

Multimodal Text Analysis

Kay L. O'Halloran and Bradley A. Smith

National University of Singapore

kay.ohalloran@nus.edu.sg

semiosmith@gmail.com

Word Count: 2,785

Reference Word Count: 338

Multimodal text analysis has become a crucial part of research, teaching and practice for a wide range of academic and practical disciplines. A variety of techniques, theoretical frameworks and methodologies have therefore evolved for such analysis. For linguists, in particular, concerned with accounting for the communication of meaning within texts, issues arising from the consideration of semiotic resources other than language, in interaction with each other and with language – such as gesture, gaze, proxemics, dress, visual and aural art, image-text relation and page-layout, cinematographic and sound design and production resources, etc – have emerged in recent decades as important challenges. Meanwhile, the emergence of multimodal studies as a distinct area of study in linguistics has also revealed a range of issues specifically relevant to the multimodal text analyst.

Multimodal analysis includes the analysis of communication in all its forms, but is particularly concerned with texts which contain the interaction and integration of two or more semiotic resources – or ‘modes’ of communication – in order to achieve the communicative functions of the text. Such resources include aspects of speech such as intonation and other vocal characteristics, the semiotic action of other bodily resources such as gesture (face, hand and body) and proxemics, as well as products of human technology such as carving, painting, writing, architecture, image and sound recording, and in more contemporary times, interactive computing resources (digital media hardwares and softwares). Different semiotic resources bring with them their own affordances and constraints, both individually and in combination, as well as analytical challenges in terms of the natures of the media, the detail and scope of analysis, and the complexities arising from the integration of semiotic resources across media.

While it has long been understood that human meaning in the round involves more than simply the (written) language studied by the early linguists (cf Saussure’s observations on the need for a semiological science (1916/1974, p. 16) “that studies the life of signs within society”), practical and theoretical challenges have meant that it is only in recent decades that a distinct field of multimodal studies has begun to emerge, and many issues remain unresolved within this field, particularly with respect to text analysis. Halliday (1985) has observed that it wasn’t until the wide availability of sound recording technology (in particular tape) that widespread empirical study of authentic natural spoken discourse began, with profound consequences during the 1960s and 1970s for linguistic theory and methodology. Yet studies of speech, as well as other modalities have long remained tied to the analysis of transcriptions rather than a direct ongoing

engagement with the source text, bringing issues of transcription to the fore that are relevant to multimodal text analysis (cf Ochs 1979 for a discussion). The reasons for this constraint are clear: in the first instance, the difficulties of accessing and annotating dynamic audiovisual media such as sound and film are manifest, and modern interactive hypermedia have added further difficulties (e.g. Lemke 2002). Secondly, and as a result, without such access and annotation capabilities a close repeated analytical attention to the source text has been difficult, meaning either that theoretical development has tended to occur without extensive grounding in empirical analysis, or when such grounding is present – where multimodal analysts derive general principles from the close sustained study of multimodal phenomena within actual text - the texts tend to be static (visual) media or such as can readily be examined and reproduced on the printed page. If the texts analysed are dynamic, the issues both of transcription, multimodal analysis and reproduction for publication are palpable, particularly in print-based approaches which offer limited resources for multimodal text analysis (e.g. Baldry and Thibault 2006; O'Halloran, in press a; O'Halloran et. al. 2010).

Two major strategies have thus emerged for dealing with these challenges, constituting two of the major approaches by which multimodal text analyses can be characterised: one is to explore theory, using text analysis as both test and illustration of the discussion of general principles; and the other is explore actual texts, working from such analyses towards generalizations. The first approach doesn't assume and indeed problematises theory, and much of the work is involved with comparing and integrating knowledge and practice from often disparate theoretical traditions. Multimodal studies suits such an approach, when one considers

the wide range of disciplinary and theoretical traditions for which multimodality is relevant. In this approach theoretical generalization is often applied to analyses across several different texts and text types, in order to derive, test and illustrate general principles. The second approach is to pay very close attention to and work from actual specific texts, where the focus is upon the text, building up detailed description, often as a form of running annotation (particularly if the text is dynamic), and usually adapting and applying an established theoretical and descriptive framework but deriving descriptive generalizations out of such text analysis and modifying theory as a result. In this approach, analytical detail is paramount, and again, the challenges of access, transcription, analysis and reproduction in publication are all too apparent.

It is interesting to note that two pioneering works in multimodal text analysis, roughly contemporaneous, can be characterised with respect to these generalizations. The first derives from, among other influences, a synthesis of social semiotics, European semiotics and critical discourse analysis traditions. The exemplar work is Kress and van Leeuwen (1996), where powerful theoretical generalizations from these different schools of thought are applied, calibrated and synthesized via consideration of a diverse range of texts. The multimodal text analysis appears as both a reference and illustration for the discussion at hand; but it is the theoretical discussion which drives the analysis and text description: generalizations are paramount; while the analysis is usually conducted and presented discursively. Thus, Kress and van Leeuwen (1996) begin a general discussion about the title of the book: highlighting the difference between a 'grammatical' and a 'lexical' approach to the semiotics of visual design from the perspective of social semiotic theory, stressing the importance of the distinction between

formal and functional theories, and highlighting the relation of specific multimodal features, structures and systems to more general social (particularly power) structures and (ideological) forces:

The dominant visual language is now controlled by the global cultural/technological empires of the mass media, which disseminate the examples set by exemplary designers, and through the spread of image banks and computer imaging technology, exert a ‘normalizing’ rather than explicitly ‘normative’ influence on visual communication across the world. Much as it is the primary aim of this book to describe the current state of the ‘grammar of visual design’, we will also discuss the broad historical, social and cultural conditions that make and remake the visual ‘language’ (Kress and van Leeuwen, 1996, pp. 4-5)

The second approach was pioneered and is best represented by the work of Michael O’Toole (1994). O’Toole (1995, p. 159) provides an “adaptation of Systemic-Functional grammar” as “at least some shared (or sharable) terminology and assumptions”, stating that:

My thesis is quite simple: Michael Halliday’s Systemic-Functional linguistics offers a powerful and flexible model for the study of other semiotic modes besides natural language, and its universality may be of particular value in evolving new discourses about art.

Here, as in later work O'Toole's work has a characteristically close analytical orientation to specific texts, explicitly working, in the first instance, analytically from the text itself rather than from perspectives outside of the text, such as historical, biographical or mythological interpretations. O'Toole (1995, p. 159) prophetically argues that "a proper semiotics will only grow out of a large body of analysis, description, interpretation and theory by people with a range of orientations".

Thus O'Toole advocates an approach that provides tools, adapted from linguistic theory and description, for working from specific texts – and the title is significant in this regard; while Kress and van Leeuwen (1996) develop an approach that explores generalizations out of the study of different types of theory and text, and from the integration of theories to apply to texts in general, including, importantly, extra-textual perspectives (e.g. related to the analysis of ideology, historical power structures etc). However, O'Toole (1994) also later works towards social semiotic generalizations, drawing upon a wider range of theoretical traditions; and Kress and van Leeuwen include a wealth of useful analysis derived from the application of a coherent theoretical framework: the difference, that is, is one of approach and aim.

In both these foundation works in multimodal studies, although reference is made to the study of dynamic audiovisual text, there is a preponderance of static visual art as data. Earlier multimodal text analysts also tend to either an analytical focus static visual text (eg Barthes' famous 1957/1972 Paris Match analysis) or, when confronting dynamic (audio, visual) texts are compelled to rely on discursive description or generalisation rather than on the presentation of

detailed empirical analysis as the foundation for discussion. In practical terms the analysis of static art is relatively feasible even within the constraints of the printed page; while the multimodal analysis of other media, as discussed earlier, points clearly to the difficulties of representing on the printed page the mass and complexity of detail involved in multimodal analysis, and of capturing the dynamism and dimensionality of audiovisual film, hypermedial navigation etc (e.g. Baldry and Thibault 2006; Lemke 2002).

Thus while the difficulties of multimodal text analysis are apparent in most works which explore this field, there has nevertheless been built up over time a corpus of detailed multimodal analysis which has provided the means for testing, exploring and illustrating ideas about how multimodal communication works. It is important to recognize in this respect that while it is only in recent decades that extensive multimodal text analysis has heralded the emergence of a distinct field of multimodal studies, in fact the study of multimodal communication and artefacts can of course be found in the long lineage of works within anthropology, archeology, art criticism and history (painting, sculpture, music, theatre, opera, film etc), computer science, engineering, psychology and all fields of research engaged with human or non-human communication. The related practical disciplines, such as the various forms of art, also provide ample material for multimodal text analysts to draw upon. In addition, in recent decades the rapid increase in sophistication and availability of technological (particularly computational) resources and techniques for analysis of multimodal text has no doubt driven the rapid increase in multimodal analyses appearing within a range of disciplines, vastly improving, as technology did for the study of speech earlier, our access to and understanding of multimodal text using, for example, multimodal annotation

software (e.g. Praat, ELAN, MacVissta, see Rohlving 2006). The ongoing development of interactive digital techniques, along with the increasingly collaborative nature of research within the twenty-first century, point to a period of further growth in coming years within this field.

It is clear that both the empiricism of detailed, exhaustive text analysis (coping with the challenges this raises) and the ongoing problematisation and exploration of theoretical generalization and abstraction are needed for the development of resources for and practice of multimodal text analysis. In addition, multimodal text analysis requires the integration of both low and high level analyses: as with early studies of intonation, the study of multimodality has been centrally concerned with the material plane, the organization of the physical (including technologically-driven) phenomena by which meaning is created, in their relation to higher level grammatical and semantic organizations of such as abstract, semiotic systems and structures. How to relate what Hjelmslev called the content and expression planes has been a central challenge for multimodal analysts, and this challenge has become more acute as sophisticated instrumental analysis – particularly computational automated techniques for feature extraction – make possible the analysis of primarily low-level features within large corpora of multimodal text. The availability of such techniques is both a great opportunity and difficulty for scholars of multimodal communication: one major challenge is to trace and make explicit the path from such automatically detected features to the socio-cultural patterns significant to multimodal analysis of text, in ways that draw upon the respective riches of computational and semiotic sciences, as well as the knowledge and practices of other disciplines such as mathematics, physics, psychology, ethnography and so on. The immediate demands of multimodal text analysis in fact

become site for the exploration of synergies between disciplinary and theoretical traditions, as we seek to understand actual texts through the kaleidoscope of the multitude of relevant perspectives.

The variety of work and approaches within this range of somewhat disparate fields of science and humanities research constitutes both a rich resource and a challenge for the multimodal analyst. One must sort through the complementarities, inconsistencies and redundancies of the different approaches and perspectives, working out which types of analysis suit which research project - some approaches being more appropriate to certain tasks than others. This has been a key issue with respect, in particular, to the adaptation and application of linguistic theory and description, which has been an important influence upon multimodal studies. Multimodal analysis must of course include analysis of language where relevant; but in the analysis of semiotic resources other than language, whether involved in intersemiotic relations with language or not within a particular text, the question of how much of linguistics can be adapted for the analysis is still an open question (Machin 2009). It is important to note in this respect the difference between the application of general theoretical principles, and the adaption of a specific description (of, for example, language or languages): as O'Toole has shown, it is the former (social semiotic theory) that drives the development of descriptions of semiotic resources other than language, while the latter (linguistic descriptions) can nevertheless provide useful materials, if appropriately adapted, for such descriptions. But the same question can be applied to work within all relevant disciplines: determining the affordances and

constraints of theories, descriptions and methodology adapted and applied from existing disciplines and traditions is a crucial challenge for multimodal text analysts.

Other issues remain for the multimodal text analyst. For example, the issue of corpus constraint continues to challenge multimodal text analysts: gathering multimodal corpora has become increasingly difficult, not the least because of ethical concerns which become acute when audiovisual recording of naturally occurring discourse is available. There are also difficulties in terms of data collection (such as the recording process itself, often involving technical expertise not always at hand for multimodal analysts), and the dynamism and hyper-textuality of contemporary interactive digital media has compounded this issue. The issues of detail, scope and complexity also continue to bedevil multimodal analysts, an issue raised as early as the 1960 study of psychiatric interview 'The first five minutes' (Pittenger et al 1960), but becoming increasingly evident with the availability of sophisticated interactive digital software applications which provide platforms for an ever-more minute analysis of multimodal phenomena. As Halliday and Greaves (2008) point out, the human analyst can never be replaced by computer-based or other technology-based approaches, but such resources increase the power of our analytical reach (for example, via low-level feature extraction, mathematical processing, visualization techniques), presenting richer but more complex and detailed phenomena to analyse. A key issue for 21st century multimodal text analysis is to find principled ways of sorting through and making sense of such complexity and detail.

The applications and value of multimodal text analysis are immense. Multimodal communication is central to human existence, and yet such is the nature of multimodality – the joining of abstraction and material in semiosis – that this is an area of the natural world that still remains relatively underexplored, and not well understood, compared with the material plane. Jewitt (2006) and others (e.g. Lemke 1998; O’Halloran 2010; Unsworth 2008) have shown how important an understanding of multimodality is to the study of classroom discourse and education; the integration of knowledge and techniques from multimodal semiotics science promises new approaches to the development and study of computational science; and multimodal text analysis has been shown to be crucial to a consideration of a wide range of fields, (e.g. Jewitt 2009). The proliferation in forms of contemporary interactive digital media and the ubiquity of their use puts demands upon scholars of human communication to keep up with wider socio-cultural developments. What is required at this stage in the development of multimodal studies as a field is the sort of empiricism of extensive text analysis such as revolutionized the study of language during the 1960s and 1970s. To do this will inevitably require that analysts of multimodality learn to use (particularly contemporary digital) multimodal resources and techniques in order to appropriately deal with the natures of such media: to paraphrase Firth (1957), to turn multimodal language back on itself.

Cross References

SEE ALSO: Multimodality and Technology; Kress, Gunther; van Leeuwen, Theo; O’Toole, Michael

References

- Barthes, R. (1957/1972). *Mythologies* (A. Lavers, Trans.). London: Jonathon Cape.
- Firth, J. R. (1957). *Papers in linguistics: 1934 - 1951*. London: Oxford University Press.
- Halliday, M. A. K. (1985). *Spoken and written language*. Waurin Ponds, Victoria: Deakin University Press.
- Jewitt, C. (2006). *Technology, literacy and learning: A multimodal approach*. London: Routledge.
- Jewitt, C. (Ed.). (2009). *Handbook of multimodal analysis*. London: Routledge.
- Kress, G. and van Leeuwen T. (1996). *Reading images: The grammar of visual design*. London: Routledge.
- Lemke, J. L. (1998). Metamedia Literacy: Transforming Meanings and Media. In D. Reinking, L. Labbo, M. McKenna & R. Kiefer (Eds.), *Handbook of literacy and technology: Transformations in a post-typographic world* (pp. 283-301). Hillsdale, NJ: Erlbaum.
- Lemke, J. L. (2002). Travels in hypermodality. *Visual Communication*, 1(3), 299-325.
- Machin, D. (2009). Multimodality and Theories of the Visual. In C. Jewitt (Ed.), *Routledge handbook of multimodal analysis* (pp. 181-190). London and New York: Routledge.
- Ochs, E. (1979). Transcription as Theory. In E. Ochs & B. B. Schieffelin (Eds.), *Developmental pragmatics* (pp. 43-72). New York: Academic Press.
- O'Halloran, K. L. (in press a). Multimodal Discourse Analysis. In K. Hyland and B. Paltridge (eds), *Companion to discourse analysis*. London: Continuum.

- O'Halloran, K. L. (in press b). The Semantic Hyperspace: Accumulating Mathematical Knowledge across Semiotic Resources and Modes. In F. Christie & K. Maton (Eds.), *Disciplinarity: Functional linguistic and sociological perspectives*. London & New York: Continuum.
- O'Halloran, K. L., Tan, S., Smith, B. A., & Podlasov, A. (2010). Challenges in designing digital interfaces for the study of multimodal phenomena. *Information Design Journal*, 18:(1), 2-12.
- Pittenger, R. E., Hockett, C. F. and Danehy, J. J. (1960). *The first five minutes*. New York: Martineau.
- Rohlfing, K., Loehr, D., Duncan, S., Brown, A., Franklin, A., Kimbarra, I., et al. (2006). Comparison of Multimodal Annotation Tools - Workshop Report. *Online-Zeitschrift zur Verbalen Interaktion, Ausgabe 7*, 99-123.
- Unsworth, L. (2008). *Multimodal semiotics: Functional analysis in contexts of education*. London & New York: Continuum.