The language-image-text – Theoretical and analytical inroads into semiotic complexity

Hartmut Stöckl

The present contribution sketches out a social semiotic and text linguistic view of the language-image-link. I start out by placing this specific kind of bi-modal link in the wider context of multimodal text theory explaining what multimodality is and how it is motivated historically and cognitively (chs. 1, 2). I then propose a model of how pictures are understood in context and in combination with language (ch. 3). Chapter 4 briefly illustrates the social semiotic view of a pictorial grammar. At the heart of my contribution is a methodology for the multi-level analysis of the language-image-link and a typological approach to the link (ch. 5). Finally, I extend the scope of my enquiry to include special types of links which are created through the pictoriality of language (ch. 6). Either figurative expressions evoke mental imagery or writing, layout and typography produce what could be called “typopictoriality” (Weidemann 1997). A conclusion looks at some requirements and directions for further research (ch. 7).

1 Intro: Why do linguists go multimodal?

It would seem to be a contradiction in terms for a linguist to deal with pictures, music, noise and other non-verbal sign systems. And, indeed, even though the acceptance of multimodal approaches to text, which seek to integrate various semiotic modes, is on the rise (cf. Kaltenbacher 2004 for a concise overview of multimodality research), linguistics is still under quite some pressure to legitimize the pictorial or multi-semiotic turn. Let me, therefore, start out with a few reasons why a consideration of pictures as semiotic artefacts seems inevitable in linguistic accounts of textual objects.

Whether printed or running across the screen, writing itself possesses a visual dimension, which the German term Schriftbild aptly captures. Layout, typography and the materiality of the page come as indispensable con-
comitants of language use (Stöckl 2005, 2004c, van Leeuwen 2005b, 2006),
just as speech cannot be separated from its natural setting and communicative situation, which integrates all kinds of non- and para-verbal elements, such as gestures, gaze, body language and, perhaps most importantly a shared visual and tactile experience. Add to this the fact that language is full of frozen images in the form of idioms and you arrive at a notion of language which is far more pictorially grounded than commonly assumed. Ultimately, language, communication and text are essentially multimodal, rather than exclusively verbal.

Historically, it has been a long way from the odd illustrated medieval manuscript to the full-scale visualisation of contents in nearly all sorts of documents today. The development of human mediated symbolic expression seems to have proceeded from pictorial forms (early cave paintings, pictographic writing systems, biblia paupera) to verbal forms. Some sceptics have misinterpreted the technologically conditioned flood of today’s images as a return to early pictorial representations (cf. Ross 2001: 382). However, on closer inspection it seems that ideally, communication always involves a division of labour between language and image. So the history of text types is — from the very outset — a history of the types of language-image-texts. Interestingly, many modern types of ‘mixed’ texts mirror practices that have been around for very long: the Bayeux tapestry anticipates the comic and Da Vinci already commanded all the techniques of linking language and image that are now the stock repertoire of technical illustration. For diachronic text-linguists there would seem to be ample scope for research on how exactly multimodality crept into the making of texts and to chart the dynamics of the development of language-image-links (cf. Eckkrammer 2005).

The prime reason for linguists to get involved with pictures, though, is textual reality. If it is true that images convey meaning over and above ornament and embellishment, then this meaning needs to be linked to the linguistic parts of text. As communication is a patterned activity, we would expect a whole set of text-type-sensitive devices and techniques to be in action for linking pictorial and verbal content. It is the task of a textual, multimodal semiotics to uncover those patterns and to model how picture and language co-operate to form a coherent whole. Cognitively, the two major semiotic modes go hand in hand anyway as understanding language often means to mentally picture things whereas understanding pictures often necessitates a knowledge of language, communicative routines and set phrases (Stöckl 2004b: 18f.). As Androutsopoulos (2000) put it, the majority of text-types are “constituted verbally but structured and organized visually”.

I will now just briefly turn towards a rough account of what we mean when we say texts are multimodal and can be studied in a social semiotic framework.
2 What is multimodality? – Transcribing content

The linguist’s pride in language often obscures the fact that communication is ultimately and always multimodal. The mono-modal text, therefore, seems as much of a fiction as is the idea, advanced in early Chomskyan days, that language is a separate mental system cut loose from other kinds of cognitive faculties. There are two obvious facts which promote multimodality as the natural mode of communication. Firstly, as humans are equipped with more than one sense, it seems fair to address all of them, if possible, in mediated communication. Secondly, what sign-users generally want from texts is the possibly perfect simulation of reality: semiotic objects are supposed to convey information in a true-to-life fashion, which reduces the discrepancy between the world depicted and the medium used. It would seem that the more senses and signing modes are employed in communicative tasks the more effectively meaning can be conveyed and negotiated.

A third, even more essential, argument for multimodality as the *conditio sine qua non* comes from modern media philosophy. Jäger (2002: 37) argues that we can never access the surrounding world immediately, but only indirectly through the use of communicative media, a phenomenon he labels “media immanence”. When we use media and the concomitant semiotic modes employed in them we seem to be permanently “transcribing” meaning from one medium/mode to another. These “transcriptions” characterize the communicative and textual repertoire of a culture. Pictures and films are commented on in language, verbal texts (mainly literary ones) are converted into pictures or music. Music again is scripted, performed, talked about and analyzed. But “transcriptions” also work within one medium: a difficult verbal text may thus be annotated and heavily discussed, a piece of music can quote or re-work well-known tunes, and a picture might take up a famous motif or style.

The reason why we engage in those “transcriptions” is the limited potential of each and every mode taken in isolation. Historically, this may mean that new media and modes came into existence as a reaction to a gap or need in the communicative landscape. Jäger (2002: 39) claims that the prime function of these “transcriptions” is to make texts “legible” and comprehensible in the first place and to extract meaning from semiotic artefacts. We can obviously only make sense of the world around us by commenting, paraphrasing and explicating one mode with the help of another. Every semiotic mode, then, is unique and makes resources available that other modes cannot. It commands its own “autochthonic” (Holly 2007: 392) semantics, which is shaped both by potentials and assets as well as by weaknesses and shortcomings. Cultural or social semantics – seen as the multiple senses and readings that can be gained from the total sets of texts circulating in a semiotic community – only comes about through interrelating
and integrating various semiotic modes. For a variety of reasons language is seen as central in those processes of transcription. It therefore is often regarded as the archetypal medium (cf. Archimedium in Jäger 2002: 34), mainly thanks to its semiotic qualities.

In conclusion, it would seem that there are at least three perspectives on multimodality, which may be insightful and beneficial for those interested in text linguistics.

1. Multimodality is the co-presence of various semiotic modes in a given overall text (compare the German term Gesamttext used by Doelker 1997, 2007). Among the major modes are: language, picture and sound (music/noise). It seems difficult to neatly distinguish modes as they frequently overlap, intermingle and combine (cf. Stöckl 2004b: 11–18). The essence of multimodality seems to be that the various modes are integrated and interrelated on a number of levels (syntactically and semantically).

2. Multimodality, more generally, relates to an all-pervasive semiotic and cognitive activity of transcribing one mode/medium/text into another for the sake of getting at meaning and making sense of a culture’s discourses. Multimodal texts become “legible” only when transcribed. A given “pre-text” (source-text) is converted into a “script” by means of transcriptions (Jäger 2002: 35). In this light, multimodality is a cultural technique, a competence which guarantees communication and mutual intelligibility. Both the production and reception of texts build upon this “transcriptive intelligence” (Jäger 2002: 35).

3. Most importantly, multimodality is necessitated and shaped by the semiotic strengths and weaknesses of the individual modes. Linking modes in complex texts as well as paraphrasing, re- ‘writing’ content from one text/medium to another draws upon techniques and conventions. So, ultimately, multimodality is a patterned semiotic activity, in both producing or understanding texts. Text linguists may reasonably expect to uncover the patterns used to link one mode to another. In what follows I hope to be able to shed some light on the methodology and underlying semiotics of one specific multimodal link, namely that of language and image.

3 How does meaning enter the image? – Understanding pictures

There are two opposing views concerning the semiotics of pictures (cf. Stöckl 2004a: 11–20, 47–86 for an overview of pictorial semiotics). Calling pictures “messages without a code”, Roland Barthes (1977: 43, 45) empha-
sized that we apparently understand pictures intuitively because they look like the real world-objects they depict. This view – based on the iconicity of pictorial signs – is still going strong in cognitive accounts of semiotics. Its merit is that looking at pictures is somehow seen as corresponding to real-world processes of vision. So, in perceiving pictures we access the same kind of knowledge, the same kind of ‘mental models’ as we do in other visual processes. And indeed, it seems that knowledge and experience greatly facilitate pictorial understanding.

The opposite view, based on the metaphor of visual ‘grammar’, claims that – just as language – the visual image is a coded semiotic object which follows rules that connect form to meaning. Most notably, Kress & van Leeuwen (1996) have developed a highly praised (cf. Kaltenbacher 2007) functional account of how we ‘read images’ which is reminiscent of linguistic grammar. The bottom line of the theory models a picture on a sentence, saying that in any one picture there are participants or actors which relate to one another spatially so as to express specific actions, processes or states. Beyond depicting (representation/ideation) pictures also address the viewer by establishing a certain interactional relationship, and they form a graphic composition, which makes additional textual meanings available (Kress & van Leeuwen 1996: 40ff.). In many ways, Kress and van Leeuwen’s account is functional grammar revisited and adapted to the visual image.

Both views have their merits and pictorial meaning probably emerges from a whole series of perceptual and cognitive operations. Understanding images involves both checking pictorial content against stored knowledge as well as following some coded rules of graphic design. In what follows, I will give my own, slightly broader account of how recipients make meaning of images in context.

It has been emphasized in Social Semiotics (van Leeuwen 2005a, Kress & van Leeuwen 2001) that every semiotic resource, every mode has an inherent meaning potential of its own which does not easily compare to that of another mode. What does it mean to tackle textual artefacts from a social semiotic perspective?

1. Social Semiotics looks at how signs (semiotics resources) are used in certain social practices. So it always involves singling out some discourse segment or genre and applying relevant theoretical conceptions to it.

2. Social Semiotics seeks to integrate various semiotic modes. It emphasizes the common principles underlying complex communicative artefacts/events.
3. Social Semiotics asks how semiotic practices are driven by social conditions and psychological needs and how they are embedded in them and emerge from them.
(van Leeuwen 2005a: 1, xi)

I am now going to apply this rationale to an account of how recipients make meaning from pictures in context. The social practice I chose is advertising (Stöckl 2004d), so my sample text is an advert integrating picture and language (cf. fig. 1). Socio-psychologically, advertising is remarkable for its strong reliance on pictures to effectively communicate minimalist commercial messages in a context of recipients’ fleeting attention or downright ignorance of the medium. The advertising genre is known to feature a rich variety of language-image-links, which mainly stems from the creative desire to break with established norms (cf. Gaede 2002) and get the audience’s attention through suspense (cf. Fill 2007: 61f., 135–149), shock, novelty, hyperbole, paradox, etc.

The model I have devised emphasizes two points. First, it conceives of visual understanding as a succession of stages at which the recipient engages in a number of essential perceptual and cognitive tasks. Although these stages are organized sequentially, no claim to a fixed ordering is intended. It is best to think of the model as an individually flexible process, a cycle which can be repeated and which can be accessed at various points. Second, every cognitive operation involves, most importantly, an act of identifying distinctive pictorial features. So at every stage in the process of understanding recipients categorize what they see and allocate it to a certain type of picture or visual pattern. My account of meaning-making in language-image-texts is as follows (cf. Stöckl 2004a: 115–129):

1. Even before viewers become aware of pictorial content, they make educated guesses as to what FUNCTION or PURPOSE the picture in question fulfills. They can do this easily as they are familiar with genre conventions, the look of an ad (“external text-type markers”, cf. Gieszinger 2000, 2001) and its location in various media. Advertising images seek to impress, to arrest the gaze, to help develop an argument or to envelop the viewer in a certain aura and mood.

2. Understanding an image presupposes seeing visual shapes (gestalts) and integrating them to form meaningful signs (of objects/entities). At this stage viewers register the QUALITY of the picture: they notice whether it is simple or complex in design, whether it can easily be read or not and whether it comes up to their or any standards of aesthetics. In our example (cf. fig. 1) only one shape emerges (simple), which turns out to pose a problem of ‘legibility’ (curvy, bulging object/car?) to be sorted out in
Fig. 1: VOLKSWAGEN New Beetle, Ogilvy & Mather Rightford Searle-Tripp & Markin, South Africa 2000 (Wiedemann 2006: 597)
context. Colour and shape seem optically extravagant and are pleasing to the eye.

3. At a next stage viewers will try and form an idea of pictorial content. This involves two things: working out what is depicted and in which context or situation. Viewers construe how the graphic configuration depicts the world and how it relates to reality (PRACTICE OF DEPICTING). They also notice the TECHNICAL/MATERIAL NATURE of the image. Being a photograph the picture denotes a real-world object. It has been manipulated, however, so as to present a view of the object (car), which we do not normally get in real-world experience (PERSPECTIVE, VISIBILITY).

4. Finally, the two semiotic modes need to be integrated to produce an overall message (SEMIOTIC COUPLING). Due date February 28th is a predication which calls for a nomination and two objects seem likely candidates: the new car advertised (cf. VW logo and picture tilted by 90° to the left) and – evoked by the set phrase and the curvy, bulging object (pregnant belly) – a baby to be born. The visual design (oscillation between two objects) suggests a metaphor, which ultimately provides the commercially relevant message: the launch of the new ‘Beetle’ is like the birth of a human with all the ramifying connotations that this analogy may spin off.

What I have outlined here is a likely reading of the advert in question. It highlights a few underlying cognitive-semantic principles as well as the socio-semiotic nature of present-day advertising (cf. Stöckl 2008). Recipients need to integrate pictorial and verbal content in context. They manage this by oscillating between the signs provided and messages suggested. The sense is generated in a quest for a likely commercial message, a process driven by working out likely nominations and predications to form a simple statement relevant in the context of announcing a new car. The semiotic modes taken separately do not produce any stable meaning – they merely offer vague meaning potentials to be activated and shaped in reciprocal integration. More specifically, in order for a coherent, homogenous textual whole to emerge, language and image will have to be integrated in perception and cognition on at least three levels of text: thematic conceptual (content), speech events and pragmatic functions, rhetorical-logical (cf. chapter 5 and Stöckl 2009). So language and image cooperate as they co-manage and interrelate concepts, functions and rhetorical-logical operations.

Rather than make any banal product claims, advertisers try and connect their goods to experiences, realities and values capable of positively affecting the commodity. The sample discussed illustrates the kind of semiotically minimalist advertising now in fashion, which some call “no-copy ads”
Here, a limited number of signs are presented to activate the recipient’s cognitive involvement (cf. “sympraxis” in Kloepfer 1987) in what is a highly indirect, connotative and metaphorical text. This technique is a brake on cognition as it slows down the path from the signs presented to the senses intended (Stöckl 2008: 193f.). More often than not, such messages are also supposed to socially provoke – is it ethically acceptable to compare giving birth to car manufacturing?

4 Are pictures coded? – A functional ‘grammar’ of the image

So far I have made no attempt to discuss the view that pictures follow the rules of a code. Instead, I concentrated on how meaning may derive from vision, cognition and context. These priorities may indicate that I am somewhat sceptical of a visual grammar. Yet, the view has its merits, which I ought not miss to explain. I will do this as briefly as possible dispensing with details (for a concise account of visual grammar, cf. Jewitt and Oyama 2001).

The core idea of functional grammar (Kress and van Leeuwen 1996: 40–42) has it that every semiotic artefact (semiotic system) operates on three levels. First, it represents ‘reality’, that is, it denotes objects, situations, actions, etc. (→ ideational). Second, it establishes a certain kind of social contact and interaction with the recipient (→ inter-personal). Third, it builds a textual structure whose parts cohere formally and content-wise (→ textual). Any rule will operate on one of those levels, which are also called “meta-functions” of semiotic objects or events. The essence of a visual grammar is the belief that there are formal configurations or patterns (i.e. spatial arrangements of signs) in pictures, which come equipped with certain, more or less stable social meanings.

On the ideational level, viewers work out which fragment of the world is represented in the picture in which fashion. If “participants” (i.e. recognizable objects in pictures) are connected to one another by “vectors” (real or imaginary diagonal lines), the picture denotes some kind of action, process or event. In contrast to those “narrative representations”, an absence of such a vector would make any image a “conceptual representation” (Kress & van Leeuwen 1996: 56ff.). Conceptual pictures show objects “in terms of their (…) stable and timeless essence, in terms of class, or structure, or meaning” (Kress & van Leeuwen 1996: 79). A conceptual picture demonstrates properties of objects, explains their parts and functioning or compares objects to one another. Ideational meaning may be enhanced by those visual objects that show context, setting or accompanying elements (→ circumstances) of the actions or concepts depicted. Similarly, the main participants are characterized by their looks, their gestures, body language and clothes.
On the interpersonal level (Kress & van Leeuwen 1996: 121ff.), viewers are positioned in a number of ways relative to what is depicted. First, in terms of a general speech act orientation ("image act") pictures can either offer information to be studied and taken on board or they can demand some action from the viewer and function as an expressive appeal. Second, depending on the size of the frame, viewers are positioned at a certain distance or closeness to what is depicted. The functional meanings range from intimacy or focus on details (close) over social distance (medium) to impersonal detachment (long). Third, the angle of the shot may create certain attitudes towards pictorial content in the viewer. The options here are dominance/superiority (high) vs. fear/inferiority (low) and involvement (frontal). Fourth, pictures have a coding orientation (ibid. 168ff.), which determines how viewers interpret the relationship between pictorial content and reality. Coding orientations may be e.g. “naturalistic” (photograph), “abstract” (x-ray), “sensory” (fashion photography), and “technological” (floor plan). They are signalled through whole clusters of form and design features.

On the textual level (ibid. 181ff.), the attention of the viewer is organized and guided during visual understanding. First, depending on where objects are positioned in the picture plane and what the overall structure of the composition is, various elements obtain different “information values”. A centre/margin structure, for instance, will position the super-ordinate concept or the main subject in the centre and all the subsidiary elements around it on the margins. Second, it is crucial for visual understanding to work out which shapes belong together and which ones do not. Spatial closeness and graphic linking are two devices that signal semantic unity. Third, certain elements may be made salient through a number of visual devices like colour, lighting, contrast, shape, size or others. Like intonation in speech, salience gives prominence to those elements that are deemed semantically central to the overall message.

I feel ambivalent as to how to assess this view of a visual grammar. What is commendable, of course, is the structure this approach provides for explaining how pictures make meanings. All categories introduced seem plausible and create an ordered system. However, we have to acknowledge that the grammar proposed is weak in the sense that there are no hard-and-fast rules and no clear-cut distinctions between the types of visual configurations delineated. More importantly, what makes me somewhat wary is that without a detailed knowledge of the objects depicted in images and without a consideration of the relevant contexts, this kind of grammar cannot really explain the whole story of understanding pictures. So, it seems the meaning of images resides, after all, not primarily in the categories suggested here, but more so in our ability to relate pictorial content to real-word vision and to the full scope of our experience.
5 Is linking a patterned activity? – Types of language-image-links

Graphic design – even if limited to one genre like advertising – seems to come up with a bewildering variety of linkage between picture and language. The sample ad discussed above picked out just one way of combining the two modes. But, hopefully, the interpretation supported the assumption that quite some cognitive work goes into reading language-image-texts. It surely demonstrated the need to study the language-image-link on the basis of a multi-layered methodology. Of course, one may suspect that, given the huge scope of human creativity both in making and reading texts, no system can ever be gleaned behind the mechanics of language-image-links. Gross (1994), for instance, has argued that “semiosis” (i.e. meaning-making) tends to be “wild” invariably, in the sense that multiple readings are available and connections can be made at random.

Notwithstanding these familiar arguments, I should like to claim here that types or patterns of language-image-links can be identified. Just like the potential for semantically linking propositions is limited (e.g. causal, temporal, consecutive, conditional, etc.), the number of logical patterns underlying the linking of picture and text will not be infinite. These patterns are the result of what might be seen as a drive towards cognitive standardization and a desire for easy orientation. To make this quite clear: I am not saying here that there is a limit to what can be expressed, thought or designed on the graphic surface. Rather, my claim is that textual interpretation and production are facilitated by recurring underlying patterns, which still leave ample scope for variation and modification on the designable surface.

If there is a system to designing language-image-links, one should be able to describe it. My suggestion earlier on was that the linking happens on a number of levels. Therefore, a typology of links ought to proceed in a multi-layered fashion, too. I propose three levels here, which may come in useful in a typological description: spatial syntax, info-content, and rhetorical logic. Any specific instance of a language-image-link may then be analysed and categorized on those three levels, which I am now going to outline briefly. The emerging typology pursues a simple objective: it makes transparent cognitive operations and design decisions in the crafting of language-image-texts.

5.1 Spatial-syntactic patterns

On a formal level it is first of all relevant how the two modes are positioned in the space of the page or layout. This criterion may metaphorically be called syntactic because it evaluates the sequence or distribution of elements. Semantically, the topographic relation of picture and text is important as their sequencing or configuration along a likely reading path determines
how sense and message are construed. It is generally assumed that scanning a layout works quite similar to reading, that is, it proceeds from left to right and top to bottom. However, access to a layout might be influenced by elements that attract visual attention by virtue of content in context, size, shape, colour, etc.

Generally, two broad syntactic linking types may be distinguished at first, namely linear vs. simultaneous. Either one mode follows the other in a sequence (linear) so they are more or less neatly delineated in space, or picture and language are spatially integrated so as to be perceived and understood as one visual and graphic entity (simultaneous). These syntactic patterns have immediate repercussions on how meaning is produced from the combination of picture and language. In a linear sequence either picture or language become the starting point, the access to an argument or a statement. Usually, picture-first linkages utilize the visual to introduce the viewer to a scene or constellation of objects, whose meaning potential is then verbally anchored, channelled and pinned down in the advertising context (cf. fig. 1). This often works out as a communicative game of guesswork, semantic manipulation, punning, humour and metaphor. Conversely, language-first linkages mainly employ pictures for illustration. As the recipient has first been confronted with a claim or a product description, the image is easily understood and related to the verbal message. A third syntactic pattern emerges when picture and language take turns in the delivery of the message. In fact, this alternating pattern is perhaps most frequent in advertising. Oscillating back and forth between the two has its own attractions, which spring from the code differences and the revealing discovery that effective messages can be produced with astoundingly small means.

In the ‘Triumph’ advertisement (cf. fig. 2), alteration between language and image also involves turning the page. The simplicity of the idea is impressive, its impact probably mildly humorous and euphemistic.

Fig. 2: TRIUMPH Lingerie, Wirz Werbeagentur, Zurich, Switzerland 1996 (Berger 2001: 106)
Simultaneity of picture and language means both modes are so close that they are perceived as a graphic and perceptual unity. A sequence is not implied here. The simultaneous syntactic pattern rather aims at flexibly integrating verbal and visual content. Two major types are possible with this linkage pattern. Either language and image are configured so that the writing is in the picture space or – and this is the really conspicuous case – writing transmutes into picture or vice versa, so that the boundaries between the two modes become hard to draw. The example in fig. 3 illustrates this transmuting type of syntactic linkage: in an appropriate (verbal) context letter shapes (or whole words and blocks of text) may make simple pictorial signs available.

5.2 Content-related patterns

On a semantic level we are looking at how informational content of picture and language link to form an overall message. Analysis and typology of the linkage are harder here for two reasons. First, as we have seen, pictorial meaning tends to be vague and comes to us rather as something like a potential to be activated and not as a stable given. Second, the linkage works like a reciprocal unity in perception, which can be approached either
way: from image to text or vice versa. So, one may judge the function of the image in relation to the verbal message or the other way round.

In order to keep things simple, I should like to introduce two broad types of content-related linkage, elaboration and extension (cf. van Leeuwen 2005a: 222ff., Martinec and Salway 2005: 349ff.). Whereas in elaboration one mode is used to explain, illustrate or specify the other, in extension new information is added in one mode, which is not co-present in the other. This distinction can, I believe, never be a hard-and-fast one, but it has played a crucial role in the study of image-text-relations since Barthes’ seminal paper (1977). In the ‘Vick Vaporub’ ad (cf. fig. 4) picture and text elaborate one another. The open box is depicted so as to create the shape of a crescent moon against the dark background thus evoking connotations of ‘quietude’, ‘peacefulness’ and ‘rest’, not to mention the more denotative meanings of ‘night’. The verbal description adds product features and the claim good night explicates what the picture merely suggests. It is – as so often in good advertising – left to the recipient to work out the causal logics: ‘Vick Vaporub’ secures a good night when you have a cold.

The image-language-link in the ad for a restaurant/bar called ‘Sopranos’ (cf. fig. 5) works noticeably differently and can rightly be termed extension. Here the baseball bat depicted against the shape of a spotlighted shoeprint is overtly at odds with the verbal message: We strongly recommend the risotto. The picture clearly extends the text or vice versa. A seemingly
paradoxical semantic clash (semantic frames: restaurant/eating – sports (?)/beating) can, however, be resolved by tapping into cultural context knowledge. Media consumers will be familiar with the television series “The Sopranos”, which is set in a social context of mafia petty-thievery and is rife with black humour. The same kind of humour results from the incompatibility of a recommendation and an implied threat (cf. Brock 1996 for a pragma-semantic theory of humour). Where the ‘Vick’ ad (cf. fig. 4) networks associations from parallelized text and image to marshal a redundant message, the ‘Sopranos’ ad (cf. fig. 5) combines different and unrelated information in text and image.

Elaboration can be further subdivided according to which mode elaborates which. So a picture may illustrate a verbal text or language may explain a picture. Here is an implication of direction, which must perhaps be seen as more of a theoretical fiction than a perceptual or cognitive fact. Extension, on the other hand, can be further differentiated according to how the information added relates to the other information present. The emerging inter-modal relationship can be one of similarity, opposition or complementarity. However, as I see it, there seems to be a cline going from elaboration to extension with no neat borderline, so there is an element of extension in elaboration and vice versa.

5.3 Rhetorical-logical patterns

Both levels of analysis discussed so far provide essential information on the kind of linkage. Yet it is only on the third level that we are asking the ques-
tion most crucial to understanding combinations of picture and language: Which logical operation or rhetorical pattern underlies the linkage of the modes? So, whereas syntax and content answer relatively general questions, rhetoric and logic are supposed to uncover what exactly motivates the link cognitively and functionally.

Again, I have opted for relative simplicity and suggest three broad types of linkage. The first co-ordinates language and image so that both enter into relatively straightforward semantic relations, which are based on likeness, contrast and spatial or temporal contiguity. The preferred cognitive operation with the co-ordinated linkage would be comparing, aligning in space and time and associating meaning from picture and language. In an ‘Audi’ ad (cf. fig. 6), a line of buttons and a zip are depicted alongside one another in a de-contextualized fashion. The text underneath the buttons reads *Conventional automatic and manual gearboxes*, the one underneath the zip *The multitronic® transmission*. Both mundane objects act as visual analogies in a co-ordinated language-image-link, designed to demonstrate the advantages of the multitronic gearing. By associating pictorial content and technical term the recipient can work out properties of the technical features advertised without special knowledge. Conventional gearboxes are step-by-step, slow and cumbersome whereas multitronic is direct, smooth and swift. In this example, analogy and contrast go hand-in-hand.
In the second type, both modes are put into a hierarchical order with one mode governing, leading or organizing the other. Here, a more complex logic emerges, which calls on the recipient to work out mode interdependencies on the basis of cause-effect, condition-consequence, part-whole or superordinate-subordinate. In an ad for knives by ‘WMF’ (cf. fig. 7), the claim in the verbal text is *unsurpassably sharp* – the picture shows a delicate little figurine apparently carved from a beetroot. The logic the recipient is supposed to establish features both a causal and a conditional element. If you use the knives advertised, you can cut things with great precision (a visual hyperbole) and because the knives are so sharp, fine cuts can be made using them.

Finally, a third type can be called playful or humorous. It does not subject the relation of image and language to any rigorous logic – co-ordinated or hierarchical – but simply uses the potential for coincidental, allusive and meta-communicative connections between the two modes. Another ‘Audi’ ad promotes the *Audi A4 with free leather interior*. To this end it uses the headline *130 horses and three cows* on a painting of a wide prairie reminiscent of the Wild West. The composition features many horses grazing peacefully with three cows interspersed in the centre. What seems astounding at first sight is the semantic parallelism between picture and headline. The irrelevance of the literal message, however, leads to an interpretation on a metaphorical/metonymic level. Horses, of course, are short for horsepower and

---

**Fig. 7**: WMF Knives, KNSK Werbeagentur, Germany 2005 (Wiedemann 2006: 60)
cows have a metonymic relation to the leather interior advertised. So here the picture literalizes the verbal message and the language-image-link provides a little humorous semiotic game. In addition, the picture can be read to symbolize freedom and space on a connotational level.

Typologies are always fraught with difficulties and none can ever be entirely clear and exhaustive. Yet, the one presented here and similar attempts at categorization (e.g. Bonsiepe 1968, Gaede 1981, Spillner 1982, Geiger/Henn-Memmesheimer 1998, Nöth 2000, Fix 2001, Martinec and Salway 2005, Doelker 2007) make for a good deal of order. The more they do so, the more they help to model what in reality is quite a complex and intricate interaction of semiotic modalities. What we can glean best from typological work is the dimensions, criteria and parameters useful to describe the linking of various codes. It also sensitizes us to the rich variety this linking can generate.

6 What other links are there? – Alternative views

At least two issues are still missing from my account of the language-image-link, which I am going to briefly address now. Both have to do with the fact that pictures cannot really be restricted to the material image itself. Language, too, is essentially pictorial and has the powers to evoke mental images. In natural languages, there is a large repertoire of expressions, often called figurative (cf. Cacciari and Tabossi 1993, Glucksberg 2001 for accounts of figurative language and its processing). Those idioms are, of course, usually used in their non-literal meanings; the original image, responsible for the creation of the idiom has mostly been obscured by time (‘demotivation’). So, *take the lid off* is such a set phrase whose idiomatic meaning ‘to tell somebody about something that was a secret’ coincides with a possible literal meaning (e.g. in cooking). Adverts like the one in fig. 8 for the ‘Volkswagen Sharan’ exploit the potential availability of both meanings thus linking a visual picture with a linguistically evoked one. The charm in this particular example is that the visual image recalls the set phrase, so it can be seen to take the phrase literally, although the expression is not materialized in the ad. At the same time, the idiomatic meaning will also be activated as this is the standard and fits the text: Launching the new car is revealing a secret.

A special case of language-image-link, then, would be texts that combine a visual image with verbal text containing figurative expressions. This is particularly interesting when a text networks a whole series of phrases geared towards stimulating the imagination. A ‘Nissan Primera’ ad shows a prototypical scene from a car race. The text that goes with the picture accumulates figurative expressions which all hinge around the image of dogs and
eating: it’s dog eat dog, like a rottweiler holds a bone, it inspires Pavlovian responses, makes you drool, gets you licking your lips, other cars are nothing more than a dog’s breakfast, whet your appetite. The semantic link between this idiomatic network and pictorial content may be established through metaphor and experiential association. The car race is an illustration.
Fig. 10: LORENZINI Men’s shirts, Claus A. Froh, Germany 1960ies (Wiedemann 1994: 151)

of the dog-eat-dog principle and the joys of driving a powerful car may be seen as analogous to the joys of eating.

Language does not achieve pictorial effects through figurative expressions only. Writing may by virtue of its visual materiality also help to boost
the pictoriality of a text. So over and above encoding linguistic messages, print and typography convey additional meanings, which inter-relate with pictorial or textual content (cf. Stöckl 2005). I should like to point out four ways in which writing and the materiality of texts may generate subtle and effective meaning. First, specific graphic properties of the typography used in a text can spin off independent meanings. So in the ‘Häagen-Dazs’ ice cream ad (cf. fig. 9), colour, size and font type of the headline communicate the kind of irregularity, spontaneity and individuality that go together well with the claim *lose control*. Second, text can be arranged in the layout so as to be portioned neatly into separate chunks. Also, headlines, margin notes or footnotes may help to organize it graphically. All this facilitates legibility and access to selected points. An example would be the older ‘Lorenzini’ ad (cf. fig. 10). Third, text – i.e. individual letters, lines of print, text bodies – may transmute to form or represent pictorial shapes (cf. fig. 3). There is great variety here and the ‘typopictorial’ subtleties possible on the graphic continuum from text to image are beyond imagination. Finally, it is the very material of a text, that is, its graphic substance and the techniques used in its production that can have semantic effects. So, the look of the ‘Schlafly’ beer ad (cf. fig. 11) suggests that pictures and text have been glued onto an old wall. This connotes tradition and well-worn publicity. The text looks as if written with an old typewriter – here again the same connotations are fostered. Add to this the provisional, make-shift character, which stems from the patchwork technique underlying the montage of the ad, and you get an idea of how powerful and subtle those meanings generated by medium and material may be.

**Fig. 11:** SCHLAFLY Beer, Core, St. Louis, Missouri, USA 1995 (Berger 2001: 173)
7 Conclusion

The theoretical reflections and practical, text analytical suggestions presented above are a study in the making. Although no longer academic virgin territory, research into the semiotics of the language-image-link is still in its infancy. What I hope to have demonstrated is that linguists have a lot to contribute to this line of enquiry, provided they take a multimodal view of text and embrace ideas evolved in semiotics, cognitive psychology and other disciplines. In future work the following guidelines may turn out to be valuable.

1. Text corpora need to be built up in order to conduct truly empirical and systematic research into the workings of the language-image-link. The more diverse the material accumulated, the more reliable and sensible the modelling of underlying processes will be (cf. Baldry and Thibault 2006, Bateman 2008).

2. The relevant terminology for describing visual phenomena and graphic design must be revised and unified. We need a suitable language to speak about multimodal texts, their structures and styles.

3. In all this research it is important to be aware of the fact that both verbal text and pictures come in types. So, instead of unduly generalizing, it will be useful to inspect various communicative practices and their genres separately, e.g. advertising, journalism, science and technology, the arts, etc. Abstraction must then follow from those empirical facts. A potentially wide perspective integrating text, picture, graphics, layout, typography, materiality and the like must also be advocated as helpful in the enterprise.

8 Bibliography


