CHAPTER 1

In between modes
Language and image in printed media

Hartmut Stöckl
Technical University Chemnitz, Germany

The present chapter addresses theoretical and text-analytical issues of the language-image-link in printed media. After modelling multimodality as a networked system of core modes, medial variants, peripheral modes, sub-modes and features, the chapter goes on to characterise the linguistic and the pictorial mode according to semiotic, cognitive and semantic criteria. Central to this chapter is the idea that modes and sub-modes shift or blend (mode overlapping) and mix (mode mixing) in multimodal communicative events. This is demonstrated on two advertising texts sketching out the levels and criteria of analysis necessary to adequately describe language-image-links. The two sample texts show that there is a strong pictorial element in language and a linguistic element in images. The chapter concludes by looking at possible semiotic principles operating across modes.

1. Multimodality – the late discovery of the obvious

Whether as the reflection of a changing communicative landscape (i.e. stronger reliance on modes other than language) or a practical tool in text and discourse analysis, multimodality is currently gaining academic ground. Beyond its mere hype-aspects fostered by an incessant quest for novel research topics, it addresses a phenomenon which is as old as representation itself and crucial to an understanding of almost all forms of communication: multimodal refers to communicative artefacts and processes which combine various sign systems (modes) and whose production and reception call upon the communicators to semantically and formally interrelate all sign repertoires present. Spontaneous face-to-face talk relies heavily on non-verbal means (e.g. gestures, posture, body language) and has our visually perceptible environment as a constant topic of communication. Often it even has the immediate manipulation of objects as its target. Written language, on the other hand, incorporates images and – through typography and layout – wields strong pictorial powers. More recent media, like film, television or computer, mul-
tightly semiotic potentials by integrating moving images, language (spoken and written), sound and music. The same can already be said of older mediated experience like theatre, drama or opera. I would go as far as to argue that the purely monomodal text has always been an exception while the core practice in communication has essentially been multimodal all along. The dominance of linguistics, however, and the concentration on language as the central mode, paired with a lack of adequate models for the analysis of other modes, made verbal monomodality appear to be the standard and dominant form of communication.

Considering that multimodality research has been around only for some 30 years, its achievements are quite remarkable. For one thing, research into multimodality has resulted in various 'grammars' of individual non-linguistic modes. For another, the social semiotic school as one major driving force behind multimodality research has succeeded in showing that 'phylogenetically' some modes seem to displace others in the development of media and genres. Social semiotics has also hinted at possible causes for this, and it has amply illustrated that the specific usage of one or the other mode is guided by socially determined intentions and realises group interests, subjective points of view or ideological stances. A third strand in multimodal research, namely studying the interrelations between various modes, is – as far as I can see – underrepresented. We seem to know more about the functioning of individual modes than about how they interact and are organised in text and discourse.

In the present sketch I want, therefore, to turn towards the interface between two prominent modes, language and image in printed media. First of all, I would like to raise the question of what a mode actually is and how many there are. This involves the tricky business of neatly delineating various sign systems. I hope to show that there is a whole network of heavily interdependent modes and sub-modes and that in textual practice modes can shift and blend into one another (Sections 2 and 4). Secondly, I seek to demonstrate how language and image can be distinguished from one another looking at their semiotic structure, their semantic characteristics and the cognitive operations they typically entail (Section 3). Thirdly, analysing two sample texts from advertising and describing various facets of inter-modal relations, I will show that there is a pictorial element in language and a linguistic element in images (Sections 4 and 5). Finally, I would like to address the question of whether there are common semiotic principles operating across modes – as Kress and van Leeuwen (2001) have suggested – and what exactly they could be (Section 6).

2. How many modes is 'multi'?

'Mode' would seem to be an easily definable term. Following Jakobson's idea of 'code' (Jakobson 1971), it is often glossed as 'sign system', 'sign-repertoire' or 'semiotic' from which communicators can pick their signs to realise their communicative intentions. Apparently, we intuitively gauge the meaning of 'code' as the quality or type of signs used in a communicative event. Along with 'mode' comes the notion of a 'grammar', i.e. signs belonging to one mode are seen to be governed by a common set of rules that state how these signs can be combined to make meaning in particular situations. In practice, however, things turn out to be less straightforward than this. Let us look at modes a little more closely.

If we take sign quality to mean the different sensory channels via which signs can be perceived and meaning made by communicators, there is justification to subdivide visual, auditory (or audial, cf. ledeama 2003:32), tactile, olfactory and gustative sign(-mode). While this seems a valid and largely uncontroversial systematisation, it is too rough to be of much use to a theory of multimodality. Language or the linguistic code, for instance, which we justly feel to be a mode in its own right, would thus fall both into the visual and the auditory category as it can be medially realised as either speech or writing. Both are governed by the grammar of language, but as different medial or material varieties of one mode they entail a number of concomitant, additional sub-modes. So speech – besides being linguistic – also employs volume, intonation, timbre, rhythm, speed or pausing, all of which are design features of language in its spoken form and are often termed para-verbal. Furthermore, speech is accompanied and crucially shaped by what has come to be called the non-verbal mode, i.e. gesture, posture and body language. Similarly, writing as the visual counterpart of speech entails typography, which can be seen as the written variant of para-verbal means. What intonation, speed and rhythm are to speech, typography is to writing. However, to complicate matters, some aspects of typography like layout or paper quality are rather non-verbal as they seem further removed from language as such but still accompany writing and contribute to its meaning beyond the linguistic. Moreover, despite being an integral part of writing, the quality of a document’s paper transcends the visual mode and has a clear tactile quality to it.

At least three points can be generalized from these brief observations. Firstly, modes cut across sensory channels, so the nature of a sign is not sufficiently characterised by looking at its path of perception. Secondly, one mode can be realised in different media thus creating medial variants of one mode (e.g. speech and writing as variants of the linguistic mode). As any one variant has distinct materiality, it in turn commands individual sets of concomitant sub-modes facilitating or accompanying the variant. This is, among others, a reason why media and
Figure 1. Network of modes, sub-modes and features in printed media

modes ought not to be confused but neatly kept apart and regarded in their interdependencies. Thirdly, and perhaps most importantly, the range of existing modes represents a hierarchically structured and networked system, in which any one mode can be seen to fall into sub-modes which in their turn consist of distinct features that make up the sub-mode. Let me exemplify these points looking at the image as another major signing mode.

Images, just like language, are not purely visual, they also have a tactile quality to them, which may be reflected in the meaning we construe from them. So while essentially visual, the material and techniques used in the production of the picture are also subject to touch. The nature of the pictorial sign can, however, not be gauged from its visual quality alone because this also pertains to written language, as we have seen. In this respect both image and language are equally visual.

Figure 2. Network of modes, sub-modes and features in TV- and film-media
Images are realised in different media, the static printed image and the dynamic moving image of film and television being the most prominent examples. Besides sharing some sub-modes (e.g. elements, vectors, distance, angle, colour etc.), both medial variants also differ in the set of sub-modes they entail. So the moving image commands such specific sub-modes as, for instance, panning or tilting, it can have narrative and has at its disposal sub-modes originating in post-production like cut or visual effects (e.g. slow-motion, time-lapses). Sub-modes in their turn can then be seen to comprise sets of distinctive features, i.e. specific aspects of the sub-mode, which are both phenomena in perception contributing to an overall gestalt as well as analytical categories which help to theoretically come to grips with sub-modes. Colour, for instance, can be decomposed into six distinctive features: value, saturation, purity, modulation, differentiation and hue (Kress & van Leeuwen 2002: 354ff.). Any concrete colour may then be specified by determining values on the scales of those distinctive features and their individual combination.

Applying this kind of systematisation of mode to textual genres in printed media and TV- or film-media, I suggest the following schematically represented network (cf. Figures 1 and 2).

The graphic representations of how modes are structured (cf. Figures 1 and 2) are necessarily formalised simplifications, which demand some comments on inherent problems and limitations.

1. While the columns of core modes contain central sign-repertoires that are deeply entrenched in people’s popular perceptions of codes and communication and can stand on their own, peripheral modes come as inevitable ‘by-products’, as inherent elements of a core mode’s specific medial realisation. This is not to say that core modes are more important than peripheral modes or more powerful in terms of their internal grammar and resulting communicative potential. Rather than being a major/minor distinction the differentiation indicates that some modes only come into being along with others and depend on them to some extent. Non-verbal means are ambivalent here as they can also function independently of language. On the one hand, then, the non-verbal is a concomitant aspect of language (cf. Müller 1998), on the other hand, it is part of a communicator’s image and thus also relevant to pictorial analysis.

2. In another sense core modes are also abstract modes that need to be instantiated in a specific medial variant. The grammar of language must be realised either in speech or in writing — it is only in these medial variants that peripheral modes pertain to them.

3. Sub-modes constitute a mode in that they provide the building blocks of a mode’s grammar. It would be wrong, I believe, to view any of the sub-modes in isolation or as dominant in the make up of a mode. It is rather all sub-modes in conjunction and their manifold interrelations that establish a mode and facilitate its realisation in a communicative event. As a gestalt in perception, the concrete materialisation of a mode in text and discourse is always more than the sum of its parts. No claim to completeness is made here, both for reasons of space as well as for the incomplete nature of relevant research.

4. If sub-modes are the building blocks of a mode’s grammar, then features are the shapes of the blocks and the patterns from which they can be combined. Features can be conceived of as property scales on which certain values can be adjusted in the design of a communicative product/event. Again, it is the sum total of all features and their values that make up the specific realisation of a sub-mode. As they are far from always being binary oppositions, it seems advisable to dispense with the attribute ‘distinctive’. For reasons of space I have confined myself to indicating features for only one sub-mode per mode. Within the features listed, again, completeness was not my objective.

5. It is my impression that some sub-modes, like colour, will allow for a relatively finely-grained differentiation into features, whereas others will only command few features or none. Those that consist of a larger variety of features seem to have a more flexible internal grammar and can - due to their complexity - function more autonomously in communication. Such sub-modes will therefore also be able to shift across modes. For instance, colour is not just relevant for pictures but also for typography. Still other sub-modes, like rhythm for instance, are so basic they pertain not just to music, but to speech, sound, animated writing and moving images. Generally, individual sub-modes can apply to more than one mode when the modes in question share some essential design feature, like in the example of rhythm a time-based structure, that is a linear sequencing as its underlying principle. Yet, it would be wrong, I feel, to grant such strong sub-modes the status of modes in their own right as they are always part of a whole configuration of sub-modes, without which the individual sub-mode, however strong, would not be able to operate.

6. Most importantly, the text is the locus where all modes, sub-modes and features are realised. So it is the dynamics of text production and reception, the complex chain from discourse and design to production and distribution (Kress & van Leeuwen 2001: 1–23) that determines how we deploy modal resources and how they in turn are construed in reception. In textual practice one mode is often made salient while others recede into the background both in perception and production. Genres or text types differ a great deal in how they foreground and background modes and sub-modes. Whereas typography is largely “automatized” (Ledema 2003: 40) in many text types, because it follows established conventions, advertising makes creative use of it and thus puts it centre-stage. Taking a textual perspective on modes and multimodality
means to also pay attention to "the socio-semiotic histories and transitions" (ledema 2003:48) of texts, i.e. to uncover how in production and reception modes shift and blend and how producers and users translate or 'transpose' meaning from one mode to the other. In personal e-mail communication or in some types of advertising, for instance, texts are medially realised in writing, while conceptually they are oral in nature. So typically oral textual characteristics normally realised with the help of para-verbal means in speech may be transposed to the written medium using typography and its sub-modes.

3. Mode distinctions

When 'reading' a multimodal text, average recipients will normally become only dimly aware of the fact that they are processing information encoded in different modes. The manifold inter-modal connections that need to be made in order to understand a complex message distributed across various semiotics will go largely unnoticed. All modes, then, have become a single unified gestalt in perception, and it is our neurological and cognitive disposition for multimodal information processing that is responsible for this kind of ease in our handling of multimodal artefacts. A theory of multimodal communication, however, has to meticulously dissect an apparently homogeneous and holistic impression. It has to sensitise us for the essential differences of the modes involved and make us aware of the textual work we invest in building inter-modal relations so crucial to understanding. Multimodal theory also needs to ask in how far there are systematic similarities and ties between the modes involved.

In what follows, I want to demonstrate that it is at least on three different levels that modes can be distinguished from one another: semiotic properties, cognitive orientation, and semantic potential. While I will here confine myself to the juxtaposition of the two core modes language and image, the three theoretical perspectives can be applied to all other modes and sub-modes.

3.1 Semiotic properties

The semiotic properties of a mode refer to its internal structure and to the general ways in which users can make meaning with a mode's signs. Language has what linguists call double articulation, i.e. discrete signs on two levels of organisation, phonemes and morphemes, which combine to form words and utterances. This design feature explains the boundless flexibility and resourcefulness of language. Images, in contrast, have no distinct signing units. There are no rules that would explain how pixels yield higher-level units when combined. What comes closest to a repertoire (alphabet) of pictorial signs is the elements that can be depicted in images and that we are familiar with from our visual experience of the world. The rules of how to combine them are tantamount to the natural order of things in worlds real or imagined. Consequently then, it is the iconic nature of signs that enables and determines their meaning. Because pictorial signs can access the same mental models that real-world objects access, we can understand pictures. It is the similarity between optical impressions gained from picture viewing and real-world vision that facilitates meaning-making in images. Whereas images thus represent an analogue code, language is deeply symbolic and digital. We need to recode visual perception into abstract sounds or graphs that bear no resemblance to the objects depicted and vice versa. Although writing and pictures share some similarity, as both use the surface area of a medium for representation, they radically differ in that writing utilizes two-dimensional arbitrary graphic forms to represent speech sounds, whereas pictures systematically evoke the three-dimensionality experienced in the perception of objects.

3.2 Cognitive orientation

These and other semiotic properties result in different cognitive operations demanded or afforded by language and images. Most importantly, language is a linear mode that calls for the successive integration of signs into phrases, whereas images are rather based on simultaneous and holistic gestalt-perception. Consequently, images can be regarded as a quick mode relative to language as they do not necessitate parsing. We know from psychological experiments that images are far more likely to be attention-getters in perception than language and can also be memorised much more easily and effectively. Both have to do with their analogue code characteristics – no recoding needs to take place and pictures can therefore be regarded as a code close to reality or – as some semioticians have argued – a 'language' without a code. The speed of pictorial perception is usually put down to the simultaneity of gestalt formation, whereas the communicative impact of images is seen in the fact that they directly tap into the emotions and provide immediate sensory input.

3.3 Semantic potential

Semiotic and cognitive characteristics determine what users of a mode can do with it in terms of specific meaning-making resources. Although debatable, it has generally been accepted that the semantics of language is less vague and polysemous than that of the image. While language provides scope for double meaning, it has conventional semantics attached to words and utterances. Images, on the contrary,
are seen to be inherently vague and ambiguous and can only be made to mean and communicate specific contents by a combination with other modes or the embedding into narrowly defined communicative situations. Most importantly, images lack a definite speech act repertoire, which is why their illocutions remain cloudy unless they are complemented by language. Language, on the other hand, counts as less rich in information than images, which carry a welter of sensory information and are particularly intense in terms of connotation. Conversely, language is at a great advantage as for its potential to communicate all sensory modalities, whereas pictures clearly are confined to visual information. Similarly, the self-referential capabilities of images are weak, whereas they are basically unlimited with language. Finally, language can be used to make just about any utterance imaginable. This huge semantic flexibility, which results from the linguistic design principle of double articulation and an elaborate set of rules, is contrasted by some obvious semantic restrictions of images. Some meaning relations like causality cannot be expressed, negation and affirmation are impossible and the utterances construable from images are usually additive. To sum up, language has its strength in the depiction of events and states-of-affairs in time, whereas images are particularly suited to the representation of objects in space and their physical characteristics (cf. Kress 1998:68ff.).

4. Mode integration: Overlapping and mixing

The comparison between language and image showed up a number of essential differences, which shape the ways these two central signing modes can be made use of in communication. The danger inherent in contrasting two modes, however, is that we tend to somehow look at one mode in terms of another. So, mostly, due to language’s dominance, we seem to be asking which linguistic properties images have. Thus we run the risk of overlooking some important design features of images which are outside the linguistic perspective. While a possible denigration of images resulting from this must be avoided, the metaphoric stance of a pictorial language is engrained in our naturally logo-centric take of communication.

More importantly, we need to realise that the mode distinctions pointed out are relative, as language and image are inseparably intertwined both in concrete forms of communication (mode mixing) as well as cognitively, semantically and historically (mode overlaps). Producing and understanding images presuppose verbal categorization of visual elements just as producing and understanding language relies heavily on conjuring up and manipulating mental images, the latter being most poignantly illustrated by the vast store of phrases that use a literal image to make metaphoric meaning. The history of alphabetic signs is a testimony to the fact that the graphic articulation used to represent speech sounds has its origins and immediate precursors in the pictorial. The strongest argument for the innate tie between language and image, after all, is their co-presence in almost all forms of communication, a symbiotic mode integration (mixing) which is guided by the principle of reciprocally balancing out limitations and weaknesses of the modes combined.

5. The textual interface between language and image

There are two basic ways in which the linguistic and the pictorial mode can come together in a text. Firstly, a verbal text can itself acquire image qualities by means of typography and layout. In this case a peripheral mode (typography) of a medial variant (writing/language) is employed for a partial transfer from one core mode (language) to another (image). Here, the carrier of the linguistic mode emulates the pictorial. Secondly, and this is the more common option, a verbal text is combined with an image. The two core modes are semantically and formally integrated so that each mode strategically employs its range of sub-modes thus unfolding the specific semiotic potential of each mode and contributing to an overall communicative gestalt. A specific type of this language-image-combination would result if the verbal text contained language that was itself pictorial or figurative and established a semantic or formal link with the accompanying visual image. Let me illustrate both types of language-image-link by discussing two sample texts drawn from the advertising genre. My aim in this will be to show on which levels such interfaces between two modes can be analysed.

5.1 The typographic image

In an advertisement of the RSPCA (the Royal Society for the Prevention of Cruelty to Animals) for free-range eggs (cf. Figure 3) the verbal text is typographically designed to yield the visual form and appearance of a supermarket receipt. Although, of course, the language contained in the text is not what we would expect to read on a receipt, conjuring up the image of a receipt is possible, because this specific text type, like many others, comes with a built-in range of graphic features that can be imitated. Wehde (2000) calls such configurations of typographical and layout properties which form a set of visual expectations tied to a particular text type (format) a "typographisches Dispositiv", which could be rendered in English as the 'typographic repertoire' of a text type. Which typographic/layout sub-modes, then, have been employed to give the impression of the typographic repertoire of a receipt?
should ban all battery cages as soon as possible. Margaret Beckett is considering it. So help her decide. Stop buying battery eggs. Farm animal welfare. It's in your hands.
www.rspca.org.uk/eggs Registered charity number 219099

The visual image of a receipt is mainly formed by narrow margins, which have been marked by lines of three stars each. The text body is heavily subdivided into small portions of variously aligned print. This has been achieved by paragraphing and spacing as well as by lines formed of the same stars as at the margins. The font clearly betrays its provenance as having been produced by the typical cash-desk printer. This is signalled through the formation of the characters from individual dots, a system also used in digital displays. Contributing to this is the blue colour of the print as well as its irregular quality and the blurred print blotches in between, which by association also indicate the low but functional quality of the paper and the printing technology. The enumerative and strongly portioned character of the receipt also materialises through bold print, tab stops in the middle of lines, capitalisation and the use of numbers so typical of receipts. Besides being heavily paragraphed, receipts also come as parts of a continuous text (paper roll) from which they are torn when handed out to the customer. This continuous character of the text is borne out by the recurring logo on either side and the centrally aligned text body right underneath, which usually communicates the name and address of the supermarket or some other standing detail. Continuity is also expressed by the cut off logos at the top and bottom of the text.

When typographical repertoires are exported from one text type (receipt) to another (advertisement), the resulting effect is not merely pictorial as in our example, where the receipt is sort of refined as a textual object. More importantly, exported or emulated typographical repertoires have a semantic impact. In the RSPCA advertisement the receipt-like character of the text adds to the meaning of the verbal text. It supports the central argument of the ad, which says that purchasing behaviour can make all the difference in the battle for more free-range farming. The receipt image of the text makes the pivotal point that it is in the supermarket where farming policies are shaped via the price of the eggs and consumer behaviour. The readiness to overcome one's own meanness and spend a few pence extra on eggs as the target of the advertisement is suitably transported in the visual image of a receipt, as it directly taps into the knowledge script of shopping and thus reduces the mediated effect of communication and makes it more direct.

5.2 The verbal image and the language-image-link

Print advertising is a textual genre whose reliance on language-image-combinations is almost obligatory. In the design of such inter-modal links creators of ads exploit the whole range of semantic, formal and pragmatic relations, which can be
see why more people are working wherever and whenever they like, call us on …
Hopefully you'll agree, Toshiba really do have portable computing canned (9)...

The first level to scrutinize is the semantic and pragmatic tie between language and image. It is obvious here that both the image and a whole network of linguistic expressions function metaphorically. The central metaphor could be spelt out as LAPTOP = FISH CAN (visual image) or BUILDING LAPTOPS = CATCHING FISH/CANNING FISH (verbal text). Whereas the visual image provides the overriding metaphor which serves as a frame of orientation for the interpretation of the verbal metaphors, the literal meanings of the metaphorical phrases (in italics) detail and structure the emerging mental images (sea, fishing). The result of this kind of visual-verbal metaphorical play is the mental mapping of source domain features (sea, fishing) onto target domain features (computer manufacturing). So here analogies are built between

(1) the sea and the market (oceans apart, create waves, take the plunge)
(2) the quality/quantity of the catch and the quality of the computers (no small fry, good haul)
(3) fish processing (canning) and quality computer assembly (have portable computing canned)
(4) fish/fishing and computer firms (rather lead the shoal than swim with it, no catch)
(5) instruments of fishing (bait) and compatibility with networking standards (take the bait from most networks).

On a second level, we could ask which cognitive operations are afforded by the design of the language-image-link. Clearly, what is intended here is the oscillation between literal and metaphorical meaning-making throughout the text, but also the successive integration of verbal phrases into a whole. On the one hand, all figurative expressions can be read literally (i.e. with recourse to sea and fishing), an interpretation facilitated mainly by the visual image, which provides the concept of a can with all its physical properties. On the other hand, of course, the phrases can be decoded in their metaphorical meanings, which support the persuasive intention of the text. Both on the literal and metaphorical plane of meaning the phrases combine to build a structured network. The network’s elements cohere because on a literal level they build paradigmatic sense relations and all add to a common mental image. On a metaphorical level the phrases are networked as they all contribute to realising typical advertising speech acts and establishing the advertising text pattern. So, no small fry and good haul describe and evaluate the advertised product, created waves, rather lead the shoal than swim with it and have portable

We've squeezed a desktop into a portable. No catch (1). The only difference between a Toshiba portable and a desktop is scale. When it comes to size and weight the two are oceans apart (2). And for power, function and features our portables are no small fry (3). Features like the best chips, Intel MMX, the largest screens, 13,3” and the fastest CD-ROM drives, mean our portables are a good haul (4). And the PC Card Slots on every Toshiba mean our portables can take the bait of most networks (5). Whilst our innovative firsts have always created waves (6) amongst desktop and portable manufacturers alike. It's because we would rather lead the shoal than swim with it (7), that makes us the world leader in portable computing. Our portables have consistently landed prizes from PC Magazine for service, reliability and technical innovation. If you'd like to take the plunge (8) and
computing canned promote the firm's image and express praise, whereas take the plunge is an appeal to the consumer to try out the advertised products.

On a third level of analysis, the overall textual structure built from language and image in our sample text is—as we have seen—one of metaphorical projection and literalisation. Verbal and pictorial text are strongly interdependent in as much as the visual image promotes the literal readings of the figurative phrases, and the metaphorical language explains the context and the motivation for the pictorial metaphor.

Finally, on a fourth level of analysis we need to enquire into the specificity of the visual image. Images in concrete communicative events always come as types, and there are a number of design features in our example that are for one thing typical of advertising and, for another, facilitate the language-image--link described so far. Most importantly, the image's metaphorical nature is realised by a morphing technique which allows for the carefully engineered blending of visual features. As a result the reader gets a very realistic impression of an imaginary object, which plays tricks on his perception. Shape, size and colour come as unifying characteristics of both objects (can, laptop) morphed into a single gestalt, while can opener and keyboard as well as the plug-in connections represent distinguishing traits of the objects blended. In advertising images single objects are often shown against a neutral background in order to bring out their salient characteristics, as has been done here. Also, there is something like a functional perspective in advertising images, which makes the perception of the objects depicted as easy and effective as possible. The can/laptop is shown from slightly on high and from a relatively short distance so we can easily take in all its important attributes. Curiosity and tension is created by the half-open state of the can with the opened part pointing away from the viewer.

6. Semiotic principles across modes

The two sample texts were to demonstrate that mode-integration may be complex, because there is both mode mixing, i.e. the calculated and complementary co-deployment of language and image, as well as mode overlapping, i.e. the collapsing of modes into one another. Overlaps of modes are seen in instances where language can be doubly pictorial. Firstly, verbal text can assume pictorial quality via typography and layout and secondly, language can be based on and evoke mental imagery. Mode-overlapping, however, also reflects in the easily neglected fact that images are to a great deal rooted in language or rather in knowledge frames and scripts which are heavily codified and structured in the form of our linguistic repertoires. So, what can be expressed and communicated in images (in production and reception) is not only dependent on our visual experience of the world or the material and technical properties of image-media, but is also crucially shaped by our stock of words, phrases and stereotypical language utterances.

When separate modes indeed so closely intertwine in multimodal texts, is it not likely, then, that "common semiotic principles operate in and across different modes" (Kress & van Leeuwen 2001: 2)? This is a central question of multimodal theory and analysis. As we have seen in the case of language and image (cf. Section 3), modes differ noticeably from one another in terms of internal sign structure, semantic potential and cognitive operations afforded. This is why individual modes need to first of all be regarded as possessing their own 'grammars', which are distinct from one another as they follow different organizing principles and make different functionalities available. This does not, however, rule out the possibility that some overriding principles govern and guide all modes simultaneously. Such trans-modal operating principles, then, would have to be sufficiently general and basic to be able to span the great variability of modes. My view on common semiotic principles across modes, therefore, is a balanced and dynamic one. I endorse the formal and functional differentiation between modes while at the same time acknowledging the trans-modal operation of very global semiotic rules governing the organisation of individual modes and their reciprocal integration. The general cross-modal principles are only instantiated in texts and communication, and I suggest to look at those principles as a means to regulate and guarantee a kind of semiotic equilibrium in a concrete multimodal text. In what follows I will roughly sketch out some of the basic semiotic principles operating across modes.

1. The three Hallidayan meta-functions (Halliday 1994) would be the first principle that can easily apply to all modes imaginable and to the multimodal text as a whole. Any mode is—to varying degrees—able to depict states-of-affairs (ideational), design some social interaction between the communicators (inter-personal) and contribute to organising and structuring the text (textual). In any one multimodal text these three functions need to be fulfilled and, more importantly, distributed across the modes present. Here, the aim must be an inter-modal balance between the meta-functions, that is their distribution across modes will be guided by how, in a given communicative event, the functions can be realised most efficiently. Consequently, modes will be positioned towards one another according to which part they play in fulfilling the meta-functions. This is a first key to the structuring of multimodal discourse.

2. Segmentation, that is the decomposability of larger sign structures or gestals of perception into their constituent elements, would seem to be a second semiotic principle operating in and across modes. All modes need to signal their internal structure as keys to the retrieval of portions or layers of meaning.
But segmentation also applies to the organisation of the multimodal text as a whole, which needs to be indicated by structurally salient signs connecting modes and specifying their semantic ties. Syntax, then, in its widest sense is a core feature of any accumulation of signs, both with respect to the internal structure of one mode as well as with respect to the overall structure of a multimodal text.

3. Meaning in texts comes on three interrelated planes. Signs can refer to concepts (denotation), they can convey concomitant, socially shared emotive or evaluative meaning (connotation), and signs can also activate and tap into purely individually valid facets of meaning (association). Modes seem to differ in their ability to provide for those three types of meaning: music, for instance, seems weak on denotative meaning, but strong on associative meaning. Language, on the other hand, has its strength in the domain of the denotative, but can also service other layers of meaning. However, it has relatively little potential on the associative plane. Again it is my claim, here, that multimodal texts need to generate and conserve a kind of balance between denotation on the one hand and connotation/association on the other. The deployment of modes in a multimodal text will seek this semantic equilibrium, and hence its structure will reflect the adherence to this inter-modal principle. So in a commercial, for instance, sound and language will usually be geared towards the denotative, whereas music clearly provides associative meaning. The image stands midway between the two as it can deliver both denotations as well as associations.

4. Another common semiotic principle of multimodal communication can be seen in the necessity of modes to build semantic relations to concomitant modes. As soon as two modes are combined they will automatically create semantic ties, which can be of a great variety. Meaning made in one mode may be repeated or paralleled in another, or it may be complemented, negated, contradicted or reinterpreted etc. Inevitably, co-deploying modes will cause them to come together in perception, and this simultaneity to the senses must be construed semantically in some way or in another.

5. Signs can make meaning in three ways. As icons they can, by virtue of their similarity with the depicted objects, simulate real or possible worlds. Indexical signs point to objects or states-of-affairs other than the one denoted by way of logical operations and mental continguities. Finally, signs can be symbols arbitrarily standing for some meaning not otherwise connected to the signs. These three signing options are universal meaning-making-principles that can apply to all modes, although, of course, modes are known to favour one or the other way of meaning-making. Language is a dominantly symbolic mode, whereas the image is predominantly iconic in nature, yet it may well act as symbol or index, while language may be used iconically or indexically. Again, using modes or individual sign complexes from one mode as icon, index or symbol and thus relating modes to one another is a major structuring device in multimodal communication.

6. Finally, I believe that recognizing gestalt similarities, i.e. inter-modal analogies on both the formal and semantic level, is the core mental operation required in multimodal communication. This applies to the perspective of the producers, who craft analogies, just as well as to that of the recipients, who have to recover those analogies and interpret them. Analogies are so central to the design and understanding of multimodal discourse because it is mainly thanks to them that individual signs from various modes cohere into a common whole and form a unified gestalt in perception (Stafford 2001).

7. Conclusion

My observations on multimodality were focused on the interface between language and image. More than anything else, it should have become obvious that inter-modal relations like those in the language-image-link are anything but simple and pose both theoretical and analytical problems. For one thing, the notion of ‘mode’ justifies the view that what we do in multimodal communication is, toolkit-like, pick from a set of signing resources according to concrete communicative intentions and an envisaged overall design of the text. For another, this perspective is too static. While it is certainly true that modes have their individual characteristics (semiotically, semantically and cognitively) which pre-determine how they can be deployed in a textual structure, the dynamics of meaning-making must be given due emphasis. In the production of multimodal texts modes and their respective sub-modes blend, shift and mix in possibly myriad ways. It is only through a multi-faceted and integrated analysis on all levels of text, which is susceptible to the dynamic processes of “inter-semiotic shifts” (Iedema 2003:42), that meaning-making can be reconstructed. The meaning recovered in such analyses is both a reconstruction of meaning deployed by the producers as well as an interpretation of the meaning construable by recipients.

Notes

1. The confusion between multimediality and multimodality can be seen as an effect of the hype surrounding the discipline, which has mainly been generated by a fascination with new technologies and their apparently boundless opportunities.
2. Iedema (2003:40) calls this kind of argument the "always already (Überhaupt) multi-semiotic nature of meaning-making".


4. Such 'grammars' of individual modes are, for instance, outlined in Kress and van Leeuwen (1996), Doelker (1997) or Stöckl (2004) (image), in van Leeuwen (1999) (sound and music), in O'Toole (1994) (displayed art, e.g. sculpture), in Stötzer (2003), Willberg and Forssmann (1999), Walker (2001) or Wehde (2000) (typography), and in Kühn (2002) (non-verbal communication). While some of them are based on an explicitly systemic functional approach, others apply different methodologies. Common to most theories of single signing modes is the (metaphorical) transfer of some kind of linguistic or semiotic pattern. This goes to show that language is seen to be central in signing practice.

5. Kress (1998:55ff.), for instance, suggests that it is first and foremost a change in social conditions, and not technology as such, which drives the shift from language to image in the communicative landscape. He pinpoints information overload as a cause for a greater reliance on the image.

6. Kress and van Leeuwen (2001) is the first attempt to outline a general theory of multimodal communication. The objective to be all-embracing has, however, been pursued, here, to the detriment of specificity.

7. Stöckl (2002) shows how the static printed image is also capable of realising narrative.

8. The terminology used for sub-modal and features here is heterogeneous. I have largely followed Kress and van Leeuwen (1996) and van Leeuwen (1999), but have also used rough labels borrowed from generally accepted vocabulary. As for typography, I have made use of Stötzer (2003) and Willberg and Forssmann (1999). For inspiration, I have also turned towards Neuenzwander (2001) and Bellantoni and Woolman (1999). As for the para-verbal, I have consulted Neuber (2002). The linguistic sub-modes follow accepted notions of levels of text-analysis. Again it needs to be emphasised that my aim was a very general but systematic scheme of things, not total completeness or precision. An application of multimodal analysis to the TV-commercial which roughly follows the lines sketched out here can be found in Stöckl (2003).

9. Kress and van Leeuwen (2002:350f.) can be understood to generally endorse the idea that colour is a mode in its own right, although they also advise caution. Their main argument for regarding colour as a mode is that "it can combine freely with many other modes" (ibid.:351). Although that seems true, colour, when combining with other modes, is part and parcel of those modes, which is why I would like to maintain that colour comes as a sub-mode.

10. The following differences between language and image are presented in more detail and with relevant bibliographical sources in Stöckl (2004:245ff.). Almost all of the arguments represent a broad consensus in semiotics and cognitive psychology.

11. A concept of images based on their iconic nature can also be maintained when applied to abstract, i.e. non-representing images. Colours and shapes in spatial combinations that do not refer to concrete objects real or imagined will mainly make 'meaning' by associations we have with these colours and shapes acquired in sensory experience.

12. By synaesthetic connections images can also communicate information other than the purely visual.

13. Stöckl (2004) is a detailed study of the language-image-link with respect to advertising and journalism.

References


Stöckl, Hartmut (2002). "From space to time into narration – Cognitive and semiotic perspectives on the narrative potential of visually structured text." In H. Drescher, W. Thiele, & C. Todenhagen (Eds.), Investigations into Narrative Structures (pp. 73–98). Frankfurt am Main: Peter Lang.


