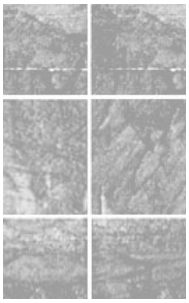
visual communication



ARTICLE

A system for image–text relations in new (and old) media

RADAN MARTINEC
London College of Communication, University of the Arts, London, UK
ANDREW SALWAY
University of Surrey, UK

ABSTRACT

This article presents a generalized system of image–text relations which applies to different genres of multimodal discourse in which images and texts co-occur. It combines two kinds of relations – the relative status of images and text, and how they relate to one another in terms of logico–semantics. Every instance of an image–text combination in the data sample is described by a selection of features from the system. The units of images and text between which the relations obtain are identified and the realizations of the logico–semantic and status relations are specified, both for the human analyst and a machine. Two application scenarios are discussed. The system should be useful for distinguishing between image–text relations for (genuinely) new and old media.

KEY WORDS

image–text relations • intersemiotic • logico–semantics • multimodality system • new media

INTRODUCTION

In this post-modern age when boundaries tend to become increasingly blurred, spurred by intermixing of cultures and the increasingly unfettered flow of information, one of the boundaries that surely deserves attention is that between arguably the most ubiquitous means of communication – text and images. This is all the more so because, due to the relentless pace of developments in information technology, text and images are increasingly coming together, creating multimodal texts. The boundary has however been explored by artists and designers for centuries, from the illustrations of biblical manuscripts to avant-garde movements between the two world wars, and including the more recent experimentation in electronic art (see, e.g., Rush, 1999; O'Donohoe, 2003).

Copyright © 2005 SAGE Publications (London, Thousand Oaks, CA and New Delhi:
http://vcj.sagepub.com) /10.1177/1470357205055928
Vol 4(3): 339–374 [1470-3572(200510)4:3; 339–374]

The boundary between text and image seems most blurred in the case of typographic art, which has been enabled by recent developments in electronic media, but which has a precedent in medieval illuminated manuscripts, and which the avant-garde artists of the earlier years of the 20th century also explored (see, e.g., Drucker, 1994). A great degree of sophistication in this exploration of the meanings and forms of such an extreme image–text fusion has been achieved by Islamic art, with its extensive elaboration of the material qualities of writing (see, e.g., Hillenbrand, 1999).

The focus of this article is on image–text relations where the degree of fusion is not as extreme, but rather the two modes appear separate yet integrated in both semantics and form. Emphasis is placed on presenting a system of semantic relations, and on specifying their perceivable realizations. The layout of the article plays a role, particularly in machine-recognizable realizations, but the more important realizations involves elements of structure of the text and images themselves. This endeavour therefore presupposes descriptions of text and images, descriptions that must to a great extent be compatible. Systemic–functional linguistics and semiotics provide such a theoretical and descriptive framework and we therefore draw on them when building our system (see, e.g., Halliday, 1985, 1994; Martin, 1992; Kress and Van Leeuwen, 1996; O'Toole, 1994).

Much has been written recently about the increasing importance of images in communication (e.g. Kress and Van Leeuwen, 1996, 2001), and there is no doubt that the new media have played a major role in the recent emancipation of the image in the western culture (in some other cultures, like the Japanese, the situation has always been more balanced, see Martinec, 2003). This is certainly in part related to the great ease with which anybody with a scanner or a digital camera, and an easily available computer program, such as PhotoShop, can manipulate images along scores of dimensions such as size, colour, brightness, and integrate them with text by means of authoring software. The resulting multimodal texts can then be distributed over networks to multiple recipients with equal ease.

But notwithstanding the enabling role of technology, there must be reasons why we would want to do something like that in the first place. This may be for more strictly functional reasons – some kinds of images may be better at creating direct emotional impact, and text may be more suited to carrying out logical analysis, for example (Arnheim, 1997). But a simple fascination with reproducing the visual part of our experience may also lie behind this, the desire to achieve transparency, or 'immediacy' (Bolter and Grusin, 2000) in our mediated communication. Such desire finds its most extreme form in the ever increasing sophistication of computer graphics that characterize the very sizeable sector of computer gaming, and of course virtual reality environments, in which the user interacts with more or less faithful replicas of reality.

Even in electronic media, however, images most often occur in combination with text and, despite the increased emphasis on their importance in

the writing of semioticians and new media theorists, not much attention has been paid to analysing the semantic relations that allow them to interact with the surrounding text and create more or less coherent, meaningful wholes that may be called multimodal texts. Most of the work that has been done on image–text relations has either been based on practitioners' brilliant intuitions (e.g. McCloud, 1993; Horn, 1998), has borne a cultural studies orientation and thus not concerned itself with the details of how text and images work together (e.g. Mitchell, 1994), or has been aimed at writers, illustrators or librarians and lacks a theory of how images and text are structured and how they function (Marsh and White, 2003).

The one theoretical framework whose followers have concerned themselves more systematically with relations between images and texts, and with multimodal texts in general, is systemic–functional semiotics. Most systemic–functional semioticians began to inquire into intersemiotic relations in the late 1990s (e.g. Lemke, 1998; Martinec, 1998a, 1998b; Royce, 1998; O'Halloran, 1999) and it is this work which still forms the basis of their study of intersemiotic relations. Two of the publications most significant for our work however appeared even earlier (Van Leeuwen, 1991; and Martin, 1994). This foundational body of work is now briefly reviewed.

Lemke (1998) is a detailed exegesis of a scientific article that combines a diagram and text. The article is for the most part programmatic and, as Lemke explains, there is no attempt to specify a system of intersemiotic relations that would integrate the two semiotic modes. A reference is however made to thematic systems, which in Lemke's earlier work (e.g. Lemke, 1990) were used to model semantic patterns in language and which could possibly be used to integrate meanings realized by verbal and visual representations. O'Halloran (1999) presents a system for analysing mathematical formalism derived from O'Toole's (1994) semiotics of images. Using Halliday's (1994) grammar, she then discusses the translation from language to mathematics by the process of semiotic metaphor – a process by which new elements are introduced as a result of different choices made in the two semiotics. Focusing on translation between semiotic modes, O'Halloran's angle on intersemiotic relations is different from ours.

Royce (1998) presents a detailed analysis of intersemiotic relations between an image and the text of an advertisement in *The Economist*. Following Halliday's (e.g. 1985, 1994) metafunctional model, he presents various ways of intersemiotically relating the ideational, interpersonal and textual aspects of the image and text in the advertisement. His ideational intersemiotic relations are an adaptation of Halliday and Hasan's (1976) lexical–cohesive relations, here renamed as sense relations (cf. Lyons, 1977), and his textual relations are for the most part identical with Kress and Van Leeuwen's (1996) layout systems. The interpersonal relations that relate images and text have to do with the reinforcement of their function of addressing the reader/viewer and with the congruence or dissonance of their attitudinal meanings. We see componential relations, which are a variant of

sense relations (see later) and layout, as realizing our system of image–text relations. As discussed in more detail in the conclusion, we would include reinforcement of address and attitudinal congruence, as well as his attitudinal dissonance under different kinds of relations in our system.

Martinec (1998a, 1998b) presents an adaptation of Halliday and Hasan's (1976) and Martin's (1992) different types of cohesion in language to embodied action. His componential cohesion, which is an adaptation of lexical cohesion to action and to relations between language and action, is similar to Royce's (1998) sense relations. His conjunction is an adaptation of linguistic conjunction to action and its relationship to co-occurring speech. Componential cohesion will be shown to realize the intersemiotic relations in our image–text system. Alongside Van Leeuwen (1991), Martinec's intersemiotic conjunction was an inspiration for our system as a whole.

Our system is based on combining Halliday's (1985, 1994) logico–semantic and status relations, developed to classify the relationship between clauses in the clause complex, with Barthes' (1977a[1961], 1977b[1964]) text relations, whose main object seemed to be newspaper photographs and, to a lesser extent, moving images and dialogue in film. Halliday's logico–semantic relations and their variant of conjunctive relations (Martin, 1983) have previously been used by systemic–functional semioticians to analyse moving images and voice-over relations in film documentaries (Van Leeuwen, 1991) and text–diagram relations in academic discourse (Martin, 1994).

Van Leeuwen's (1991) pioneering work in particular has been an inspiration to us since he combined Martin's (1983) and Barthes' (1977a[1961], 1977b[1964]) relations in one system. Our system differs from his on three main counts. First, we use Halliday's (1985, 1994) logico–semantic relations, which have the advantage over Martin's conjunction of having the system of projection needed to account for projected meanings and wordings in diagrams, comic strips and similar image–text types. Second, Van Leeuwen considers Barthes' categories as having to do with directionality of the image–text relationship and does not include Barthes' 'relay' in his system. We consider Barthes' relations to do with relative status and 'relay' plays an important role in our system. Status in any case implies directionality since the subordinate item modifies the superordinate one, and the directionality of equal items is both ways. Finally, we specify the realizations of both status and logico–semantic relations between images and texts and of the units which they link.

Martin's (1994) account of projection between chunks of text, and between those and some abstract images, such as figures and diagrams is inspiring. However, he does not deal with units in any detail but rather considers projection to relate whole images and texts. In addition, he does not make a distinction between equal and unequal status, and his projection relations lack realizations. The work presented here, on the other hand, identifies the units of both status and logico–semantics, and specifies the realizations of both.

Apart from Van Leeuwen's (1991) conjunction, the previously mentioned approaches have come nowhere near developing a generalized semantic system of image–text relations that would map out how images and text interact. Van Leeuwen's conjunctive relations, however, do not have perceivable realizations. Most of Royce's (1998) intersemiotic relations, which have the potential to be applied to different image–text genres, are at the level of what we consider realizations and are applied to a single textual instance. We have analysed electronic encyclopaedias, print advertisements, news websites, online gallery sites, anatomy and marketing textbooks, and made quick forays into other genres in which image–text combinations occur. On the basis of this research, we have built a generalized system of image–text relations. Our system aims to account, in a principled manner and in some detail, for all the image–text relations in both new and old media. The system may need modifying as our sample of image–text combinations increases; however, even if the relations that we are writing about can be further subclassified and genre-, or register-specific realizations added, we surmise that the outline of the basic system will probably stay as it is.

In addition to explicating our system of image–text relations and identifying their perceivable realizations, in this article we also aim to specify, at least briefly and selectively, realizations which could reliably be recognized by a machine. In the conclusion, we suggest two new media scenarios for their applications.

RELEVANT THEORIES OF TEXT AND IMAGES AND OF THEIR RELATIONS: BARTHES AND HALLIDAY

Barthes' (1977a[1961], 1977b[1964]) foundational study of image–text relations is based on a simple logic of three possibilities of how images and text relate to one another and relies on penetrating observations rather than on any specific realizations. Barthes identified three possible image–text relations: text supporting image ('anchorage'), image supporting text ('illustration'), and the two being equal ('relay'). We argue that two kinds of relations can be discerned in Barthes' classification: logico–semantics and status.

After describing the ways in which anchorage guides the viewer in describing and interpreting an image, Barthes (1977b) says 'In all these cases of anchorage, language clearly has a function of elucidation' (p. 40), and in relation to illustration, he writes that the image elucidates, or 'realizes' the text (1977a[1961]: 25). Barthes' 'elucidation' and 'realization' could reasonably be interpreted as the logico–semantic relationship of elaboration (see Halliday, 1994: 225–9). As for relay, Barthes (1977b) says:

> While rare in the fixed image, this relay-text becomes very important in film, where dialogue functions not simply as elucidation but really does advance the action by setting out, in the sequence of messages, meanings that are not found in the image itself. (p. 41)

Here, Barthes' text advancing the action by setting out new meanings, sounds very much like the logico–semantic relation of extension and perhaps also enhancement (see Halliday, 1994: 230–9).

Some of Barthes' descriptions of illustration and anchorage, however, also lend themselves to being read as if concerned with the relative status of image and text. For example, he writes of anchorage that: 'Firstly, the text constitutes a parasitic message designed to connote the image ... in other words ... the image no longer illustrates the words; it is now the words which, structurally, are parasitic on the image' (Barthes, 1977a[1961]: 25). And he says of relay: 'Here text (most often a snatch of dialogue) and image stand in a complementary relationship; the words, in the same way as the images, are fragments of a more general syntagm and the unity of the message is realized at a higher level' (Barthes, 1977b: 41).

A similar but much more explicit and systematic distinction between status and logico–semantic relations was made by Halliday (1985, 1994) in order to map out the relationships between clauses in the clause complex. Halliday keeps the two dimensions clearly separate, and the options in each combine independently. The status of the clauses in the clause complex is thus equal or unequal and, at the same time, they are related by logico–semantic relations of expansion and projection.

The status between two clauses is considered to be equal when they are joined on an equal footing and they can both stand on their own. Their status is unequal when one cannot stand on its own and is dependent on the other. Expansion is further divided into three types: elaboration, extension and enhancement. A clause elaborates on the meaning of another by a more detailed description of it. One clause extends the meaning of another by adding further, related information. Finally, a clause enhances another by qualifying it in terms of time, place, cause, and other such circumstantial meanings. Projection is subclassified into two types: locution and idea, where locution is a projection of wording, usually by a verbal process, and idea a projection of meaning, most often by a mental process.

Halliday (1985: 306–7) has shown how logico–semantic relations recur throughout the lexicogrammar, and Martin (1992) extended them to model relations between discourse units. The relations appear abstract enough to be generalized to images and text as well, as has been demonstrated by Van Leeuwen (1991) and Martin (1994). They thus form one part of our system. The other part, status relations, have not been applied to relationships between images and text, but in our opinion they should be, most importantly because they appear to have realizations different from the logico–semantic ones. We draw on Halliday for the logico–semantic relations and in part also on image–text relative status. His clause complex system, however, provides only for independent and dependent clauses: when the status is equal, the related clauses are both independent; when it is unequal, one clause is dependent on another. In our image–text system, we incorporate Barthes' complementary status, since images and text can be

interdependent, or mutually dependent, as well. We thus divide equal status further into independent status and complementary status.

A SYSTEM FOR IMAGE–TEXT RELATIONS

Our system for the semantics of image–text relations thus has two subsystems that combine independently, status and logico–semantics. In this article we focus on explicating them both, and on exemplifying the permitted combinations of equal status. We believe that equal status combinations are especially useful to new media, although their potential has so far not been fully exploited. Much of the time, in the old media, unequal status relations have been copied rather mechanically instead. In a later paper, we will exemplify combinations of unequal status with logico–semantics.

Status

Just like the relationship between clauses in a clause complex, images and texts are considered to be unequal in status when one of them modifies the other. The modifying element is considered to be dependent on the modified one. Equal status between images and text is further divided into independent and complementary. An image and a text is considered independent and their status equal when they are joined on an equal footing and there are no signs of one modifying the other. When an image and a text are joined equally and modify one another, their status is considered complementary.

When the relative image–text status is equal, a whole image is related to a whole text. The exact nature of a whole text and a whole image is discussed in more detail in the next section and in a later section on logico–semantic relations. We only mention at this point that in the image, it tends to be a process or a combination of processes. Van Leeuwen (2005) conceptualizes such a combination of processes as a process complex and O'Toole (1994) as an episode. It can however be a smaller unit as well, such as O'Toole's (1994) 'member' or a still smaller unit. In the text, it can be a clause, group or phrase, or even a word, and their complexes. The largest textual unit that we deal with in this article is the paragraph, but it may be possible that an image can relate to a larger textual unit as well.

When an image and a text are independent, they do not combine to form part of a larger syntagm (cf. Barthes, 1997b[1964]), but rather the information they provide exists in parallel – they each form their own processes. An example of such an independent relationship is the following image–text combination from a screen of a CD-ROM encyclopaedia for children.[1]

The image in Figure 1 is a symbolic attributive process (Kress and Van Leeuwen, 1996: 108–9), with the map as the Carrier and the cross-hatched band (orange in the original) as the Attribute. The Attribute identifies the area of the map as the one where the moray eel lives. The text consists of two

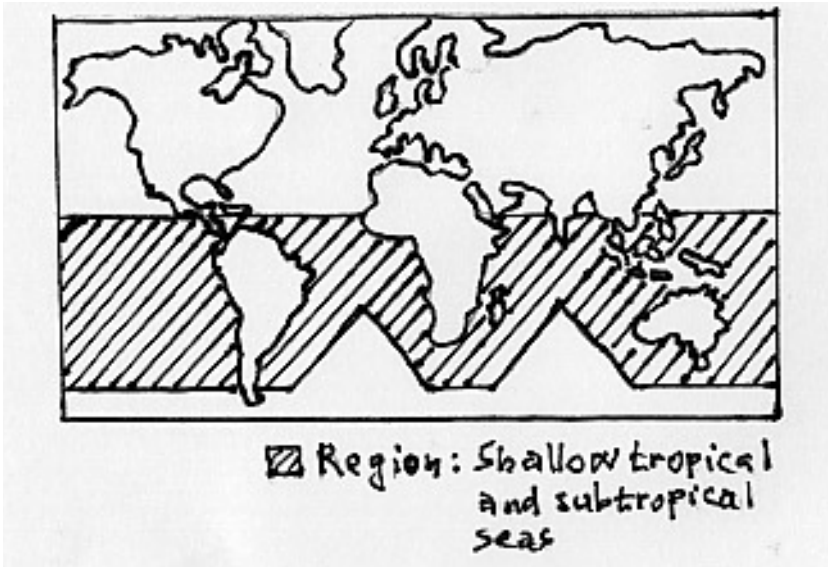


Figure 1 Example of independent image–text relationship. Drawing after a screen from Dangerous Creatures (Microsoft Corp., 1994).

relational, intensive identifying processes (Halliday, 1994: 122–8). In one of them, *region* (in which the moray eel lives) is the Identified and *shallow tropical and subtropical seas* the Identifier. The process itself is replaced by a colon. This whole process is embedded (Halliday, 1994: 188), by playing the role of the Identifier, in another intensive identifying process, in which the Identified is the little cross-hatched (orange) square. The little cross-hatched (orange) square is related by the componential relation of repetition (Martinec, 1998a, 1998b) to the cross-hatched (orange) band in the image.[2]

When the relative status of an image and a text is complementary, this is realized by them combining to form part of a larger syntagm. It seems that in most cases, this means that they play the role of participants in a type of process. The process itself, normally realized in language by a verbal group, is most often implicit. An example of such a syntagm is in Figure 2.

The two S&M (sado-masochistic) teddy bears in this ad play the role of a Carrier in a relational, intensive attributive process, in which the nominal group complex *Sweet. But not too sweet* functions as an Attribute (see Halliday, 1994: 120–2). There is another instance of the same, complementary, status relationship between the cereals package and the nominal group complex. The 'cute but naughty' meaning of the S&M teddies is transferred onto the cereals, which should appeal to the target audience of young, urban professionals that the product is aimed at.

In independent and complementary status, a whole image is thus related to a whole text. In contrast, when an image is subordinate to a text, the image is related to only a part of the text. An example of such a relationship is in Figure 3.

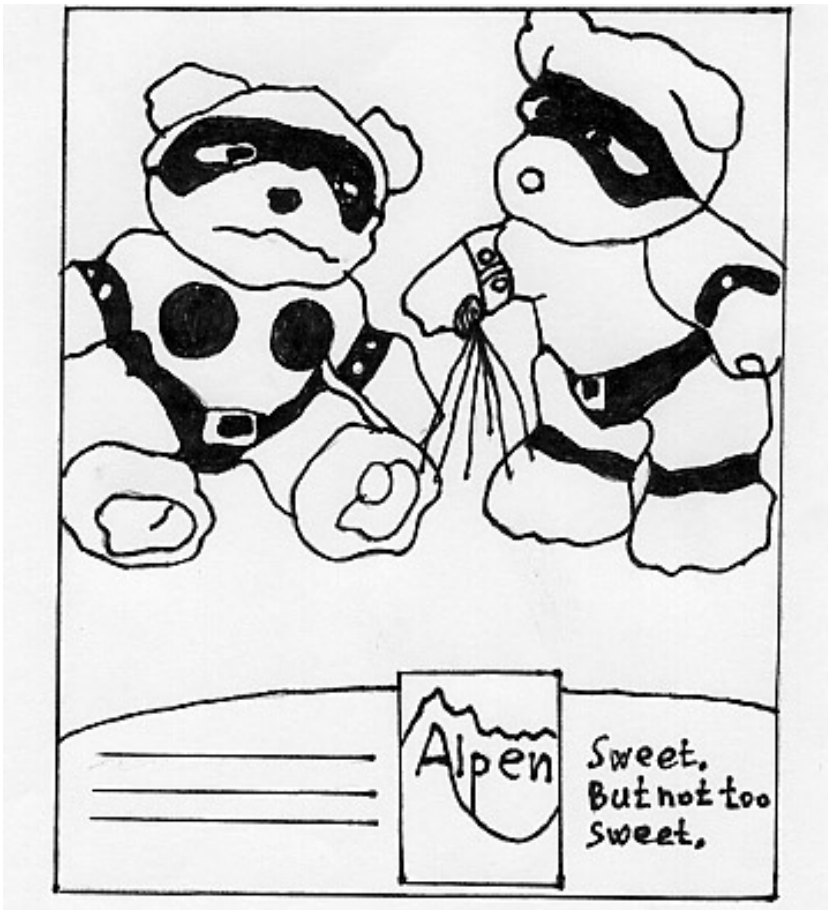


Figure 2 Example of complementary image–text relationship. Drawing after an advertisement in *The Guardian*, December 2004.

The image of starfish in this example relates to only some of the content of the text. In particular to *Starfish have five to forty arms arranged around a central area that contains the mouth. A starfish's arms are lined with hundreds of tiny, rubbery tube feet. Each one is tipped with a suction cup to help the starfish to hold onto slippery surfaces.*

There are four ranking processes with their participants and circumstances in this part of the text (as well as an embedded one *that contains the mouth)*. Apart from these, the text consists of three other ranking processes, and two embedded ones, with their participants and circumstances: *Some starfish eat the corals that form reefs* and *They kill their prey by pushing their stomachs out over their victims to let digestive juices flow over them*.

When a text is subordinate to an image, the text may well be related to only a part of it, but this is not the only possibility. Art criticism texts, for example, tend not only to describe what is in the whole image, but they often

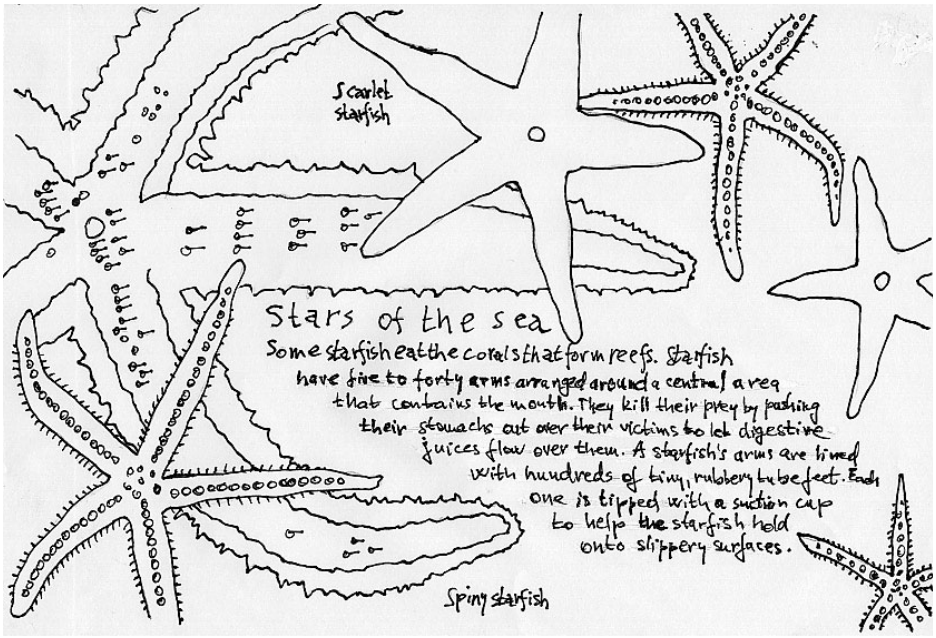


Figure 3 Example of image–subordinate-to-text relationship. Drawing after a screen from Dangerous Creatures (Microsoft Corp., 1994).

also include information about the historical background of the persons, objects and locations represented in the image, as well as information about the painter and the context of the particular work in his or her life and the overall oeuvre. A more reliable indication of text subordination is the presence of implicit devices that need to be decoded by reference to an image. An example of a text being subordinate to an image, with numerous instances of such textual reference (e.g. *this, the back, the inscriptions, the work, the sitter,* etc.), is in Figure 4.

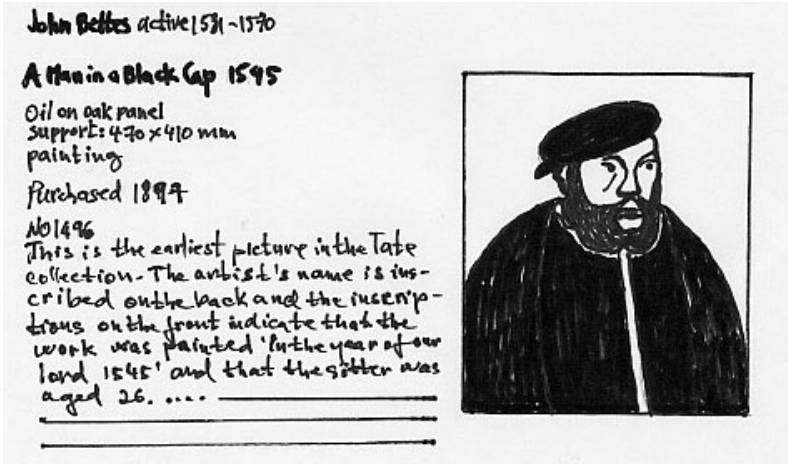


Figure 4 Example of text–subordinate-to-image relationship. Drawing after <http://www.tate.org.uk>

UNITS AND REALIZATIONS OF STATUS RELATIONS

It was remarked earlier that the relative status of an image and a text is considered independent when the whole image is related to the whole text. An image is considered subordinate to a text when it only relates to a part of it. But what do the 'whole image', and the 'whole text' and 'a part of it' mean? The question of the image and text units that are related by status is now addressed in more detail.

Despite the pioneering work of O'Toole (1994) and Kress and Van Leeuwen (1996), much less effort has been spent on identifying units of analysis in images than in text. The issue is dealt with in more detail in the section on logico–semantic relations, where we go at least some way towards specifying the units, or ranks, in images that appear relevant to logico–semantics. In the case of status relations, the situation is simpler, however, because the units that are related are either a whole image and a whole text, or a whole image and a part of a text.

As mentioned earlier, it depends of course on what the 'whole text' means. The largest unit of text that is related to images in our data is the paragraph; so we will consider the 'whole text' to mean a paragraph or smaller units, such as clause complexes, clauses, and even words, whenever these are the units that images relate to. This happens with news photo captions, image titles, etc. and in these cases, it is the whole caption, title and so on, that relate to the image by status. It is quite likely that images relate to larger textual units than the paragraph, such as sections. And the realizations will have to be extended to include those as well. In any case, the relatedness between images and texts is realized by componential cohesion (see Martinec, 1998a, 1998b), which relates participants, processes and circumstances, or 'components' in images and texts.

How can we tell whether a whole text or only a part of it is related to an image? It depends on its size. When the text is a paragraph, the related units are either independent clauses or hypotactic clause complexes, with their processes, participants and circumstances. In this case, the componential–cohesive relations must relate the processes in the images and those in the text, the latter realized in the unmarked case by verbal groups.[3] If all the processes in independent clauses in a paragraph are related to an image, the image may be said to relate to the whole paragraph. If some are not, the image is said to relate to a part of the paragraph. The clauses in the paragraph that only relate to images by cohesion between a participant or a circumstance are not considered to be related at this level. This is the case with the clauses and clause complexes in Figure 3 that were previously said to be unrelated.

If a text is the length of a clause or a clause complex, for these to relate to an image, it is enough that there be a componential–cohesive tie between the image and a participant or a circumstance. The related textual unit in such cases is the independent clause (or clauses) in which the participant or

Figure 5 Example of image–text status realization. Drawing after <http://news.bbc.co.uk>, 24 January 2004.

Figure 6 Material process combined with present tense – text subordinate to image. Drawing after <http://news.bbc.co.uk>, 23 January 2004.

circumstance occurs and any clauses dependent on it (or them). An example of a text and an image related in this way is in Figure 5.

As stated earlier, one realization of text being subordinate to image is the presence of implicit devices to be decoded by reference to the image. Although this may be the most frequent realization of text subordination, there appears to be an alternative which is common in, for example, news photograph captions. This is the combination of material or behavioural processes (Halliday, 1994) with simple present or present progressive tense. The processes in this case describe what is going on in the image (see Figure 6).

Material and behavioural processes in the past tense, on the other hand, do not have this effect – see Figure 7, where the image is subordinate to the text.

Figure 7 Material process combined with past tense – image subordinate to text. Drawing after <http://news.bbc.co.uk>, 24 January 2004.

This functioning of tense in the realization of status seems puzzling at a first glance, but can perhaps be explained by Halliday's (1994) conceptualization of tense as deixis. Seeing tense in this way places it in the same domain of meaning as the reference items mentioned earlier, which are also a form of deixis, or pointing, at the image and its parts. Present tense can thus perhaps be interpreted as pointing at the action taking place in

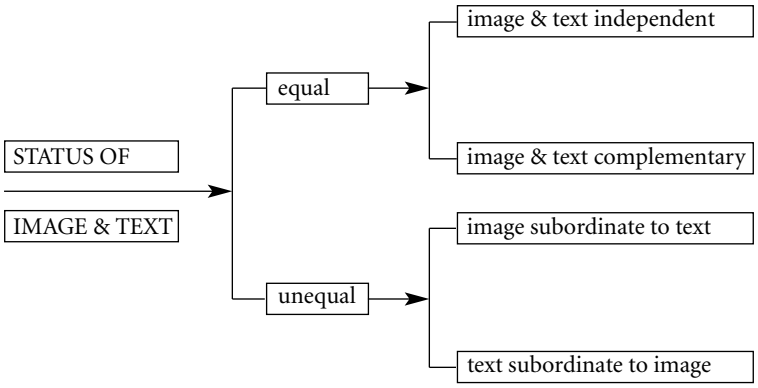


Figure 8 System of image–text status relations.

the image, and so subordinating the text to the image, whereas past tense pointing away from it, and thus not making it so.

To summarize: the units related by status relations are the whole image and the whole text, or the whole image and a part of the text. What exactly the 'whole text' and 'a part of the text' means depends on how large the text is – a paragraph, a clause or a clause complex, etc. The relatedness in question is relatedness by componential cohesion. Independent and complementary status are realized by the whole image being related to the whole text. Image subordination is realized by the image relating to a part of the text. Text subordination is realized by deixis from text to image, either by reference items or present tense combined with material or behavioural processes.

This section has dealt with the units and realizations specific to the relative status of images and text in image–text combinations. Since this issue has, to our knowledge, not been addressed since Barthes' (1977a[1961], 1977b[1964]) initial classification, we consider it one of the main contributions of our article. The different kinds of status we have argued for are grounded in semiotic theory and based on observable realizations. The diagram in Figure 8, a system network commonly used in systemic linguistics to model linguistic systems, summarizes the different kinds of status and their realizations.

LOGICO–SEMANTIC RELATIONS

We use both expansion and projection, the two main types of logico–semantic relations in Halliday's grammar, to model image–text relations. The main difference between the two is that, while expansion deals with relations between represented events in the non-linguistic experience, projection deals with events that have already been represented (Halliday, 1994: 252–3). In language, the already represented experience has either been

said or thought. If it was said, verbal processes are generally used to project it and the exact words are quoted, as in *Mary said 'John turned the tap off.'* If the experience was a thought, it tends to be projected by a mental process and the meanings rather than exact words are reported, e.g. *Mary thought that John turned the tap off.*

Projection is useful to account for cases when content that has been represented by text or images is re-represented in the other mode. The most obvious case are diagrams that summarize texts. They usually select the most important meanings of the texts and re-express them in a visual, diagrammatic form. Projections of meaning and wording also frequently occur in comic strips.

Expansion

As for expansion, all three of Halliday's main types – elaboration, extension and enhancement – relate images and texts. We have identified two kinds of elaboration between images and texts: exposition and exemplification (see Halliday, 1994: 226).[4] In exposition, the image and the text are of the same level of generality, whereas in exemplification the levels are different. An example of exposition is the earlier combination of the moray eel habitat image and text (Figure 1). When the level of generality of the image and the text is different, either the image or the text can be more general. An example of an image being more general is in Figure 9.

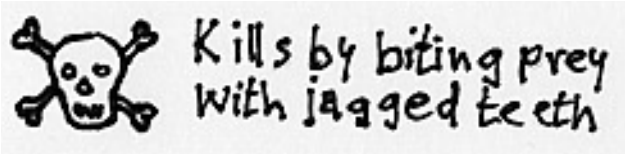


Figure 9 Image more general than text. Drawing after a screen from Dangerous Creatures (Microsoft Corp., 1994).

The skull and crossbones in this image–text combination is a generally recognized symbol representing death. The text makes its meaning more specific in this context, i.e. how the moray eel kills its prey.

An example of a text more general than an image is in Figure 3. *Starfish have five to forty arms arranged around a central area that contains the mouth* is information about the makeup of starfish in general. *Starfish* stands for the whole class of creatures called starfish, whereas the different kinds of starfish (*scarlet*, *spiny*) in the image are examples of that class. The rest of the sentence is exemplified by the images of each particular starfish, which enable the reader/viewer to actually see what the arms look like (thicker towards the middle and thinner towards the end, sometimes more and other times less so), how they are arranged around the central area, and what the mouth looks like, in the middle. The same is true of the other information in the clauses that relate to the image, e.g. *hundreds of tiny, rubbery feet*, etc.

Extension is a relationship between an image and a text in which either the one or the other add new, related information. An example of a text that adds information to an image is in Figure 4. *This is the earliest*

picture in the Tate Collection adds new information to the content of the image. We consider the information an addition because it goes beyond what is represented in the image, beyond its participants, processes and circumstances. The same is true of *The artist's name is inscribed on the back* because one cannot see the back of the picture. Yet other examples of extension in this passage are *Originally this portrait was larger, and would have had a blue background similar to the colour often used by Holbein* and *Due to long exposure to light, the pigment (smalt) has changed to brown.*

Finally, when an image and a text are related by enhancement, one qualifies the other circumstantially. We have so far identified circumstantial relations of time, place and reason/purpose. For a text to be considered enhancing an image or vice versa, it has to be related to its ideational content. An example of an image enhancing a text by place is the photograph and caption in Figure 10, where the image specifies the place where the woman arrived too late.

Figure 10 An example of enhancement by place. Drawing after <http://news.bbc.co.uk>, 24 January 2004.

In the following example (Figure 11) however, *during the war* in the first sentence does not enhance the image because it is not related to it. It rather presents information about the painter (Beckmann). *Following a nervous breakdown,* on the other hand, enhances the image because it gives information about Beckmann's pictures of which the one that is accompanied by this text is an instance; the non-finite clause situates the painting in time by reference to a period in the painter's life.

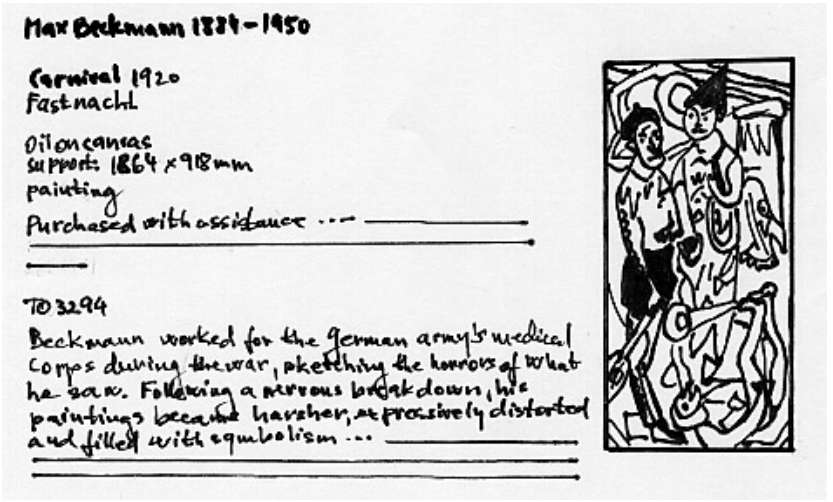


Figure 11 Example of enhancement by time and of image-unrelated text. Drawing after <http://www.tate.org.uk>

Figure 12 Example of enhancement by reason/purpose. Drawing after <http://news.bbc.co.uk>, 23 January 2004.

An example of enhancement of an image by reason/purpose in a text is *to help the starfish hold onto slippery surfaces* in Figure 3, which gives the reason for starfish's feet to be tipped with suction cups, as shown in the close-up of one of the starfish in the image.

A different example of enhancement by reason/purpose is in Figure 12. Here the image enhances the text. The dead bodies lying on the floor are the result of *a short circuit set fire to the hall's thatch roof*.

The different kinds of expansion between images and texts are summarized systemically in Figure 13.

Projection

Having outlined and exemplified expansion between images and texts, it is time to turn to projection. As remarked earlier, there are two main kinds of projection, depending on whether an exact wording is quoted or an approximate meaning is reported. Halliday (1994: 220) calls these 'locution' and 'idea'. Projection is a logico–semantic relation that mainly seems to appear in two image–text contexts: in comic strips and in combinations of text and diagrams, such as those found in textbooks, and scientific and

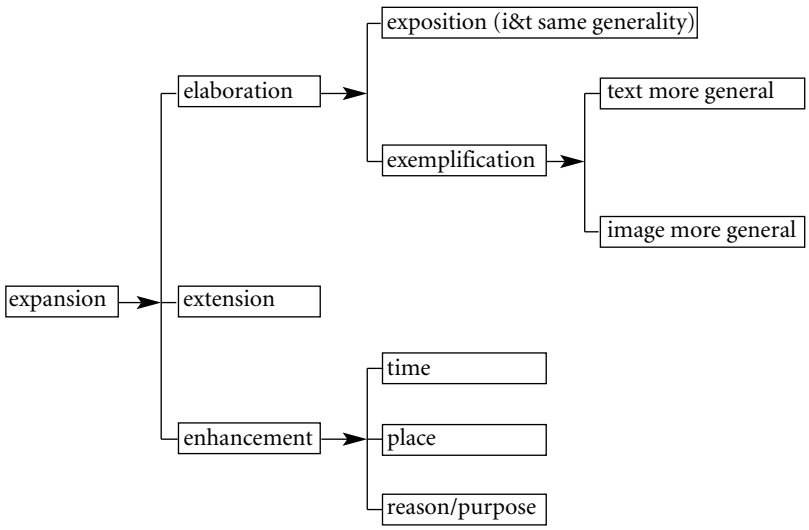


Figure 13 System of expansion for image–text relations.

Figure 14(a) and (b) Examples of projection of wording and meaning in comic strips. Drawings after Adams (1995).

similar publications. Distinguishing between locution and idea, or projection of wording and meaning in comic strips is straightforward because there are developed conventions for doing so – locutions are enclosed in speech bubbles and ideas in thought bubbles. Examples of such projection of wording and meaning are in Figure 14(a) and (b).

In order to deal with projection in combinations of texts and diagrams, a short digression into the semiotic of images is needed. What is at issue is whether the text that is often part of diagrams should be regarded as part of the image or as a text in its own right, that is in a relationship with the image. Kress and Van Leeuwen (1996) consider such text to be part of the image. They subsume such diagrams, and other images of the same kind that often contain labels for parts of the image, under the category of 'analytical' images (p. 89ff). We take the following approach. If the image or its parts provide the ideational content of the image and the text only provides labels for that content, we consider the labels a text in its own right

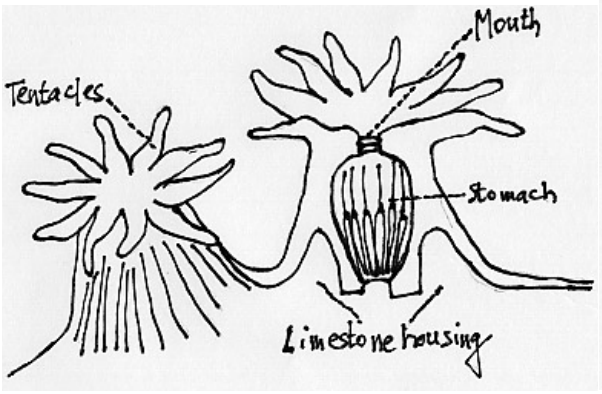


Figure 15(a) and (b) Examples of [exposition] and [text more general] image–text relations. Drawings after (a) a screen from Dangerous Creatures (Microsoft Corp., 1994) and (b) 'Battle Gear' (National Geographic, 2000).

and the relationship between the text and the image (or its part) is either [exposition] or [text more general]. We distinguish between them on the basis of the generality and abstraction of the labels and the image. If the labels are generic and the image of abstract (or technological) coding orientation (see Kress and Van Leeuwen, 1996: 170), the relationship is [exposition]; if the labels are generic and the image of naturalistic orientation, it is [text more general]. Examples of such relationships are in Figure 15 (a) and (b).

If, on the other hand, the ideational content of the image is provided by the text, and the graphics only consist of lines that enclose the spaces in which the labels are, for example, the image parts, we treat the whole as an image. An example of such an image is in Figure 16.

The diagram combines with the following text:

> In looking at the commonalities among the three disciplines, design and marketing tend to both focus on desirability of a product – the brand and lifestyle images, ease of use, and costs to take into account the aesthetics. Marketing and engineering both focus on usefulness of a product – the functional features, platform upon which the product is built, safety and reliability issues, and production costs. And design and engineering both focus on usability of a product – the ergonomics, interface with the product, the integration of the different features and associated costs, the selection of material, and manufacturing. Each overlap is secondarily also concerned with the other two value attributes, but the primary driver of interaction is as indicated. The point is that the usefulness, usability, and desirability of the product stem directly from the interaction between the disciplines. Thus, it is the overlaps between disciplines that define the value of the product to the consumer, the value that leads to success in the market and profit for the company (as shown in Figure 6.2). (Cagan and Vogel, 2002: 139–140)

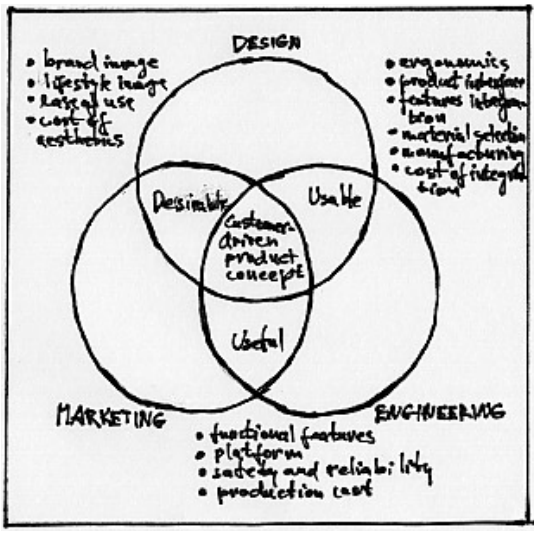


Figure 16 Example of text as ideational content of image. Drawing after Cagan and Vogel (2002).

This image–text combination is an example of a projection of meaning. Although many of the words in the image are the same as the words in the related paragraph, the form of the image as a whole is of course different

from the formal patterns in the text. There are several projections in this image–text combination.

> design and marketing tend to both focus on desirability of a product – the brand and lifestyle images, ease of use, and costs to take into account the aesthetics.

projects the two intersecting circles of *DESIGN* and *MARKETING*, as well as the epithet *desirable* at their intersection, and the four points in the top-left corner of the image.

> Marketing and engineering both focus on usefulness of a product – the functional features, platform upon which the product is built, safety and reliability issues, and production costs.

projects the two intersecting circles headed *MARKETING* and *ENGINEERING*, and the epithet *useful*, as well as the four points at the bottom of the image.

> And design and engineering both focus on usability of a product – the ergonomics, interface with the product, the integration of the different features and associated costs, the selection of material, and manufacturing.

projects the final two intersecting circles of *DESIGN* and *ENGINEERING*, the epithet *usable*, and the six points in the top-right corner of the image.

The three projections are all analytical processes. The last one consists of *DESIGN* and *ENGINEERING* as the Carrier, the two intersecting circles as processes and *usable, ergonomics, product interface, features integration, material selection, manufacturing* and *cost of integration* as the Attribute. The preceding two processes have a similar structure. All three are the constituents of an overarching, higher-rank analytical process with *ENGINEERING*, *MARKETING* and *DESIGN* as the Carrier, the three intersecting circles as the process, and the space with *customer-driven product concept* as the Attribute. The process is projected by *Thus, it is the overlaps between disciplines that define the value of the product to the consumer, the value that leads to success in the market and profit for the company.*

The relationships of projection of the different parts of the image by the different parts of the text are realized by lexical and componential cohesion. Taking just the first projection as an example, *design* and *marketing* in the text are related by repetition to *DESIGN* and *MARKETING* in the image, and so are *desirability* in the text and *desirable* in the image, *brand and lifestyle images* in the text and *brand image* and *lifestyle image* in the image, etc. At the same time, the two intersecting circles are synonyms of (*tend to) both focus on.*[5]

UNITS AND REALIZATIONS OF LOGICO–SEMANTIC RELATIONS

Apart from helping to decide what the units are between which status relations hold, componential cohesion is also crucial to determining the units between which logico–semantic relations obtain. The units in the text are of different sizes again. The largest is a hypotactic clause complex, but it can also be an independent clause, and a group or a phrase, and even a word.

When the related unit is an independent clause or a hypotactic clause complex, cohesive ties link the processes and their participants and circumstances in the clause (or the main clause in the case of the clause complex) and in the image. An example of this is in Figure 3. *A starfish's arms are lined with hundreds of tiny, rubbery tube feet* is an independent clause, which is related to the blown-up image of a starfish, where the tiny, rubbery feet can be seen. *Each one is tipped with a suction cup to help the starfish hold onto slippery surfaces* is a hypotactic clause complex, which is also related to the blown-up image of the starfish where not just the feet but also the suction cups can be seen.

This example brings up the question of the units in images that are linked by componential–cohesive and logico–semantic relations. Logico–semantic relations link whole processes, including their participants and circumstances (and any clauses in the text that hypotactically modify the linked main clause). Componential cohesion relations link processes, participants and circumstances themselves. One is then left with the question of how many processes there really are in the 'starfish' image. On superficial analysis, there would seem to be only one process, classificational, covert taxonomy, with the starfish as participants and their symmetrical arrangement as the process itself. It is, however, not this process that is linked logico–semantically to the processes in the paragraph. It is rather linked to *stars of the sea*, which is the title of this whole encyclopedia entry, or screen. At least for the purposes of image–text relations, *stars of the sea* is thus a text separate from the following paragraph. The logico–semantic relation between the image as a whole and this text is [exposition: text more general] – the starfish in the image function as instances of the general class of starfish, or 'stars of the sea'. This is in fact made explicit by labels for the two kinds of starfish in the image, scarlet starfish and spiny starfish.

The conclusion seems to be that there are other, embedded processes in the image (see Kress and Van Leeuwen, 1996: 112–14). And there seem to be three more levels, or ranks, at which they function. O'Toole's (1994) ranks of work, episode, figure and member are suited to analysing narrative representations, but do not seem to fit relational processes such as the starfish one very well. There seems to be a need for a more type-of-process neutral rank-scale, but this is not the time to go into a general discussion of ranks in images (see Martinec, 2005, for a brief, initial attempt). We therefore limit ourselves to a brief analysis of this particular image.

The logico–semantic relations between this image and text suggest four ranks, or units of different size related by constituency. There is, first of all, the overall classificational process, or covert taxonomy, which is related to *stars of the sea*. The participants in this process are the six Subordinates, i.e. the six images of starfish. Each one of the Subordinates can itself however be analysed as a process, this time analytical (Kress and Van Leeuwen, 1996: 89ff). And each one of the analytical processes functions as one of the Subordinates in the classificational process and is thus embedded in the process structure.

The Carrier in each of the analytical processes is the starfish as a whole, and the Attributes are its arms, the central area and the mouth. The part of the text that is related logico–semantically to this first-level embedded analytical process is *starfish have five to forty arms arranged around a central area that contains the mouth*. The relation is [exposition: text more general] because each image/process of a starfish is an instance of the way the text describes how starfish look in general (cf. the generic *starfish*). The componential–cohesive relations link the following participants: *starfish* and each image of a starfish, *five to forty arms* and the arms of the starfish in each image, *central area* and the central area of the starfish in each image, and *mouth* and the mouth in each starfish image. The relationship is hyponymy (Halliday and Hasan, 1976; Martin, 1992; Martinec, 1998a, 1998b) in each case, with the message parts (Martin, 1992) in the text being the Superordinates and the components in images the Subordinates. The componential–cohesive relations also link the process *have arranged around* and the part of the image that shows the arrangement, or position of the arms (around the central area). They finally link *contains* and the position of the mouth relative to the central area of the starfish.

A second-level embedded analytical process is in the blown-up image of the bottom side of one starfish, with its arms as the Carrier and the feet as the Attribute. The process functions as one of the Attributes in the next-higher level analytical process in the image, i.e. the starfish's arms. The process is related to *a starfish's arms are lined with hundreds of tiny, rubbery tube feet* in the text by [exposition: text more general] again, and the realizations of the logico–semantic relationship by cohesive relations are similar to those of the previously discussed embedded process.

Finally, there is also a third-level embedded analytical process, in which the Carrier is the feet of the starfish in the blown-up image and the Attribute is the suction cups by which they are tipped. The entire process functions as the Attribute of the next-higher level analytical process in the image, in particular the starfish's feet. It is related by [exposition: text more general] to *each one is tipped with a suction cup to help the starfish hold onto slippery surfaces* in the text, and the realizations of the relationship are similar to the previous ones.

There thus appear to be four levels, or ranks, in the whole 'starfish' image. The highest rank is the classificational process of the image as a

whole. The other three ranks are all realized by embedded analytical processes.

EQUAL-STATUS AND LOGICO–SEMANTICS COMBINED

Drawing together all the image–text status and logico–semantic relations that we have discussed results in the network in Figure 17 (the curly bracket indicates that the systems of status and logico–semantics are to be chosen from simultaneously).

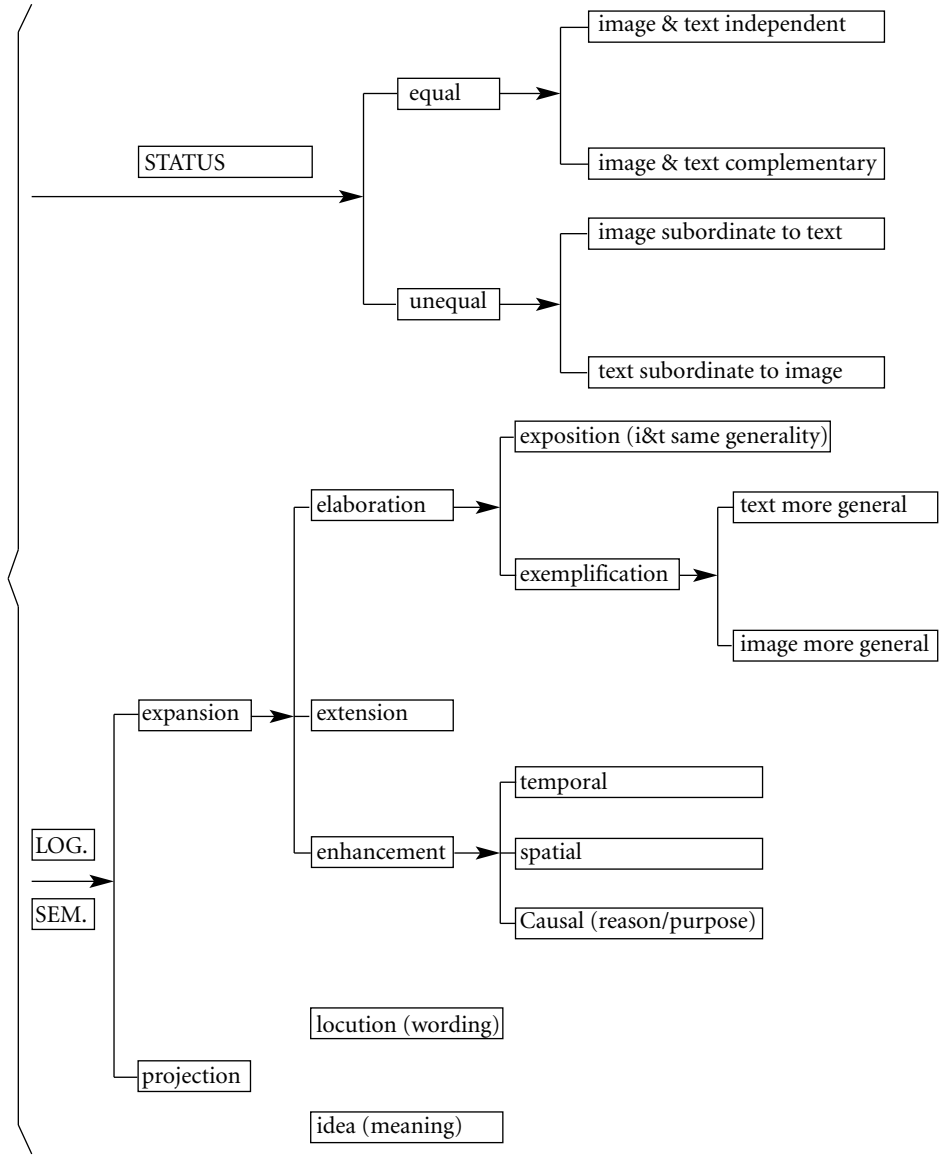


Figure 17 Network of combined status and logico–semantics.

In this section, we focus on exemplifying the combinations of equal-status and logico–semantic relations. Such combinations are found especially often in new-media products for children, in advertisements, comics and other products that often have an entertainment aspect to them. The entertainment aspect of some of the image–text combinations is simply due to what seems to be an enlivening part played by images in a syntagm that in other products would be fully realized by language. The rest of image–text combinations seem to put the onus on the reader/viewer to figure out the implicit connection between the image and the text. Equal status combinations lend themselves especially well to both.

Image and text independent, exposition

The image and the text are independent, which is realized by the whole image being related to the whole text. The logico–semantic relation is exposition, i.e. the level of generality of the components in the image and the text is the same, which is realized by them being related by synonymy. An example of this image–text combination is in Figure 1.

Image and text independent, text more general

In this image–text relationship, an image functions as an example of a text. This is realized by hyponymy between the two, with the superordinate element in the text and the subordinate in the image. The image and the text have an independent status, which is realized by the whole image being related to the whole text and by each forming separate processes. An example of this relationship is *Remember when total freedom came in a box?* combined with an image of three children jumping and a fourth rolling in a cardboard box in Figure 18.

In this image–text combination, *total freedom* in the text is cohesively related by hyponymy to the three children laughing and jumping, and to the fourth one rolling in a cardboard box (these appear to be two specific representations of total freedom in the US culture). *A box* relates by synonymy to the cardboard box that one of the children is rolling in.

The image–text combination is followed by 'Introducing the ORINOCO™ Wireless Networking Kits', which is related to the text above rather than the image. The relationship is mediated by an image of a box containing the Orinoco wireless networking toolkit, which is next to the main text in the bottom half of the advertisement, and by the main text itself. In particular *a box* is synonymous with the image of a box containing the Orinoco wireless networking toolkit, and *in a box* relates by repetition to *in a box* in the main text.

Image and text independent, image more general

Here a text further specifies an image, which is often, but not always, realized by hyponymy between a superodinate item (or items) in the image and a

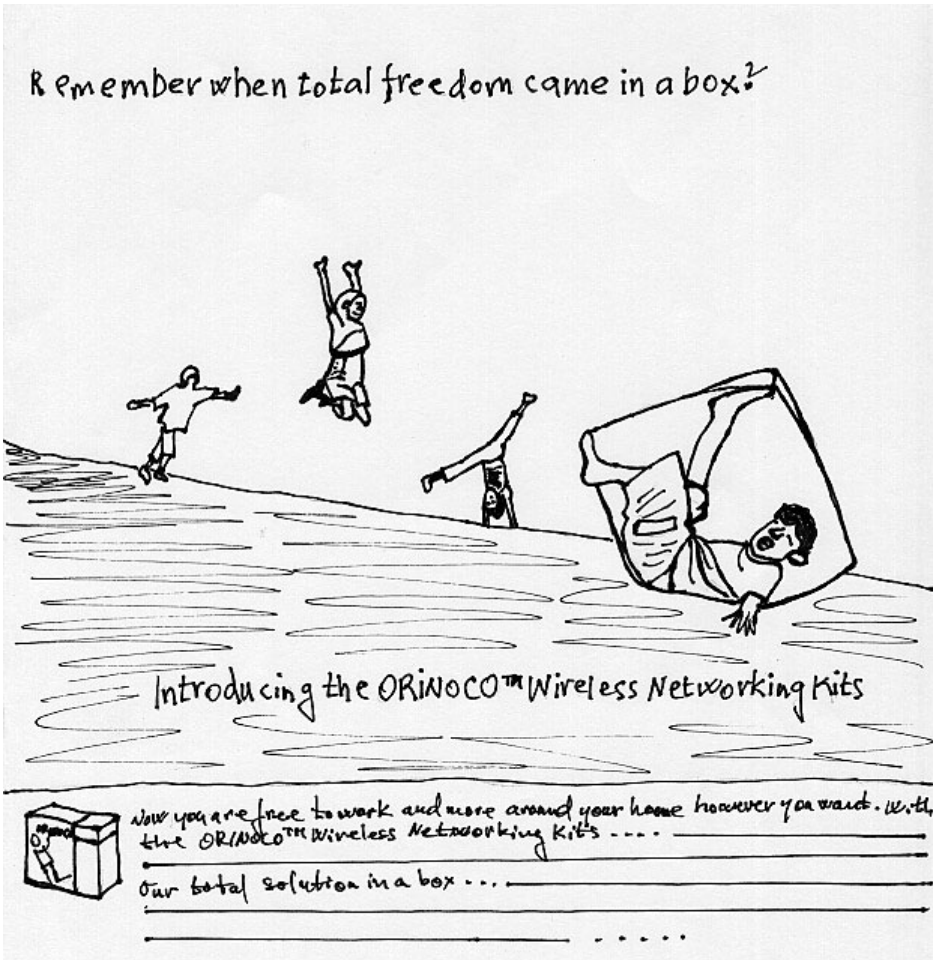


Figure 18 Example of image and text independent, text more general. Drawing after an advertisement from *Wired*, November 2001.

subordinate item (or items) in the text. The whole text and image are related again and form separate processes. An example of this image–text relationship is in Figure 9. The skull and crossbones is a fairly generally recognized symbol of death, or killing. Synonymy relates *kills* to the death symbol. In the case of this example, it is *by biting prey with jagged teeth,* which is a Circumstance of manner (Halliday, 1994), that further specifies the meaning of the image.

Image and text independent, extension

An example of this combination of status and logico–semantics is in Figure 19.

We consider the advertisement to consist of two images – the large image of the vat in which the potatoes, water and yeast mixture is fermenting, and a smaller one of the bottle of vodka with the ribbon

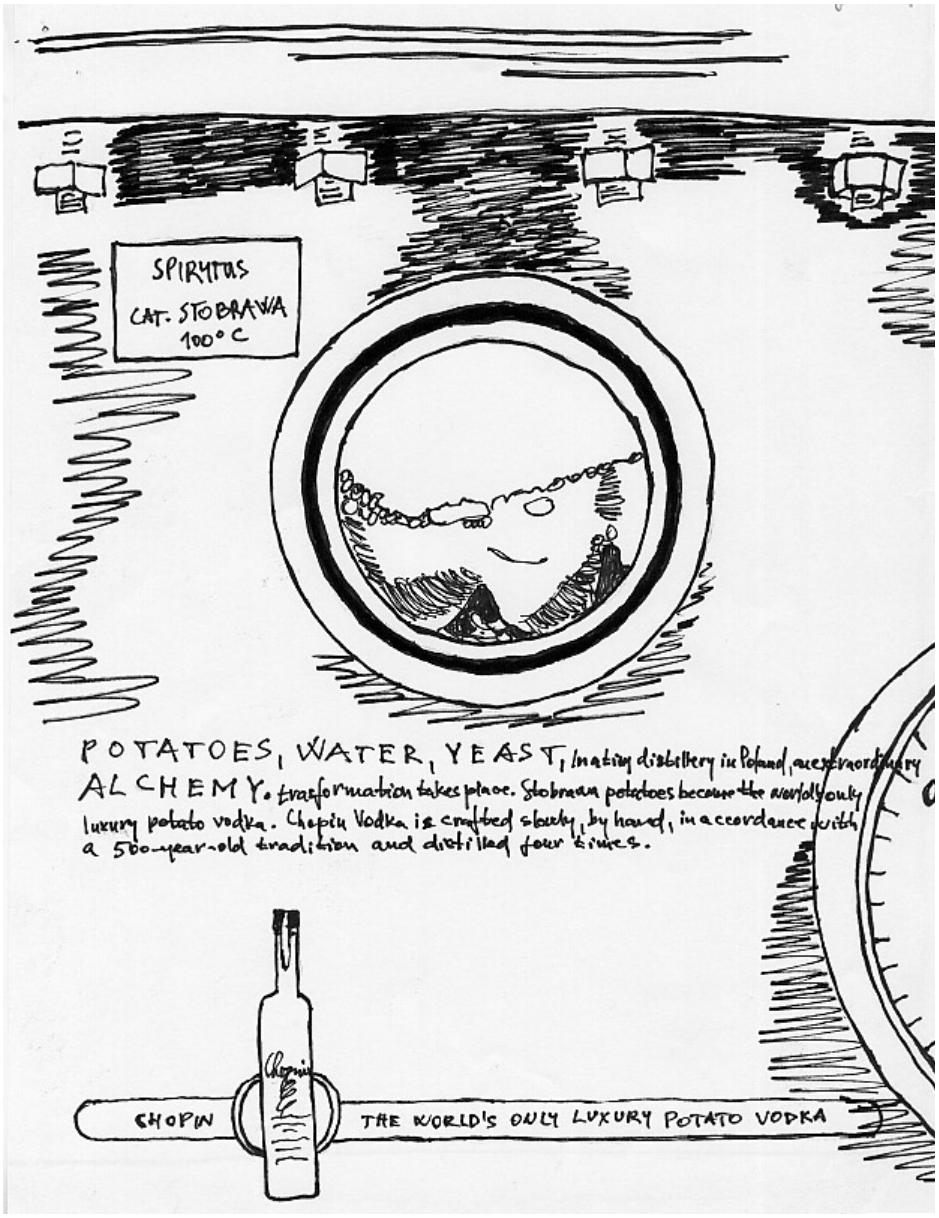


Figure 19 Example of image and text independent, extension. Drawing after an advertisement from *Wired*, November 2001.

containing the title *Chopin – the world's only luxury potato vodka*. The handwritten *Spirytus*, *Cat. Stobrawa* and *100°C* are part of the large image because the information they contain is not in the image and cannot be deduced from it. Only the large image–text relationship is relevant and is analysed here.

The whole text is related to the whole large image. *Potatoes*, *water*,

yeast are related by synonymy to the contents of the vat that can be seen through the porthole. *Alchemy* and *transformation* relates by synonymy again to the process of fermentation that is taking place in the vat. The fermentation can be observed since the potato mixture seen through the porthole has bubbles in it. Repetition links *Stobrawa* in the image and *Stobrawa* in the text. *Become* in the second sentence is a synonym of *transform*, and so related to fermetation, and *potato* is a synonym of the potatoes in the vat. The last sentence is related to the connotative rather than denotative meaning of the image. The old-looking vat (including the hand-written label) connotes old times, when things were *crafted slowly, by hand* rather than mass-produced. *500-year-old tradition* is also related to this connoted meaning and so is *distilled four times*.

At least some of the ways the text extends the image are as follows. *Alchemy* adds a touch of 'magical' meaning to the process of fermentation observed in the image. *Extraordinary transformation* adds a further, hyperbolic, component of meaning to the humble fermentation. *Luxury* also adds meaning to the image, since it is not clear from looking at it that Stobrawa potatoes are luxury potatoes.

Image and text independent, enhancement

There is an example of this combination in Figure 19 – *in a tiny distillery in Poland* enhances the image by providing information about where the vat is.

Image and text independent, locution

This combination is not permitted by our system because the text and the image would have to be formal 'word-by-word' re-representations of one another, yet they have different forms.

Image and text independent, idea

An example of this combination is in Figure 16, following which the realizations of both status and logico–semantics were discussed.

Image and text complementary, exposition

The image and the text are interdependent, they each play a role in a relational identifying process (Halliday, 1994). The image and the text are both at the same level of generality or abstraction. The text tends to be realized by a nominal group and the process itself tends to be implicit. An example of such an image–text combination is in Figure 20.

The caption in this image–text combination functions as the Identifier and the image as the Identified in a relational identifying process. The example comes from an online course in anatomy, the image has an abstract coding orientation and the text is generic.

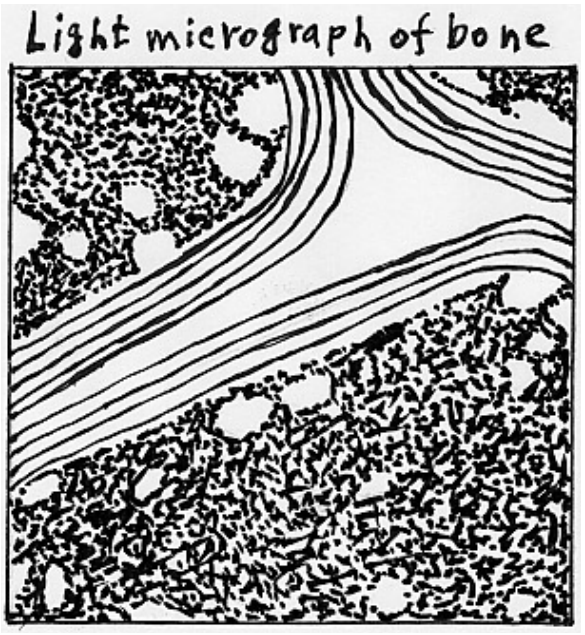


Figure 20 Example of image and text complementary, exposition. Drawing after 'The Biology Project', University of Arizona (1997), <http://www.biology.arizona.edu>

Image and text complementary, text more general

The image and the text are interdependent. The text represents a class to which the content of the image belongs. The image functions as the Carrier and the text as the Attribute in a relational attributive process. The process tends to be implicit. Two examples of this image–text combination are in Figure 2. Both the teddies and the advertised box of cereals are members of the 'sweet but not too sweet' class.

Image and text complementary, image more general

This combination of image and text seems to occur quite frequently in, for example, company logos. The image is usually of a fairly abstract modality and the text further specifies what it means. One such logo is in Figure 21.

The image represents, in an abstract form, juice swirling in a blender. The image–text combination seems best interpreted as assigning *jamba juice*, i.e. a particular brand of juice, to a class of fresh, fun juices, which are full of energy. The process is thus relational, intensive attributive, with *jamba juice* as Carrier and the image as Attribute. By implication, the company, too, is assigned to a fresh, fun and energetic class of companies.

Figure 21 Example of image and text complementary, image more general. Drawing after Miller and Brown (1998).

Image and text complementary, extension

The image and the text are interdependent and the text adds new information to the image or vice versa. The image and the text both play a role in the structure of a material or behavioural process. In the example in Figure 22, the process itself, which in language is realized by a verbal group ('eats'), is realized by a well-known symbol of a knife and a fork.

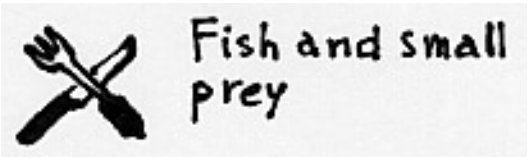


Figure 22 Example of image and text complementary, extension. Drawing after a screen from Dangerous Creatures (Microsoft Corp., 1994).

In Martin's (1992) re-interpretation of Halliday's (1985, 1994) expansion, the participants in material and behavioural processes are considered related by extension, which fits well with the above analysis.

Image and text complementary, enhancement

The image and the text are interdependent and the text qualifies the image by circumstantial information or vice versa. The image and the text play different roles in the structure of various types of process. An example of this kind of image–text relation is in Figure 10, where the image functions as a circumstance in a material process.

Image and text complementary, locution

The image and the text are interdependent, which is realized by each of them playing a role in a verbal projection. The image projects the text or vice versa. Image–text combinations of this kind are a staple of comic strips as has already been mentioned, see Figure 14(a). The person in the image has the function of the Sayer, the speech balloon realizes the verbal process, and the text plays the role of projected wording.

Image and text complementary, idea

The image and the text are interdependent again, which this time is realized by them both playing a role in a mental projection. Comic strips again provide copious examples of this kind of image–text combination, see Figure 14(b). The image of the person in this case plays the role of the Senser, the thought bubble realizes the mental process and the text functions as the projected thought.

TOWARDS MACHINE-RECOGNIZABLE REALIZATIONS

In this section, we make some suggestions about text and image features that could be used in automatic classification of the image–text relations. The observations are only preliminary and a more in-depth analysis is needed,

especially of realizations specific to different genres. Furthermore, some of the realizations involving layout and formatting may not be as reliable as those involving componential cohesion that have been discussed so far. However, since we eventually aim to make the system ready for various multimedia applications, we believe it worthwhile to make a start along this path. We are currently compiling a corpus of webpages to test some of these hypotheses.

Features of interest to automatic classification of image–text relations include:

- Page layout and formatting: the relative size and position of the image and the text, font type and size
- Lexico–grammatical references in text; for example, 'This image shows ...', 'See Figure 1', 'on the left', 'is shown by'
- Grammatical characteristics of the text: tense – past/present, process type, quantification – single/many, full sentences or short phrases/ groups
- Modality of images: a scale from naturalistic to abstract coding orientation – a function of depth, colour saturation, colour differentiation, colour modulation, contextualization, pictorial detail, illumination and degree of brightness – may correlate with choice of either GIF or JPEG image compression formats
- Framing of images: for example, one centred subject or no particular subject

Features to classify status relations

Two features that seem most relevant to machine-recognizable realization of status relations are page layout and formatting, and deixis (including lexico–grammatical reference and verb tense). As for layout, three systems seem to be most indicative of relative image–text status: Given and New, Ideal and Real, and Salience. Given and New is a layout structure derived from the left-to-right writing of European languages (see Kress and Van Leeuwen, 1996: 186–92). The left-most position in the clause is occupied by information that tends to be already known whereas the information on the right tends to be yet unknown. The New thus forms a certain kind of prominence. The known information however also tends to be thematic, in the sense of realizing what the clause is about (see Halliday, 1994: 37–67), so it forms another kind of prominence. At the beginning of a text, the two prominences coincide, which signals an especially prominent, or important item.

Texts on our pages tend to start at the top and continue downwards, and another layout structure, Ideal and Real, seems to have been influenced by this aspect of writing. The Ideal is the more prominent half; it tends to

contain the big ideas, the promise of what is to follow. The Real is less prominent, featuring the more detailed, down-to-earth information (see Kress and Van Leeuwen, 1996: 193–202). The top-left corner thus seems to be where three different kinds of prominence coincide: the Theme, the New and the Ideal. It is thus a very important place in the layout structure indeed. On websites, for example, that is the place where the company logos are. On the two kinds of websites that we have looked at in some detail – news sites and online galleries – it is also where the title of the article, the name of the painter and the title of the image are.

A fourth kind of prominence, salience, is in principle independent of the other three, and is realized by the visual 'weight' of an item, the most important component of which seems to be its size (see Kress and Van Leeuwen, 1996: 212–14). A more important item tends to be of a greater size than a less important one. Salience combined with the other layout dimensions plays a role in signalling the relative status of image and text.

In the two website genres that we have looked at in some detail – news sites and online galleries – we found different patterns in the positioning of images and the bulk of the text, indicating their different relative status. On news websites, the first paragraph of the superordinate main text is in the top-left corner, just under the headline, and it is in thicker, bold-print font. The rest of the text takes up most of what remains of the page. The subordinate news photographs that accompany the text are small in size, positioned on the right and lower or at the same level as the first paragraph they are related to.

In online galleries, on the other hand, where text is subordinate to image, the pattern consists of images that are most often of greater size than the text that accompanies them and, although positioned on the right, are considerably higher up on the page than the main text.

Finally, the centre and margin pattern (Kress and Van Leeuwen, 1996: 203–7) also has much to do with status realization. Positioning items centrally on a page indicates their greater importance relative to items that are placed around it, on the page margins. A good example is the 'starfish' image–text combination (Figure 3), whose layout supports the interpretation of status we have given it on the basis of componential cohesion realizations.

As for lexicogrammatical reference, reference items such as 'this painting' or 'Figure X shows' realize text subordination. However, lexical references like '(see Figure X)', especially when they occur towards the end of the text suggest equal status.

Present progressive tense in combination with a material or behavioural process signals that a text is subordinate to an image. This certainly seems to be the case in image descriptions, such as those found in news photograph captions and in gallery sites.

Features to classify logico–semantic relations

Determining logico–semantic relations involves a comparison between what is depicted in the image and what is referred to by the text. If the same participants, processes and circumstances are depicted and referred to, then there is elaboration. If new but related things are referred to or depicted, then there is extension. If related temporal, spatial or causal information is provided, then there is enhancement. The question is how such comparison can be computed. For text, information extraction techniques can recognize proper and common nouns, work out who is the subject of a story, and determine what kind of event or state is being referred to in a text. For images, image processing techniques can detect faces, indoor vs outdoor scenes, whether the represented subject is centred and in focus, and framing. Working out, for example, whether a text refers to the same number of people as depicted in an image (one or many) would involve analysing quantifiers in the text and detecting numbers of faces in the image. This would obviously have implications for determining the type of image–text relation.

In the case of a single face, it would be important to analyse whether it was a portrait of a specific person, realized by the subject's being centred, with head and face framed to fill the photograph (a close shot), naturalistic coding orientation. A more anonymous character, on the other hand, is more likely to be represented as non-centred, and not filling in the frame, perhaps decontextualized, slightly out of focus, with some degree of abstraction. In news websites, an image that elaborates a text is often a portrait, and the text tends to repeat the name of the depicted person, most often in the thematic position. The enhancement relation of cause and effect is realized by the image depicting a process, while the text refers to a state or vice versa. The image tends to normally be a general scene, rather than a closely framed photograph with one main subject.

One or more nominal groups on their own rather than full clauses may signal a text that elaborates an image, such as image captions in scientific publications.

CONCLUSION

The system for analysing image–text relations that we have presented combines Halliday's (1985, 1994) logico–semantic relations with Barthes' (1977a[1961], 1977b[1964]) foundational classification of image–text relations. The system is generalized, applicable to many different image–text genres and its categories are based on perceivable realizations. It is based on our analysis of contemporary image–text combinations, and in our account we have focused on image–text relations of equal status, which we think are especially suited to new media. It may well be that the system will need to be modified as new image–text genres evolve or that at least the realizations of existing categories will change.

We made some observations about machine recognizable realizations and it may be worthwhile to briefly speculate about possible application scenarios for automatic classification of the relations. When using correlated text as a source of terms for indexing images (see Salway and Ahmad, 1998; Salway et al., 2003), for example, we predict that priority should be given to a text that is subordinate to an image. In this way, more of the text is likely to be related to the image content, so keywords extracted from the text should facilitate more precise retrieval. When an image is subordinate to a text, then the image is only about part of the text, so we would hypothesize that there would be a greater chance of extracting erroneous words from the text.

Our system could also be used for making explicit the types of links that connect text and images in hypermedia systems. Users browsing through such systems may lack an intuitive feel for what to expect at the end of a link. The idea of a typed hypertext link was proposed for a digital library of scientific papers (Trigg, 1983) and this could perhaps be extended to typed hypermedia links between texts and images. A user viewing an image may thus be offered links to: (1) information about what can be seen in this image; (2) information about what this image could mean; and (3) information about the history of this image. The options relate to our elaboration, extension and enhancement, respectively. Making the nature of the hypermedia links explicit in this way both to the user and to the machine would be useful because it can help the user navigate the information space more efficiently and because the machine can mine information from the ways in which the nodes are connected. Other application scenarios are possible as well, including multimedia generation.

All the examples of image–text combinations we have presented related ideational meanings in images and texts. But it may be that the same kinds of relationships apply to interpersonal meanings as well. One such image–text combination we have encountered is on the same children's encyclopaedia screen as the examples in Figures 1 and 9. It consists of an image of a large, red exclamation mark next to the following text: *The ancient Romans so admired moray eels that they kept them in special pools, dressed them in jewels, and even fed slaves to them!* The image obviously realizes a kind of speech function (an exclamation), whose meaning relates to the whole text. We thus consider the status of this text and image combination to be independent, and the logico–semantic relation as elaboration: exposition. One might say, in Royce's (1998) terms, that the speech function of the image reinforces the speech function of the text.

Advertisements and similar multimodal texts that address the viewer directly by both (or either) an image or a text can be analysed similarly. In Royce's (1998) advertisement from *The Economist*, the speech function of demand, realized by the image of a young woman looking directly at the viewer, elaborates the speech function of the headline and of the first sentence of the main text: *Does your environmental policy meet your granddaughter's expectations? Is your business community among the millions*

of customers across the world already using our environmental services? The meanings of equality and involvement realized by the girl's being represented at an eye-level angle and in a close social to personal distance can also be said to be in an elaboration relationship with similar meanings realized by the direct address in the text, realized by the second person *you* and *your,* which personally involve the reader. Royce's attitudinal congruence and dissonance can be articulated in our system respectively as elaboration and extension.

Although our classification system was developed for text and images, we surmise that similar relations that we identified for image–text combinations will be useful for mapping out relations between other semiotic modes as well. Gestures and speech in everyday interaction are a case in point, and so are action and dialogue in film. The relations may also be identifiable at more abstract levels of semiosis, such as between different story lines in narratives.

NOTES

1. Examples of image–text combinations are presented in the form of line drawings. These are faithful reproductions of the originals, to each of which we have provided a detailed reference. The line drawings do not in any way alter the image–text relations that were identified in the originals.
2. The text is considered to include the cross-hatched (orange) square because it is part of the overall image caption. This solution seems preferable to considering the cross-hatched (orange) square an image, because its function is closer to the word 'orange' than to any kind of image. Even if the cross-hatched (orange) square were considered an image, this would still not interfere with the proposed realization of independent status. The whole caption would still be a process, in this case realizing a combination of complementary image–text status with the logico–semantic relationship of exposition (see further below). This process would run in parallel with the process in the main image and would not combine with it to create another process. The cross-hatched (orange) square is considered to be in a relationship of repetition with the cross-hatched (orange) band in the image because it is purely the colour itself that forms the cohesive link.
3. When the processes are realized metaphorically (see Halliday, 1994: 342–67) for grammatical metaphor), a congruent rewording that involves a verbal group is used.
4. Halliday's (1994: 226) exposition and exemplification are categories arising out of combining elaboration with equal status (parataxis), whereas for us they are subcategories of elaboration only.
5. The question is how to analyse the relationship to the diagram of the remaining two sentences in the paragraph:

> Each overlap is secondarily also concerned with the other two value attributes, but the primary driver of interaction is as indicated. The point is that the usefulness, usability, and desirability of the product stem directly from the interaction between the disciplines.

The second sentence is a generalization based on the diagram as a whole and on the parts of text that project it. The whole of it thus relates to the whole diagram. The first sentence is also a generalization, and its second clause again relates to the whole diagram, i.e. to the value attributes of desirability, usefulness and usability and how they arise out of the interaction of marketing, design and engineering. The first clause is a little more puzzling, but we surmise it to mean that, while the overlap between each two disciplines results in one of the primary value attributes, it secondarily concerns the other two value attributes through the central intersection of all three disciplines.

REFERENCES

Adams, S. (1995) *Always Postpone Meetings with Time-Wasting Morons*. London: Boxtree.

Arnheim, R. (1997) 'The Two Sources of Cognition', in T. Sebeok and J. Umiker-Sebeok (eds) *The Semiotic Web*, pp. 253–9. Berlin: de Gruyter.

Barthes, R. (1997a[1961]) 'The Photographic Message', in R. Barthes (ed.) Image–Music–Text, pp. 15–31. London: Fontana.

Barthes, R. (1997b[1964]) 'Rhetoric of the Image', in R. Barthes (ed.) Image–Music–Text, pp. 32–51. London: Fontana.

Bolter, J.D. and Grusin, R. (2000) *Remediation: Understanding New Media*. Cambridge, MA: MIT Press.

Cagan, J. and Vogel, C.M. (2002) *Creating Breakthrough Products*. Upper Saddle Rover, NJ: Prentice Hall.

Drucker, J. (1994) *The Visible Word: Experimental Typography and Modern Art, 1909–1923*. Chicago: University of Chicago Press.

Halliday, M.A.K. (1985) *An Introduction to Functional Grammar*. London: Arnold.

Halliday, M.A.K. (1994) *An Introduction to Functional Grammar*, 2nd edn. London: Arnold.

Halliday, M.A.K. and Hasan, R. (1976) *Cohesion in English*. London: Longman.

Hillenbrand, R. (1999) *Islamic Art and Architecture*. London: Thames & Hudson.

Horn, R.E. (1998) *Visual Language*. Bainbridge Island, WA: MacroVU.

Kress, G. and Van Leeuwen, T. (1996) *Reading Images*. London: Routledge.

Kress, G. and Van Leeuwen, T. (2001) *Multimodal Discourse*. London: Arnold.

Lemke, J. (1990) *Talking Science*. Norwood, NJ: Ablex.

Lemke, J. (1998) 'Multiplying Meaning: Visual and Verbal Semiotics in Scientific Text', in J.R. Martin and R. Veel (eds) *Reading Science*, pp. 87–113. London: Routledge.

Lyons, J. (1977) *Semantics*, Vol. 1. Cambridge: Cambridge University Press.
Marsh, E.E. and White, M.D. (2003) 'A Taxonomy of Relationships between Images and Texts', *Journal of Documentation* 59(6): 647–72.
Martin, J.R. (1983) 'Conjunction: The Logic of English Text', in J.S. Petofi and E. Sozer (eds) *Micro and Macro Connexity of Texts*, pp. 1–72. Hamburg: Helmut Buske Verlag.
Martin, J.R. (1992) *English Text*. Amsterdam: Benjamins.
Martin, J.R. (1994) 'Macro-genres: The Ecology of the Page', *Network 21*, pp. 29–52, available online at <http://minerva.ling.mq.edu.au/Resources/Network/Network.html>
Martinec, R. (1998a) 'Cohesion in Action', *Semiotica* 120(1/2): 161–80.
Martinec, R. (1998b) 'Cohesion in Language and Action', paper presented at the Sociolinguistics Symposium 12, Institute of Education, University of London, UK.
Martinec, R. (2003) 'The Social Semiotics of Text and Image in Japanese and English Software Manuals and Other Procedures', in T. van Leeuwen and C. Caldas-Coulthard (eds) *Critical Social Semiotics* (Special Issue) 13(1): 43–69.
Martinec, R. (ed.) (2005) 'Topics in Multimodality', in R. Hasan, J. Webster and C. Matthiessen (eds) *Continuing Discourse on Language*, Vol. 1. London: Equinox.
McCloud, S. (1993) *Understanding Comics*. New York: HarperCollins.
Miller, A.R. and Brown, J.M. (1998) What Logos Do and How They Do It. Gloucester, MA: Rockport Publishers.
Mitchell, W.J.T. (1994) *Picture Theory*. Chicago: Chicago University Press.
O'Donohoe, E. (2003) 'Between Image and Text', MPhil thesis. School of Art and Design, Swansea Institute of Higher Education, UK.
O'Halloran, K.L. (1999) 'Towards a Systemic Functional Analysis of Multisemiotic Mathematics Texts', *Semiotica* 124(1/2): 1–29.
O'Toole, M. (1994) *The Language of Displayed Art*. Leicester: Leicester University Press.
Royce, T. (1998) 'Synergy on the Page: Exploring Intersemiotic Complementarity in Page-Based Multimodal Text', JASFL Occasional Papers 1(1): 25–50.
Rush, M. (1999) *New Media in Late 20th-Century Art*. London: Thames & Hudson.
Salway, A. and Ahmad, K. (1998) 'Talking Pictures: Indexing and Representing Video with Collateral Texts', in D. Hiemstra, F. de Jong and K. Netter (eds) Proceedings of the 14th Twente Workshop on Language Technology – Language Technology for Multimedia Information Retrieval, pp. 85–94. ISSN: 0929–0672.
Salway, A., Graham, M., Tomadaki, E. and Xu, Y. (2003) 'Linking Video and Text via Representation of Narrative', AAAI Spring Symposium on Intelligent Multimedia Knowledge Management, Palo Alto, 24–6 March, pp. 104–12. ISBN 1–57735–190–8.
Trigg, R. (1983) 'A Network-Based Approach to Text Handling for the Online

Scientific Community', PhD dissertation, University of Maryland, USA. <http://www.Workpractice.com/trigg>
Van Leeuwen, T. (1991) 'Conjunctive Structure in Documentary Film and Television', *Continuum* 5(1): 76–114.
Van Leeuwen, T. (2005) 'Rank in the Analysis of Images', in R. Martinec (ed.) 'Topics in Multimodality', in R. Hasan, J. Webster and C. Matthiessen (eds) *Continuing Discourse on Language*, Vol. 1. London: Equinox.

BIOGRAPHICAL NOTES

RADAN MARTINEC is Senior Lecturer at the London College of Communication, University of the Arts, London. His current research interests include intersemiotic relations, multimodal systems and texts, non-linearity and interactivity in new media, epistemology and the semiotics of authenticity. He has published in *Semiotica*, *Leonardo*, *Functions of Language*, *Social Semiotics*, and the *Journal of Consumer Research*.

Address: London College of Communication, Elephant and Castle, London SE1 6SB, UK. [email: radan@martinec.demon.co.uk]

ANDREW SALWAY is Lecturer in the Department of Computing, University of Surrey. His research interests include intelligent multimedia information systems and computational theories of narrative, and of image–text combinations. He has published at AAAI Symposia, ACM Multimedia and IEEE ICME.

Address: Department of Computing, University of Surrey, Guildford, GU2 7XH, UK. [email: a.salway@surrey.ac.uk]