RESEARCH ARTICLE

Analysing multimodality in an interactive digital environment: software as a meta-semiotic tool

Bradley A. Smith[a]*, Sabine Tan[b], Alexey Podlasov[b] and Kay L. O'Halloran[b]

[a]*Independent researcher;* [b]*Multimodal Analysis Lab, Interactive and Digital Media Institute, National University of Singapore, Singapore*

The present paper discusses issues arising from the use of an interactive digital software tool to analyse multimodal communication. The focus is on the ways in which such technical resources and associated techniques enable the analyst of social semiosis to apply different types of analysis, and provide the site for critical reflection upon the results of such analyses. The aim is to present an argument for the use of an interactive digital software application as a meta-semiotic tool. Three major challenges for scholars engaged in multimodal social semiotics are addressed in pursuing this aim, specifically with reference to the development of digital interactive analytical resources. Firstly, scholars are faced with the task of not only accounting for an increasing range of semiotic resources, but also for the way different semiotic phenomena interact to produce meaning. Secondly, the multimodal semiotician has a broad range of analytical approaches potentially relevant to any task to draw upon. Thirdly, the nature of contemporary media forms offer challenges themselves in terms of access, analysis and presentation of analysis.

Keywords: interactive digital software; multimodality; multimodal semiotics; meta-semiosis

1. Studying multimodal semiosis: theoretical and technical challenges and resources

A major challenge for scholars concerned with multimodal semiotic phenomena – that is, semiotic acts and discourse involving the interaction of more than one semiotic resource such as language, gaze, gesture, vocal features, proxemics, graphic display, cinematography, page-layout, and so forth – has been to not only develop an account of the range of semiotic phenomena and their affordances discernable in multimodal texts, but also show how such phenomena (can and do) work together inter-semiotically to produce meanings within texts. Scholars drawing upon Halliday's social semiotic approach to language (for example, Halliday 1978), known as systemic functional linguistic theory (Halliday and Matthiessen 2004; Martin 1992), have pointed the way to a holistic account of multimodal semiosis by urging attention to sign-making as a motivated social process, rather than simply as a product and (isolated, self-contained) sign system, highlighting Hodge and Kress' (1988, 1) focus on the "functions and social uses of semiotic systems, the complex interrelations of semiotic systems in social practice" (cf. also Kress and van Leeuwen 1996, 2001; O'Halloran 1999, 2004; Baldry and Thibault 2006; Bateman 2008; Jewitt

---

2009; Lemke 2009). Such an approach to the study of multimodal discourse makes it possible to show how different semiotic resources collaborate to produce complex but coherent discourses within and as constitutive and creative of social contexts.

A similar challenge for the emerging field of multimodal semiotics is understanding and relating the multiplicity of analytical approaches that have been drawn upon for or are relevant to the development of a comprehensive and integrated account of multimodal discourse. Studies dealing with human language (spoken and written), gesture and other bodily display (in humans and animals), gaze, proxemics, static and dynamic visual and aural art, architecture, and so on, from within traditions and backgrounds as diverse as (the various schools and disciplines of) linguistics, mainstream semiotics, social semiotics, psychology, anthropology, animal behaviour, cultural studies, media studies, and theory, criticism and practice of visual and aural art (painting, photography, music, film, theatre, sculpture, architecture, etc.) are all potentially directly relevant to (any particular and the general) study of multimodality. Just as on the one hand a challenge for the study of multimodal semiotics is accounting for the particular characteristics, affordances, roles and relations of different semiotic resources within unified discourse acts, so another issue is identifying the affordances and constraints of different analytical perspectives and techniques (within and between particular theoretical traditions and frameworks) applicable to multimodal studies.

A third challenge for the study of the contemporary multimodal discourse is the nature of the media themselves through which such discourse is realized. Access to and analysis of the significant phenomena of, for example, dynamic audiovisual media (video and film), online virtual worlds and other (interactive) web-based media requires appropriately sophisticated tools and techniques that reflect the nature of and so attend adequately to the particular characteristics of such media and their texts. The predominantly page-based approaches evident in much of the discourse about and analysis of multimodal text constrain to a significant extent analysts' ability to adequately (re)present the often complex nature of and relations between the phenomena under study (O'Halloran, forthcoming). Such constraint can lead towards a tendency to theoretical discursiveness and abstraction with a lack of grounding in analysis of actual instances of text; or, otherwise, to an analytical focus on static visual media (see, for example, Bateman 2008; Machin 2007). Where detailed analysis of audiovisual data is presented (as in, for example, the work of Baldry and Thibault 2006; Bateman 2007), the difficulties both of the analysis process and the presentation of its findings are evident.

In the present paper we address these issues in the study of multimodal semiotics with respect to a project at the National University of Singapore[1] in which the authors are currently engaged, to develop and apply software resources designed to facilitate the study of multimodal discourse and its phenomena. We propose the contemporary interactive digital environment as a suitable one for relating both various phenomena contributing to semiosis within a multimodal text and various analyses of such (see Goldman, Dong, and Lansiquot 2009 for a similar perspective). Such software applications of course do already exist (e.g. ELAN, EXMARaLDA, ORION; see Rohlfing et al. 2006 for a comparison of multimodal annotation tools), but there is clearly a need for further development of more sophisticated and powerful technical resources adapted for and thus able to manage the study of complex, abstract semiotic phenomena. It is particularly important that such resources are adapted to the analytical concerns of social semioticians and others

applying and relating detailed high-level and low-level analyses of signs and their social formations and functions. Contemporary software resources developed predominantly for the physical sciences remain in general underexploited by and unadapted for semiotics science (O'Halloran, in press).

In O'Halloran et al. (2010) we focus on the software development itself, showing how different functionalities and configurations of analyses within the software analytical interface bring differing semiotic affordances and constraints in terms of the analytical and explorative processes. In this paper we explore the use of the software: the ways in which such technical resources and associated techniques not only enable the analyst of social semiosis to apply different types of analysis, but crucially provide the site for critical reflection upon the results of such analyses.

The aim here is therefore to present an argument for the use of an interactive digital software application as a meta-semiotic tool. In developing a multimodal analysis software tool we are aware that software resources themselves constitute a multimodal semiotic potential (see O'Halloran, in press), in that they expand the analysts repertoire for creating (multimodal) discourse about (multimodal) discourse. The interactive digital environment becomes a site for creating and representing analysis, and therefore for more consciously reflecting upon the results of one's analyses. Analysts are thus empowered to apply critical social semiotic perspectives to their own analysis, in a self-reflexive and recursive process of meta-semiosis. We thus consider social semiotic theory to offer an appropriately holistic framework for exploiting the technical potential of such contemporary interactive digital resources, as it offers a theoretical perspective on the process of the social formation of both signs and meta-semiotic discourse about signs.

In order to develop this argument we first present a brief overview of the crucial aspects of the software (see O'Halloran et al. 2010 for an extended discussion), still in its present early stage of development. We then illustrate how such software resources may facilitate a variety of analyses of a video text, and also assist in developing a holistic social semiotic view on the text. These then form the basis for our consideration of a critical social semiotic perspective on the interactive digital environment itself, as we discuss how such a perspective can assist in the development and particularly in the use of such software.

2. Affordances of interactive digital software for multimodal study

We see the software we are developing as having three major affordances for the study of multimodal discourse:

(1) access to the audio, visual, somatic, and other data, in a variety of formats;
(2) the capacity to annotate the data within the same environment in which one accesses the data, via an annotation interface, and store these analyses in a database; and, as a result,
(3) the capacity to retrieve, interrogate, present and share one's analyses in a variety of ways. Particularly important in this respect is the capacity to visualize/auralize analyses in a variety of relations to one another in the template interface.

These affordances will be further discussed in the next three subsections.

2.1 Accessing and exploring multimodal data

Contemporary digital software resources can provide multiple sites and viewpoints for the exploration of the text and its phenomena. This access is critical to the study of multimodal phenomena: Halliday (1985) has remarked on the importance of widely available sound recording technologies for the study of spoken discourse, particularly tape-recorders. We believe that the interactive digital resources available today are important in a similar way to the study of multimodal discourse. For example, an interactive digital interface allows one to study the minutiae of multimodal discourse by accessing the source data through: a *filmstrip*, which decomposes a video into a series of frames/stills; a *single mobile/dynamic frame viewer*, which shows the particular frame at a location in time within a video according to where the user points the mouse; a *movie viewer window*, by which one plays a video with sound; *extracted static frames and overlay editor*, which allow one close-up viewing of individual frames; and *sound playback* (through, of course, headphones/speakers) or visual display such as frequency-intensity or pitch graph (see Figure 1 numeric annotations).

This access to the data in itself, although hardly novel, is significant to the study of multimodal texts in that it encourages repeated and detailed "close-up" investigation of multimedia data from a variety of perspectives, and in this sense any increase in sophistication of access resources is important. Functionalities to support this major affordance include: slow-motion and reverse playback; the ability to track across the text quickly using the dynamic single-frame viewer, which can be "grabbed" and moved along the timeline, with the time-relevant frames appearing successively as it is moved; the ability to extract frames, from the window viewer, filmstrip or dynamic single frame viewer, at any point, which can then be studied, as a static image, in large-screen mode or inserted in an analytical strip to provide reference points for particular analyses along the timeline; the ability to view the data within specific time-stamps via selected node or drag-and-select functions; a range of sound playback facilities commonly associated with sound access such as pause, looping, equalizer/sliders, and a suitable interface window design specially adapted for sound; and the application of various representations of audiovisual and somatic source text data, such as, for example, waveform and spectrograph representations of sound and pitch (F_0) extraction, and visual and aural filters (e.g. pitch or frequency bandwidth or window length for sound) that manipulate the way in which the signal is accessed by the user.

2.2 Annotation/analysis of data

Two major advantages afforded by digital technology are the capacity to analyse, via a variety of annotations, within the same environment as the text one is studying, and to store such annotations for later retrieval, interrogation, display and export. These annotations can be human-generated or computer-generated. Human-generated analyses are written text and graphic overlays, or categorical and gradient systemic analyses applied by "point-and-click" and "slider" interface resources; computer-generated analyses can be solely automated, as in video, shot detection, optical character recognition and gesture recognition, or semi-automated, requiring human checking and editing, such as for speech recognition annotations (see Figure 2).
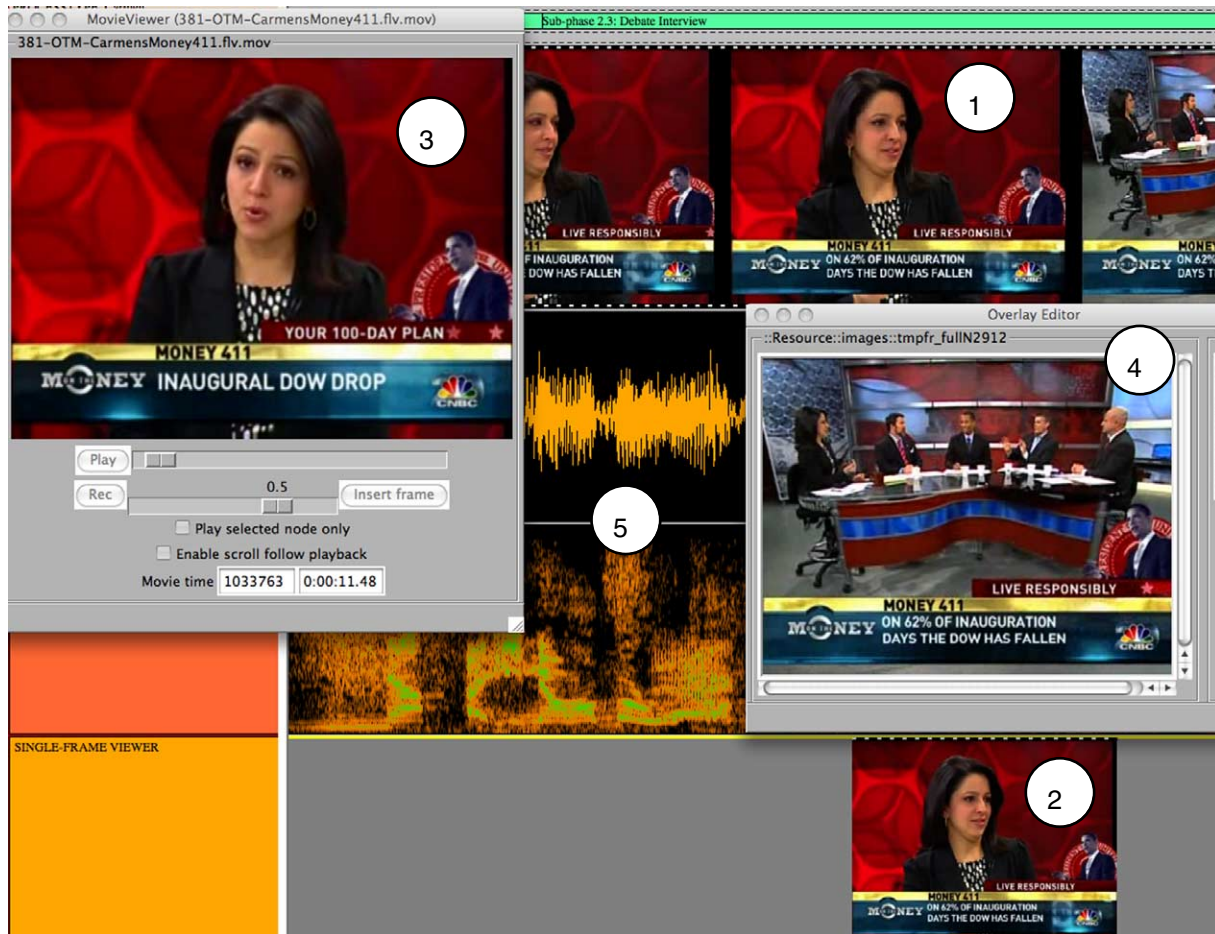
Figure 1.   Accessing data: (1) filmstrip; (2) single mobile/dynamic frame viewer; (3) movie viewer window (4); extracted static frames and overlay editor; (5) sound playback.
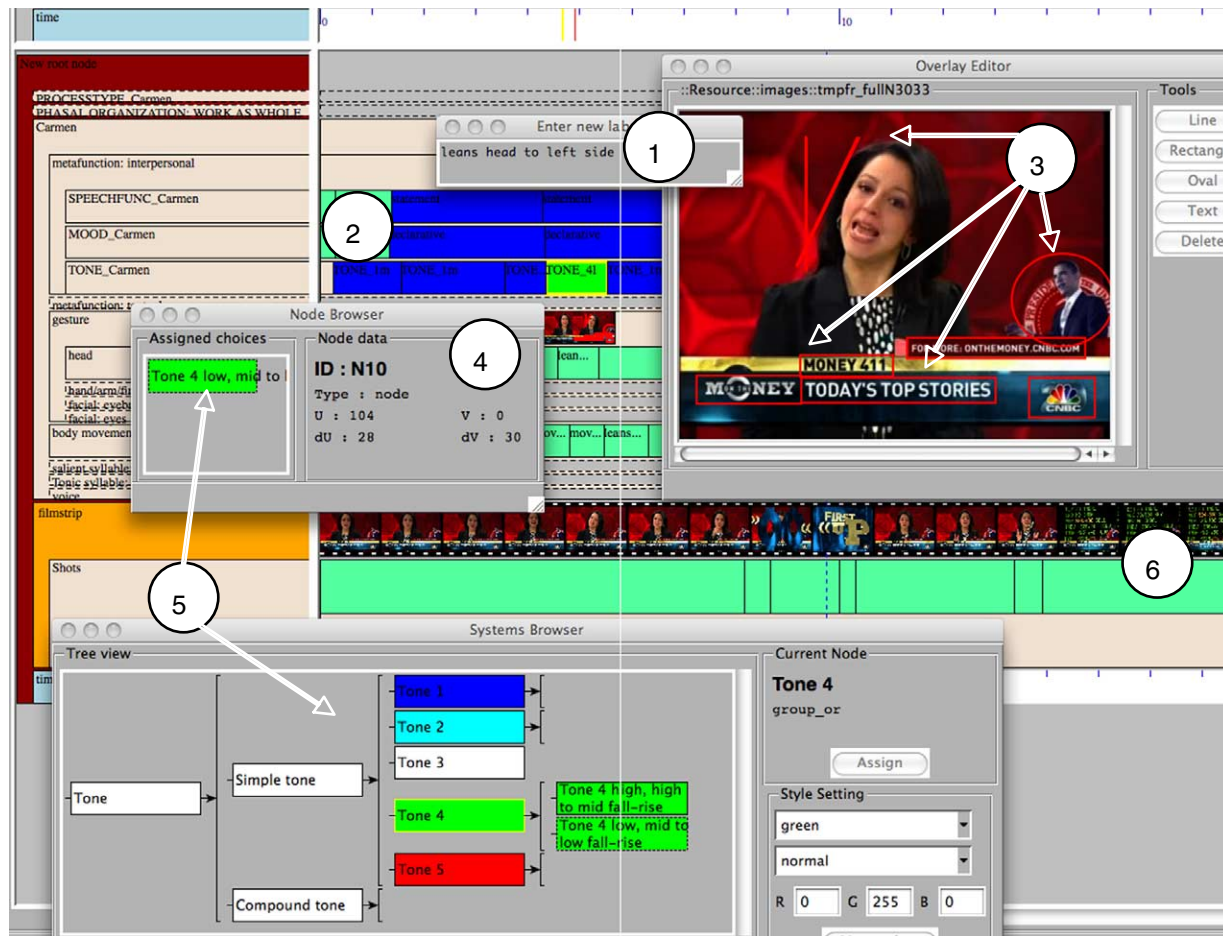
Figure 2. Manual, semi-automated and automated analyses: (1) type-in text annotation; (2) annotation nodes; (3) graphical annotations; (4) node browser: categorical choice; (5) systems browser: categorical one-click analysis; (6) shot detection.

Human-generated annotations are made within a node aligned with a particular part of a (dynamic or static) text, either (depending on the type of text) in annotation strips, configured in relation to one another in the interface within an analysis template, or in an "Overlay Editor" window (see Figures 2 and 4, for example). Categorical (systemic) and gradient/parametric annotations (see van Leeuwen 2009) are stored in a database for further processing (see Section 2.3). Computer-generated analyses can be performed on the data, yielding results that may not readily be accessible for the human analyst, particularly across large datasets. This type of analysis is usually of low-level features; challenges to semiotics science are to on the one hand find ways to exploit such existing techniques in semiotic terms, while on the other hand feeding back to computer scientists' ideas for new types of automated analyses that may yield useful results for semiotics, and for computational science. Importantly, in this respect, multimodal analysis software allows for the comparison of low-level, computational analyses with higher-level, human-generated analyses.

2.3 Search, retrieval, interrogation, visualization/auralization and exporting of annotation/analysis

The advantage of systemic annotation over written and graphic annotation is that each annotation from a system is stored as a unique value in the database, which means that it can be retrieved and visualized/auralized, interrogated and compared in relation to all other selections within this or any other analysis stored in the relational database, and shared/exported. The visualization/auralization component of the software enables the user to (re)present the various user-generated and automated analyses stored in the database as a variety of two-dimensional and three-dimensional visual and aural metaphors, with different analyses configured in relation to one another according to the analyst's wishes in the interface template. This component is yet to be developed in our current state of development of the software, except for the application of colour and border styles to options in a categorical analysis; but with these already one can illustrate some of the semiotic potential of visualization resources.

3. Multimodal analysis of a video text

In Section 2 we reviewed three major affordances of the software we are developing. In this section we will present an analysis that we have conducted, using the software, of a particular text. Although the translation from the digital software environment to the printed page brings with it constraints in terms of our capacity to illustrate the use of the software, we will show how different types of analysis can be applied and related within the software interface environment, and discuss issues arising from the interactive digital study of the multimodal discourse in question.

Our sample data are drawn from the discourse domain of broadcast business news. It is an extract from a video entitled "Carmen's Money 411", which is part of a one-hour personal finance programme *On the Money*, aired daily by CNBC weeknights at 10:00 pm eastern time. The video clip in question was first streamed on CNBC.com on 21 January 2009; that is, the day after US President Obama's inauguration.[2] Drawing on the classifications established by Montgomery (2007) and Clayman and Heritage (2002), the discourse-type or genre represented by the news

video under investigation is a debate interview with certified experts, which is commonly used to elucidate the event or topic under discussion by providing background information, spelling out the implications of a development, or providing independent comment from different perspectives, including ideological viewpoints. However, our text is not the typical formal debate interview that viewers may be familiar with from mainstream newscasting: it is, as we will show through various analyses and their integration within a social semiotic perspective, essentially a hybrid that incorporates elements of conventional news presentation (see Montgomery 2007; Clayman and Heritage 2002), conversationalization and informalization (see Fairclough 1995), which betrays the exercising of a definite editorial ideological stance despite having the appearance of genuine open debate.

The immediate research questions that we are concerned with here are: to what extent is the text informalized; what (interactions of) semiotic resources and discourse structures are drawn upon in its creation; and what ideological representations can be discerned and by what means are they created? According to Fairclough (1995, 45–45), ideological representations are generally implicit rather than explicit in media texts: ideologies are usually not "adopted" but taken for granted without recourse to overt rhetorical devices. We will show, however, how a detailed multimodal analysis can reveal higher-level ideological patterning realized across different semiotic resources in the text. We will also briefly analyse how this correlates with instantiations of viewer engagement, as realized by certain semiotic modes and resources in the news video, as well as explore other analytical perspectives on the spoken discourse and other semiotic phenomena that together contribute to the construction of the text and its meaning(s).

We begin with the observation that text's overall thematic development unfolds in periodic, wave-like stages or phases, which are instantiated multimodally and multiplicatively. Phasal shifts may, according to Thibault (2000), be indicated by a change, a break, or a pause in the rhythm of movement (in our text, of the camera, dynamic graphic displays and/or the human body) or by cutting between shots; or they may be signalled by a corresponding shift in the visual or linguistic thematics or the interpersonal orientation (see Lemke 1988).

The notion of information unfolding in periodic wave-like phases derives from Halliday's (1994) discussion of a clause in its character as a message. Part of the meaning of every clause lies in the part of the message that is chosen as Theme, which Halliday defines as "that with which the clause is concerned" (1994, 37), or the point of departure of the message. Information structure is generated by the interplay between Theme, and what Halliday terms "New" information, which is the structural complement of Theme, characterized by Halliday (1994, 298) as "information that is presented by the speaker as . . . not recoverable (New) to the listener". This textual structure in the clause manifests as a wave-like move from thematic prominence at the beginning of the unit to new informational prominence at the end.

Martin (1992) proposes that a similar patterning can be observed across a text, involving larger scale phases of discourse. In a paragraph, for example, a topic sentence can seen as functioning as a higher-level Theme – a hyper-Theme; with a hyper-New sentence functioning in the same, complementary way as New within the clause. Global textual patterns of Theme and New are termed macro-Theme and macro-New. These higher-order Themes and News help establish expectations about what the text is all about, how it will unfold, and the "point" of a text (see Martin

1992; Martin and Rose 2007). Analysis within the interactive digital software environment enables us to identify and visualize multiple layers of thematic development as they unfold in relation to one another and across the video text, affording us an enhanced perspective on ideological patterning as realized across different semiotic resources and at different levels of analysis.

In the video under investigation, the text's overall thematic structure appears to be organized around the day's main events; namely, the Obama inauguration. This overall macro-Theme is presented verbally by the show host at the onset of the programme, providing the overall thematic frame for the segment that is to follow, and visually through the image or emblem of President Obama which stays fixed on the screen throughout the entire duration of the programme (see Figure 2, window (3)). The Obama inauguration can thus be seen as representing the programme's highest-order Theme, providing the frame within which the programme is to be interpreted. Further macro-Themes and hyper-Themes are realized in the visual representations in the "lower third" section of the screen; that is, the text and images presented in the fixed or running bars (see Figure 2, window (3)), whereby their hierarchical order is also signalled typographically: for example, higher-order macro-Themes appear to be accorded a larger font size than lower order hyper-Themes.

While in page-based analysis the onus will be on the analyst to conceptualize this hierarchical order textually or typographically for the reader (see Figure 3), a digital platform allows us to approach this task interactively, multimodally and topographically by drawing on vertically ordered annotation strips within the analysis interface template, in addition to horizontally time-aligned colour-coded annotation "nodes", which together represent precisely how the thematic choices unfold sequentially along the timeline and in relation to one another (see Figure 4, window (1)).

In the video under investigation, the multimodal and inter-semiotic realizations of the text's overall thematic structure can perhaps be best observed at points of

Overall Theme: THE OBAMA INAUGURATION

⇨ Macro Theme 1: INAUGURAL DOW DROP

   ↜ Hyper Theme 1a: THE DOW HAD BIGGEST INAUGURATION DROP FALLING MORE THAN 300 POINTS, OR 4%

   ↜ Hyper Theme 1b: ON 62% OF INAUGURATION DAYS THE DOW HAS FALLEN

⇨ Macro Theme 2: THE $150M INAUGURAL

   ↜ Hyper Theme 2a: OBAMA'S INAUGURATION PROJECTED TO COST $150–$170 MILLION

   ↜ Hyper Theme 2b: MOST OF INAUGURATION COSTS ARE DUE TO INCREASED SECURITY

   ↜ Hyper Theme 2c: OBAMA RAISED OVER $41 M IN PRIVATE INAUGURATION DONATIONS

   ↜ Hyper Theme 2d: OBAMA SET A $50,000 LIMIT FOR PRIVATE INAUGURATION DONATIONS

⇨ Macro Theme 3: MERCHANDISING OBAMA

   ↜ Hyper Theme 3a: OBAMA FRANCHISE ESTIMATED TO BE WORTH $2.5 BILLION IN '09 SALES

   ↜ Hyper Theme 3b: 2009 PROJECTED OBAMA-RELATED FOOD SALES RANGE FROM $100–$200M

   ↜ Hyper Theme 3c: THE TOURISM INDUSTRY WILL SEE $300M IN '09 SALES DUE TO OBAMA

   ↜ Hyper Theme 3d: OBAMA MERCHANDISE PROJECTED TO GENERATE $2B IN '09 SALES

Figure 3.   Page-based representation of macro-thematic and hyper-thematic development.

Figure 4.   Themes and their expression: (1) horizontally time-aligned annotation nodes; (2) Template 1.

transition. For example, the introduction of the macro-Theme "TODAY'S TOP STORIES" at the beginning of the text is realized through both on-screen graphic display (see Figure 2, "Overlay Editor" annotation) and speech, "*where we give you the money download on today's top stories*", with a marked circumstantial Theme introducing the actual themes to be explored, "*and today, it's all about the inauguration and your tax-payer dollars*". All of this is executed in direct address to camera, and, by extension, the viewer. These themes are assigned further textual highlighting via the combination of a pronounced gesture on the part of the host, Carmen – a marked tilting of her head to one side as she says "*and today*" – and a separate tone group being assigned to this marked circumstantial Theme, with an exaggerated "tone 4" (fall–rise pitch contour) choice from the tone system of English encoding subordinate status to this information unit, a sense of "more to come" (see Halliday and Greaves 2008). Meanwhile, a choice of marked New information status to the personal pronoun (addressee) "*your*" is also accompanied by a pointing gesture by the host straight at camera, that is, to the projected viewer (see Figure 4, "MovieViewer"). These marked language choices are then succeeded by a highly dynamic graphic display of the written text "FIRST UP" (see Figure 4, "Frame Viewer"), employing various visual semiotic resources such as a dynamic revolving, appearing and disappearing text display, gold colour and large three-dimensional font size and type, and a dynamic "breaking up" of the text as it disappears. In combination, that is, several semiotic resources co-contextualize and add textual emphasis to the higher-order themes around which the discourse to come (especially, the debate) is to be organized as coherent text.

We may further note at this point that higher-order themes are consistently co-contextualized visually (or rather tele-visually) in the text as experiential meaning through actuality footage. For example, at this point in the text the host next elaborates on the macro-Themes of "the inauguration and your tax-payer dollars" by reintroducing a hyper-Theme first introduced by Carmen "*at the top of the show*", that the Dow Jones index had fallen more than 300 points on inauguration day. Carmen now adds that this is "*the biggest inaugural day drop . . . in the history of the stock market*", again with marked use of intonation (several tone groups) to add textual focus to the elements of this nominal group; while the lower third on-screen text shows two messages in succession, "INAUGURAL DAY DROP" and "THE DOW HAD BIGGEST INAUGURATION DROP FALLING MORE THAN 300 POINTS, OR 4%". However, we note that at this point the Dow inauguration-day drop is also represented visually as a material, dynamic display through a televised replay showing the Dow figures actually dropping in visual "continuous present" time (see Figure 4, "Overlay Editor"). The effect of this dynamic actuality footage is, of course, different from that of the representations just mentioned above, or for example from that which would have been provided by an even more abstract chart representation. The meaning, realized multimodally, is not simply "the Dow has dropped" but also "the Dow is dropping", with an orientation to the actual and dynamic rather than the abstract and static. The choice of expressive medium, in this case, is itself significant to an understanding of the meaning of the text.

While the Dow display continues, Carmen asks a question regarding this theme (the inaugural Dow drop), which may also be seen as part of the overall macro-Theme of "the inauguration and your tax-payer dollars": "*what do regular consumers take away from this?*". Crucially, in terms of our ideological analysis of the text, at

this point Carmen, thus specifically contextualizing the debate to immediately follow, adds her own answer to this rhetorical question, thereby clearly stating her own position on the issue she has just raised: "*Guys, I say nada, nothing*". This move can in fact be seen as a macro-New for the text as a whole, a message that is played out throughout the text in a variety of ways, by both Carmen and the senior programme editor Tyler Mathison, as well as in the on-screen graphics and other semiotic resources: that, regarding the macro-Theme of the Dow drop and its effect on the viewers' finances, viewers should "*not pay attention to the day-to-days*", and a general minimizing of the significance to viewers of this event (Tyler, for example, later makes reference to large Dow fluctuations over an extended time period). The lower third, it should be noted, during the visual display of the Dow drop also carries two messages in smaller-type in succession: "YOUR 100-DAY PLAN", and "LIVE RESPONSIBLY". The first can clearly be related to the macro-New just mentioned; the second is a more obscure, but the fact of its temporal relation to the other macro-Themes might suggest it is also indicating an editorial viewpoint: that viewers should pay no attention to the (daily) fluctuations of the stock market but instead should concentrate on their overall long-term financial (100-day) plan and fiscal responsibilities.

While the visual display of the Dow dropping is on-screen and Carmen is finishing this vocal turn, the host has meanwhile changed her gaze and postural direction so that she is now orientated towards her co-participants. This indicates both a shift in phase, from the introduction/macro-Theme orientation phase to the debate phase proper, and in her interpersonal orientation, from the viewer to her co-interactants:[3] with this move she confronts and challenges her co-interactants with her own viewpoint on the issue she has just raised for debate. We may also note that the introduction of spoken macro-Themes is the prerogative of the show host: it is the anchor's responsibility to direct the discussion (Themes) and speaker turn-taking. This is done verbally as well as through body orientation and gaze, hand gestures and intonation, as when she uses a Vocative – "*Ulzheimer*" – with a rising tone, spoken with additional loudness, direct gaze and body leaning towards the addressee, to command a particular interactant to respond (see Figure 5).

While macro-Themes are always co-contextualized verbally by the show host as the agenda for discussion, the textual hyper-themes in the lower thirds may not necessarily be presented or taken up in the spoken discourse as topics for debate – this multimodal division of semiotic labour making the latter potential carriers of implicit ideological positioning by the programme editors. Participants who digress from the projected thematic frame, or the conventions of institutional power and authority generally adhered to in the news video under investigation, find their contributions either vigorously disputed, quickly dismissed, or ignored by their co-participants as well as the televisual apparatus (i.e. who is put "on camera"). One speaker, Jeffrey Sonn, infringes not one but multiple discourse practices as established by the televisual apparatus (see Caldwell 1995), thereby violating his role as expert interviewee: not only does he usurp Carmen's anchor role by selecting his own speaker turn via an interjection, overriding comments by both Carmen and the senior business editor Tyler Mathison, he also introduces his own agenda through his statements (see Figure 6: "*You can't be surprised about this. Today Obama said we're gonna punish the greed on Wall Street*"), and does so in direct address to the camera (see Figure 6, window (3)).

Figure 5.   Multimodal command – "*Ulzheimer!*": (1) Template 2.

Figure 6. Convention and transgression: (1) convention: institutional authority; (2) convention: expert interviewee; (3) transgression: expert interviewee.

As observed by Budd, Craig, and Steinman (1999, 125), in televisual news discourse direct address signifies institutional authority: only the anchor, and correspondents introduced by the anchor are allowed to address the audience directly (in this case, Carmen and the business editor, Tyler; see Figure 6, window (1)), either by gaze or body postural orientation; all others either are told not to look at the camera, a convention that is scrupulously observed by the other expert interviewees in this text – even if framed by the camera frontally in a "head-on" shot (see Figure 6, window (2)) – or they appear in indirect address as if the camera was invisible. This, of course, has further implications for audience address. As Masterman elaborates:

> as the news opens, we are addressed by a newsreader who looks directly at the camera and delivers "the facts". Each viewer is given the role of direct addressee. We cut to a filmed interview. Our position changes. We are no longer directly addressed, but eavesdrop, watch and judge. (Masterman 1985, 229–230 cited in Ellsworth 1997, 25)

The ways audiences are positioned and expected to "experience" news and discussion are thus enacted through the conventions of the televisual apparatus, which becomes a semiotic resource for the production team to position the viewers in certain social relations to the interactants and to the (multimodal) discourse being presented in the video text.

The preceding discussion raises the issues mentioned earlier: the extent to which the debate is informalized, and by implication whether an open exchange of viewpoints is being engendered or in fact a particular ideological position is being constructed by the editorial team, host, senior business editor and production team. Although the discussion has at certain points the appearance of a genuine and somewhat fierce debate of differing viewpoints, Carmen and the show's editorial team effectively control the debate (including speaker turns, hyper/macro-Theme and macro-New development, camera positioning, etc.), including how the issues being discussed are to be interpreted by the viewer as well as the other interactants: whether to take seriously or even attend to a particular viewpoint, which Themes are to be taken up, and indeed which conclusions to draw from the issues under discussion. It is therefore not surprising that Sonn's indiscretion does not draw the expected response of acknowledgement from his co-participants but rather the discretionary alternative, contradiction (see Halliday, 1994) – even though his style of delivery mimics the type commonly used by anchors for news presentation (see Tolson 2006; Ellis 2000).

However, it is also an issue whether such transgressions as practiced by Sonn in the example above are in fact acceptable or not to the programme editors or the host. It could be argued that such transgression is allowed and even encouraged to some extent by the anchor, in a controlled fashion, as part of the spectacle of this debate. This would certainly contribute to the text appearing to be an authentic debate, important if there is in fact a particular covert editorial bias being constructed. At several points in the debate, Carmen seems happy to let the experts argue over the top of one another in what seems like a heated and uncontrolled interaction, and sometimes in fact has trouble regaining control of the discourse – at one point, after a failed attempt at topic-shift during a heated exchange between two experts, addressing the viewer with the resigned and ironic exclamation: "*ahhh love it*". But in

shifting the discourse forward onto new ground each time not only does Carmen reassert her anchor role but, ultimately, hers is the final word on each issue being hotly contested (see Figure 7, overlaps and Carmen re-taking the floor): as in the discussion about the appropriateness of the high cost of Obama's inauguration, where Carmen clearly indicates her view that it is "*money well spent*" before moving on the new topic of Obama's ("considerable", in terms of the framing by Carmen) worth to the economy.

On the basis of this short analysis, we can conclude that although the vigorous and occasionally rancorous debate seems chaotic and more alike to informal contexts – such as after-work discussions in a public bar or at a dinner party – the programme in fact follows a formal and controlled structure, both in terms of thematic development and viewer engagement, both of which are closely monitored and policed by the anchor and editorial team. Manifestations of informalization, realized predominantly through verbal discourse (including vocal resources such as loudness, pitch height, etc.), speaker turn assignment and overlap, address to viewer and instances of dynamic graphic displays, mask a range of semiotic strategies by which the segment's creators produce what is in fact a highly controlled text that exhibits strong underlying coherence in terms of its ideological orientation. The differing views represented by each of the experts in fact do little to disturb the editorial thread running through the text, the "take-home message" of the segment, which is that viewers should disregard daily stock market fluctuations and concentrate on their own "100-day plan".

4. Multimodal discourse and meta-discourse in the digital environment

In the previous section we have tried, within the constraints of the printed page, to show how interactive software is a resource for analysis not only of lower-level expression plane phenomena such as graphic display and intonation, as well as written text and speech, but also of higher-level meanings such as the thematic and
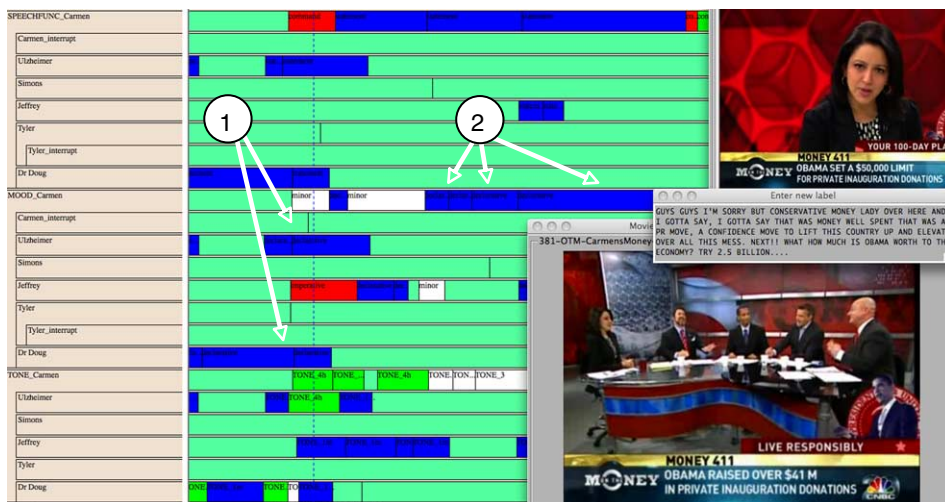
Figure 7.   Controlling the discourse: (1) overlap; (2) Carmen retaking the floor.

phasal organization of the text as a whole. We showed how such phenomena at different levels (in systemic functional theory, "strata") can be related to one another in the analysis of an overarching editorial bias within the text, as realized through/evidenced in such phenomena. Software tools such as we are developing are resources for studying the characteristics and functions of an array of semiotic resources within multimodal discourse, the ways they work together, and for linking highly abstract and critically oriented interpretations of social semiotic phenomena to a grounding in empirical discourse analysis. The software brings together within the same environment various types of analysis with the source texts which those analyses seek to study and explain, making it a powerful resource for relating theory and its application within actual analytical tasks.

Users of the software are encouraged to be explicit about the analytical tools they employ. Within the interactive digital environment, although a discursive (manual) analysis may also be conducted, the distinctive affordance is the capacity to conduct categorical and parametric analyses, the results of which are stored in a database and thus capable of being interrogated and visualized, shared and compared. As such, the digital environment is a resource for exploring, testing, comparing and calibrating multiple analytical perspectives with respect to each other and to data. The user may develop a variety of perspectives through the design of different template configurations that correlate different analytical views on the source text in different ways; and these representations of analyses, as configured in the interface template, may become themselves the site of further investigation as patterns and relations between different analyses are revealed in the interface representation.

This crucial aspect of analysis within the interactive digital environment has been recognized by Goldman, Dong, and Lansiquot (2009, 32), who identify the need for a framework that "addresses the need for interacting with the artifacts or representations that are continually being created in the process of communication about meaning". The "perspectivity framework" that they propose:

> acts as a conceptual scaffold to address the journey from bits and segments (video-data-in-the-small) into meaningful stories and valid results (video-interpretations-in-the-large) . . . [and] is open, flexible, and inclusive of the diverse theories and methodological approaches that have emerged (and will continue to emerge) as researchers use digital video into their research. (Goldman, Dong, and Lansiquot 2009, 32)

While for Goldman, Dong, and Lansiquot the concern is with developing such a conceptual framework for conducting such a meta-semiotic process, our concern is with the development and use of software resources themselves: in particular in the present paper, the way in which interactive digital resources may in general terms facilitate the development of multiple such frameworks and analytical perspectives.

For example, our earlier discussion was derived primarily from the analyses, initially conducted independently, by two of the present authors; analyses that were then compared within a single interface template design (compare, for example, Templates 1 and 2 in Figures 4 and 5). These analyses drew on manual textual, graphic and systemic analyses, and automated analysis techniques. In terms of the latter, the shot detection algorithm, for example, revealed cinematographic phenomena – for example, points in the text at which there were few shot changes as compared with points where there were several – that could be correlated with other

potentially automated analyses (e.g. speaker turns, camera positioning) and to higher-level social semiotic aspects of the text: revealing, for example, that the anchor or senior editor was often afforded unbroken shot continuity as compared with the rapid transitions during the debate phases, were almost always afforded on-camera status if talking, and were usually presented in close-up as compared with the expert panel who were more often pictured at a greater "social distance" to the viewer. It is feasible that correlation of such analyses with manually generated analysis may inform the development and application of more sophisticated automated analyses (e.g. face and pitch detection, event recognition) across larger multimodal corpora.

In terms of manual analysis, one author focused on the semiotics of the visual display: on-screen text, images and dynamic graphic displays, in terms of the thematic and phasal structuring of the text. The purpose here was to explore the thematic design of the text with reference to particular semiotic resources, applying also meta-functional perspectives to these analyses. The second analyst focused primarily on spoken discourse, particularly the inter-stratal relations of speech function, mood and tone systems within language (i.e. interpersonal aspects of the discourse), vocal and gestural phenomena, and turn-taking strategies. There were areas of overlap in these different analyses – for example, Theme as realized through visual and aural semiotic resources (e.g. on-screen text, icons, and spoken language) – as well as areas of complementarity between them.

What was fascinating to see, once the separately-conducted analyses were shared and compared within the interface, was not only how both different aspects of the text were revealed, but also how the various analytical views showed a consistency in terms of the higher-level ideological bias evident throughout the text. The many semiotic resources we studied all seemed to play a concerted role in shaping the way in which the viewer was being positioned with respect to the issues under discussion, as the discussion in Section 3 illustrates. This orchestration effect of different semiotic resources working together to make meaning within multimodal discourse is of course a finding that in itself is not novel. What is exciting to us is that the software resources we are, in the early twenty-first century, able to employ made a significant contribution to exploring through empirical analysis what was for us an early intuitive reading of an editorial bias in the text, deriving from our training in social semiotics, and relating this to multiple lower-level and higher-level analytical evidences in a way that is capable of being (re)presented as visual patterns and relations between patterns within the software interface (note: such visualizations may also reveal help "outliers", significant instances that depart from patterns).

This affordance is illustrated, for example, in the visualization via nodes of turn-taking analysis for the different speakers (e.g. Figure 7), revealing patterns of speaker turn and overlap, which patterns can be correlated with other significant analyses (e.g. tone choices), in a subsequent study of the visualizations of existing analyses. It can also be used to illustrate relations between phenomena as at different levels of analysis; for example, the division into clauses in relation to the division into tone groups (see Halliday and Greaves 2008); choices in mood to choices in tone; and how these relate to speech function choices (e.g. a statement and/or clause distributed over several tone groups; minor clause vocative with a rising tone as command – "*Ulzheimer!*"). We are also afforded a view on the relations of the various visual semiotics resources to each other, and to their semantic functions as the co-realization of (often implicit) meanings. The various template configurations and the

analytical views they afford reveal and relate different views upon and phenomena within the text, which views then become the site for higher-level analysis, in our earlier discussion, for example, of an over-arching ideology reflecting editorial bias.

The storage of the results of one's analyses within a database structure also means that we may later develop more sophisticated ways of visualizing analyses: both in terms of the visualizations themselves (e.g. three-dimensional, animation, etc.), and in terms of the processing (search, retrieval, mathematical processing, etc.) that can be performed on the analyses, the results of which may then themselves be visualized. We intend also to make use also of auralization in this respect. That is, instead of using colour, for example, to represent systemic choices in the visual interface, the analyst would hear some aural phenomenon, such as the tone system of English (falling, rising, etc.), as the representation of systemic choices for any analysis or secondary processing of such.

In this respect the social semiotic approach, which has been shown to be so powerful in accounting for sign-making processes and their socially-constructed nature, is also relevant to the ways in which we exploit the potential of such software resources for producing and, importantly, critiquing semiotics (meta-)discourse. The software environment is a resource for analysing, comparing and discoursing about our own (meta-)semiotic practices, encouraging critical self-reflexion – an active application of one's analytical process back upon one's own analyses. The fact of being able to store, retrieve, share, interrogate and represent in a variety of ways (visualizations and auralizations) the results of one's analysis means that a semiotician can conduct a variety of analyses, and then explore the range of such analyses as, in effect, secondary (meta-)texts created by the configuration of such analyses within a particular template design. One's own abstract analyses, given visual or aural expression through the affordances of contemporary computer techniques, thus become texts, phenomena, for further analysis, an analytical recursiveness with great potential power. This is the sense in which we use the term "meta-semiotic tool" to describe our software. Thus the same social semiotic principles we apply to any discourse can also apply to our own, multimodal discourse within the interactive digital medium.

It is important in this respect to highlight an inherent danger that this affordance brings. Although any analysis in fact creates a new meta-text as commentary upon its object text, within the interactive digital environment the capacity to multiply, recursively, analysis upon analysis creates the potential to reify one's own interpretation, treating it as though it was the source text itself. However, the capacity to analyse within the environment of the source text – something difficult or impossible within a page-based analysis of a video text, for example – means that the analyst is always encouraged to return to the source text, as it remains immediately accessible within the software interface environment. The capacity to create multiple analyses also makes more conscious one's choice of analytical approach, and the capacity to visualize one's analytical results also encourages the analyst towards critical self-reflection. Different analyses and perspectives upon analysis are encouraged, so that an analyst may produce multiple interpretations of a text. At all points the analyst is reminded of the relations of data and analytical interpretation, and of the status of their own interpretations. The presentation, comparison and sharing of (digital) analyses (see Goldman, Dong, and Lansiquot

2009 on the use of such resources for social action) thus encourages the analyst to detach from any particular analytical view.

Contemporary cultures (Kress and van Leeuwen 2001, 1) "have begun to use an increasing variety of materials and to cross boundaries between the various art, design and performance disciplines, towards multimodal Gesamtkunstwerke, multimedia events and so on" (what Jewitt 2009 discusses as the "turn to the multimodal"), requiring for semiotics (Kress and van Leeuwen 2001, 1) "a theoretical framework applicable to all semiotic modes" and capable of offering an adequate account of multimodal semiosis. It is clear that we need to develop the sort of multimodal approach to analysing contemporary discourse phenomena as is exhibited in the texts we study: that the means of study must be suited to the object of study. It may be some time before publishing conventions and practical constraints (technological – e.g. publishing excerpts from dynamic audiovisual data as well as their digital analysis – and institutional – the latter involving, for example, issues of intellectual property and privacy constraints on the reproduction of source data) catch up with multimodal practices within the wider culture, but the more such analyses are conducted the greater pressure there will be to provide the forum within which such studies may be appropriately presented.[4]

As an appropriate response to such shifts within contemporary culture, interactive digital semiotics offers the potential not only for dealing appropriately with contemporary multimodal social practices, but also for constructing a multiplicity of multimodally-constructed meta-discourses and integrating these within specific analytical tasks. The application and correlation of a variety of analytical and theoretical views may lead towards richer, more heteroglossic and ultimately holistic perspectives upon the political economy of contemporary social semiosic and semiotic practices.

Notes

1. The research for this article was supported by the research grant Events in the World: Developing and Using Interactive Digital Media for Multimodal Discourse Analysis (MDA) (Project No. NRF2007IDM-IDM002-066, Principal Investigator Kay L. O'Halloran) awarded by the Interactive Digital Media Program Office under the National Research Foundation Singapore. For further information, please see the Multimodal Analysis Laboratory website (http://multimodal-analysis-lab.org) or contact Kay O'Halloran (kay.ohalloran@nus.edu.sg).
2. http://www.cnbc.com/id/15840232?video=1007243273.
3. Textually, we may also note, at the point that Carmen wraps up this introductory monologic turn – "*guys, I say nada, nothing*" – the melodic aspects of her speech: the syllables of this utterance form a melodic progression from a B note on "*guys*", to two D notes of half the temporal value of the B on "*I say*", and then A and E notes also of half the temporal value of the B for "*nada*", repeated for "*nothing*". What does this mean? Further research is needed to explore this interesting area, but any musician will see this as a "final cadence" in the key of E scale, from the Dominant fifth, up to the "unstable" seventh note, and down through the fourth resolving to the Tonic: an instantly recognizable

musical gesture of "finality" that fits nicely with this ending of the introductory phase of the discourse.

4. For example, *The Public Journal of Semiotics* [http://www.semiotics.ca/] is a free online journal that publishes research articles in domains relating to semiotics, and is available in a Flash version with video files embedded.

Notes on contributors

Bradley A. Smith is an Independent Researcher and former Research Fellow for the Multimodal Analysis Laboratory of the National University of Singapore. He is responsible for the integration of resources for the study of speech, music and other organized sound into the software applications being developed by the laboratory team. Bradley completed his Bachelor of Arts degree at Macquarie University, Sydney, Australia, where he graduated with 1st Class Honours in 2004. In 2008 he completed his PhD thesis, entitled "Intonation and register: A multidimensional exploration".

Sabine Tan is a PhD Research Scholar at the National University of Singapore, pursuing her PhD in English Language Studies within the Events in the World project at the Multimodal Analysis Laboratory. Sabine's primary research interests include social semiotics, visual communication, and multimodal discourse analysis. She is particularly interested in applications of systemic functional theory and its derivatives to the analysis of business discourses, corporate television advertisements, corporate web-pages, televisual and Internet-based news discourse, and other emergent multimodal discourse genres.

Alexey Podlasov is a Research Fellow at the Multimodal Analysis Laboratory primarily responsible for the technological development of the software application. Alexey graduated with MSc in Applied Mathematics from Saint-Petersburg State University, Saint-Petersburg, Russia in 2002, and MSc and PhD in Computer Science from the University of Joensuu, Joensuu, Finland in 2004 and 2008, respectively. His research interests are in image processing and compression, machine vision and multimodal analysis.

Kay L. O'Halloran is the Director of the Multimodal Analysis Laboratory, Interactive and Digital Media Institute and Associate Professor in the Department of English Language and Literature at the National University of Singapore. She is an internationally recognized scholar in the field of multimodal research, and she has published extensively on multimodal approaches to mathematics discourse and the use of digital technology for multimodal analysis.

References

Baldry, A.P., and P.J. Thibault. 2006. *Multimodal transcription and text analysis*. Oakville, CT: Equinox Publishing.

Bateman, J. 2007. Towards a grande paradigmatique of film: Christian Metz reloaded. *Semiotica* 167: 13–64.

Bateman, J. 2008. *Multimodality and genre: A foundation for the systemic analysis of multimodal documents*. Basingstoke: Palgrave Macmillan.

Budd, M., S. Craig, and C. Steinman. 1999. *Consuming environments: Television and commercial culture*. New Brunswick, NJ: Rutgers University Press.

Caldwell, J.T. 1995. *Televisuality: Style, crisis and authority in American television*. New Brunswick, NJ: Rutgers University Press.

Clayman, S., and J. Heritage. 2002. *The news interview: Journalists and public figures on the air*. New York: Cambridge University Press.

Ellis, J. 2000. *Seeing things: Television in the age of uncertainty*. London: I.B. Tauris.

Ellsworth, E. 1997. *Teaching positions: Difference, pedagogy and the power of address*. New York: Teachers College Press.

Fairclough, N. 1995. *Media discourse*. London: E. Arnold.

Goldman, R., C. Dong, and R. Lansiquot. 2009. Software design principles for video research in the learning sciences and CSCL: Two studies use the perspectivity framework and Orion™. In *Proceedings of the 9th international Conference on Computer Supported Collaborative Learning – Volume 2*, June 8–13, in Rhodes, Greece. http://portal.acm.org/citation.cfm?id=1599503.1599548.

Halliday, M.A.K. 1978. *Language as social semiotic*. London: Edward Arnold.

Halliday, M.A.K. 1985. *Spoken and written language*. Waurn Ponds, Victoria: Deakin University Press.

Halliday, M.A.K. 1994. *An introduction to functional grammar*, 2nd ed. London: Edward Arnold.

Halliday, M.A.K., and C.M.I.M. Matthiessen. 2004. *An introduction to functional grammar*, 3rd ed. London: Arnold.

Halliday, M.A.K., and W.S. Greaves. 2008. *Intonation in the grammar of English*. London: Equinox.

Hodge, R., and G. Kress. 1988. *Social semiotics*. Cambridge, UK: Polity Press in association with Oxford: Basil Blackwell.

Jewitt, C. 2009. *Handbook of multimodal analysis*. London: Routledge.

Kress, G., and T. van Leeuwen. 1996. *Reading images: The grammar of visual design*. London: Routledge.

Kress, G., and T. van Leeuwen. 2001. *Multimodal discourse: The modes and media of contemporary communication*. London: Arnold.

Lemke, J.L. 1988. Discourses in conflict: Heteroglossia and text semantics. In *Systemic functional approaches to discourse: Selected papers from the 12th International Systemic Workshop*, ed. J.D. Benson and W.D. Greaves, 29–50. Norwood, NJ: Ablex.

Lemke, J.L. 2009. Multimodal genres and transmedia traversals: Social semiotics and the political economy of the sign. *Semiotica* 173: 283–97.

Machin, D. 2007. *Introduction to multimodal analysis*. London, Hodder Arnold and New York: Oxford University Press.

Martin, J.R. 1992. *English text: System and structure*. Amsterdam: John Benjamins.

Martin, J.R., and D. Rose. 2007. *Working with discourse: Meaning beyond the clause*. London, New York: Continuum.

Masterman, L. 1985. *Teaching the media*. London: Comedia.

Montgomery, M. 2007. *The discourse of broadcast news: A linguistic approach*. New York: Routledge.

O'Halloran, K.L. 1999. Interdependence, interaction and metaphor in multisemiotic texts. *Social Semiotics* 9, no. 3: 317–54.

O'Halloran, K.L. 2004. Visual semiosis in film. In *Multimodal discourse analysis: Systemic-functional perspectives*, ed. K.L. O'Halloran, 109–30. London: Continuum.

O'Halloran, K.L. In press. Multimodal analysis and digital technology. In *Interdisciplinary approaches to multimodality: Theory and practice. readings in intersemiosis and multimedia*, ed. A. Baldry and E. Montagna. Palladino: Campobasso.

O'Halloran, K.L., S. Tan, B.A. Smith, and A. Podlasov. 2010. Challenges in designing digital interfaces for the study of multimodal phenomena. *Information Design Journal* 18, no. 1: 2–12.

Rohlfing, K., D. Loehr, S. Duncan, A. Brown, A. Franklin, I. Kimbarra, et al. 2006. Comparison of multimodal annotation tools – workshop report. *Online – Zeitschrift zur Verbalen Interaktion Ausgabe* 7: 99–123.

Thibault, P.J. 2000. The multimodal transcription of a television advertisement: Theory and practice. In *Multimodality and multimediality in the distance learning age: Papers in English linguistics*, ed. A. Baldry, 311–85. Campobasso, Italy: Palladino Editore.

Tolson, A. 2006. *Media talk: Spoken discourse on TV and radio*. Edinburgh: Edinburgh University Press.

van Leeuwen, T. 2009. Parametric systems: The case of voice quality. In *The Routledge handbook of multimodality*, ed. C. Jewitt, 68–77. London: Routledge.