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 etc - has been to not only develop an account of the range of semiotic phenomena and their affordances discernable in multimodal texts but to show how such phenomena (can and do) work together inter-semiotically to produce meanings within texts. Scholars drawing upon Halliday’s (e.g. 1978) social semiotic approach to language, 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 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. The term ‘orchestration of meaning’ is thus a relevant one for multimodal semiosis as seen within the holistic perspective afforded by social semiotic theory.

A similar challenge for the emerging field of multimodal semiotics is understanding and relating the multiplicity of theoretical, analytical and methodological approaches which 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 theoretical and 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 in press 2009). 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 (cf. 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.
There is clearly a need for a multimodal semiotics that draws upon the resources both of state-of-the-art digital multimedia technologies and theoretically informed descriptions to adequately manage the task of exploring the multimodal discourse practices of contemporary human culture. Contemporary software resources developed predominantly for the physical sciences remain in general underexploited by and unadapted for semiotics science (O’Halloran in press 2009). While some notable exceptions do exist (e.g. ELAN, EXMARaLDA, cf. Rohlfing 2006 for a comparison of multimodal annotation tools), there is clearly a need for the development of more sophisticated technical resources that are adapted for and thus able to manage the study of complex, abstract semiotic phenomena, and that are not only of use to the analytical concerns of social semiotics but adaptable to the application, testing and integration of a variety of analytical approaches from within different theoretical traditions and disciplines concerned with multimodal communication.

In the present paper we discuss issues in the study of multimodal semiotics with respect to a project at the National University of Singapore in which the authors are currently engaged, to develop a software application 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 multimodal semiosis and various analyses of such, and social semiotic theory as offering an appropriately holistic framework for exploiting this technical potential. In developing our multimodal analysis software we have found that different functionalities and designs of the analytical interface bring differing semiotic affordances and constraints in terms of the analytical and explorative processes (O’Halloran et al. in press 2010): that is, digital software resources themselves constitute a semiotic potential, potentially both enabling and constraining, for making (Firth 1968, 19) ‘statements of meaning’ about semiotic phenomena, and as such can be subjected to the application of the same social semiotic theoretical principles as have been applied to other discourse practices.

We will first discuss some of the features and functionalities of the software we are developing at its present early state of development; and then present an analysis of a video as an illustration of the way in which such software resources may be deployed in the study of multimodal text, from both a variety of analytical perspectives and from a holistic view that reveals aspects of interest to social semioticians. We consider, from the perspective of social semiotic theory, what it means to conduct analyses within the interactive digital environment, showing how a critical social perspective can assist in the development and 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;
3) as a result of 2), 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/auralise 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: (1) a filmstrip which decomposes a video into a series of frames/stills; (2) 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; (3) movie viewer window, by which one plays a video with sound; (4) extracted static frames and overlay editor, which allows one close-up viewing of individual frames; and (5) 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 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 (F₀) extraction, and visual and aural filters (e.g. pitch or frequency bandwidth or window length for sound) which 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- 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). 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 & 3, for example). Categorical (systemic) and gradient/parametric annotations (cf. van Leeuwen 2009) are stored in a database as numeric values 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 data sets. 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.
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.
Figure 3: Themes and their expression: (1) horizontally time-aligned annotation nodes; (2) template 1.
2.3 Search, retrieval, interrogation, visualisation/auralisation 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/auralised, 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/auralisation 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 2- and 3-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 for mapping the interactions of different systems within the same or different semiotic resources, and the unfolding selection patterns and departures from such within texts.

3. Multimodal analysis of a text

In Section 2 we reviewed three major affordances of the software we are developing. In this section we will present an analysis which 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 is 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 program ‘On the Money’, aired daily by CNBC on weeknights at 10 pm ET. The video clip in question was first streamed on CNBC.com on 21st January 2009, that is, the day after US President Obama’s inauguration [http://www.cnbc.com/id/15840232?video=1007243273]. 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 by 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 (cf. Montgomery, 2007; Clayman and Heritage, 2002), conversationalization and informalization (cf. Fairclough 1995), that 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 ideological patterning realized across different semiotic resources in the text. The analysis will draw upon Martin’s (2001, 2004) ‘method of development’ of Theme in discourse, according to which introductory paragraphs, titles, and subtitles, may function as macro-Themes for a text, its semantic structure, and which may, in turn, predict lower-order Hyper-Themes which contextualize particular paragraphs/sections of a text, and so on. We will also briefly analyse 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 (cf. Lemke 1988). In the video under investigation, the text’s overall thematic structure is 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 program, 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 program (see Figure 2 (3)).

Table 1. Page-based representation of macro- and hyper-thematic development

<table>
<thead>
<tr>
<th>Meta Theme: THE OBAMA INAUGURATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>➜ Macro Theme 1: INAUGURAL DOW DROP</td>
</tr>
<tr>
<td>⊳ Hyper Theme 1a: THE DOW HAD BIGGEST INAUGURATION DROP FALLING MORE THAN 300 POINTS, OR 4%</td>
</tr>
<tr>
<td>⊳ Hyper Theme 1b: ON 62% OF INAUGURATION DAYS THE DOW HAS FALLEN</td>
</tr>
<tr>
<td>➜ Macro Theme 2: THE $150M INAUGURAL</td>
</tr>
<tr>
<td>⊳ Hyper Theme 2a: OBAMA’S INAUGURATION PROJECTED TO COST $150-$170 MILLION</td>
</tr>
<tr>
<td>⊳ Hyper Theme 2b: MOST OF INAUGURATION COSTS ARE DUE TO INCREASED SECURITY</td>
</tr>
<tr>
<td>⊳ Hyper Theme 2c: OBAMA RAISED OVER $41 M IN PRIVATE INAUGURATION DONATIONS</td>
</tr>
<tr>
<td>⊳ Hyper Theme 2d: OBAMA SET A $50,000 LIMIT FOR PRIVATE INAUGURATION DONATIONS</td>
</tr>
<tr>
<td>➜ Macro Theme 3: MERCHANDISING OBAMA</td>
</tr>
<tr>
<td>⊳ Hyper Theme 3a: OBAMA FRANCHISE ESTIMATED TO BE WORTH $2.5 BILLION IN ’09 SALES</td>
</tr>
<tr>
<td>⊳ Hyper Theme 3b: 2009 PROJECTED OBAMA-RELATED FOOD SALES RANGE FROM $100-$200M</td>
</tr>
<tr>
<td>⊳ Hyper Theme 3c: THE TOURISM INDUSTRY WILL SEE $300M IN ’09 SALES DUE TO OBAMA</td>
</tr>
<tr>
<td>⊳ Hyper Theme 3d: OBAMA MERCHANDISE PROJECTED TO GENERATE $2B IN ’09 SALES</td>
</tr>
</tbody>
</table>

Drawing on the post-modernist concept of grand- or meta-narratives (cf. Hodgson, 2001; Knoespel, 1991), the Obama inauguration can be seen as representing the program’s higher-order grand- or meta-theme, which provides the context within which the program is to be interpreted by both the expert interviewees in the studio and the potential viewer. Further macro- 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 (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 Table 1), 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’, that together represent precisely how the thematic choices unfold sequentially along the timeline and in relation to one another (see Figure 3 (1)).

The multimodal and inter-semiotic realizations of the text’s overall thematic structure can perhaps be best observed at points of 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’ (cf. 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 3: ‘MovieViewer’). These marked language choices are then succeeded by a highly dynamic graphic display of the written text “FIRST UP” (see Figure 3: ‘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 three hundred 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%”. Drawing on van Leeuwen’s (2008) framework for the representation of social action, we can note that at this point the Dow inauguration-day drop is also represented visually as a material, non-transactive action through a televised replay of a display showing the Dow figures actually dropping in visual ‘continuous present’ time (see Figure 3: ‘Overlay Editor’). The effect of this dynamic actuality footage is, of course, different than that of the ‘live’ real-time inserts of the Dow that are commonly seen in business news broadcasts, a more abstract chart representation, or spoken discourse: the meaning is not ‘the Dow has dropped’ but ‘the Dow is dropping’, and an orientation to the actual and dynamic (and of course to the visual) rather than the abstract and static, in the graphic display used for the construal of this meaning. The choice of expressive medium, that is, is itself significant to an understanding of the meaning of the text (cf. Kress and van Leeuwen 1996 on visual modality).

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, in thus specifically contextualising 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 program 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 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: 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 4).

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’).
Figure 4: Multimodal command - “Ulzheimer!”: (1) template 2.
Figure 5: Convention and transgression: (1) convention: institutional authority; (2) convention: expert interviewee; (3) transgression: expert interviewee
One speaker, Jeffrey Sonn, infringes not one but multiple discourse practices as established by the televisual apparatus (cf. 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 5: “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 5: bottom-right image).

As observed by Budd et al. (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 5: top-left image), either by gaze or body postural orientation; all others either are told not to look at the camera, a convention which 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 5: top-right image) - or they appear in indirect address as if the camera were invisible. This, of course, has further implications for audience address. As Masterman (1985) 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, macro-Themes, macro-New, 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 (cf. Halliday, 1994) – even though his style of delivery mimics the type commonly used by anchors for news presentation (cf. 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 program 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 6: 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 program 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.
Figure 6: Controlling the discourse: (1) overlap; (2) Carmen retaking the floor.
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 ideological thread running through the text, the ‘take-home message’ of the segment.

4. Multimodal discourse and metadiscourse in the digital environment

In the previous section we have tried, within the constraints of the printed page, to show how interactive software resources may be employed in the 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 phasal organization of the text as a whole; and how these phenomena at different levels (in systemic functional linguistic terms, ‘strata’) can be related to one another and to the analysis of an overarching ideological 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 studies of social semiotic phenomena to a grounding in empirical discourse analysis. The software brings together within the same environment the user’s various analyses 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.

As such, we are aware that the digital environment is also a resource for exploring, testing, comparing and calibrating multiple theoretical and 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 which 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. For example, the discussion
presented above was derived primarily from the analyses, initially conducted independently, by two of the present authors, analyses which were then compared within a single interface template design (compare, for example, Templates 1 & 2 in Figures 3 & 4). These analyses drew on textual, graphic, systemic 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 such analyses in combination might reveal significant aspects of larger multimodal corpora of such text types. For example, in the present study, one analysis 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 metafunctional perspectives to these analyses. The second analysis focused on spoken discourse, particularly the inter-stratal relations of SPEECH FUNCTION, MOOD and TONE systems within language (that is, interpersonal aspects of the discourse), vocal and gestural phenomena, and turn-taking strategies. There were areas of overlap in these two sets of analysis, for example, Theme as realized through visual and aural semiotic resources: e.g. on-screen text, and speech, 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
within the multimodal discourse is of course, as we mentioned at the beginning of this paper, 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 make a significant contribution to relating what was for us an early intuitive reading of the ideological bias of the text, deriving from our training in social semiotics, to multiple lower- and higher-level analytical evidences in a way that is empirically grounded, meaningful, and capable of being (re)presented as visual patterns and relations between patterns within the software interface.

It is clear that we need to develop the sort of multimodal approach to analyzing contemporary discourse phenomena as are exhibited in the texts we 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 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. Furthermore, 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 discourse. That is, the software environment is a resource for analyzing, comparing and discoursing about our own (meta-)semiotic practices, encouraging critical self-reflection and critique as well as enabling greater meta-discursive social action through the presentation, sharing and comparison of (digital) analyses.

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 which are stored as unique, numeric values and thus capable of being interrogated and visualized, shared and compared. Such analysis imposes upon the analyst the necessity to move from abstract theory and discursiveness to actual systemic analysis, moving, for example, from a description of the minutae of a variety of facial gestures and body movements to a systematization of these: from Pike’s (1954) ‘etics’ to ‘emics’. The fact of being able to store, retrieve, share and represent in a variety of ways (visualizations and
auralisations) the results of one’s analysis also means that the digital semiotician need not remain confined to one particular analytical or theoretical orientation. A semiotician can conduct a variety of analyses, and then view the range of such analyses as, in effect, secondary (meta-)texts created by the configuration of such analyses within a particular template design.

There is thus an interplay set up between (meta-) sign and signifier: one’s own abstract analyses are given visual or aural expression through the affordances of contemporary computer techniques, which thus become texts for further analysis, an analytical recursiveness with great potential power. This affordance was illustrated, for example, in the visualization via nodes of turn-taking analysis for the different speakers (e.g. Figure 6), revealing patterns of speaker turn and overlap. It can also be used to illustrate such phenomena as inter-stratal relations of interpersonal semantics, lexicogrammar and phonology, revealing, amongst other things, the distribution of the clausal information into tone groups (cf. Halliday and Greaves 2008) and non-congruent grammatical choices for speech functions (minor clause for command, for example); and the relations of the various visual semiotics resources to each other and to their semantic functions as the realization of implicit ideological meaning. The various template configurations and the analytical and theoretical views they afford (drawing upon systemic functional linguistics, social semiotic theory and multimodal description, phasal analysis etc) reveal and relate different views upon and phenomena within the text (cf. Halliday 2008 on complementarities in language and linguistics), which views then become the site for higher-level analysis, for example of ideology.

It is in this sense that we offer the title of the present paper. 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 the ‘orchestration effect’ of multimodal discourse. As an appropriate response to such a move within the wider culture and within social semiotics, 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 & Using Interactive Digital Media for Multimodal Discourse Analysis (MDA) (Project No. NRF2007IDM-IDM002-066, Principal Investigator Kay L. O’Halloran) awarded by Media Development Authority (MDA) under the 1st Interactive Digital Media Research and Development (IDM R&D) Grant Call by National Research Foundation (NRF) Singapore. For further information, please see the Multimodal Analysis Lab website (http://multimodal-analysis-lab.org) or contact Kay O’Halloran (kay.ohalloran@nus.edu.sg).

2. 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 5th, up to the ‘unstable’ 7th note, and down through the 4th resolving to the Tonic: an instantly recognizable musical gesture of ‘finality’ that fits nicely in with this ending of the introductory phase of the discourse.

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

References


