

The Multimodal Analysis Laboratory Interactive Digital Media Institute (IDMI) National University of Singapore

- **Director** Associate Professor Kay O'Halloran
- Department of English Language & Literature.
- National University of Singapore

© Kay O'Halloran. Not to be reproduced without permission.
Email: ellkoh@nus.edu.sg

Read More:

One pressing problem for industry, business and government is the analysis, storage and retrieval of information from the visual images and video texts which have proliferated through the rapid advance of digital technology. Efficient practices for managing, searching and retrieving data from text-based documents have been developed. However, static visual images (e.g. photographs, drawings and diagrams) and video texts remain elusive with regards to management, coding and searching for information. The problem intensifies as the movie clips and video-blogs in video-sharing websites such as “YouTube”, the company recently purchased by Google Inc. for 1.65 billion U.S. dollars, become the standard forum for exchanging information. Moreover, the information on such web sites arises from the multimodal integration of spoken/written language and static/dynamic visual imagery and video clips (see Figure 1), a phenomenon which requires theoretical concepts and approaches which clearly extend beyond those developed for text-based data.

The Multimodal Analysis Lab involves a major research programme where computer scientists and social scientists collaborate in a series of projects to develop new perspectives and approaches for analysing visual imagery and video texts in the complex multimodal sites afforded by interactive digital technology. The research programme has significant implications and practical applications for industry, business and government because the problem of understanding and accessing

information in interactive digital sites is critical, especially in the fast-approaching era of 'the virtual reality experience' where the 'object-based' static environment of old technology (such as the printing press) is replaced with 'process-based' dynamic interactive multimodal environment of digital technology.

The shift to dynamic process-based environment calls for new theoretical perspectives and analytical practices for understanding, coding and retrieving information. Simultaneously, and somewhat ironically, digital technology functions to constrain the meanings which can be made, a disjunction which is exploited in the development of new technologies which open up further potential for meaning making (e.g. the ongoing development of mobile phones).

One theoretical foundation for developing new approaches to the analysis, storage and retrieval of information from visual images, video texts and interactive digital sites is 'multimodal social semiotics' because *interactive digital technology is multimodal semiotic technology*. That is, digital technology provides the platform for semiotic resources (e.g. language, visual imagery, gesture, facial expression, movement, use of space, music, sound and so forth) to combine and unfold in new and innovative ways. A theoretical framework for understanding the potential meanings which can be made, and those meanings which are actually made in the visual images, video texts and interactive digital environments is required. Such an approach involves understanding the functions of each semiotic resource, and the ways in which the different semiotic choices integrate and combine to make meaning. That is, the *experiential meanings* and *logical relations* of semiotic choices within and across visual images, video texts and interactive digital sites must be analysed and understood, in addition to the *interpersonal meanings* which function to persuade or direct users to undertake certain courses of action. The *textual organisation* of semiotic choices and the compositional arrangement of items in the interactive digital sites enable particular experiential, logical and interpersonal meanings to be made. The multimodal social semiotic approach provides a comprehensive theoretical framework for understanding the events taking place in the process-based dynamic

multimodal environment of interactive digital technology, and the nature of the experiential, logical, interpersonal and textual meanings which are subsequently made.

The research programme in the Multimodal Analysis Lab requires the collaboration of social scientists and computer scientists to develop new platforms for manual, semi-automatic and fully-automated techniques and procedures for image, video and complex digital site analysis using multimodal social semiotic principles. It is envisaged that the research collaboration between social scientists and computer scientists in the Multimodal Analysis Lab, each sharing, developing and expanding their own knowledge base, will revolutionise existing approaches to interactive digital media analysis, leading to the development of patentable commercial software for multimodal data analysis, search and retrieval within and across visual images, video texts and interactive digital sites. This vision is developed in the ensuing discussion of the research programme for the Multimodal Analysis Lab.

OVERVIEW OF THE MULTIMODAL ANALYSIS LABORATORY

The major aims of the research programme for the Multimodal Analysis Lab are (a) to develop new approaches to the analysis, storage and retrieval of information from visual images, dynamic video texts and complex integrated digital environments, and (b) to produce prototype software for multimodal semiotic data analysis, search and retrieval. Significantly, the research will help solve important problems of industry, business and government, and furthermore, it will give rise to new interdisciplinary approaches to computer science and humanities based research. Lastly, the theoretical concepts and analytical procedures developed in the Multimodal Analysis Lab will enhance the quality, functionality and understanding of the digital technologies developed in IDMI. The Multimodal Analysis Lab therefore provides an integrating horizontal structure within the IDM Institute, as displayed in Figure 2.

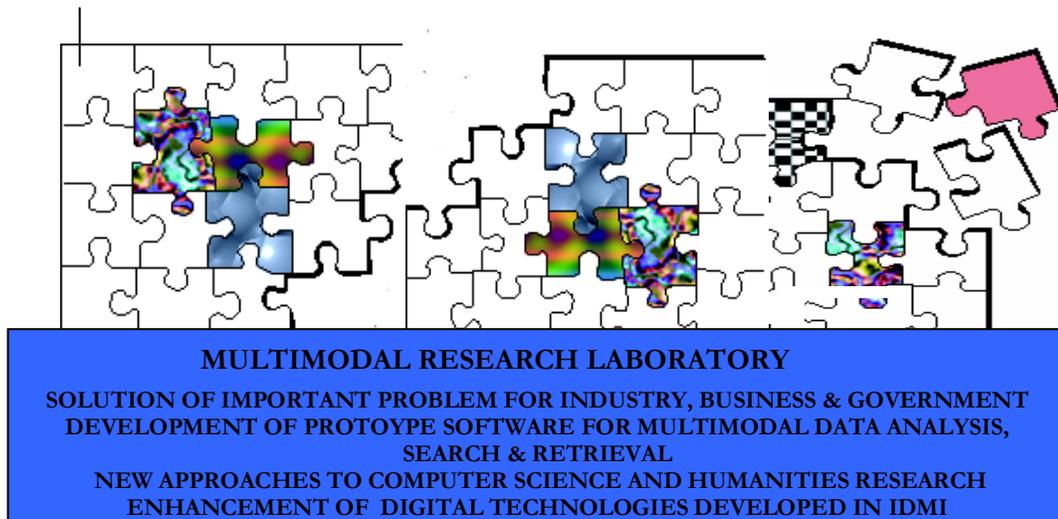


Figure 2 Integrating Horizontal Structure of Multimodal Analysis Laboratory

The specific objectives of the Multimodal Analysis Lab are:

- (1) To help solve important problems for industry, business and government involving multimodal data analysis, storage and retrieval of information from visual images, video texts and complex interactive digital environments.
- (2) To develop prototype software for multimodal data analysis, storage and retrieval with downstream applications for industry R&D.
- (3) To develop new interdisciplinary approaches for computer science and the humanities research using multimodal semiotic theory and interactive digital technology.

The predecessor unit, A/Prof Kay O'Halloran's Semiotics Research Group (SRG), based in her Laboratory for Research in Semiotics (LRS) in the Department of English Language & Literature NUS, has a proven track record with regards to developing commercial software for linguistic analysis (Judd & O'Halloran, 2002); attracting and retaining research scholars and publishing their research (O'Halloran, 2004a) and developing a strong international network of researchers in the field of multimodal social semiotics. Leading international scholars in the field have expressed interest in collaborating with the Multimodal Analysis Lab.

A brief introduction to multimodal social semiotic theory which underlies the development and use of IDM technology for multimodal discourse analysis is given below.

MULTIMODAL SOCIAL SEMIOTIC THEORY

Multimodal social semiotics is concerned with the theory, analysis and interpretation of the communicative practices of a culture which are regarded as dynamic processes and objects for analysis (e.g. day-to-day interactions, movies, theatre productions, cities and other 3-D sites, art installations, interactive gaming, educational practices and virtual realities). Multimodal social semiotics draws upon a variety of traditions which include systemic functional theory, critical theory, cultural studies and critical discourse analysis to provide theoretical frameworks for conceptualising the complex array of semiotic resources which are used to create meaning, and practices for analysing the integrated use of those resources. For example, Figure 3 is a still shot from a dynamic analysis in film format (O'Halloran, 2004b) which shows how interpersonal emotion is heightened through the integrated use of semiotic choices such as gaze, proxemics, movement, gesture and language, in addition to other semiotic choices in the film production (e.g. lighting, music, items of clothing, props, framing and camera angle, camera distance and camera movement). Multimodal social semiotic analysis takes into account the abstract systems (the micro-sign systems) from which semiotic choices are made (the textual instance) within different situational, cultural and global contexts (the macro-social/cultural system).

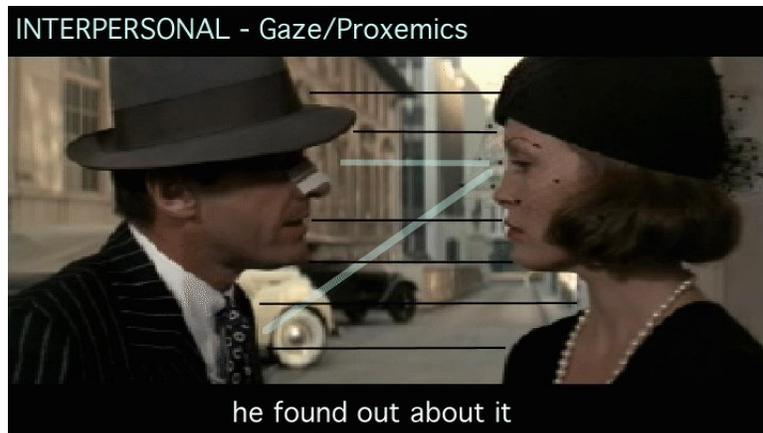


Figure 3 Visual Analysis of Gaze and Proxemics (by O'Halloran, 2004b)

Multimodal social semiotics is concerned with “the way people use semiotic ‘resources’ both to produce communicative artefacts and events and to interpret them ... in the context of specific social situations and practices” (van Leeuwen, 2005: preface). The notion of *semiotic resource* in social semiotics is critical. Van Leeuwen (2005: 3) explains “[i]t originated in the work of [Michael] Halliday who argued that the grammar of a language is not a code, not a set of rules for producing correct sentences, but a ‘resource for making meanings’ (1978: 192)”. Social semiotics differs from other theoretical approaches in that the major concerns include the ‘meaning potential’ of semiotic resources (i.e. the grammar) and the ‘actualised potential’ of semiotic choices which function to construct discourse, situational contexts and, more generally, culture. Halliday’s (2004) identifies four types of semiotic meaning potential, which he calls the metafunctions of language: (a) ‘experiential meaning’ for constructing our experience of the world (b) ‘logical meaning’ for establishing logical relations in that world) (c) ‘interpersonal meaning’ for enacting social relations and (d) ‘textual meaning’ for organising the message. The analysis of spoken and written discourse involves the study of the four metafunctional-based meanings of the grammatical choices which are made. The aim is to critically interpret discourse with respect to the situational and cultural context of the text. Halliday’s social

semiotic theory therefore necessarily encompasses larger sociological and cultural perspectives of meaning making.

Halliday's (2004) systemic functional approach has been extended to other semiotic resources, which are conceptualised as communicative resources for making meaning in a culture; e.g. displayed art (O'Toole, 1994), visual images (Kress & van Leeuwen, 1996), music (van Leeuwen, 1999), architecture and three dimensional space (O'Toole, 1994; Stenglin, 2004), gesture and action (Martinec, 2000, 2004) and mathematical symbolism (O'Halloran, 2005). The study of meaning arising from the interaction of semiotic resources has been investigated in sites such as cities, hotels, and museum exhibitions, and printed, digital and film texts (e.g. Baldry & Thibault, 2006; O'Halloran, 2004a, 2005; Ventola et. al., 2004; van Leeuwen, 2005). The research has been collectively called *multimodality*, where 'multimodal' refers to the multiple modes (e.g. spoken, written, printed and digital media, embodied action, and three-dimensional material objects and sites) through which social semiosis takes place. Recent research in multimodality has been concerned with multimodal transcription (e.g. Baldry & Thibault, 2006) and corpus-based approaches to multimodal discourse analysis (Baldry 2004; Bateman et. al., 2004).

The interdisciplinary approach developed in the Multimodal Analysis Lab will involve the multi-dimensional dynamic modelling and visualisation of the metafunctional meanings arising in multimodal social semiosis through the platforms afforded by interactive digital technology (O'Halloran, forthcoming). Such an approach necessarily requires the expertise of computer scientists in the Multimodal Analysis Lab. The impact of technology on the advancement of knowledge in mathematics and science is briefly discussed in Section 4, with the aim of demonstrating how the use of interactive digital technology has the potential to revolutionise multimodal social semiotic data analysis of visual images, video texts and interactive digital sites found on the internet.

INTERACTIVE DIGITAL TECHNOLOGY FOR MULTIMODAL ANALYSIS

The relationship between advances in technology and the development of mathematics and science has been documented (Galison, 2002; O'Halloran, 2005; Swetz, 1987). The printing press, for example, explains the increased popularity of the Hindu-Arabic numerical system in early arithmetic books. Manual forms of computing, such as using counters and the abacus, together with visual representations of these computational strategies, could not compete with the more efficient semiotic form of the Hindu-Arabic system. Furthermore, printed mathematical texts contributed to the close study and development of arithmetical algorithms and the standardisation of mathematical procedures, techniques and symbols which in effect paved the way for the development of symbolic algebra. Eisenstein (1979: 467) explains: '[c]ounting on one's fingers or even using an abacus did not encourage the invention of Cartesian coordinates'. Swetz (1987: 284) agrees that:

'[p]rinting ... forced a standardization of mathematical terms, symbols, and concepts. The way was now opened for even greater computational advances and the movement from a rhetorical algebra to a symbolic one'.

The ongoing development of the integrated use of mathematical symbolism, visual imagery and language led to the scientific revolution of the seventeenth century. More recently, digital technology is impacting on the development of theory, knowledge and practices in mathematics, sciences, and other disciplines which include engineering, medicine and architecture. Galison (1997: 689), for example, explains, '[W]ithout the computer-based simulations, the material culture of late-twentieth-century microphysics is not merely inconvenienced - it does not exist'. The same could be said for other new paradigms of research, such as dynamical systems theory in mathematics.

The humanities disciplines, however, have been slow to use digital technology for the creation of new knowledge, especially in those areas of research involving multimodal social semiotic data analysis. The collaborative research between social scientists and computer scientists in the Multimodal Analysis Lab will revolutionise the way in which multimodal social semiotic data analysis is undertaken through the development of manual, semi-automated and fully automated software applications.

For example, the interactive digital technology for multimodal social semiotic analysis developed in the Multimodal Analysis Lab, may potentially involve the dynamic visual display of patterns which track the metafunctional orchestration of semiotic choices over time. The interactive digital platform will consist of multiple tracks and overlays which can be annotated in various ways to code the multimodal social semiotic analysis. Such an approach to multimodal discourse analysis has been attempted using commercially available video-editing software applications such as *Adobe Premiere* (O'Halloran, 2004b). For example, Figure 6 is a screen shot of the film tracks and overlays used in *Adobe Premiere* to analyse a movie extract. However, it soon becomes apparent that the complexity of multimodal social semiotic analysis requires dedicated platforms for entering metafunctional analysis to map the flow of meanings of semiotic choices as they combine over time. Drawing manual overlays to map semiotic flow is simply too difficult and time-consuming; it is not an option.

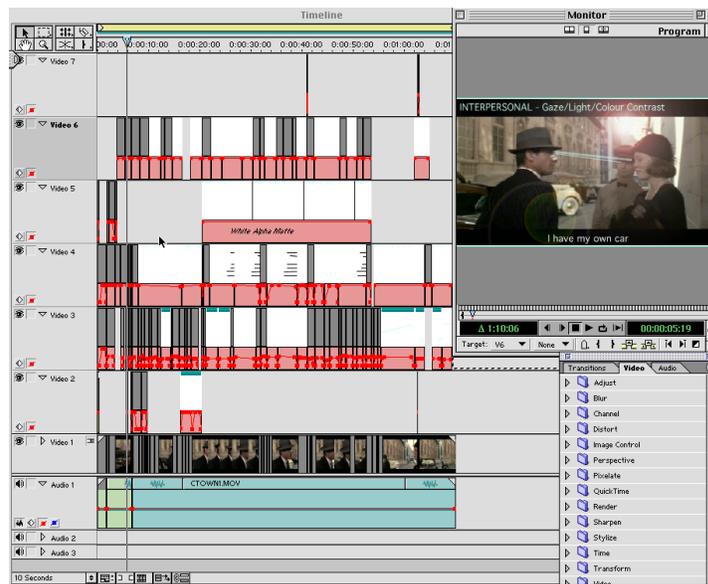


Figure 4 Visual Tracking of the Metafunctional Orchestration of Meaning

Advances in computer science, especially those in the field of visual computing, mean that the interactive dynamic modelling of complex multimodal relations has become a possibility. For example, changes in facial expressions can be dynamically mapped and coded using 2D/3D graphics for real-time face detection, recognition and tracking and 3-D face modelling. Similarly, gestures can be dynamically traced using 2D/3D human tracking, image-based modelling, simulations of virtual human motion and 3-D reconstruction methods. The social scientists in the Multimodal Analysis Lab will work with computer scientists to investigate such advances in visual computing technologies in the quest to develop process-based interactive digital technologies for multimodal social semiotic analysis.

The challenges for researchers in the Multimodal Analysis Lab include:

- (i) Modelling, tracking and coding the dynamic metafunctional orchestration of meaning of semiotic choices in written, printed and digital media, including transitions and phases where the flow is

temporarily suspended and/or disrupted into new orbits and constellations.

- (2) From (1), developing new theories and practices for multimodal social semiotics involving perspectives from other fields such as mathematics, science, computing architecture, music, sociology, anthropology and cognition.
- (3) Developing manual, semi-automated and fully automated techniques and analytical practices for visual image, video text and integrated digital site data analysis, tagging, coding, searching and retrieval.
- (4) Developing new research paradigms for modelling and analysing multimodal social semiotic phenomenon.

The Multimodal Analysis Lab will explore theoretical and practical approaches which contribute to research fields (1)-(4) with the aim of producing prototype software for multimodal social semiotic discourse analysis, search and retrieval. The present gaps in knowledge include developing a dynamic approach to theorising and analysing intersemiotic phenomena from an interdisciplinary perspective using digital technology. That is, theories and practices from computer science and other disciplines (e.g. dynamical systems theory) for analysing dynamic phenomenon may be included in the approaches which are developed. The Multimodal Analysis Lab will become a leading international centre for innovative research in multimodal and multimedia analysis, search and retrieval, and the research projects will have downstream applications with regards to R&D aimed at solving important problems for industry, business and government.

REFERENCES

- Baldry, A. P. and Thibault, P. J. (2006). *Multimodal Transcription and Text*. London: Equinox.
- Baldry, A. P. (2004). 'Phase and Transition Type and Instance: Patterns in Media Texts as Seen Through a Multimodal Concordancer', in K. L.

- O'Halloran (ed) *Multimodal Discourse Analysis*. London & NY: Continuum, pp. 83-108.
- Bateman, J., Delin, J. and Henschel, R. (2004), 'Multimodality and Empiricism: Preparing for a Corpus-based Approach to the Study of Multimodal Meaning-Making', in E. Ventola, C. Charles and M. Kaltenbacher (eds), *Perspectives on Multimodality*. Amsterdam: John Benjamins, pp. 65-87.
- Eisenstein, E. L. (1979). *The Printing Press as an Agent of Change*. Cambridge: Cambridge University Press.
- Galison, P. (1997). *Image and Logic: A Material Culture of Microphysics*. Chicago : University of Chicago Press
- Galison, P. (2002). 'Images Scatter into Data, Data Gathers into Images', in B. Latour, and P. Weibel (eds), *Iconoclash: Beyond the Image Wars in Science, Religion, and Art*. Karlsruhe: ZKM, pp. 300-323.
- Halliday, M. A. K. (1978). *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.
- Halliday, M. A. K. (2004). *An Introduction to Functional Grammar* (3rd ed. revised by C. M. I. M Matthiessen). London: Arnold.
- Judd, K. and O'Halloran, K. L. (2002). *Systemics 1.0*. [CD ROM]. Singapore: Singapore University Press.
- Kress, G. and van Leeuwen, T. (1996). *Reading Images: The Grammar of Visual Design*. London: Routledge (2nd ed. 2006).
- Martinec, R. (2000). 'Types of Processes in Action'. *Semiotica*, 130-3/4, 243-268.
- Martinec, R. (2004). 'Gestures that Co-Concur with Speech as a Systematic Resource: The Realization of Experiential Meanings in Indexes'. *Social Semiotics*, (14 (2), 193-213.
- O'Halloran, K. L. (ed) (2004a). *Multimodal Discourse Analysis*. London & NY: Continuum.
- O'Halloran, K. L. (2004b), 'Visual Semiosis in Film', in K. L. O'Halloran (ed) *Multimodal Discourse Analysis*. London & NY: Continuum, pp. 109-130.
- O'Halloran, K. L. (2005). *Mathematical Discourse: Language, Symbolism and Visual Images*. London & NY: Continuum.

- O'Halloran, K. L. (forthcoming). Mathematics, Science and Technology: A Multimodal Perspective. In A. Baldry and E. Montagna (eds.) *Interdisciplinary Perspectives on Multimodality: Theory and Practice*. Campobasso: Palladino.
- O'Toole, M. (1994). *The Language of Displayed Art*. London: Leicester University Press.
- Stenglin, M. K. (2004). *Packing Curiosities: Towards a Grammar of Three Dimensional Space*. PhD Thesis. Linguistics Department, University of Sydney.
- Swetz, F. J. (1987). *Capitalism and Arithmetic*. La Salle, Illinois: Open Court.
- van Leeuwen, T. (1999). *Speech, Music, Sound*. London: Macmillan.
- van Leeuwen, T. (2005). *Introducing Social Semiotics*. London: Routledge.
- Ventola, E., Charles, C. and Kaltenbacher, M. (eds) (2004). *Perspectives on Multimodality*. Amsterdam: John Benjamins.