Multimodal literacy

THEO VAN LEEUWEN, PROFESSOR, INSTITUT FOR SPROG OG KOMMUNIKATION, SYDDANSK UNIVERSITET

After defining the concept of multimodality and contextualizing multimodality studies as a form of inquiry with roots in linguistics, the article reviews applications of multimodality to education and exemplifies how multimodality studies approach the analysis of the meaning-making potential of multimodal resources and the way they are integrated in multimodal texts.

Throughout, implications for multimodal literacy are discussed, emphasizing that multimodal literacy requires textual as well as contextual knowledge, that is, an understanding of, for instance, visual grammar as well as an understanding of the rules or conventions that govern its use in specific contexts. Multimodal literacy also includes an aesthetic dimension, an ability to produce and appreciate the aesthetic uses of layout, colour and typography that are, today, not only found in art, but also in many forms of everyday written communication. And finally it includes a critical dimension, particularly with regard to the way contemporary digital technology favours, or even imposes, modes of communication that suit the purposes of corporate culture but may have drawbacks in other contexts.

What is multimodality?

For several centuries there was, in Western culture, a tendency towards 'monomodality'. The culturally most highly valued genres of writing (literary novels, academic treatises, official documents, etc.) were characterized by graphically uniform, dense pages of print, and carried no illustrations. Paintings nearly all used the same support (canvas) and the same medium (oils), whatever their style or subject. In concert performances, all musicians dressed more or less identically and only the conductor and soloist were allowed some bodily expression. The academic study of such different modes of expression was equally monomodal: one discipline for language (linguistics), another for visual arts (art history), yet another for music (musicology) and so on, each with its own methods, its own assumptions, its own technical vocabulary, its own strengths and its own blind spots.

More recently, the dominance of monomodality has weakened, although it still persists in some practices, for instance in the writing and publishing of academic papers. The trend towards multimodality began with the avant-garde experiments of the early twentieth century, when 'concrete poetry' began to express itself not only through words, but also through typography, and when visual artists used new kinds of materials and sought to produce Gesamtkunstwerke which would combine as many forms of expression as possible. From the 1920s onward, the mass media, too, became increasingly multimodal. Magazines acquired colour illustrations and sophisticated layout and typography, and film, of course, became the multimodal art form par excellence, especially when the 'talkies' added speech, music and other sounds to its palette. More recently, formerly austere genres such as textbooks and documents produced by corporations, universities and government departments have also become multimodal, and the multimodal affordances of ubiquitous digital technologies such as Word and PowerPoint have made multimodal text design accessible to all.

Linguists have only gradually realized this. During the 1930s and 1940s the Prague School began to extend linguistics into the visual arts and the non-verbal aspects of theatre (cf. Garvin, 1964; Matêjka and Titunik, 1976). Veltruský (1964 [1940]), for instance, wrote about the theatre as a "multiple sign system", discussing sets, costumes and props as signs that provide setting and characterization and also take part in the action. And Bogatyrev (1971[1934]) studied dress as a language conveying what we would now call demographic information such as age group, place of residence, marital status, religion and occupation.

Multimodality is therefore the study of how meanings can be made, and actually are made in specific contexts, with different means of expression or 'semiotic modes'.

The second wave was the 1960s Paris School of structuralist semiotics, which used linguistic concepts and methods to understand communicative modes other than language, for the most part in analyses of popular culture and the mass media, rather than of the arts. Roland Barthes was its most crucial and influential proponent, writing, among other things, about the use of image and text in advertising and the press, and about the language of fashion (e.g., 1967, 1977, 1983). During roughly the same period, American linguists developed an interest in non-verbal communication, for instance in therapeutic interviews (Pittenger et al., 1960).

Most linguists, however, continued to occupy themselves with the grammar of sentences. Only when, in the early 1970s, the emphasis changed from analysing sentences to analysing texts did they begin to realize, first, that language is used differently in different contexts, and that literacy should therefore be understood as the ability to use language in ways that are appropriate to context, and, second, that all texts are multimodal and cannot be adequately produced or analysed, unless this is taken into account. Face-to-face speech, for instance, is also multimodal, involving not only language but also voice quality, intonation, facial expression, gestures etc. Even the densely printed pages I referred to earlier as 'monomodal' are in fact multimodal, as they use spacing, typography, punctuation, etc. (cf. Norgaard). Multimodality is therefore the study of how meanings can be made, and actually are made in specific contexts, with different means of expression or 'semiotic modes' – whether these are articulated with the body (speech, facial expressions, gestures and so on) or with the help of tools and materials (writing, drawing, making music and so on). And it is also the study of the ways in which multiple semiotic modes can be integrated into coherent multimodal texts. It is therefore not restricted to 'multimedia' in the sense of contemporary digital media but applies equally to face-to-face communication and other non-digital types of text. Multimodal literacy is therefore the ability to use and combine different semiotic modes in ways that are appropriate to the given context, both in the sense of the context-bound rules and conventions that may apply, and in the sense of the unique demands made by each specific situation. Such a form of literacy must be based on a knowledge of what can be done with different semiotic modes and how and of the ways in which they can be integrated into multimodal texts; however, it also, and equally importantly, requires an understanding of communicative contexts and an ability to respond creatively to the unique demands of specific situations.

Multimodality and education

The New London Group (Gunther Kress, James Gee, Allan Luke, Mary Kalantzis and others) stimulated an interest in applying multimodal text analysis to education (New London Group 1996). This led to four kinds of studies: studies of the development of multimodal literacy in very young children, often leading to a call for integrating multimodal literacy into the curriculum; studies of the affordances and learning potentials of specific semiotic modes; studies of multimodal classroom interaction; and studies of multimodal learning resources, including textbooks, toys, and the Internet.

Gunther Kress' *Before Writing* (1997) initiated the study of the development of multimodal literacy, investigating how very young children use the

affordances of whatever materials they have at hand, or whatever techniques they have mastered, on the basis of 'interest', that is, what is of crucial importance to them at the given moment. In one of his key examples, a three-year-old child draws a car as a series of circles ('wheels'). Having mastered the drawing of circles, the child now uses circles as a means of expressing what, to him, is a crucial characteristic of cars. As a semiotic resource the circle has many possible meanings, but the one the child selects is motivated by his interest at the moment, his interest in thinking about cars. Thus, learning to draw and learning to understand the world around him go hand in hand. But, as Kress said, "As children are drawn into culture, 'what is to hand' becomes more and more that which the culture values and therefore makes readily available" (1997: 13). This work inspired other studies of the way young children use a range of materials and techniques to create representations of the world around them (e.g., Ivanic and Ormerod).

Closely related is the study of the affordances and learning potentials of different semiotic modes. In *Literacy in the New Media Age* (2003: 52–7), Kress studied the use of different semiotic modes by junior high school students learning about blood circulation. Analysing two specific examples, he showed how one student used language, writing a kind of travel diary with a red blood cell as its protagonist, making a voyage through the body, while another drew a concept map, with boxes representing the heart, the blood vessels, the lungs, and so on, and arrows representing the movement of blood from one 'box' to another. The linearity of the story, Kress said, was an apt signifier for the blood moving from organ to organ, and language allowed the expression of causality, but the use of many different words for the idea of movement ('leave', 'come', 'squeeze through', 'enter' etc.), while stylistically desirable, diminished the generality which the scientific genre normally requires. The diagrammatic elements of the concept map (boxes and arrows), on the other hand, did provide scientific generality, but since all the arrows in the student's concept map radiated from a central 'blood' box, the concept of circularity was less clearly expressed, and since a visual convention for expressing causality does not, or net yet, seem to exist, causality was not expressed at all. Each mode, Kress concluded, has its own epistemological affordances and limitations, and understanding these is fundamental for creating effective multimodal texts.

Studies of classroom interaction have also moved from the traditional emphasis on linguistic exchange structures to strong contextualization and detailed attendance to non-verbal communication and setting, e.g., to the way classrooms are arranged, what is hung on the walls, the technological resources available and so on. Kress et al. (2005) , for instance, described one classroom as realizing a 'transmission' pedagogy, with individual student tables lined up in rows, another as realizing a 'participatory/authoritarian' pedagogy, with tables put together to create teams of four or five students facing each other ('participation'), yet also arranged to allow the teacher total visual control from the front of the classroom, which constrained the posture of the students, at least if they wanted to see the teacher and follow the lesson ('authoritarian')

Visuals, too, can express static and dynamic processes, they just do it differently.

Finally, many studies have analysed multimodal learning resources, from textbooks to computer games. Jewitt (2006), for instance, studied a computer game called *Playground*, designed to help children learn the basics of physics. When learning to understand 'bouncing', for instance, children could choose a 'behaviour' (a particular kind of bounce, represented by pictures of a spring, a ball etc.) and attach it to an object (a 'bullet') which could then bounce off bars. This, as Jewitt argued, was at times confusing. Can bullets be bouncy? Isn't the behaviour of 'bouncing' the property of the bars the bullets bounce against rather than of the bullets? Nevertheless, Jewitt concludes, games of this kind do allow children to explore the rules of mechanics systematically, interactively and multimodally, practically without any verbal input.

Multimodal literacy

The study of multimodal ways of making meaning has been inspired by linguistic concepts and methods, not because it was assumed that all semiotic

modes work the way language does, but because in order to produce and analyse multimodal texts, we need at least some concepts and methods that can apply across different modes. From the point of view of production we must be able to ask: How shall I express, say, a feeling of optimism? Visually, verbally or both? Can optimism be expressed both verbally and visually, and if so, how? From the point of view of analysis, we must be able to ask: How is agency expressed verbally and visually in this multimodal text? And to answer that question, we must know whether agency can be expressed both visually and verbally, and if so, how.

A recent attempt at finding such a common language has been the 'visual grammar' developed by Kress and Van Leeuwen (2006) and further elaborated by others, e.g., by Boeriis (2009). To give just one example, in Hallidayan linguistics (Halliday, 1994), the 'process' of a clause is expressed by a verb or a verbal group. Two broad categories of processes are recognized, those which denote a more or less permanent state of affairs, a characteristic attribute of something (relational processes) or the overall identity or meaning of something (identifying processes), and those which denote some kind of action or event, whether 'material', 'verbal' or 'mental'. The former are expressed by stative verbs or verbal groups, most typically by 'have' and 'be', the latter by dynamic verbs or verbal groups, such as 'walk', 'talk', 'think' and thousands of others. There are no visual equivalents of verbs or verbal groups; in other words, there can be no equivalence between language and visual communication in terms of their forms. However, there are equivalences in terms of their meanings and functions. Visuals, too, can express static and dynamic processes, they just do it differently. They express dynamic processes by means of a vector, a dynamic element formed either by a (usually oblique) element of what is being depicted, for instance a hand raised in a gesture of greeting, or by an abstract element that has a sense of direction, for instance an arrow. And they express static processes through the lack such a dynamic element. Kress and van Leeuwen exemplified this with two pictures that were placed side by side in an Australian social studies textbook for primary school children. One depicted Aboriginal artefacts (including a wooden sword), the other the weapons of the English settlers. The former were depicted in a static, symmetrical arrangement, against a blank background, the latter dynamically, in a picture of English settlers raising their guns as they stalk up to a group of Aboriginal people seated around a fire. This is why language and visual communication can be seen as resources for representation and interaction. They offer choices for how to represent people, places, things and events. If a different choice had been made, with the Aboriginals shown as attacking the settlers (which they did) and the British guns as museum pieces, a wholly different view of history would emerge. This kind of choice is not tied to language or to visuals. It is a cultural choice that derives from a tradition, which includes both the Hebrew god, who acts in history and the Greek god who exists in Heaven. Not all cultures have this choice. In Wintu, a now critically endangered Canadian native language, you cannot say, for instance, 'This is bread'. You have to say the equivalent of "It looks-to-me-bread" or "I think-it-to-be-bread" (Lee, 1954: 51). The distinction between objective and subjective does not exist here.

Figure 1 below is the main part of a screen definition from *Mathletics*, a commercial online resource for learning mathematics that is used by more than four million students the world over. Both the verbal definition and the picture contain a dynamic element. In the verbal definition, the dynamic element is the verb 'arrange'; in the picture, it is the steam coming from the locomotive, representing, together with the slight slant of the locomotive itself, the idea of 'motion' rather than 'arranging'. True, the rails are neatly arranged, but they are less salient than the train, and this may cause confusion. A picture of men laying the track might have been more appropriate.

Figure 1: "To arrange in a line" (from the Mathletics site)

So far we have looked at the verbal and the visual separately. But how do they relate to each other? Here, too, there are choices. They were described in outline by Barthes (1977) and worked out in greater detail by others (e.g., Martinec and Salway, 2005), departing from the linguistic theory of conjunction, which charts how clauses can be related to each other, for instance temporally ('and then') or causally ('and therefore'), or conditionally ('if'). A key general choice, first introduced by Barthes, is the choice between elaboration, the case in which the image 'says the same thing' as the text (or the text 'the same thing' as the image), and extension, the case in which the image adds information that is not in the text (or vice versa), so that text and image complement each other, or perhaps contrast with each other.

In Figure 2, again from *Mathletics*, the picture shows an example of the to-be-defined term (2D Shape) and therefore elaborates the definition, But it also provides additional information about the components of 2D shapes ('sides' and 'corners') and itself contains elaborative text and images relations in which the image shows things and the text labels them. The question, however, is whether 2D shapes, as defined here, always have corners and sides. There are of course cases where text and image relations are deliberately ambiguous, open to multiple interpretations. But perhaps that is less than appropriate in the case of mathematics.

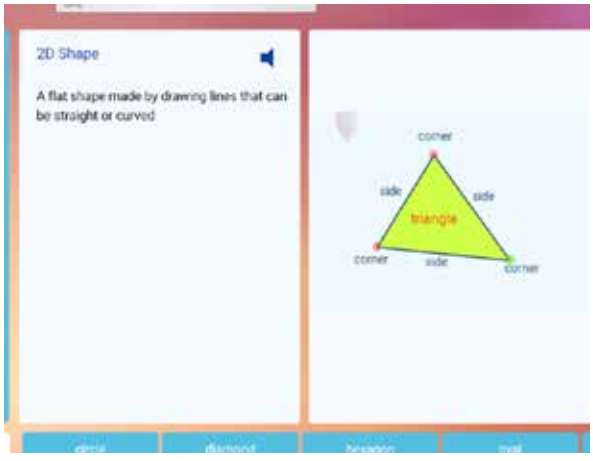

Figure 2: '2D Shape' (from the Mathletics site)

To give another example, 'mind maps' were originally created by advertisers as a resource for brainstorming: an idea is placed in the centre of the visual space and then linked, with lines, to whatever association comes to mind.

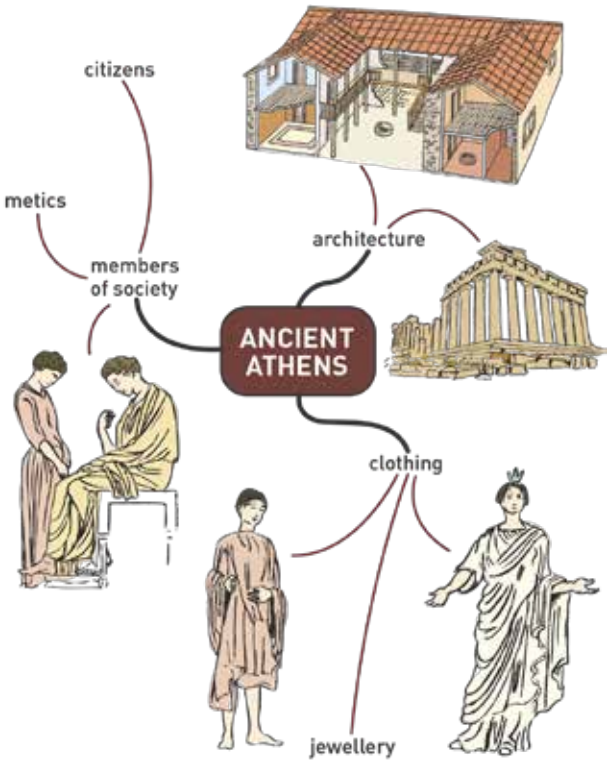

Figure 3: Mind map or concept map? (Reproduced with permission from Addison, P. Bickham, C. Melissas, S. Wood, J, Pearson History 7 Student Book © 2011 Pearson Australia, page no. xix.)

As a whole, this is a visual structure, although the linked elements may be either verbal or visual. 'Concept maps', on the other hand, require more precise relations. In the example above, taken from a junior high school history textbook (Addison, 2011), 'citizens' and 'metics' are depicted as kinds of 'members of society' in Ancient Greece. In other words, the image text relation is one of hyponymy, the 'kind of' relation traditionally expressed in tree diagrams. However, the inventory of citizens is not complete (slaves have been left out, for instance), and the picture of the third kind of 'members of society' is not labelled. Such ambiguities (is this a 'mind map' or a 'concept map'?) may impede effective learning, and resources for explicit multimodal analysis are needed to spot them.

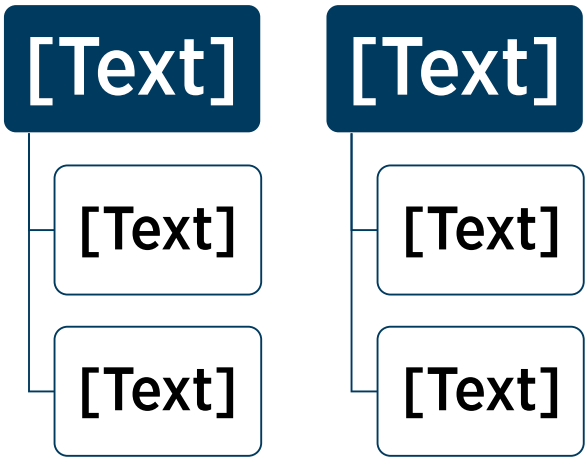

Figure 4: SmartArt template

The SmartArt template in figure 4 is, according to Microsoft, to be used "to show hierarchical relationships progressing across groups" and "can also be used to group or list information". However, it may be argued that hierarchies and lists are not the same – all the items in a list are, in principle, equal, of the same kind, even though the items at the top of the list may be prioritized. Also, hierarchies and lists are static and do not include any element of 'progress'. When such contradictions occur, there may be an ideological snake under the grass. Power is played down here, and corporate structure becomes a model for understanding the world. As Kvåle (2016: 269) has explained in the conclusion of a study of the use of SmartArt in the work of students in higher education:

SmartArt [is] a historically and socially evolved resource for defining organizational charts. (...) Because of the profound status of Microsoft Office today, the company's templatized idea of the most "effective" visual style of organizational charts is infused into all social practices, including education (...)

Clearly, multimodal literacy should also be a form of critical literacy.

Aesthetics

In the more or less 'monomodal' past, the visual was seen as embellishment, an optional extra. Today, verbal and visual communication work together (often in tandem with other means of expression) with respect to all three of the main functions of communication – creating representations of the world, enabling interactions and the relationships that go with them and forming coherent texts that can be recognized as performing a communicative job of some kind in the world. This does not mean, however, that the aesthetic element has gone away. On the contrary, as pioneered by advertising, a field whose communicative strategies are now a major influence in many other forms of communication, the aesthetic has entered many domains where it formerly played no role. Writing of all kinds now has to 'look good', and not only in the work of professionals: "If you think a document that looks this good has to be difficult to format, think again!" says a Word template for company brochures, and "add professional quality graphics which automatically match the look of your report" (cf. Van Leeuwen, 2015). Functional and aesthetic communication have now merged, and many non-marketing forms of communication are expected to follow the lead of marketing in combining functionality (providing information, persuading, instructing, applying for jobs or leave and even invoicing) with the communication of the values and meanings that make up the identity of the communicators, whether they are individuals or corporations. This, too, has to form part of multimodal literacy, which perhaps suggests that art and communication must begin to merge in the curriculum, as it already has in the world.

The aesthetic has entered many domains where it formerly played no role.

Concluding remarks

Not long ago, multimodal text design was a specialized professional skill. Today, digital technology has brought resources for multimodal text design within reach of anyone who has a computer and introduced multimodal text design into many areas that were previously 'monomodal', for instance in workplace documents, such as invoices, reports, presentations, organization charts and workflow documents, brochures, newsletters.

As a result, the study of multimodality has grown rapidly, devising detailed grammars that show how different semiotic modes such as layout, colour and typography make meaning and combine into coherent multimodal texts. Multimodal literacy has become an important work area and life skill, requiring not only a working knowledge of multimodal ways of making meaning but also of the rules and conventions for using multimodal meaning making that have emerged, with greater or lesser prescriptiveness, in specific contexts.

Multimodal literacy therefore also needs to be critical literacy, especially in relation to the way technological tools favour certain forms of discourse.

Multimodal literacy has also added an aesthetic dimension to everyday forms of communication that were formerly purely functional and monomodal. It fosters forms of communication that combine functionality, a focus on the job that a given piece of communication is intended to perform, with the use of aesthetic design ideas meant to communicate corporate or personal identities and values.

Finally, many of the digital resources that provide wide access to multimodal text design favour specific forms of communication that are built on specific values and serve specific aims. PowerPoint, for instance was originally designed by Bell Laboratory engineers to pitch ideas to management and is still particularly suited for the concise presentation of a number of 'selling points'. However, it less easily facilitates other forms of discourse that may, for instance, be important in education, such as extended arguments and narratives (cf. Tufte, 2006). Multimodal literacy therefore also needs to be critical literacy, especially in relation to the way technological tools favour certain forms of discourse.

References

Addison, P. (2011) *Pearson History Student Book 7*. Melbourne: Pearson

Barthes, R. (1967) *Elements of Semiology*. London: Cape

Barthes, R. (1977) *Image, Music, Text*. London: Fontana

Barthes, R. (1983) *The Fashion System*. Berkeley and Los Angeles: University of California Press

Boeriis, M. (2009) *Multimodal Semiotik og Levende Billeder*. Odense: Syddansk Universitet, Humanistisk Fakultet

Bogatyrev, P. (1971[1934]) *The Function of Folk Costume in Moravian Slovakia*. The Hague: Mouton de Gruyter

Garvin, P. L. (1964) *A Prague School Reader on Esthetics, Literary Structure and Style*. Washington, DC: Georgetown University Press

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

Jewitt, C. (2006) *Technology, Literacy and Learning: A Multimodal Approach*. London: Routledge.

Kress, G. (1997) *Before Writing: Rethinking the Paths to Literacy*. London: Routledge

Kress. G. (2003) *Literacy in the New Media Age*. London: Routledge

Kress, G, Jewitt, C., Bourne, J, Franks, A. Hardcastle, J., Jones, K. and Reid, E. (2005) *English in Urban Classrooms: A Multimodal Perspective on Teaching and Learning.* London: Continuum

Kress, G. and van Leeuwen, T. (2006) *Reading Images – The Grammar of Visual Design*. 2nd edition. London: Routledge

Kvåle, G. (2016) 'Software as Ideology: A multimodal critical discourse analysis of Microsoft Word and SmartArt'. *Journal of Language and Politics* 15(3): 259–274

Lee, D. (1954) 'Notes on the conception of the self among Wintu Indians', in E. Carpenter and M. McLuhan, eds. *Explorations – Studies in Culture an Communication, Vol 3*, p.44–57

Martinec, R. and Salway, A. (2005) 'A system for image-text relations in new (and old) media'. *Visual Communication* 4(3): 337–372

Matêjka, L. and Titunik, I. R., eds. (1976) *Semiotics of Art: Prague School Contributions*. Cambridge, MA: MIT Press.

New London Group (1996) 'A pedagogy of multiliteracies: designing social future' *Harvard Educational Review* 66: 60–92

Pittenger, R. E., Hockett, C. F., and Danehy, J. H. (1960) *The First Five Minutes*. Ithaca, NY: P. Martineau

Tufte, E. R. (2006) *The Cognitive Style of PowerPoint: Pitching Out Corrupts Within.* Cheshire, CT.: Graphics Press LLC

Van Leeuwen, T. (2005) *Introducing Social Semiotics*. London: Routledge

Van Leeuwen, T. (2015) 'Looking good: aesthetics, multimodality and literacy studies' in J. Rowsell and K. Pahl, eds. *The Routledge Handbook of Literacy Studies*, pp. 426–439

Veltruský, J. (1964[1940]) 'Man and object in the theatre', in P. L. Garvin, ed. *A Prague School Reader on Esthetics, Literary Structure and Style*. Washington, DC: Georgetown University Press, pp. 83–89