* and *user-controlled*. Just as the greatest change in the movement from spoken to written language was that the text could be *stored* – and thus preserved and transported – the greatest change in the movement from analogue to digital

media, is that storage and presentation are separated. The information is *stored* as nulls and ones in a "concealed" storage unit while it is *presented* as understandable texts on the screen. The presentation format is significantly detached from the storage format. The stored units of content can be retrieved, combined and presented on the screen with a high degree of flexibility – and both the designer and the final user can be active in this process.[29]

In the analogue, written media the linearity of speech is to a large degree maintained. Most of the established written genres are based upon sequential "genre schemes". Content structures that to a high degree rely upon the logic of sequential progression, such as the *narrative* and the *syllogism*, therefore occupy a hegemonic position in our text culture.[30] With digital media the technological premises are different. Content units are stored in a "flat" structure in the database, or hierarchally in a filing system. These storage media possess an almost infinite capacity, and the analogue format's capacity problem has to a large extent been overcome. On the other hand, the computer screen as a presentation surface has a limited size and is relatively unpleasant to read. In combination with the possibilities of the linking technology, this creates new technological premises for text formats based upon a (large) number of shorter text units organised in non-linear structures.

These are formats that are poorly suited to the hegemonic written forms. Manovich in this connection believes that databases have developed into something more than a storage technology. He asserts that it can be understood as a *cultural form*, characterised by a rich selection of optional units. And as such it has taken up the struggle against the strictly organised, causal form of narrative:

> As a cultural form, the database represents the world as a list of items, and it refuses to order this list. In contrast, a narrative creates a cause-and-effect trajectory of seemingly unordered items (events). Therefore, database and narrative are natural enemies. Competing for the same territory of human culture, each claims an exclusive right to make meaning out of the world. [31]

The premises for Manovich's conclusion can of course be discussed, both in terms of to what extent databases can be regarded as an independent cultural form, and to what extent the anti-structural character of the databases is of decisive significance, as long as the content units normally are connected and organised in the design of the inter-face.[32] However, the database's particular characteristics as a storage technology are a significant premise for the development of digital formats. Manovich believes that the collage form, i.e. a *spatial* organisation of a number of content units, is for digital media a natural replacement for the sequential forms characteristic of analogue media.

User studies appear to support this theory: the assumption that computer users expect opportunities of choice and control of interaction with media content is supported in the empirical findings by Nielsen (2000), Murray (1997) and Engebretsen (2000).

If it is maintained that *choice* is one of the web medium's most important communicative principles, this means – in short – that any unit of content that requires the continual attention of the user over a longer period of time, must be regarded as media alien. Units that can be absorbed in short time and in a form and content that are compatible with a series of other units, are correspondingly media adapted.

A consequence of this is that the web medium cannot attain the same level of "transparency" as for example television, even though it can apply sound and living images.[33] Criteria of choice presuppose a *distanced* overview of the

possibilities of choice and over the totality of content structure. This is in clear contrast with TV's invitation to the viewer to allow themselves to be *seduced* by the illusion of reality. For this reason Bolter and Grusin use the term *hypermediacy* for digital, user-controlled media.[34] While *transparency* means that the user tends to "forget" that the content is technologically mediated, *hypermediacy* introduces a stronger focus on the mediating technology and active use of its possibilities. One is not *seduced* by an inter-face based upon menu lists and navigation windows.

When writing and video are integrated in a medium where *seduction* is replaced by *user-control* as a strategic goal, this has consequences for how the respective media texts are edited and for how they are mixed in the design of a common inter-face. Long, "dense" verbal texts should be replaced by a (possibly large) number of shorter texts, made visually overviewable (scanable) with a frequent marking of paragraphs, in-between titles and summaries. Longer video sequences with an independent dramaturgy should likewise be replaced by a number of shorter video clips that carry out particular, specialised tasks. These tasks could in particular be related to supporting the multimodal totality with *immediacy, authenticity* and *dynamism*.

The inter-medial interaction

How can elements of video and written text be co-ordinated in the design of a common, integrated inter-face in such a manner that the different content units complement and support each other in the most appropriate fashion? This is probably the most important – and difficult – question in the discussion about digital text-and-video formats. Only when this question is answered can the detailed editing of the single text and video units be planned or evaluated. Kress and van Leeuwen point out that it is in the "mix" of the different "voices" that a jointly tuned interaction is created – in front of the computer screen, as in the music recording studio.[35]

The integration of text and video elements in a multimodal web format can be discussed from perspectives that focus upon respectively *distribution of content, visual co-ordination and sequencing*. The last mentioned might appear to be a contradiction in relation to the demand for optimal choice. However, optimal choice must always be balanced against the demand for coherence. And for the reader to discover the inner relations constituting the *totality* of a multimodal material, these relations must be made clear. In practice this means signals about recommended – or forced – reader sequence at certain points in the material. How strongly the reading path should be steered is a question that must be related to the target group's presumed needs and the intention behind the communication.

Distribution of content

What kind of content should be expressed through writing and what should be left to video? Looking at the structural particularities of the two types of media provides some indications: Writing is especially suited to general summaries, precise assertions and the presentation of certain abstract issues. Video is especially suited to a more detailed description of people and situations, in addition to providing aesthetic information; with tableaus and atmospheres. Beyond this, it must be added that video, as a carrier of verbal speech, is also highly suited to providing more detail on topics treated in more general terms in the written parts. If for example values connected with closeness to and identification with people are to be realized, the people acting on the video must preferably say something. To *show a voice* is just as important in human

communication as to *show a face*. And what they say, ought to stand in clear relation to what is expressed in the written text.

Furthermore, this means that use of *speak* to communicate general information, often has little point in such a context. Such information should normally be placed in the written parts, while video's verbal elements should be reserved for the people performing in the video image. In cases where there is no focus on individual actors in the video sequence, it will generally be more appropriate to use speak to anchor the visual information and/or supplement the over-ordinate written based information.

If the two media types are to supplement each other in an optimal manner, content should be distributed in such a manner that each element of content is given a meaningful context by the remaining elements. It normally works well when the written elements provide an account of the video scene's time and place, as well as indicate who is participating and which role they play in the totality of events portrayed. The over-arching topic, according to which the video sequence is to be interpreted, must in a similar manner be established in writing. The video elements for their part must provide a situational context to the written information, and thereby provide the reader with layers of meaning concerning identification and emotional experience.

Sequencing

Which medium should introduce a reading session? When should the reader change focus? For how long can one demand the reader's attention in each unit of content, or in each modality?

These are vital questions in the integration of the media types in a multimedial presentation. In the preceding paragraph we began to answer the question about what media element should introduce a presentation. The point of view expressed was that introductory writing is especially suited to giving the user a precise and well-focused introduction to the topic to be treated. At the same time, such an introduction provides a foundation for the interpreting of one or more following video sequences. It is on the other hand immediately obvious that the question of the introductory media type must be related both to topic, target group and context of use. In some situations the introductory video will undoubtedly function to motivate and awaken interest. At the same, one must keep in mind how web users tend to search for writing before images, also within genres where from the paper media users are accustomed to the opposite movement, images before writing. [36] The explanation is that web users are "restless" readers who search for an immediate indication that the page has sufficient value and relevance before devoting time to a more detailed examination of its content. And to gain a quick impression about what the web page can offer, attention is directed towards the opening text rather than the accompanying images (which are often relatively small and claim little attention).

A short opening text sequence will therefor often constitute a good starting point for the reader's further investigation of a material that consists of several units involving several modalities. Normally, 2-5 sentences will be adequate when it comes to giving elements that follow - both writing and video – the required thematic framework.

When should a reader's attention be turned from a text unit to a video unit? And should such a point be marked in the text? George Landow maintains that there should be natural "points of departure" in content units that lead to other units.[37] In the same way the unit to which one moves, ought to have a "point of arrival" – an entrance to the content which makes the transition feel natural independent of which unit one departed from. Clear demarcation of

the points of transition between text and video can most likely be of interest in certain contexts, while in others it will be experienced as excessive "control" of the reader's freedom. Brackets around the word "video" for example, can indicate where in the text the video will be experienced as most relevant; and at the same time, the reader will see that the verbal text also continues.[38] Additionally, video elements can be equipped with a verbal caption, which provides a kind of "arrival information" that will enhance the coherence. This provides the reader with an indication as to what the video sequence is about, as well as connecting the video explicitly to the topic in the verbal text.

How long should a single content unit be in a multimodal web presentation? The answer depends of course on the genre and on the situation of the user. A literature student searching for material about contemporary authors will in all probability be open to longer written units and longer video clips than a secondary school pupil who has 20 minutes to disposition to write key words about the Stone Age. But some of the premises are stable. First, writing is presented in space, video primarily in time. This means that writing allows far greater user control than video. The demand for brevity conciseness and is therefore even more relevant in video than in writing.

Second, the earlier described distribution of content between writing and video, means that quite short video sequences will be adequate for completing the tasks to which it is most suited. The experience of truth, immediacy and movement is not necessarily enhanced if the video sequence is extended from 20 seconds to 2 minutes.

Besides, it is important to underline that the media's recommendation for fragmentation and brevity does not necessarily lead to superficial and "tabloid" presentations as a consequence. It is possible to attend to both breadth and depth when several, short content units around the same topic are made accessible in an integrated design.

Visual co-ordination

If written elements and video elements are to be experienced as parts of a whole, the *semantic* integration must be signalled through a *visual* integration.[39] In graphic design, gestalt psychology's basic thesis is readily followed: visual totality is perceived through the visual values of *proximity, likeness* and *continuity* (and their inverse values *distance, contrast* and *discontinuity*). This means that units that are positioned close to each other, are similar to each other (with respect to form, colour, size or other visual variables) or are positioned on a common axis, are considered to be associated with each other.

In many of today's digital text-video presentations, the video sequence is visually separated from the text elements. This means that they are considered to be independent content units, such that considerable effort on the part of the users is required to see possible connections between the written and the video-based elements. By comparison, it is possible to imagine a textbook with all the images collected in a single section, at a distance from the text to which they belong. Much of the book's educational force would as a consequence be lost.

Furthermore, visual co-ordination is not meant to signal just totality, it should also signal a recommended starting point and possible crossroads in the material. To achieve this, the design should follow the ruling conventions for how web pages normally are constructed and read. There are many indications that conventions from newspapers and books to a large degree have been transferred to the net media.[40] This means that within a given field, elements will be read from left to right, and from top to bottom. Accordingly, an

established convention on net pages is that information on a macro level (menus of accessible sections and services) lies to the top and to the left of the screen, such that the screen's central area, with an opening downwards and towards the right, becomes the field in focus when exploring a particular topic. The reader thereby seeks orientation on the basis of three organising conventions: left-right, upper-lower and centre-periphery.

A second point that is relevant for those developing visually integrated text-video formats, is the reader's need for *continuity*.[41] This need has both a diachronic and synchronic component. The diachronic component deals with the need to recognise the genre, as one is accustomed to seeing it in traditional formats. An educational text on the web must – at least in the introductory phase – have a certain likeness to textbooks, pupil notebooks etc. In a phase when a new, net-based text-video format is introduced in an educational context, it can be appropriate to adopt a textbook convention when it comes to the visual co-ordination of written elements (titles, indents, captions and body text) and visual elements (images and video clips). These are conventions developed over hundreds of years in order to achieve optimal interaction between written and visual elements on a flat surface, and there are few reasons to believe that they will not function in a screen based reading medium.[42]

The synchronic component in the need for continuity deals with the interaction between the different modalities realised on the screen. If the visual modality of writing is toned down, so the writing becomes "dense and grey", the transition to video will be experienced as disturbing. Correspondingly, the video elements' relation to the user-controlled and spatially realised elements of written text, will be strengthened if the video is "spatialised" by being cut in short pieces, each presented with a still image (for example using the video's opening image) in close proximity to relevant places in the text.[43] Such a presentation of video content will also raise the degree of user control.

A design example

Web-pages that today combine writing and video in the communication of educational content, do so in quite different ways. Occasionally, there is talk of so-called web-TV, i.e. web-based distribution of complete or fragmented TV programs, using topic based menu systems. On such occasions the written components are reduced to a form of navigational aid, and there exists little real interaction between the two media types. On other occasions, certain sequences of text are equipped with links to video elements, but without any video windows being integrated in the design of the page. When the links are activated, a video player opens in a new window on top of the original one. Even though this solution might appear flexible because the readers can move the video window around the screen according to their own wishes, little integration and synergy is achieved between the different modalities. Seen from the reader's point of view the unattached window will conceal and create a distance to the text from which it originates. Seen from the author's and the designer's points of view, such a solution will weaken the possibility of thinking holistically about the visual presentation of the multimodal content.

In the stylised design model below, existing format solutions, which retain the idea of semantic and visual integration, are taken as the point of departure. Traditional, paper-based article formats with titles, indents, captions and body text have for decades shown itself to be a well-functioning format with respect to multimodal interaction. If one or more image fields are replaced by a video field, this format presents itself as a good starting point in the development of new multimedial text formats within certain genres. This is a format with which both the creators of text and its readers are familiar; and even though

the mixing of media is new, many user conventions can be transferred from for example textbooks and printed newspapers.
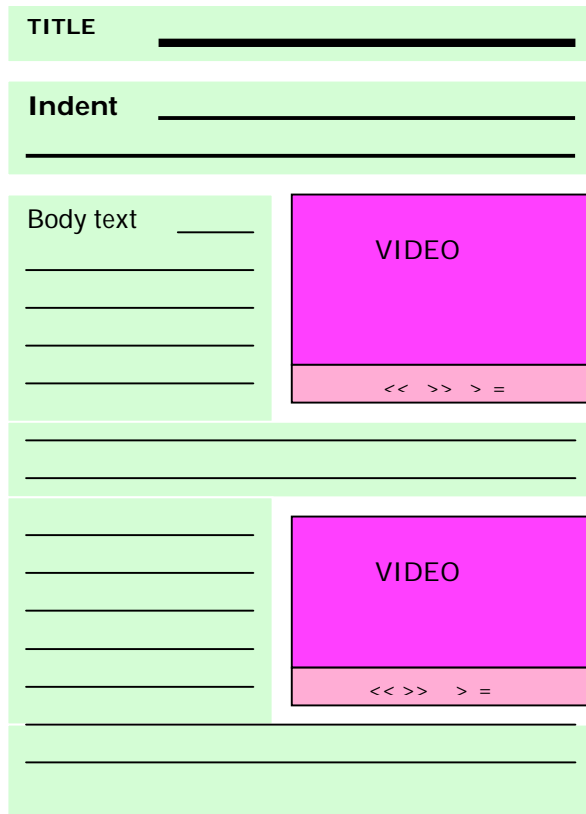
Fig: Example of a design model for an integrated text-video format, where well-known user conventions from paper media are adopted and developed.

This design signals that the verbal and the video-based units of content belong to a common thematic superstructure, specified according to the content of the title and the indent. It is also possible to assume that the reader's attention is divided between the two media types more or less in accordance with the division of content between the two. The video elements are scaled down to a size that makes it possible to physically integrate them in the text layout, and at the same they clearly occupy a subordinate position to introductory verbal elements. This means that the power of fascination that the video represents can be reduced to such an extent that it does not "steal" all initial attention.

It is important to underline that the sketch is only meant as an example, not as a genre independent standard. Within genres where the visual experience is very significant, the video window should be larger, and perhaps not so strongly enclosed in the "world" of writing.

The era of the semi-dynamic text?

Large parts of the coming generation have a more intimate and relaxed relationship to the dynamic text forms of television and computer games than to the more static types of text they meet in books and newspapers. At the same time, we see that the static types of texts possess characteristics that the dynamic ones lack; characteristics connected to important processes such as exploring, re-reading and scanning. In a time when text technology is becoming more and more flexible and the patterns of media use are in transition, there should be good reason to challenge established genre conventions with semi-dynamic text formats, i.e. formats that integrate

elements of writing and image with elements of video and sound. If such format experiments shall achieve the intended communicative effect with a reasonable use of time and resources, different competences must necessarily be combined. Technical competence has to go hand in hand with competences within text theory, genre and design. On this common arena of multimedial interest and new-orientation, it is probably true that specialists within text and genre have been the latest ones to arrive. When their team is fully equipped and ready to play, it may well be that we slowly move into the era of the semi-dynamic text.

References

Barthes, R. (1993). *Mythologies*. Selected and translated from the French by Annette Lavers. London: Vintage.

Berefelt, G. (1995). *A B Se: om billedpersepsjon*. (A B C: Image Perception) Oslo: Altera.

Bergström, B. (2001). *Effektiv visuell kommunikation*. (Effective visual Communication) Stockholm: Carlson bokförlag.

Bertin, J. (1983). *Semiology of Graphics.* Madison, Wisconsin: University of Wisconsin Press.

Bolter, J.D. & Grusin, R. (1999). *Remediation: Understanding New Media.* Cambridge, Mass: MIT Press.

Bostad, F. (1998). *Teknologiens uunværlige kontekst. Et semiotisk perspektiv på meningsskaping i hypertekst.* (Technology's Unavoidable Context. A Semiotic Perspective on the Creaiton of Meaning in Hypertext) Unpublished PHD dissertation. Norwegian Science and Technology University, Trondheim, Norway

Engebretsen, M. (2000). *Nettavisen og brukerne. En panelstudie med fokus på nettavislesernes vaner, preferanser og medieforståelse.* (The Net Newspaper and its Users. A Panel Study with a Focus on Net Readers Habits, Preferences and Understanding of Media) Kristiansand: IJ-forlaget.

Engebretsen, M. (2001). *Nyheten som hypertekst. Tekstuelle aspekter ved møtet mellom en gammel sjanger og ny teknologi.* (The News and Hypertext. Textual Aspects in the Meeting Between an Old Genre and New Technology) IJ-forlaget, Kristiansand

Engebretsen, M. (2002). Nye nyheter - ny fordeling av fortellermakt? (New News – New Distribution of Narrative Power?) In Slaatta, T. (Ed.) *Digital makt.* (Digital Power) Oslo: Gyldendals Akademisk.

Engebretsen, M. (2005). *Konvergens i tekst. En studie av et tekstformat som kombinerer skrift og video.* (Convergence in Text. A Study of a Text Format that Combines Writing and Video) Kristiansand: Høyskoleforlaget

Gentikow; B. (1999). Attracting attention in a competing media environment. A rhetorical perspective. Paper presented at the 14 Conference of Scandinavian Mass media Research, Kungälv, Sweden, August 14-17, 1999.

Hillesund, T. (2002). Many Outputs — Many Inputs: XML for Publishers and E-book Designers. In *Journal of Digital Information* (JoDI) Volume 3, Issue 1.

Jensen, B. K. (1997). Reflektioner fra en elektro-muse. Om videoens produktions-, værk-, og receptionsæstetiske egenart. (Reflections from an Elector-Muse. The Characteristics of Video Production, work and Reception) In Juel, H. (Ed.) *Multimedieteori: om de nye mediers teoriudfordringer (pp. 124-134).* (Multimedia Theory: About the Theoretical Challenge of the New Media) Odense: Odense universitetsforlag.

Kjeldsen, J.E. (2003). Magten mellom billedet og øjet. (Power Between the Image and the Eye) In Berge, K. L. et al. *Maktens tekster.* (Texts on Power) Oslo: Gyldendal.

Kjeldsen, J.E. (2004). *Retorikk i vår tid. En innføring i moderne retorisk teori.* (Rhetoric in Our Time. An Introduction to Modern Theories of Rhetoric) Oslo: Spartacus.

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

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

Landow, G.P. (1997). *Hypertext 2.0.*

Baltimore, Md.: Johns Hopkins University Press.

Liestøl, G. (1994). Tekstkulturen og den multimediale utfordring. (Text Culture and the Multimedial Challenge) In Schwebs, T. (Ed.) *Skjermtekster (pp.73-84).* (Screen Texts) Oslo: Universitetsforlaget.

Liestøl, G. (1999). *Essays in Rhetorics of Hypermedia Design.* PhD-thesis. Oslo: Department of Media & Communication, University of Oslo.

Manovich, L. (2001). *The Language of New Media.* Cambridge, Mass.: MIT Press.

Metz, C. (1974). *Film Language: a Semiotics of the Cinema.* New York: Oxford University Press.

Monaco, J. (2000). *How to Read a Film. The World of Movies, Media, and Multimedia: Language, History, Theory.* 3rd edition (first ed. 1981). New York: Oxford University Press.

Mullet, K. & Sano, D. (1995). *Designing Visual Interfaces: Communication Oriented Techniques.* California: SunSoft Press.

Murray, J.H. (1997). *Hamlet on the Holodeck: the Future of Narrative in Cyberspace.* New York: Free Press.

Neuman, S. B. (1995). *Literacy in the Television Age: the Myth of the TV Effect.* Second edition. New Jersey: Ablex.

Neuman, S. B. (1989). The Impact of Different Media on Children's Story Comprehension. In *Reading Research and Instruction*, 28 (4), 38-47.

Nielsen, J. (2000). *Designing Web Usability.* Indianapolis: New Riders.

Nisbett. R & Ross. L. (1980). *An Introduction to Visual Culture.* London, New York: Routledge.

Ong, W. J (1982). *Orality and Literacy: the Technologizing of the Word.* London: Methuen.

Pettersson, R. (1997). *Verbo-Visual Communication: Presentation of Clear Messages for Information and Learning.* Publications from VALFRID; no. 9. PhD-thesis, Göteborg Universitet.

Potter, W.J. (2001). *Media Literacy.* Thousand Oaks, California: SAGE.

Rogers, E. (1994). *Diffusion of Innovations.* New York, London, Tokyo: The Free Press.

Salomon, G. (1984). Television is "Easy" and Print is "Tough": The Differential Investment of Mental Effort in Learning as a Function of Perceptions and Attributions. In *Journal of Educational Psychology* 1984, Vol 76, No 4, 647-658.

Tønnesen, E. S. (2002). *Kommunikasjon og læring ved skjermen. Rapport om en resepsjonsstudie av multimediesystemet MigraNorsk.* (Communication and Learning on the Screen. Report on a Reception Study of Multimedia System MigraNorsk) Publication Series nr. 87. Kristiansand: Agder University College., Norway.

Wilden, A. (1987). *The Rules are no Game: the Strategy of Communication.* London: Routledge & Kegan Paul.

---

[1] I adopt a wide conceptual understanding of rhetoric in this article, as Jens Kjeldsen has formulated it, 'intentional and influential communication' (Kjeldsen 2004:23).

[2] The essay builds upon an earlier work by the same author, *Konvergens i tekst* (Engebretsen 2005).

[3] *Format* in this context is understood as a wide, media and genre dependent form of presentation, constituting an overall framework for possible designs. (The concept of design will be discussed later in the article.)

[4] A more detailed discussion of the relationship between the concepts *multimodality* and *multimediality* can be found in Engebretsen 2005. It is noted that *multimodality* is a concept directed towards semiotic resources used in the act of communication, while *multimediality* correspondingly focuses upon the material, media technological resources used in the realization of modalities. This means for example that *writing* can be considered both as a semiotic modality and as a type of media.

[5] Cf. the concept *media literacy*, which refers to a culturally dependent understanding of the grammar, syntax and metaphor system of media texts (Potter 2001).

[6] I adopt a broad concept of *text* in this work. It will be evident from the context when the concept refers to written verbal texts and when it refers to presentations that draw upon other types of semiotic resources.

[7] The concept *affordance* denotes the possibilities and limits offered to participants in a communicative act (see Gibson 1986, Norman 2002). Such participants must attend to media technological, semiotic, cultural and situational factors, each representing particular sets of affordances.

[8] Kress & van Leeuwen (2001).

[9] Plato wrote about mimesis and diegetics in his work the *State*, in book III and X.

[10] Media transparency is discussed in Kress & van Leeuwen (1996) og Bostad (1998).

[11] Bolter and Grusin (1999).

[12] Tønnessen (2002).

[13] Gentikow (1999).

[14] When Anthony Wilden (1987:299) writes that "video is rich in meaning, but low in precision", it is possible to reply in polemic fashion that the adversative "but" could favourably be replaced by the causal "and thereby". Semiotic (multimodal) richness is not easily combined with a high level of precision.

[15] A study of reception connected to the use of a multimedial program for language learning showed that all informants saw all of the video sequences in the program, while the degree of reading was much more variable when it came to the verbal text material (Tønnessen 2002).

[16] See Berefelt (1995:57).

[17] Jensen (1997) and Manovich (2001).

[18] Kress & van Leeuwen (1996) and Bertin (1983).

[19] Metz (1974) and Monaco (2000).

[20] Kress & van Leeuwen (2001).

[21] Liestøl (1994), Kress & van Leeuwen (2001).

[22] Berefelt (1995:28).

[23] In linguistics such hierarchies of assertions (propositions) are called the *macro-structure* of the text.

[24] Manovich (2001).

[25] Jensen (1997).

[26] *Speak* is one of several terms used to describe the commentating voice placed on top of the video image.

[27] See among others Liestøl (1994).

[28] Salomon (1984).

[29] In XML-circuits the validity of the expression "one input, many outputs" is discussed — to what extent can a content (in a database) be presented through several forms of media expression (see Hillesund 2002). In this discussion it is correct to distinguish between semiotics concept of content, which presumes an unbreakable connection relation between content and expression, and a concept of content that covers the stored units in a database. This data can undoubtedly be presented in different versions and combination — but each presentation will of course constitute a particular semiotic expression with an accompanying semiotic content.

[30] Ong (1982).

[31] Manovich (2001:255).

[32] Manovich does note that the database, prom a purely technological perspective, can support a narrative form of presentation — but he maintains that this form stands in opposition to the "pure" form of the database.

[33] Bolter & Grusin (1999).

[34] Bolter & Grusin (1999:4).

[35] Kress & van Leeuwen (2001:94).

[36] See the report from The Eyetracking Project (http://www.poynterextra.org/eyetrack2004/main.htm), undertaken between 1990 and 2003 in three phases, through a co-operation between Stanford University and The Poynter Institute of Journalism. The informants were asked to read news and features on web sites, while all eye movements were registered.

[37] Landow (1997).

[38] In a user study only four of ten informants followed sequence indications in a precise manner. Others passed the instructions with a few lines before starting the video. Most were of the opinion that marking created a stronger sense of connection in the presentation (Engebretsen 2005).

[39] Bergström (2001).

[40] Tønnessen (2002).

[41] See for example Rogers (1994) about the user's need to experience continuity between the old and the new, when different types of innovation are dispersed in a society.

[42] It is important to underline that we are talking about conventions for the visual design of a given number of written and image-based elements on a surface — and not the organisation of over-arching content structures. When it comes to the last mentioned, web media's data base and link technology suggest other solutions than those found in textbooks.

[43] Liestøl (1999).