

O'Halloran, K. L., Tan, S., Smith, B. A., & Podlasov, A. (2010; accepted for publication). Challenges in Designing Digital Interfaces for the Study of Multimodal Phenomena. *Information Design Journal*.

**Kay L. O'Halloran**

**Sabine Tan**

**Bradley A. Smith**

**Alexey Podlasov**

**Multimodal Analysis Lab**

**Interactive & Digital Media Institute (IDMI)**

**National University of Singapore**

## **Challenges in Designing Digital Interfaces for the Study of Multimodal Phenomena**

The following paper discusses the challenges faced in designing a flexible, easy-to-use, cross-platform software tool for modeling, analyzing and retrieving meaning from multimedia data. In a project currently being undertaken at the Multimodal Analysis Lab, Interactive and Digital Media Institute (IDMI), at the National University of Singapore (see <http://multimodal-analysis-lab.org/>), social scientists and computer scientists collaborate to produce an interactive digital software tool for studying multimodal communication (that is, communication using and integrating multiple semiotic resources such as language, visual imagery, gesture, music, dress and so on), in media such as written text, pictures, sound and video. This software is projected to have application across a variety of academic disciplines, theoretical traditions and professional vocations.

The paper gives particular consideration to two important and related issues in the emerging field of multimodal study: the theoretical foundations upon which to base the study of multimodal discourse; and the theoretical implications of the development of

technologies, in particular software interfaces (and their functionalities), for the visualization, exploration and analysis of multimodal phenomena. We first outline the aims of the software development against the backdrop of developments and issues in the study of multimodal communication, and introduce our own social semiotic theoretical perspective on multimodal semiotics. We then address some of the challenges that software developers face in designing digital resources, in particular in the design of interfaces and their functionalities, with respect to the dialectal relations between theory and application.

## **1. Context of software development**

The twentieth century was a time of rapid change and growth in the study and understanding of human meaning systems. Whereas scholars prior to the mid twentieth century were on the whole in the relatively safe position of being able to restrict themselves to the study of (for the most part written) language, the ongoing revolution in multimedia design and digital technology within contemporary society at large has led to a proliferation of multimodal documents (using media such as graphics, digitized photographs, audio and video texts, contemporary digital media, in particular web-based media, and three-dimensional virtual objects). Meanwhile, the twentieth century has seen a profusion of disciplinary and theoretical perspectives upon human communication in general and multimodal communication in particular.

The modern scholar of semiosis must thus cope with multiple forms of communication, as well as multiple theoretical perspectives upon which to draw in the study of human meaning-making activity. Furthermore, within academia, as part of the move to a multimodal orientation within the culture in general (Kress and van Leeuwen 2001, p. 1), increasingly sophisticated and powerful (particularly digital) technical resources have become available for the study of human semiotic activities. Not only have advances in digital technology opened up new areas in terms of the objects of research, they also impacted the ways in which users collect, transcribe, and analyze multimodal data (see LeVine & Scollon, 2004; Jewitt, 2006).

However, while efficient and sophisticated interactive digital technologies have been developed for the design, manipulation and dissemination of multimedia texts and artifacts (e.g., such as Adobe™, Picasa™, YouTube, etc.) which have been quickly adopted by the mainstream society, software tools for the scholarly analysis of multimodal data have tended to be less developed and less widely used by those most concerned with the study of human communication. Software applications that integrate visualization techniques, annotation, coding and interpretation in interactive interfaces are often targeted at a specific branch of social science or industry, thus requiring a significant investment on the part of their users, who need to spend time and effort in adapting the software to suit their individual needs and requirements (see Rohlfing et al., 2006). In addition, these annotation tools maybe based on frameworks that are more descriptive in nature rather than analytical and systemic, offering mostly ways for “describing effects rather than creating inventories of the ways that precise design decisions can contribute to the overall meaning” of multimodal objects and events (Machin, 2007, p. viii). The framework one applies in the development of such software applications is thus crucial to the design and functionalities of that application, both in terms of descriptive adequacy, and applicability to a range of research tasks and theoretical orientations of the users.

To develop effective and flexible digital resources for tracing the complex meaning making processes of multimodal phenomena, we therefore propose to employ multimodal social semiotics as the underlying theoretical foundation for our own project. This is because, through the prism of this theory, we are enabled to treat interactive digital technology as both the virtual and physical embodiment of multimodal meaning (c.f. Chignell, 1993, for a discussion of the relations of physical, conceptual and task layers in a model of interface design). That is, firstly, from a social semiotic perspective, “meaning is produced and reproduced under specific social conditions, through specific material forms and agencies. It exists in relations to concrete subjects and objects, and is inexplicable except in terms of this set of relationships” (Hodge & Kress, 1988, p. viii). Thus, the design and use of digital resources are themselves meaning-bearing activities, with all the implications that a social semiotic theory of meaning and its expression

affords. Secondly, like Baldry & Thibault (2006, p. 19) we believe that, in practice, texts of all kinds are always multimodal, making use of, and combining, the resources of diverse semiotic systems to facilitate both generic (i.e. standardized) and specific (i.e. individualized, and even innovative) ways of making meaning. Interactive digital technology, in particular (e.g., such as film, hypertext, video games, etc.), routinely employs a mix of traditional and new media, which allows for a multiplicity of semiotic modes and resources to unfold simultaneously on-screen.

A multimodal social semiotic approach to the study of communication offers the descriptive means to account, in both detailed and holistic views, for the multiple and innovative ways in which semiotic resources are both co- and/or cross-deployed within and across various modes of communication (i.e., visual, aural, and somatic) to fulfill certain social-semiotic functions or objectives, and thus offers a useful framework and set of guiding principles upon which to base our software interface design. As the underlying basis for the project under development, social semiotic theory offers the promise of a systematic, comprehensive and integrative analytic tool for the study of multimodal phenomena “for the many people in different disciplines who deal with different problems of social meaning and need ways of describing and explaining the processes and structures through which meaning is constituted” (Hodge & Kress, 1988, p. 2).

## **2. Theoretical foundation and developments in the study of multimodality**

### **2.1 Social semiotics**

Social semiotics (the term ‘social semiotic’ derives from Halliday, e.g., 1978) has its origins in the two independent, but complementary, branches of semiotics as developed by the American philosopher Charles S. Peirce and the Swiss-French linguist Ferdinand de Saussure. Charles S. Peirce essentially understood the process of meaning making (semiosis) as encoded in the triadic relation between a signifier, a thing signified, and an interpretant “created in the mind of the interpreter” (Bishara, 2007, p. 81). For Peirce,

A sign, or representamen, is something which stands to somebody for something in some respect or capacity. It addresses somebody, that is, creates in the mind of that person an equivalent sign, or perhaps a more developed sign. That sign which it creates I call the interpretant of the first sign. The sign stands for something, its object. It stands for that object, not in all respects, but in reference to a sort of idea. (CP 2.228) (cited in Bishara, 2007, p. 81)

In Saussure's semiotics, the "science that studies the life of signs within society" (Saussure, 1916/1974, p. 16) is seen as ensconced in the unifying relations between a signifier (the psycho-sensorial 'soundimage') and the signified (the more abstract 'concept') (see Thibault 1997, pp. 211-216). Saussure's notion of a dyadic meaning-making system informed structural theory as pursued by the Russian Formalists, and the Prague and Paris schools of semiotics (see Djonov, 2005) which, in turn, inspired Roland Barthes' notion of the layering of meaning in visual semiotics. According to Barthes, denotation, the first layer of meaning refers to the 'what, or who, is being depicted here', while connotation, the second layer, encodes 'what ideas and values are expressed through what is represented, and through the way in which it is represented' (van Leeuwen, 2001, p. 94; c.f. Machin, 2006).

According to van Leeuwen (2005, p. 3) the "sign was considered the fundamental concept of semiotics". More recently, the term 'semiotic resource' has been developed as a core concept of social semiotics. The notion of sign as resource is influenced by developments in systemic functional (SF) theory as proposed by Halliday, "who argued that the grammar of a language is not a code, not a set of rules...but a 'resource for making meanings'" (1978, p. 192; cited in van Leeuwen, 2005, p. 3). Halliday's theory is largely based on Firth's system-structure theory, but it also assimilates more abstract principles, such as Hjelmslev's view of meaning making as both stratified in terms of content and expression, and instantiating in terms of system and instance (see Iedema, 2003, p. 31). These principles underpin the dimensions of stratification and instantiation in SF theory: the first relates semantics, lexicogrammar, phonology and phonetics as ordered levels of abstraction via a dialectal relation of realization; the second relates the

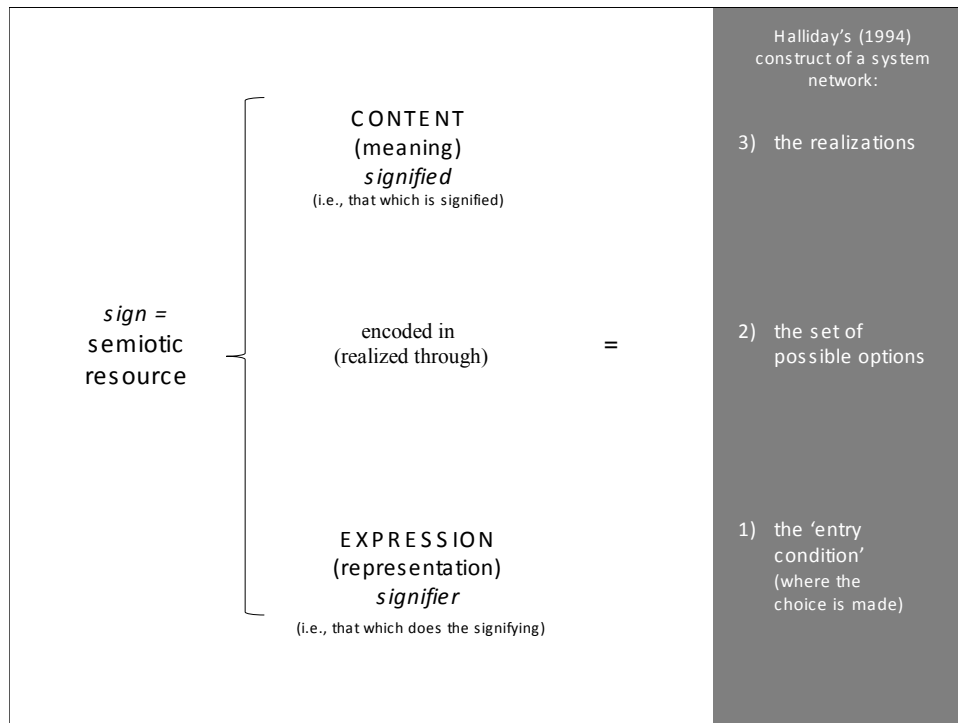
meaning potential available to members of a culture to actual choices from that potential in instances of text, with an intermediate perspective on what is called ‘register’ (e.g., Halliday, 1978), which represents the constraints observable in particular types of texts on the total potential operative within particular types of context, for example doctor-patient interviews, or telephone sales (c.f. also various discussions of ‘genre’, e.g., Martin & Rose 2007).

Stemming from the idea of language as a social resource is Halliday’s (1973, p. 34) insight that “language is as it is because of what it has to do”, reflecting in its internal organization its social functions. This internal organization realizes three kinds of meanings simultaneously, referred to as ‘metafunctions’, i.e.,

- the ideational metafunction, which comprises both an experiential and logical element, the first of which is concerned with how we represent experience, the second with logical relations between such experiential representations;
- the interpersonal metafunction, which expresses the speakers’ or writers’ role relationship with their audiences, and their attitude towards the subject matter; and
- the textual metafunction, which is concerned with how the text is organized and made relevant to its context.

Halliday’s theory is essentially a theory of ‘meaning as choice’: a language, or any other semiotic system, is interpreted as a network of interlocking options, “not a conscious decision made in real time but a set of possible alternatives” (Halliday, 1994, pp. xiv-xxvi) from which choices are made in actual texts. As van Leeuwen (1999, p. 29) points out, the principle of ‘choice’ should not be misinterpreted as ‘intentional choice’. While it may well be intentional in certain contexts and situations, “it may also result from a convention followed unthinkingly, a habit acquired unreflectively, or an unconscious impulse” (van Leeuwen, 1999, p. 29).

Stemming from the principle of resource as choice is the chief organizing concept of SF theory, the ‘system’, or more specifically the ‘system network’. According to Halliday (1994, p. xxvi), the system includes (1) the ‘entry condition’ where the choice is made, (2) the set of possible options, and (3) the ‘realizations’ (see Figure 1).



**Figure 1: Underlying Principles of Semiotic Systems**

Halliday’s systemic functional theory, with important additions and modifications, has been especially significant for the study of multimodal text, which study is explored in greater detail in section 2.2.

## 2.2 Developments in multimodal studies

As suggested in Section 1, enquiry into the nature of multimodal discourse has been one of the primary challenges to students of human communication through the past century, beginning with the shift from an exclusive focus on single modes of communication, e.g., written text or image, to the study of multimodal forms of communication, particularly in the digital age. Valuable theories and methodologies for studying communicative

phenomena other than written text have been developed long before their treatment within the emerging field known as multimodal studies. Although multimodality has been of particular interest to formerly ‘language’-based disciplines (see Machin, 2007, p. x), the study of multimodal semiotics inevitably involves the study of domains of meanings beyond the domain of the institutional discipline of linguistics.

Study of various visual art forms, music and architecture, for example, have of course long established traditions of scholarship dealing with non-linguistic forms of communication. For other established disciplines such as film and media studies, communication through the visual and aural modes and their integration have been central focuses of academic enquiry, and as noted by Machin (2007, p. x) with respect to the primacy of the visual mode, “many excellent books brim with ideas and methods for its analysis”, a comment which has application to a wide range of academic (and non-academic) fields dealing with the various forms and media of human semiotic activity. Work within the Prague and Paris schools of semiotics mentioned in the previous section are especially important to any study of the multifarious forms of human sign-making. In addition, practitioners in many fields – for example, film production – are also a potentially rich source of understanding and insight into the production of multimodal texts.

However, the development of the study of multimodal communication as a distinct area of research, at least within a social semiotic perspective, may be first found in the pioneering works of O’Toole (1994) and Kress & van Leeuwen (2006 [1996], 2001). The theoretical advances found in these works have been variously applied, adapted and expanded for the analysis of, for example, music and other sonic phenomena (van Leeuwen, 1999), static, printed text (O’Halloran, 2008a; Baldry & Thibault, 2006, Bateman 2008), dynamic media such as film and video (Baldry & Thibault, 2006; Kress & van Leeuwen, 2006; van Leeuwen, 2005; Baldry, 2004; O’Halloran, 2004; Iedema, 2001; Thibault, 2000), interactive hypertext (Baldry & Thibault, 2006; Djonov, 2005; Lemke, 2003, 2002), and studies of embodied social (inter)action and gesture (Norris, 2004; Martinec, 2004, 2001; Moore, in press). While some of these approaches, such as




O'Toole's, are closely situated within Halliday's systemic functional tradition, following its principles of metafunctional and rank-based organization, other approaches are "less bound to the systemic origins of multimodality" (Constantinou, 2005, p. 603; c.f. Jewitt, forthcoming for a review).

Consideration therefore of the development of the field of multimodal study shows that it is inherently an inter-disciplinary exercise, with a multiplicity of theoretical, methodological and analytical approaches. Yet, whilst those from a linguistics background are often ignorant of both theoretical and practical approaches outside their specific academic field, "the opposite lack of engagement in linguistics, in systematic procedures for both textual and visual modes, is true of many media researchers" (Machin, 2007, p. x-xi). For the former group, there is the danger that much of significance from other domains of knowledge that might help in the understanding of multimodal phenomena might be lost – a kind of 're-inventing the wheel' syndrome; while for those who are used to the more structured, systemic linguistic approaches for the study of multimodal text there can be the perception that "that much work in media and cultural studies is impressionistic and carried out without any real analytical tool kit" (Machin, 2007, p. xi).

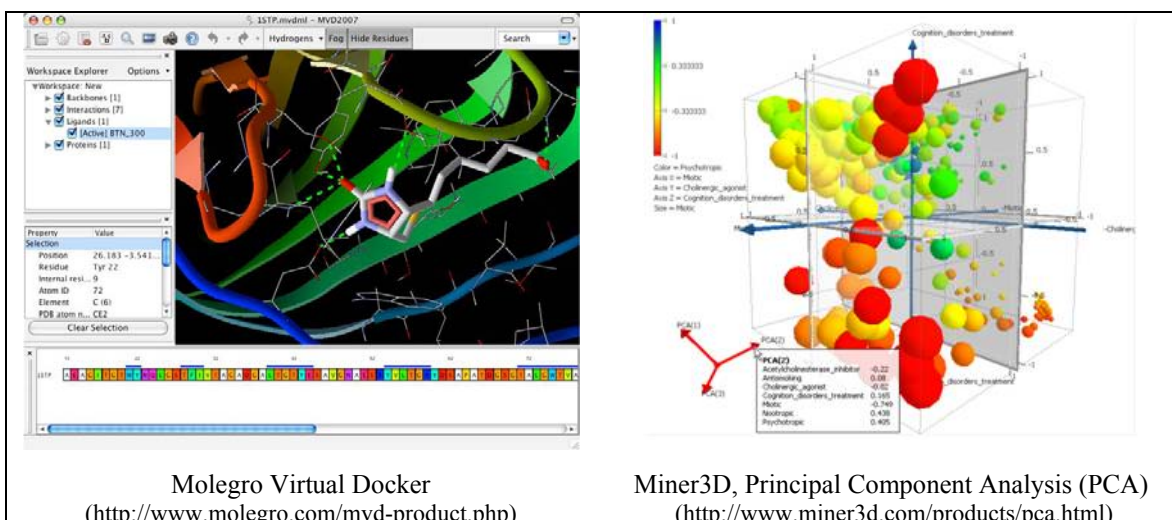
These issues have a direct bearing upon the development of software resources for the study of multimodal communication. One's theoretical orientation, the types of analysis one does, and one's ways of engaging with multimodal data will ultimately determine to a large degree the sort of structures and functionalities assigned to software interface design. Yet despite the progress that has been made in the development of multimodal theory and description, and the availability of increasingly sophisticated and powerful digital resources for modeling and analyzing complex phenomena, most approaches to the transcription and analysis of multimodal data continue to be implemented with 'low-tech', largely page-based, methodologies, which are not only extremely laborious and time-consuming to construct, but which severely constrain "the analyst's ability to display, describe and analyse the complexity of the multifunctional

interplay of semiotic choices, especially in the case of the video text and interactive digital sites” (O’Halloran, 2009a, p. 10) (see Table 1 for elucidation).

**Table 1: Example of Page-Based Analysis**

Phase/Sub-phase	2c – <i>The monkey</i>
Visual Frame	SEQUENCE 9 SHOT 66  Frame 379    Frame 380    Frame 381    Frame 382    Frame 383    Frame 384
Sound: k	Soundtrack
Music	🎵 drums, rock music Volume: (f), Tempo: F
Song	Rocksinger: (SCREA:::M)
Speech	---
Verbal Description	(Pet monkey raises paw to its head, covers its eyes). Plaque beside monkey reads “O JEITO CER.../DE FAZER CER...”, i.e., the skill/way... to create/make/produce...
Narrative Representations	Participant: I; Vector: Y:gaze:off-screen:engaged:viewer + Movement: Y:directed at self:body part; Process: Circumstance of Means
Conceptual Representations	Visual Metaphor, Humour, Irony
Mood	Direct Address: Y:demand; Size of Frame: medium long shot; Social Distance: close social; Angle/Power: HP:frontal:involved, VP:median; CM:stat
Modality	Colour: naturalistic S/D; CX: low; Depth: shallow:centeral;
Composition	Saliency: Figure:Monkey:focus+placement+contrast; Framing
Graphic/Rhythmic / Spatio-Temporal Relations	↔Graphic Conflict: colour+lighting↔ ↔Rhythmic/Dynamic Match: CM ST-Discontinuity: FX:shock cut ↔Temporal Conjunction: Simultaneity
Intersemiotic Relations	---

Such page-based analyses are not only time-consuming but also make difficult the tasks of relating the results of different analyses to one another and discerning patterns across different systems and semiotic resources, as well as assessing the results of applying differing theoretical perspectives. The digital environment is an enabling one for such tasks, as the performance of a variety of analyses and the visualization of their results can be managed through the resources of interface design, database storage and retrieval, and annotation and navigation functionalities. Yet although there are other disciplines, particularly physical sciences such as physics, chemistry, geology and medicine, and also mathematics, which have developed sophisticated digital resources and techniques of analysis and visualization for the representation and study of complex phenomena (e.g., Illustration 1), such resources of technology and technique remain underutilized in the social sciences (O’Halloran, 2009a; Manovich, 2009).



**Illustration 1: Software Applications for Physical Sciences**

### 3. Multimodal interface design: Challenges and opportunities

Following on from the discussion in the previous sections, a major challenge in developing software resources for the study of multimodal phenomena is to draw upon the knowledge of a range of disciplinary, theoretical and technical traditions of scholarship and practice. At the same time, while offering researchers the opportunity to apply and explore particular theoretical and analytical perspectives to the interactive digital study of multimodal phenomena the software interface must be made accessible and easy-to-use for users with a range of tasks and levels of computer literacy. For an undergraduate, Masters or PhD student, for example, working within and applying a particular theoretical tradition, the interface should readily offer an opportunity to apply a particular analytical schema. Meanwhile, others may wish to start with a different framework, or perhaps with a *tabula rasa* approach, particularly, in the latter case, if developing new perspectives or working with phenomena not already described (as in some PhD and much postdoctoral research). The needs of teachers and practitioners must also be taken into consideration. Furthermore, even for those working within a particular theoretical tradition, there may be different ways of approaching the exploration and analysis of data that may require different interface designs.

Within the systemic functional tradition of multimodal studies discussed in Section 2.1, for example, theoretical dimensions such as metafunction and stratification offer powerful organizing concepts for designing interfaces with which to interact with multimodal texts, but their understanding and perhaps value is restricted to those schooled within systemic functional theory. Furthermore, scholars within this tradition may have particular dimensional orientations: for example, one may wish to work within a particular metafunctional framework, analyzing choices in either ideational, interpersonal or textual systems in a text; or one may wish to adopt a stratal perspective, moving from consideration of the phenomena of the expression plane in the first instance (e.g., phonetics, or graphetics), with metafunctional classification a secondary and optional step; or utilize a combination of these perspectives during the exploration and analysis phases of one's research.

Meanwhile, even within such traditions, there is often debate about the use of terms and concepts. If one wishes to develop a software tool for a wide variety of users with differing specialist terminologies or interpretations of terms, being clear about the use of one's terms in the interface is therefore crucial. However, so too is accessibility, and a host of 'foreign' terms and a complex conceptual structure – even with detailed glossaries and explanation – can offer a forbidding introduction to the 'front gate' of a software application, and thus can easily cause difficulties in the human-computer interaction (c.f. Chignell, 1993).

The challenges arising from the acknowledgement and application of differing theoretical frameworks, perspectives and terminologies in our software design are discussed in detail in the next section, followed by our proposals for dealing with these challenges.

### **3.1 Challenges**

As an example of the issue of the use of terminology in our software design, we could consider the use of the terms 'mode' and 'medium' within multimodal studies. Without

wishing to engage in the ongoing debate amongst researchers/theorists about what constitutes a mode and/or medium, the authors are aware that these concepts are often understood differently, even by researchers working within the same tradition of social semiotics (c.f. for example Bateman, forthcoming). Whilst most multimodal researchers and theorists agree that the term *media* refers to the “physical stuff” of communication (Constantinou, 2005, p. 611), there appears to be far less agreement about the use of the term *mode*.

For Halliday (1985) the term ‘mode’ refers specifically to the difference between spoken and written language, but in a more general sense to the role of language within a situational context (Halliday, 1978). Kress & van Leeuwen (2001, p. 21), however, distinguish between *mode*, which is on the ‘content’ side of the theoretical division into content-expression, and *medium*, which is on the ‘expression’ side: they see modes as ‘semiotic resources’, while media are defined as the “material resources used in the production of semiotic products and events, including both the tools and the materials used” (Kress & van Leeuwen, 2001, p. 22). Other researchers see modes more generally as means of representing, and media as means of disseminating (Constantinou, 2005, p. 609; c.f. LeVine & Scollon, 2004, p. 2).

Constantinou (2005) is of the express opinion that the concepts of *mode* and *media* can never be absolutely defined or bounded and would need a sufficiently open definition that includes not only the tools and technologies of dissemination, but its practices and infrastructure too (Constantinou, 2005, pp. 607- 611). Furthermore, the term ‘mode’ is closely related in many uses to the term ‘modality’, as in the comment by Baldry & Thibault (2006, p. 4), that “[d]ifferent semiotic modalities make different meanings in different ways according to the different media of expression they use”. The term ‘modality’ has another specific use within both Halliday’s grammar – as a system of the interpersonal metafunction – and more generally to refer to language and other semiotic systems having similar semantic potential, for example music and other sonic sign systems (e.g., van Leeuwen 1999).

One may keep in mind in this respect Firth's insight into the development of a language of linguistic science, that "[t]echnical terms and phrases...are, so to speak, defined operatively...In operational terms, they mean what they do" (Firth, 1968, p. 33). The challenge is not to discover and define an immanent meaning for any term, nor the scope of its use within either semiotic sciences or what Firth (1957, p. 140) referred to as the "language of the common sensual life": in everyday social use the term 'mode' can refer simply to a way of doing things. However, as Firth warned, part of the business of those engaged in the development of theory is to develop well-defined terminology. Decisions about such terms become particularly important in terms of interface design: the terms assigned to the interface must be reasonably accessible and transparent so as to make the interface navigable.

These are not merely issues of terminology, of course: as in any science, different terms and uses of terms within social semiotics reflect both differing interpretations of prevailing conceptual frameworks or differing models or frameworks themselves. The discussion of the use of the terms 'mode' and 'medium', for example, centers primarily around what in systemic functional theory is the stratification dimension: the relations of signifier and signified, in Saussure's terms, or content and expression in Hjelmslev's model. Here again, one finds a diversity of perspectives on the interpretation and modeling of this dimension within the field of multimodal studies.

According to Kress & van Leeuwen (2001, p. 20), the "basis of stratification is the distinction between the *content* and the *expression* of communication, which includes that between the signifieds and the signifiers of signs used". As a result of the invention of modern communication technologies, they propose that the content stratum could be further stratified into discourse and design, while the expression stratum could be stratified further into production and distribution. Baldry & Thibault (2006, p. 224), for whom expression and content represent 'two sides of the same semiotic coin', interpret the stratification dimension in terms of *display* and *depiction*. According to Baldry & Thibault (2006, pp. 224-225), the *expression stratum* of a video text consists of visual resources such as lines, dots, the interplay of light and shade, colour, and so on (Baldry &

Thibault, 2006, p. 224). The *expression stratum* of visual semiosis is based on the *display* of visual invariants and their transformation, while the *content stratum* is based on the *depiction* of a visual scene consisting of actions, events, persons, objects and so on in the depicted world. *Display* and *depiction* therefore pertain to the expression and content strata, respectively, they explain.

In O'Halloran's (2008a) model for the analysis of a static printed text, in terms of language, the *content stratum* consists of discourse semantics (paragraph and text) and the lexicogrammar (word group, clause, clause and clause complex), while the *expression stratum* consists of phonology and typography/graphology for spoken and written language (O'Halloran, 2008a, p. 449). However, the systems for visual imagery are not the same as those for language, "which is an obvious point given the differences between the two semiotic resources" (O'Halloran, 2008a, p. 449). They thus require different descriptive categories and analytical approaches, she claims. The systems of the different semiotic resources – language, visual imagery and symbolism – can however be theoretically integrated (e.g., Table 12.1, O'Halloran, 2008b, p. 234).

O'Halloran's use here of the term 'semiotic resources' is similar to that employed by Baldry & Thibault (2006, p. 18), who claim that "[a] *semiotic resource system* is thus a system of possible meanings and forms typically used to make meanings in particular contexts" [italics in original], and can also be likened to the use of the term 'semiotic modes' by others. However, this usage also invokes the concept of categorization in terms of 'types' of semiotic systems – specifically, in the case of Baldry & Thibault, types functioning within specific social contexts, such that the classification of texts as member of a particular category of 'semiotic resource/modality' is a malleable one dependent upon the specific use within a particular context. According to Turner (1994, p. 121), however, the '*sign*' is the basic unit of communication, and it can be a photograph, a word, a sound, an object, a piece of film, in other words, anything that might be deemed significant in a certain context. Common-sense designations such as 'language', 'music', 'visual imagery' and 'photograph' are in fact pervasive in multimodal studies, and while perhaps useful as a step towards the exploration of

phenomena, need also to be problematized and thus made part of the investigation, instead of being treated as pre-theoretical categories.

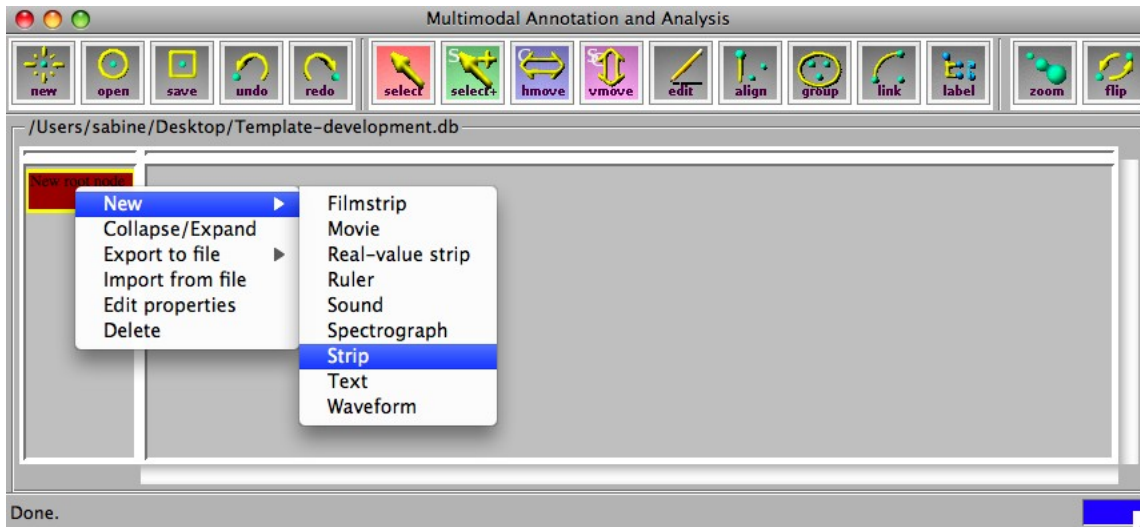
In deciding upon which terms and frameworks to use for particular interface design roles we are in fact engaging in theoretical exercise: different terms and uses of terms, and the theoretical frameworks (and their interpretations) that underpin such uses, constitute decisions that have relevance to the development of multimodal theory, description and modeling. That is, these are clearly not only issues of practical significance for our team, but are important in terms of the development of multimodal studies. In terms of software interface design, to choose one particular interpretation of these terms and a particular framework or model is as much constraining upon as it is enabling of researchers, and seems to run counter to the spirit of intellectual enquiry, diversity and debate which has characterized the short history of this emerging field of study. As Constantinou (2005, p. 604) rightly observes, terminological and conceptual agreement between different approaches to multimodality would further aid their complementarity or their ‘working relationship’, and the development of a software application might seem the appropriate forum to promote such collaborative consistency. Furthermore, for many the application of existing frameworks and descriptions is sufficient for their research, teaching or other needs, so that the provision of tried and tested analytical resources should also be a priority for the software application.

There are tensions, therefore, between developing sophisticated templates for users working within specific theoretical orientations, developing new perspectives or studying new phenomena, and making the software resource usable and accessible to the wide range of people interested in the study of multimodal phenomena. This tension presented us with one of the major challenges to our early project planning, but also served as the inspiration for a solution that draws as much upon the fundamental principles of social semiotic theory as upon the skills and techniques of contemporary software development.



### 3.2 Opportunities

We determined that the tensions outlined above would be resolvable if, on the one hand, we made the design of the interface, in terms of the terminology and template used, to a large extent at the command of the user, while on the other hand supplying sets of default templates, set up by our team according to a variety of perspectives and for application to various tasks. This would allow a user to start ‘from scratch’, developing new interface templates with their own terminology, systems etc. Illustration 2 and 3 give examples of functionalities in operation that allow users to develop both their own templates and constituent conceptual and terminological frameworks (Illustration 2), as well as annotation systems within such templates (Illustration 3, the ‘systems creator’ function\*).



**Illustration 2: Template Design In-Action**

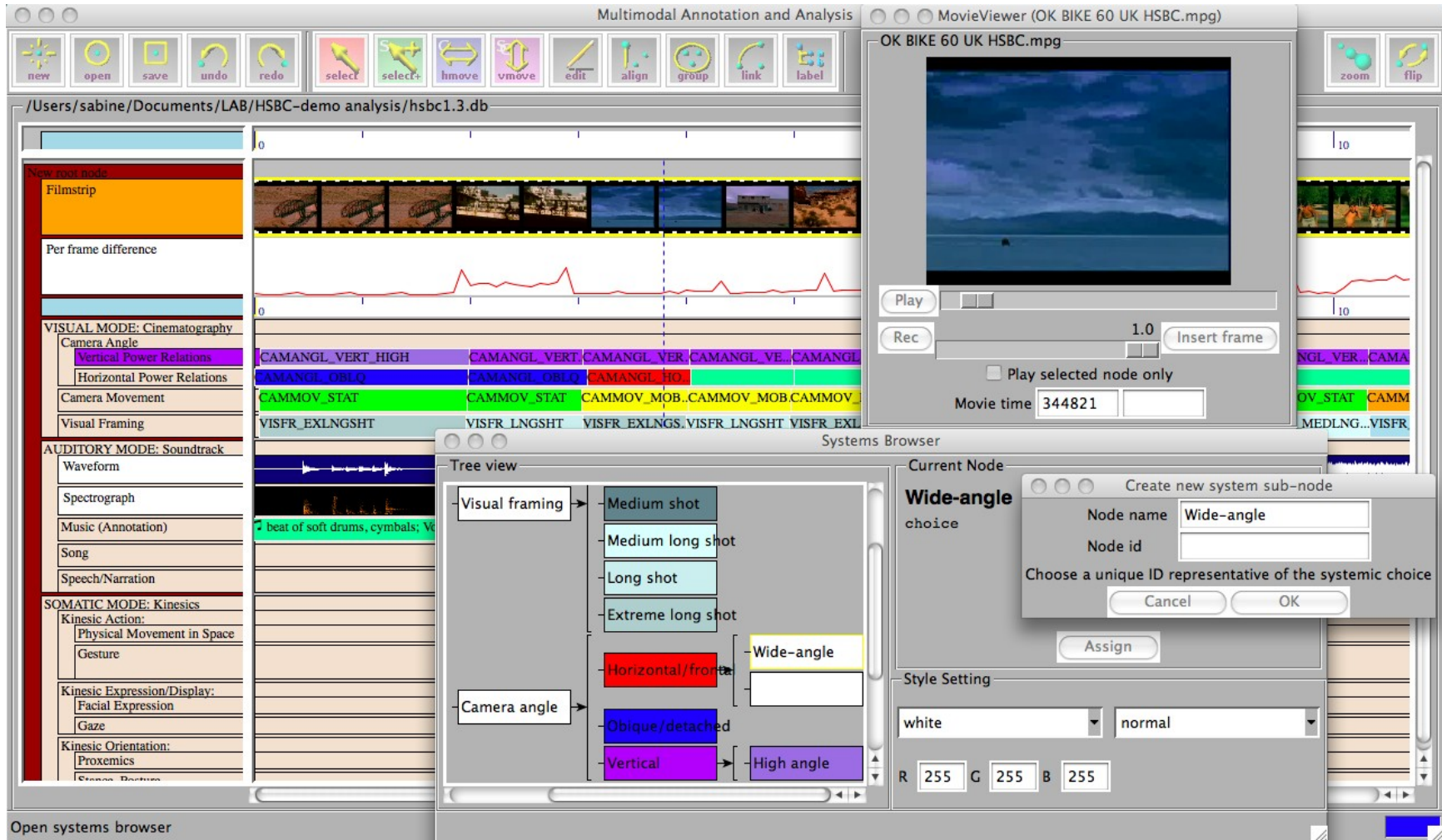
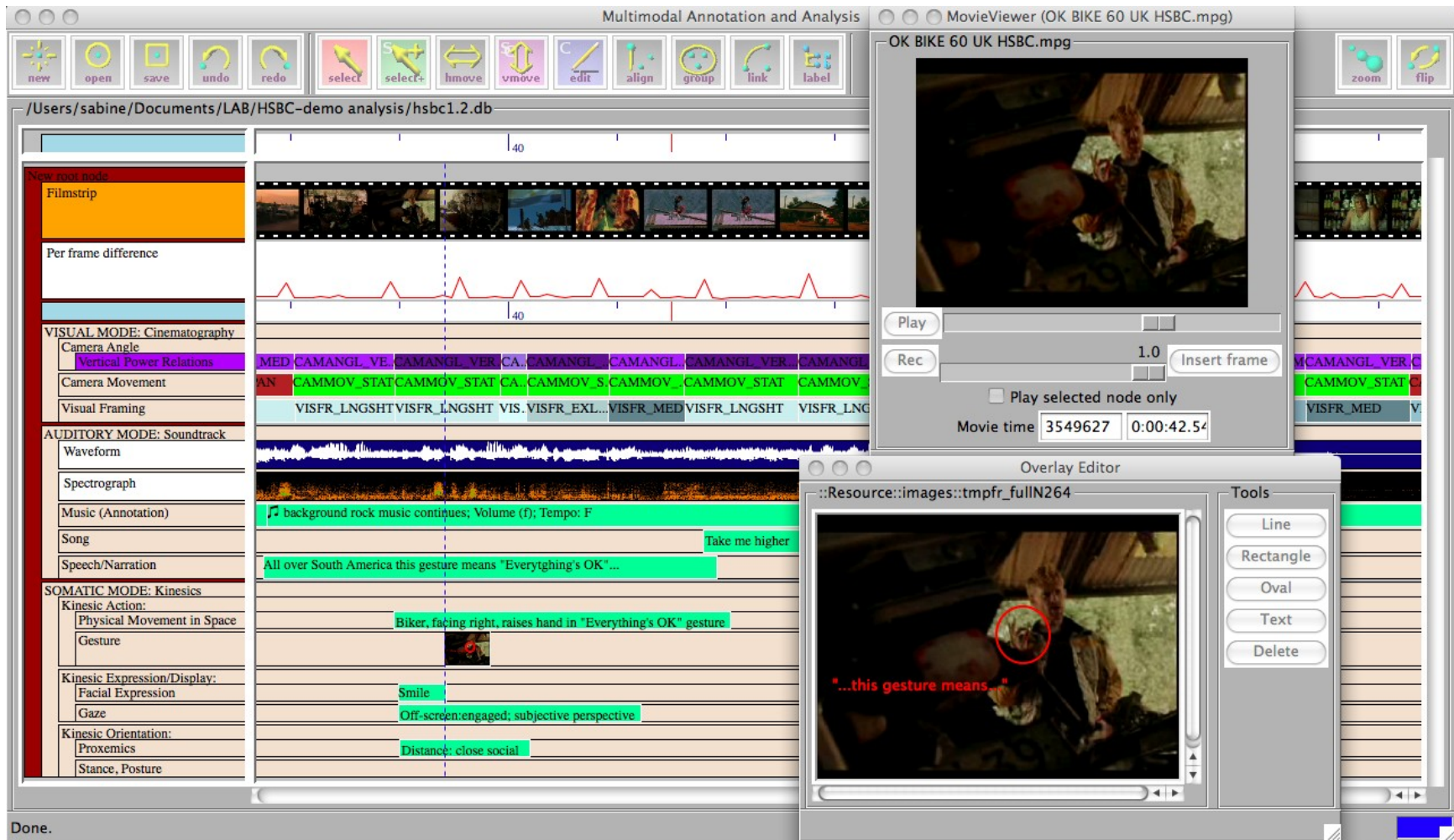


Illustration 3: Systems-Creator

The provision of default templates, however, would give those wishing to use existing frameworks the opportunity to apply these without having to construct them. These defaults, however, would themselves be manipulable and adaptable according to the desires of the user. As the templates a user creates or adapts can be saved as part of the database (becoming the user-generated default templates), these may be developed over time and shared with others. The digital age thus affords those engaged in the development of theory an important resource, that is: the use of interface design as a way to apply, explore and experiment with different ways of studying semiotic phenomena, and for communicating with others such developments in a transparent (and digitally multimodal!) way. The provision of a functional capability for the user to apply their own interface terms and schema is one important outcome of the ongoing collaboration between computer scientists developing the software and social scientists as representative users.

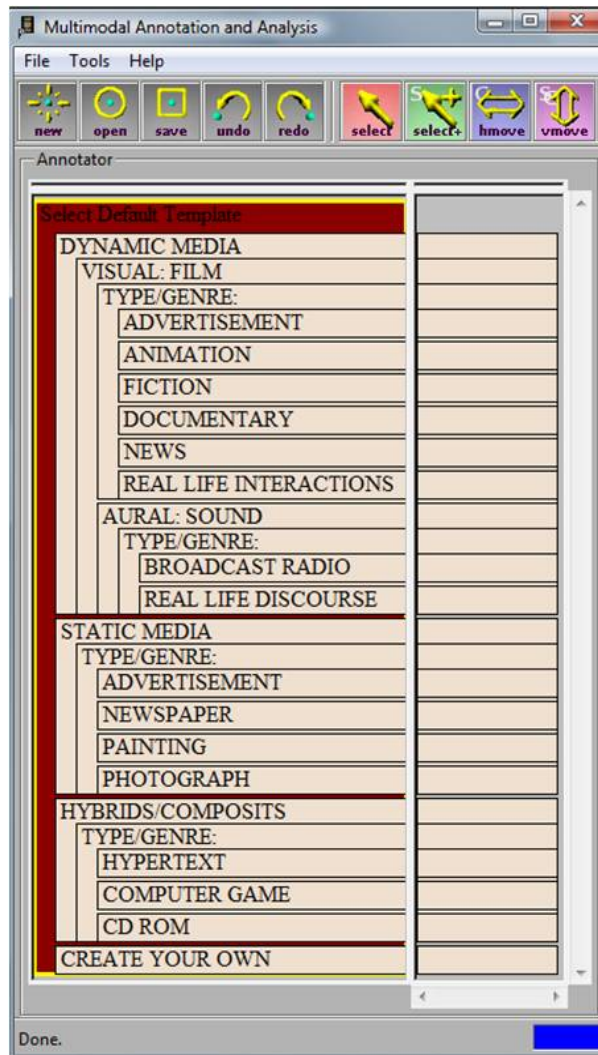
For example, to return to the earlier discussion of the use of the terms ‘mode’ and ‘medium’, we may propose a software interface design, as a default template, in which the term ‘mode’ categorizes primary sensory experiences (as categorized in the interface structure), the template thus comprising a systemic potential of *visual*, *auditory*, and *somatic* modes (see O’Halloran, 2009b) as a first level in the template design. The latter mode pertains to sensory systems which are instantiated by the human subject — or ‘non-human’ actors, such as animals in nature documentaries, for example, or the fictional ‘avatars’ prevalent in *Second Life* and computer games — through the semiotic resources of kinetic action or movement, stance, posture, gesture, haptics (touch), facial expression, and so on, which may thus interface between studies done in the functional and cognitive sciences. Within the context of these designations, the term ‘semiotic resources’, following O’Halloran (2008a), then covers all the commonly categorized forms of communication, such as (spoken and written) language, other ‘languages’ such as the symbolic language of mathematics and music, gesture, gaze, dress, architecture, etc. Illustration 4 shows this proposed template with sample analyses conducted within some of the ‘strips’ generated by this template, and a movie viewer window for viewing the text under analysis.



**Illustration 4: Screen Shot of Prototype Template 'Semiotic Modes'**

This design not only proffers a clear and accessible distinction between the terms *mode* and *semiotic resource* but forms the stepping stone for building a repository of options from which users can then select, within the interface template, that are relevant to the respective phenomenon under analysis. However, although LeVine & Scollon (2004, p. 2) posit that “there can be no mode [our ‘semiotic resource’] that does not exist in some medium”, not all modes will be utilized in all types of media. Furthermore, as Baldry & Thibault (2006, p. 4) observe, “[d]ifferent semiotic modalities make different meanings in different ways according to the different media of expression they use”. Static media such as a painting, a photo, a page-based advertisement, or a printed newspaper front-page do not draw on the auditory mode to make meaning. Similarly, analysts interested in tape-recorded telephone conversations or radio broadcasts will have no need for the visual and somatic (although it needs to be acknowledged that with advances in media technology, such as internet podcasts for example, traditional modal boundaries are constantly being transgressed and transcended), whilst an analysis of composite media like internet web-pages, or real-life cultural artifacts such as baby pram rattles (see van Leeuwen, 2005; 2008), will involve all three (visual, auditory, and somatic).

We may therefore, considering the different types of media mentioned above, instead propose to differentiate between *static* and *dynamic* forms of media as a first step in another default template design. These considerations have an impact on the type of analysis one does, at what stage, and the types and relations of different ‘spaces’ for analysis in the interface. For example, while horizontal, ‘musical score’ type templates may be more practical for capturing the rhythmic and temporal characteristics of dynamic multimodal texts, such as in music and film (e.g., see Martinec, 2007; Baldry & Thibault, 2006; Rohlfig et al., 2006; Tan, forthcoming), static media such as paintings may perhaps be best analyzed in an overlay editor that allows for annotations to be inserted directly onto the semiotic object. In this proposal we have also drawn upon ‘type/genre’ as an organizing concept for the template design: Illustration 5 shows this proposed template for different media types and genres.



**Illustration 5: Mock-up of Prototype Template ‘Media and Genre’**

The provision of a variety of templates can address the many dimensional perspectives and debates around these within multimodal social semiotics. For those proceeding from a genre-based (c.f. Bateman, 2008) or register-based perspective, for example, the provision of templates customized for different genre or register types of text, such as print advertising or telephone sales, might be appropriate. With regards to metafunctional organization, as Martinec (2007, p. 157) observes, for some scholars coming to multimodal studies from a systemic functional background there can be a working assumption that all semiotic modes and resources express ideational, interpersonal and textual meaning simultaneously. Thus a metafunctionally-based



approach (e.g., O’Toole, 1994; Kress & van Leeuwen, 2006 [1996]; O’Halloran, 2004, 2008a) to template design can enable an exploration and analysis of multimodal phenomena in terms of their roles in realizing these higher-order meanings. However, one may not wish to choose metafunctions as the overriding principle of organization, but rather focus on the realizational properties of the various semiotic modes and resources and their capacity for meaning-making (as in van Leeuwen, 1999; Thibault, 2000; Baldry, 2004; Baldry & Thibault, 2006).

The types of interface designs one develops and employs will both enable and constrain the scope of such exploration, results of such analysis and form of such presentation. A metafunctional-based entry point to the study of texts can be useful where descriptions or understandings already exist, or where the type of multimodal resource being studied has become codified within a culture, as in language. An alternative approach would be to allow users to explore the metafunctional orientation of a particular phenomenon without a providing a ‘fixed’ preconception of what that metafunctional orientation is, by offering ‘inventories’ of putative realizational phenomena for certain categories of multimodal media to which users can add their own interpretations. This would empower users to gain insights into phenomena that might otherwise escape their attention, and – at the same time – aid the search for other potentially meaningful distinctions in semiotic resources that have not yet been explored in detail. As van Leeuwen (1999, p. 193) suggests this stratally ‘bottom up’ approach (working from the ‘material’ base of the media signal/s) is best in many respects, particularly in multimodal studies:

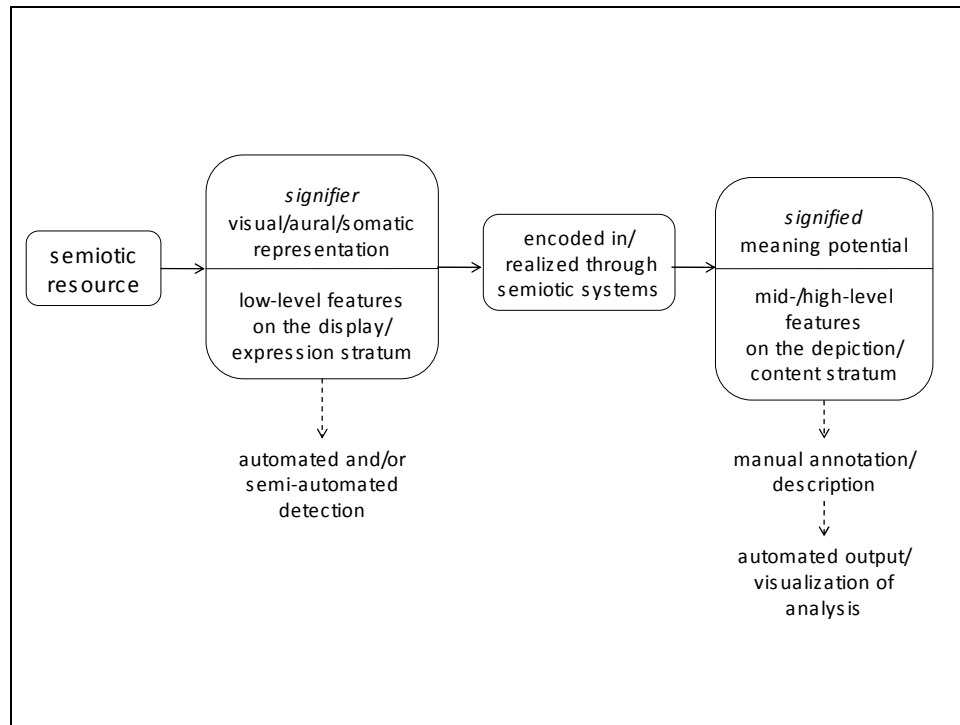
“[o]ne thing is clear. The movement is from the bottom up...if semiotic articulation and interpretation are not to stagnate in eternal repetition, they have to be able, from time to time, to go back to the source, to reconnect with the meaning potentials that are opened up by our physical experience of materiality...”

The bottom up approach, and the template designs appropriate to such an approach, can also be seen in terms of the application of various manual, semi-automated and automated annotation resources. In the digital interface, the

representations of observable phenomena on the display/expression stratum are synonymous with low-level features that can be detected by computer-assisted technology (see Smith & Kanade, 2005), such as pattern recognition, object detection, histograms, Gabor filter banks, etc. According to Smith & Kanade (2005, p. 2), “low-level and mid-level features describe the content according to the level of semantic understanding. Low-level features simply represent statistical content such as color, texture, audio levels, together with the detection of on-screen text, camera motion, object motion, face detection, and audio classification”. Mid-level features attempt to interpret semantic content or meaning, whereas high-level features inevitably involve some form of output display or application (Smith & Kanade, 2005, pp. 2-4).

Utilizing computer-assisted technology to detect low-level features (which has been applied successfully in the area of video mining, video characterization and summarization: see Rosenfeld et al., 2003; Smith & Kanade, 2005) provides the starting point for moving away from manual annotation-based analysis, essentially freeing the analyst to attend to the higher-level aspects of interpretation. These higher-level semiotic choices, while realized through expression plane phenomena, will invariably involve user-annotation, but computer-assisted annotation (e.g.; one-click systemic annotation), output or translation can be of great benefit to researchers (see Table 1 for an example of page-based analysis), speeding up the process of manual analysis considerably. Furthermore, such analysis will ultimately feed back into the development of algorithms for automated analysis, via a significant increase in the available corpora of higher-level analytical data in their (realizational) relations to low-level (expressive) features (see Figure 2).





**Figure 2: Proposed Model for Framework Interface Development**

Van Leeuwen also observes that the notion of rank, the dimension that theorizes compositional/constituent relations (e.g., words are composed of constituent morphemes), “is not always necessary in the analysis of images and that the choice of modelling semiotic systems in terms of multiple ranks as opposed to flatter hierarchies may in any case be related to the hierarchical or more levelled structure of the social system that happens to contextualise the semiotic analysis” (cited in Martinec, 2007, p. 162). Martinec (2007, p. 162) believes that the choice of having ranks or not may in fact be determined by methodological aspects such as the size and nature of the phenomena under investigation. For example, analysts interested in unraveling the meaning-making potential of a single, page-based advertisement, artwork or painting, may choose to benefit from the close analysis that a rank-based organization can afford, whilst researchers concerned with identifying patterns of style or ideology in large corpora of complex, multimodal data may not see the need for it nor have the luxury of attending to such matters. Other researchers may reject a rank-based organization on account of the

principle of ‘discreteness’. Iedema (2003), for example, observes that in the analysis of dynamic multimodal texts, the boundaries amongst the different semiotic dimensions of representation, in other words, the rules as to ‘what goes with what’ and ‘what can signify what’, are inherently fluid and constantly shifting (Iedema, 2003, pp. 33-38; c.f. Jewitt, 2006; Jewitt, forthcoming).

One may also wish to take into account what Baldry & Thibault (2006) subsume under the notion of the ‘resource integration principle’. According to Baldry & Thibault (2006, p. 18), “multimodal texts integrate selections from different semiotic resources to their principles of organisation. [...] These resources are not simply juxtaposed as separate modes of meaning making but are combined and integrated to form a complex whole which cannot be reduced to, or explained in terms of the mere sum of its separate parts”. The meaning of cultural phenomena, objects and events, they explain, is the composite product of this combination, rather than the mere addition of one mode to another (see Baldry & Thibault, 2006, p. 83). Consequently, as noted by Iedema (2003, p. 31), “semiosis not analysed in terms of discrete building blocks or structures, but in terms of socially meaningful tensions and oppositions which could be instantiated in one or more (structural) ways”. The attempts to develop templates that take account of such important considerations to the field of multimodal studies will go a long way towards solving the problem of how to model theoretically the interaction of various semiotic resources and modes in holistic acts of social-semiotic communication.

#### **4. Multimodal studies in the interactive digital age**

Such debates as have been discussed in this paper imply the question as to whether or not a single theoretical framework can in fact adequately account for the different semiotic systems that multimodal meaning making entails and that multimodal analysis and transcription seeks to describe (see Baldry & Thibault, 2006, p. 1). Such a question can seem an insoluble dilemma for those seeking to develop software applications of practical use to a range of projected users wishing to study multimodal meaning and its creation in multimedia texts. However, we have shown how differing approaches and theoretical

positionings may be applied, tested and contested within the interactive digital environment in a way that is not readily available otherwise. In fact, interactive digital resources can enable rather than constrain both the progress of theoretical development and the application of theory to analytical, explorative, pedagogic and other applied tasks.

Van Leeuwen makes a distinction between a social context of use of semiotic resources that “may either have rules or best practices that regulate how specific semiotic resources can be used, or leave the users relatively free in their use of the resource” (van Leeuwen, 2005, p. 4). We signaled in Section 2 that social semiotic theory, with its fundamental principle of choice – signs as resources – has relevance not only to the study of multimodal communication, but to the development and application of interactive digital resources for the study of the physical and semiotic phenomena of communication. We have shown how the resource-based approach allows us to see the development of software, particularly interface design, as a resource for theoretical modeling, exploration and development as well as application. Interactive digital interface design is semiotic creation and communication. In the same way that semiotic resources such as language and music form meaning potentials for communication through these forms, so too the interactive digital environment provides a practically inexhaustible semiotic potential for choice, and thus for making ‘statements of meaning’ (Firth, 1968, p. 19) about multimodal communication of a wide variety of types. Both best practices with and free use of the technical resources are enabled by and recommended with our software.

We have also earlier characterized multimodal studies as an interdisciplinary field. Van Leeuwen points out that one of the key contributions semioticians can make to interdisciplinary research projects is “inventorizing the different material articulations and permutations a given semiotic resource allows, and describing its semiotic potential, describing the kinds of meanings it affords”, which entails building inventories that are not made with an immediate, urgent purpose in mind (see van Leeuwen, 2005, p. 6), and including the “meanings that have not yet been recognized, that lie, as it were, latent in the object, waiting to be discovered” (van Leeuwen, 2005, pp. 4-5). Such a task will be

better served by flexible, easy-to-use yet sophisticated software platforms and resources for the exploration, analysis, storage and digital sharing of multimodal texts, their analyses and the interface design templates that enable and guide their study.

### **Acknowledgements**

The multimodal analysis software tool is being developed in the *Events in the World* project in the Multimodal Analysis Lab, Interactive & Digital Media Institute (IDMI) at the National University of Singapore. The research project is funded by the Singapore National Research Foundation (NRF) Interactive Digital Media R&D Program, under Research Grant NRF2007IDM-IDM002-066.

\* Thanks are due to Arun Nagarajan for developing the algorithm for applying the ‘systems creator’ concept in the software.

### **References**

- Baldry, A. P. (2004). Phase and Transition, Type and Instance: Patterns in Media Texts as seen through a Multimodal Concordance. In K. L. O’Halloran (Ed.), *Multimodal Discourse Analysis: Systemic-Functional Perspectives* (pp. 83-108). London; New York: Continuum.
- Baldry, A. P. & Thibault, P. J. (2006). *Multimodal Transcription and Text Analysis*. Oakville, CT: Equinox Publishing.
- Bateman, J. A. (2008). *Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents*. Basingstoke; New York: Palgrave MacMillan.
- Bateman, J. A. (forthcoming). The Decomposability of Semiotic Modes. In Kay L. O’Halloran (ed.), *Multimodal Representation and Knowledge*, Routledge Studies in Multimodality Book Series.
- Bishara, N. (2007). “Absolut Anonymous”: Self-Reference in Opaque Advertising. In W. Nöth & N. Bishara (Eds.), *Self-Reference in the Media* (pp. 79-92). New York: Mouton de Gruyter.

- Chignell, M. H. (1993). Cooperative Human-Machine Reasoning: Communication through the User Interface. In R. J. Jorna, B van Heusden & R. Posner (Eds.) *Signs, Search and Communication: Semiotic Aspects of Artificial Intelligence* (pp. 348-368). Berlin; New York: Walter de Gruyter.
- Constantinou, O. (2005). Multimodal Discourse Analysis: Media, Modes and Technologies. *Journal of Sociolinguistics*, 9/4, 602-618.
- Djonov, E. (2005). *Analysing the Organisation of Information in Websites: From Hypermedia Design to Systemic Functional Hypermedia Discourse Analysis*. Unpublished doctoral dissertation, University of New South Wales.
- Eggs, S. (1994). *An Introduction to Systemic Functional Linguistics*. New York: Continuum.
- Firth, J. R. (1957). *Papers in Linguistics: 1934 - 1951*. London: Oxford University Press.
- Firth, J. R. (1968). *Selected Papers of J. R. Firth: 1952-59*. F. R. Palmer (ed.) London; Harlow: Longmans.
- Halliday, M. A. K. (1973). *Explorations in the Functions of Language*. London: Arnold.
- Halliday, M. A. K. (1978). *Language as Social Semiotic*. London: Edward Arnold.
- Halliday, M. A. K. (1985). *Spoken and Written Language*. Warrn Ponds, Victoria: Deakin University Press.
- Halliday, M. A. K. (1994 [1985]). *An Introduction to Functional Grammar*. Second Edition. London: Edward Arnold.
- Hodge, R. & Kress, G. (1988). *Social Semiotics*. Cambridge, UK: Polity Press in association with Basil Blackwell, Oxford, UK.
- Iedema, R. (2001). Analysing Film and Television: A Social Semiotic Account of Hospital: An Unhealthy Business. In T. van Leeuwen & C. Jewitt (Eds.), *Handbook of Visual Analysis* (pp. 183-204). London; Thousand Oaks; New Delhi: Sage Publications.
- Iedema, R. (2003). Multimodality, Resemiotization: Extending the Analysis of Discourse as Multi-Semiotic Practice. *Visual Communication*, 2/1, 29-57.
- Jewitt, C. (2006). *Technology, Literacy and Learning: A Multimodal Approach*. Routledge Falmer.
- Jewitt, C. (forthcoming). *Handbook of Multimodal Analysis*. London: Routledge.

- Kress, G., & van Leeuwen, T. (2001). *Multimodal Discourse: The Modes and Media of Contemporary Communication*. London: Arnold.
- Kress, G., & van Leeuwen, T. (2006 [1996]). *Reading Images: The Grammar of Visual Design*. London: Routledge.
- Lemke, J. L. (1998). Multiplying Meaning: Visual and Verbal Semiotics in Scientific Text. In J. R. Martin & R. Veel (Eds.) *Reading Science* (pp. 87-113). London: Routledge.
- Lemke, J. L. (2002). Travels in Hypermodality. *Visual Communication*, 1/ 3, 299-325.
- Lemke, J. L. (2003). Texts and Discourses in the Technologies of Social Organization. In G. Weiss & R. Wodak (Eds.), *Critical Discourse Analysis Theory and Interdisciplinarity* (pp. 130-149). Basingstoke, Hampshire [England]; New York: Palgrave Macmillan.
- LeVine, P. & Scollon, R. (2004). *Discourse and Technology: Multimodal Discourse Analysis*. Washington, D.C.: Georgetown University Press.
- Machin, D. (2007). *Introduction to Multimodal Analysis*. London: Hodder Arnold; New York: Oxford University Press.
- Manovich, L. (2009). Cultural Analytics: Visualizing Cultural Patterns in the Era of “More Media” (<http://lab.softwarestudies.com/2008/09/cultural-analytics.html>)
- Martin, J. R. & Rose, D. (2007). *Working with Discourse: Meaning Beyond the Clause*. Second Edition. London; New York: Continuum.
- Martinec, R. (2000). Types of Process in Action. *Semiotica*, 130, 3/4, 243-268.
- Martinec, R. (2001). Interpersonal Resources in Action. *Semiotica*, 135, 1/4, 117-145.
- Martinec, R. (2004). Gestures that co-occur with Speech as a Systematic Resource: The Realization of Experiential Meanings in Indexes. *Social Semiotics*, 14/2, 193-213.
- Martinec, R. (2007). Topics in Multimodality. In R. Hasan, C. Matthiessen & J. Webster (Eds.), *Continuing Discourse on Language: A Functional Perspective* (pp. 157-181). London; Oakville: Equinox Pub.
- Moore, A. (in press). Surgical Teams in Action: A Contextually Sensitive Approach to Modelling Body Alignment and Interpersonal Engagement. In A. Baldry & E. Montagna. (Eds.), *Interdisciplinary Perspectives on Multimodality: Theory and Practice*. Campobasso, Italy: Palladino.

- Norris, S. (2004). *Analyzing Multimodal Interaction: A Methodological Framework*. New York; London: Routledge.
- O'Halloran, K. L. (2004). Visual Semiosis in Film. In K. L. O'Halloran (Ed.), *Multimodal Discourse Analysis: Systemic-Functional Perspectives* (pp. 109-130). London; New York: Continuum.
- O'Halloran, K. L. (2008a). Systemic Functional-Multimodal Discourse Analysis (SF-MDA): Constructing Ideational Meaning using Language and Visual Imagery. *Visual Communication*, Vol. 7 No. 4, 443-475.
- O'Halloran, K. L. (2008b). Inter-Semiotic Expansion of Experiential Meaning: Hierarchical Scales and Metaphor in Mathematics Discourse. In E. Ventola & C. Jones (Eds.), *New Developments in the Study of Ideational Meaning: From Language to Multimodality* (pp. 231-254). London: Equinox.
- O'Halloran, K. L. (2009a). Multimodal Analysis and Digital Technology. In A. Baldry & E. Montagna (Eds.), *Interdisciplinary Perspectives on Multimodality: Theory and Practice. Proceedings of the Third International Conference on Multimodality*. Campobasso: Palladino. Retrieved from: <http://multimodal-analysis-lab.org/events/publications.html>.
- O'Halloran, K. L. (2009b, in press). Historical Changes in the Semiotic Landscape: From Calculation to Computation. In C. Jewitt (Ed.), *Handbook of Multimodal Analysis*. London: Routledge.
- O'Toole, M. (1994). *The Language of Displayed Art*. London: Leicester University Press.
- Rohlfing, K., Loehr, D., Duncan, S., Brown A., Franklin, A., Kimbara, I., Milde, J.-T., Parrill, F., Rose, T., Schmidt, T., Sloetjes, H., Thies, A., & Wellinghoff, S., (2006). Comparison of Multimodal Annotation Tools – Workshop Report. *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion*, 7, 99-123. Retrieved from: <http://www.gespraechsforschung-ozs.de/heft2006/tb-rohlfing.pdf>
- Rosenfeld, A., Doermann, D., & DeMenthon, D. (2003). *Video Mining*. Boston: Kluwer Academic.
- Saussure, F. (1916/1974). *Course in General Linguistics* (trans. Wade Baskin). London: Fontana/Collins.

- Smith, M. A., & Kanade, T. (2005). *Multimodal Video Characterization and Summarization*. Boston: Kluwer Academic Publishers.
- Tan, S. (forthcoming). A Systemic Functional Framework for the Analysis of Corporate Television Advertisements. In E. Ventola & A. J. Moya Guijaro (Eds.), *The World Told and The World Shown: Multisemiotic Issues*. Basingstoke: Palgrave Macmillan.
- Thibault, P. (1997). *Re-reading Saussure: The Dynamics of Signs in Social Life*. London: Routledge.
- Thibault, P. J. (2000). The Multimodal Transcription of a Television Advertisement: Theory and Practice. In A. Baldry (Ed.), *Multimodality and Multimediality in the Distance Learning Age: Papers in English Linguistics* (pp. 311-385). Campobasso, Italy: Palladino Editore.
- Turner, G. (1994). Film Languages. In D. Graddol & O. Boyd-Barrett (Eds.), *Media Texts, Authors and Readers: A Reader* (pp. 119-135). Clevedon, England; Philadelphia: Multilingual Matters in association with The Open University.
- van Leeuwen, T. (1999). *Speech, Music, Sound*. Basingstoke; London: Macmillan Press.
- van Leeuwen, T. (2001). Semiotics and Iconography. In T. van Leeuwen & C. Jewitt (Eds.), *Handbook of Visual Analysis* (pp. 92-118). London; Thousand Oaks; New Delhi: Sage Publications.
- van Leeuwen, T. (2005). *Introducing Social Semiotics*. New York: Routledge.
- van Leeuwen, T. (2008). *Discourse and Practice: New Tools for Critical Discourse Analysis*. New York: Oxford University Press.