

**O'Halloran, K. L., Tan, S., Smith, B. A., &  
Podlasov, A. (2009; submitted for publication).  
Challenges in Designing Digital Interfaces for  
the Study of Multimodal Phenomena.  
*Information Design Journal.***

**[Unpublished Manuscript.  
Please do not reproduce without  
permission from the authors.]**

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

## Introduction

The following paper discusses the challenges faced by a team of developers in designing a flexible, easy-to-use, cross-platform software tool for modeling, analyzing, and retrieving meaning from multimodal data. This task is currently being undertaken at the Multimodal Analysis Lab, Interactive and Digital Media Institute (IDMI), at the National University of Singapore, where social scientists and computer scientist collaborate to produce user-friendly and theoretically sound digital interface for analyzing multimodal media resources, such as text, pictures, sound, and video, for application across a variety of academic disciplines and professional vocations. The paper gives particular consideration to two important overlapping aspects in the emerging field of multimodal study: the theoretical foundations on which to base the study of multimodal text, and their impact on the development of a computer-based tool for the exploration, annotation, and analysis of complex multimodal data. The paper first outlines the aims of software development against the backdrop of social semiotic theory and current developments in the field of multimodality, and then proceeds to address the many challenges that developers face in designing the digital interface.

### **1. Aims of software development: A social semiotic approach to the analysis of multimodal text**

The twentieth century was unquestionably a time of rapid change and growth in the study and understanding of human meaning systems. The increasingly powerful resources, both technological and theoretical, developed over recent decades for studying multimodal communication (that is, communication through various and multiple semiotic modes and resources), have led to progress in this field of study, but also offered challenges as a result of the availability of these new resources. Whereas researchers were in the relatively safe position in the past to concern themselves, for the most part, with the linguistic aspects of communication, the ongoing revolution in multimedia design and digital technology has led to a proliferation of multimedia texts and artifacts (e.g., graphics, digitized photographs, audio and video texts, three-dimensional objects in hyperspace) which routinely draw on multiple modes and semiotic resources (e.g., language, visual imagery, gesture, movement, music, sound, and so on) to make meaning. As Kress & van Leeuwen (2001, p. 1) point out,

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

this is not simply a result of change in the practices of academia and scholarship: the wider society of which academia is a part has undergone rapid change and development towards a multimodal orientation within the culture at large. People tend to utilize multiple means for communicating now to a much larger extent than was possible or customary even only a few decades ago. The modern scholar of multimodal texts and artifacts must thus cope with multiple modes of communication, as well as multiple theoretical perspectives upon which to draw in the study of human meaning-making systems.

Not only have advances in digital technology opened up new areas of research, they also impacted the ways in which users collect, transcribe, and analyze multimodal data (see LeVine & Scollon, 2004; Jewitt, 2006). These two aspects of the development of a multimodal semiotics – technology and theory – are thus inextricably intertwined and interrelated. The development of a theory of multimodal communication – for example, the study of music, film, website design, etc – has relied heavily if not been entirely dependent upon the technological means by which the objects of its study – the particular texts from which are derived the general theory – can be accessed and observed. The greater this access – as for example with the increasingly powerful software resources for studying audiovisual phenomena – the more powerful the capacity for the theorist to both develop descriptions of such phenomena as well as to ground such descriptions in the analysis of actual text.

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, existing software tools for the analysis of multimodal data 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 part of the 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).

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

To develop an effective digital interface for tracing the complex meaning making processes of multimodal phenomena we propose to employ multimodal social semiotics as the underlying theoretical foundation for our project, because we see interactive digital technology as both the virtual and physical embodiment of multimodal social semiotic communication. 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). Like Baldry & Thibault (2006, p. 19) we too 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 social semiotic approach to the study of multimodal phenomena is thus concerned with 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 functions or objectives in particular contexts in the process-based environment of interactive digital technology. Multimodal social semiotics 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), and thus forms the underlying basis for the project under development (see <http://multimodal-analysis-lab.org/>).

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

### **2.1 Background to social semiotics and its application to the study of multimodal text**

Social semiotics has its origins in the two independent, but complementary, branches of semiotics as developed in the past century by the American philosopher Charles S. Peirce and the Swiss-French linguist Ferdinand de Saussure. In Saussure’s semiotics, the “science of the life of signs in society” (Hodge & Kress, 1988, p. 1) is seen as ensconced in the unifying

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

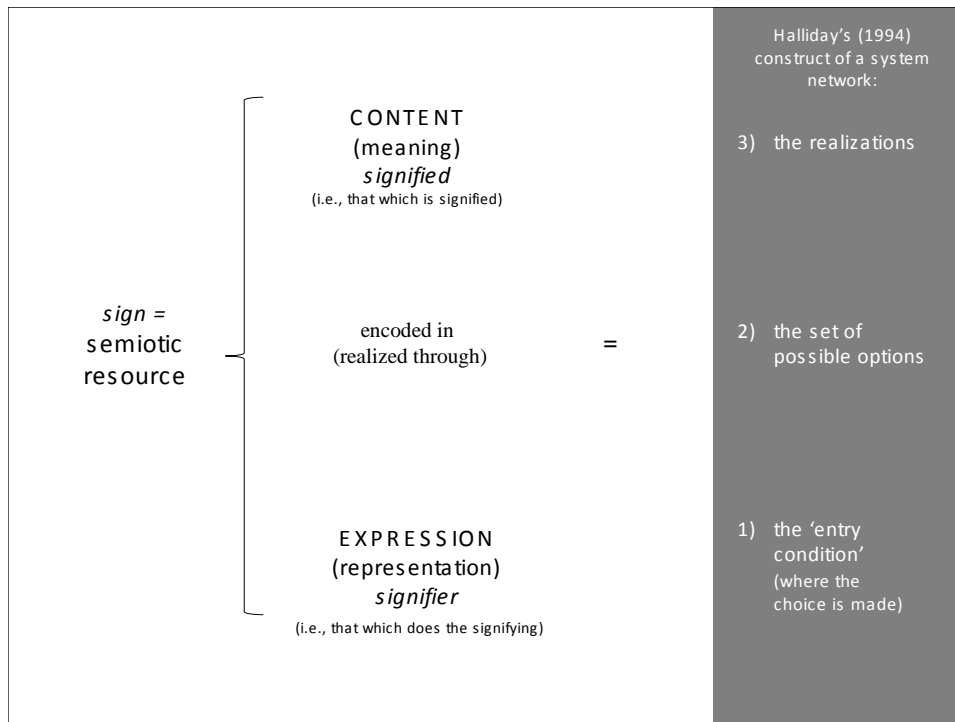
relations between a signifier (the psycho-sensorial ‘soundimage’) and the signified (the more abstract ‘concept’) (see Thibault 1997, pp. 211-216). Saussure’s notion of a dyadic meaning-making system informed structural theory as pursued by the Russian Formalists, and the Prague and Paris schools of 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), whilst 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)

According to van Leeuwen (2005, p. 3) the “sign was considered the fundamental concept of semiotics”, although more recently the term ‘semiotic resource’ has come to be preferred [references?]. The notion of ‘semiotic resources’ as the unifying principle for meaning-making is also influenced by developments in systemic functional 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 functional 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). The chief organizing principle of Halliday’s functional theory is 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’.

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



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

A systemic functional theory is essentially a theory of 'meaning as 'choice', by which 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). 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).

Another influential factor, especially for researchers working within a social semiotic tradition, is Halliday's metafunctional principle, which posits that language, or any other kind of semiotic system, is structured to make three kinds of meanings simultaneously, explained in terms of 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 experiential relations of interdependence;
- the interpersonal metafunction, which expresses the speakers' or writers' role relationship with their audiences, and their attitude towards the subject matter; and

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

- the textual metafunction, which is concerned with how the text is organized.

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 3.2 below.

### 2.2 Developments in the study of multimodal text

To some extent the enquiry into the nature of multimodal discourse has been one of the primary challenges through the past century, beginning with the shift from an exclusive focus on one mode of communication, e.g., written text, to the interest in spoken text, with the invention of phonographic recording technology in the late nineteenth century. However, as multimodal study inevitably involves the study of domains of meanings outside of linguistics, we need to duly acknowledge that other established disciplines have been dealing with these fundamental questions long before linguists began grappling with the concept of multimodality. For instance, in fields such as film and media studies, the visual mode has long been held as the central focus of academic enquiry, and, as Machin (2007, p. x) points out, "many excellent books brim with ideas and methods for its analysis". Many research traditions outside the field of social semiotics have established valuable theories and methodologies for studying non-textual phenomena long before their treatment within the emerging field known as multimodal discourse analysis, which has been of particular interest to formerly 'language'-based disciplines (see Machin, 2007, p. x). The most notable and established advances in the development of frameworks and theories for the analysis of visual images within a social semiotic/systemic functional perspective are perhaps those pioneered by the works of Kress & van Leeuwen (2006 [1996], 2001) and O'Toole (1994). These theories have been variously adapted and expanded for the analysis of static, printed text (O'Halloran, 2008a; Baldry & Thibault, 2006), 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). While some of these approaches are more closely situated within Halliday's systemic functional tradition, following its principles of



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

metafunctional and rank-based organization, other approaches are “less bound to the systemic origins of multimodality” (Constantinou, 2005, p. 603; c.f. Jewitt, forthcoming).

One issue that emerges out of the theoretical development of multimodal studies is the multiplicity of approaches available for the study of multimodal phenomena. Thus, the study of multimodal data is inherently an inter-disciplinary exercise. This is reflected in many ways in the way in which multimodal analysis proceeds, and according to what frameworks and theories. Whilst linguists 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), with the effect that those who are used to more structured, systemic approaches for the study of multimodal text may be led to believe “that much work in media and cultural studies is impressionistic and carried out without any real analytical tool kit” (Machin, 2007, p. xi).

Despite the progress that has been made in the conceptualization of multimodal theory, many frameworks and theories for 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<br><br>Frame 379    Frame 380    Frame 381    Frame 382    Frame 383    Frame 384 |
| Sound: Soundtrack                          |    |
| Music                                      | 🎵 drums, rock music<br>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: 1; 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:central;   |
| Composition                                | Saliency: Figure:Monkey:focus+placement+contrast; Framing  |
| Graphic/Rhythmic/Spatio-Temporal Relations | ↔Graphic Conflict: colour+lighting↔<br>↔Rhythmic/Dynamic Match: CM<br>ST-Discontinuity: FX:shock cut<br>↔Temporal Conjunction: Simultaneity  |
| Intersemiotic Relations                    | ---  |

Ongoing innovations in media design and technology, however, constantly give rise to new and different forms meaning making processes, which allow for a multiplicity of (alternative and innovative) system choices to be accessed and displayed simultaneously on-screen. Of course, as Machin (2007) aptly points out, digital technologies can explain how this has become so much easier to do, but they do not allow us to explain what kinds of predictable patterns can be found in multimodal objects and events (Machin, 2007, p. 20), or how they combine to make meaning in multiplicative ways.

The effective mapping of multimodal phenomena across modes and media will thus be dependent upon the development of an efficient, integrative software application that allows for the array of possible system choices to be displayed and analyzed simultaneously on-screen. As such, our goal will be to develop a multimodal database that will allow for realization patterns to be identified, traced, and displayed across different media and modes of communication. Ultimately, our project endeavors to contribute to the development of new tools and approaches that give due regard to the interactive characteristics of multimodal media, by fusing social semiotic theory with computer-based techniques and methodologies.

### 3. Multimodal interface design: challenges in bridging the gap between theory, design, and application

One of the many challenges software developers face in designing a digital framework interface that is theoretically explicable, principled and consistent, and – at the same time – accessible to a wide variety of research fields and applications, including the novice to multimodal social semiotics, is the formidable task of assimilating these two potentially competing interests or motivations. On the one hand, theoretical consistency requires that one specify, in the scientific manner, exactly how one is deploying certain terms – such as ‘mode’ and ‘semiotic resource’ for instance. On the other hand, one is developing a software tool for use by a variety of social users from (potentially) a variety of academic fields and disciplines, as well as practitioners, with differing academic registers and specialist terminology. Being clear about the use of one’s terms is clearly a necessity; but so too is accessibility of interface functionality, and a host of ‘foreign’ terms – even with detailed glossaries for their explanation – can offer a forbidding introduction to the ‘front gate’ of a software application.

#### **3.1 Developing a shared terminology**

##### 3.1.1 Definition of mode/medium/semiotic resource

Without wishing to engage in the ongoing debate amongst researchers/theorists about what constitutes a mode and/or medium, the authors are acutely aware that these concepts are often understood differently, even by researchers working within the same branch of social semiotic traditions. Nonetheless, 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’. However, whilst most multimodal researchers and theorists see *media* as the “physical stuff” of communication (Constantinou, 2005, p. 611), there appears to be far less agreement about the term *mode*. Kress & van Leeuwen (2001, p. 21), for example, distinguish between *mode*, which is on the ‘content’ side, 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

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

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).

However, rather than conflating these different dimensions of semiosis, in designing an effective digital interface developers may choose the alternative option of treating modes as primary sensory experiences, comprising the *visual*, *auditory*, and *somatic* mode (see O’Halloran, 2008b); the latter pertaining to sensory systems which have to be instantiated by the human subject<sup>1</sup> through the *semiotic resources* of kinetic action or movement, stance, posture, gesture, haptics (touch), facial expression, and so on.

The above definition not only proffers a clear distinction between the terms *mode* and *semiotic resource*, but forms the stepping stone for building a repository of options from which users can then choose to select those that are relevant to the respective phenomenon under analysis. Although LeVine & Scollon (2004, p. 2) posit that “there can be no mode that does not exist in some medium”, not all modes will be utilized in all types of media. As noted by Baldry & Thibault (2006, p. 4), “[d]ifferent semiotic modalities make different meanings in different ways according to the different media of expression they use”. That is why we propose, firstly, to differentiate between *static* and *dynamic* forms of media; considerations that will also have an impact on other design factors such as template layout<sup>2</sup>, and secondly, in terms of the semiotic modalities deployed (see Illustrations 1 and 2).

---

<sup>1</sup> It should be noted, however, that this semiotic mode can also be embodied by ‘non-human’ entities, such as animals, for example, who may be considered the primary ‘actors’ in nature documentaries, or the fictional ‘avatars’ prevalent in *Second Life* and computer games.

<sup>2</sup> 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 represented by sound, music and film (e.g., see Martinec, 2007; Baldry & Thibault, 2006; Rohlfing et al., 2006; Tan, forthcoming), static media may perhaps be best analyzed in an overlay editor that allows for annotations to be inserted directly onto the semiotic object.

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



**Illustration 1: Mock-up of Prototype Entry-Page**

Static media, such as a painting, a photo, a page-based advertisement, or a printed newspaper front-page, for example, 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).



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

(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”.

This, in turn, gives rise to the fundamental question – at least in terms of modeling the theoretical interface – of how the different semiotic modes and resources should be configured. For example, should written and spoken language be considered as separate semiotic resources, or rather ‘intra-semiotic phenomena when they involve just one mode of communication (e.g., the auditory, as in the case of tape-recorded conversations<sup>3</sup>), and inter- or cross-modal phenomena when they utilize more than one mode? At the same time, should we treat these phenomena as ‘mono-semiotic’ when they involve just one semiotic resource (such as speech, for example, in the case of taped conversations), and ‘multi-semiotic’ when they utilize two or more semiotic resources, as in the case of written text, which may draw on both language and typography for meaning making (see O’Halloran, 2008a; 2009c).

These issues strikes at the heart of software interface design: how to set up the interface so as to provide an intuitive but also theoretically consistent way to analyze and represent audiovisual and multimodal data. Is initially thinking of or approaching multimodal texts in terms of their perceptual categories – audio, visual and somatic – or their common social designations – written, spoken, gesture, body language – an effective and theoretically justified way to proceed? There is a undoubtedly a tension between the need to develop theoretically consistent models of such phenomena and the need to devise accessible categories and definitions that will enable the users of the software tool to feel that the terms and concepts they are being asked to negotiate as they interact with the interface are familiar and transparent.

---

<sup>3</sup> Many researchers and theories propose that all texts are inevitably ‘multimodal’. So, although research may be interested in just one semiotic mode or resource, such as speech in the case of tape-recorded conversations, the data itself will be a ‘de-contextualized’ record of a phenomenon that involves more than just one mode of communication. Anecdotal evidence suggests, for example, that many call centres require their front-line staff to use a mirror to ensure they are smiling when engaged with customers over the phone.

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

## 3.2 Principles of Organization

Other problems in interface design are bound up with principles of structural organization. For example, several of the existing social semiotic frameworks for the study of multimodal text draw on different dimensional principles of organization, including, but not limited to, metafunctions, stratification, rank-scale, delicacy, etc.

### 3.2.1 Metafunctional Orientation

As Martinec (2007, p. 157) observes, it seems to be a generally-held belief on the part of social semioticians who are influenced by Halliday's systemic functional theory that all semiotic modes and resources express ideational, interpersonal and textual meaning simultaneously. However, while some of these approaches adopt a metafunctionally-based framework as the primary organizational principle on the premise that metafunctions realize a higher order of meaning across modes and resources (O'Toole, 1994; Kress and van Leeuwen, 2006 [1996]; O'Halloran, 2004, 2008a), others, while firmly rooted in the tenets of social semiotic/systemic functional theory, do not 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 (van Leeuwen, 1999; Thibault, 2000; Baldry, 2004; Baldry & Thibault, 2006). The challenge of developing a theoretically sound and consistent digital interface is further compounded by the fact that there seems to be no common agreement on the descriptive terminology adopted by researchers and theorists, as Halliday's original metafunctional labels have been variably adapted and remodeled.

It is clear that the way in which we design our technical and semiotic interface for the exploration, analysis and presentation of multimodal data will both enable and constrain the scope of such exploration, results of such analysis and form of such presentation. For instance, if we were to organize the interface around the metafunctional dimension – in other words, assign all semiotic resources a place within the metafunctional framework – it would undoubtedly be seen as a powerful way of both analyzing and describing meaning potentials within established domains of enquiry, such as those based on language and visual design, and, in addition, it may also stimulate a certain type of exploration within less well described semiotic resources (such as music, for example). On the other hand, assigning metafunctional

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

roles to particular meaning-making acts inevitably represents a ‘top-down’ approach that describes just one of the effects of the realization process. Moreover, metafunctional realizations are inherently context-bound and not universally applicable across media, genres, and meta-discourses.

An alternative approach would be to allow users to explore the metafunctional orientation of a particular phenomenon without providing a ‘fixed’ preconception of its metafunctional orientation (that is, what role/s a semiotic resource is playing in terms of its metafunctional meaning potential), 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.

### 3.2.2 Rank-Scale Differentiation

In addition, developers are faced with the dilemma whether the digital interface should be organized in terms of a constituent rank-scale, as proposed by Halliday (1994, p. 35), which operates on (1) the principle of ‘exhaustiveness’ on the premise that every expression/representation fulfils some function at every rank; (2) the principle of ‘hierarchy’ based on notion that elements of any given rank are constructed out of elements of a rank below; and (3) the principle of ‘discreteness’ on the assumption that each structural unit has discrete boundaries.

Whilst researchers working more closely within a systemic functional tradition appear to favor a rank-based organization, this approach has been questioned by others, chiefly for pragmatic reasons. Van Leeuwen maintains, for instance, that the notion of rank “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), whilst 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,

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

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; forthcoming).

### 3.2.3 Models of Stratification

Another problem that confronts the researcher-developer is the question of whether to model the digital interface in terms of stratification, based on Hjelmslev’s model for language, which appears to be the preference of certain analysts working within a systemic functional tradition (such as represented by Thibault, 2000; Baldry, 2004; Baldry & Thibault, 2006; and O’Halloran 2008a). 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.

For Baldry & Thibault (2006, p. 224), on the other hand, who interpret the stratification model in terms of *display* and *depiction*, expression and content represent ‘two sides of the same semiotic coin’ in visual analysis. 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). Whereas the *expression stratum* of visual semiosis is based on the *display* of visual invariants and their transformation, 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.

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

Alternatively, 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 be theoretically integrated as illustrated in Table 2.

**Table 2: An Example of a Stratification Model for Mathematical Discourse (O’Halloran, 2009b)**

|                        |   |                                       |   |              |
|------------------------|---|---------------------------------------|---|--------------|
| <b>IDEOLOGY</b>        |   |                                       |   |              |
| <b>GENERIC MIX</b>     |   |                                       |   |              |
| <b>REGISTERIAL MIX</b> |   |                                       |   |              |
| ↑<br><b>CONTEX</b>     | ← <b>INTER-SEMIOSIS</b> →<br><b>MINI-GENRES, ITEMS, COMPONENTS and SUB-COMPONENTS</b> |                                       |   |              |
| <br><b>CONTENT</b>     | <b>LANGUAGE</b>   | <b>MATHEMATICAL<br/>VISUAL IMAGES</b> | <b>MATHEMATICAL<br/>SYMBOLISM</b>             | <b>OTHER</b> |
|                        | ← <b>INTER-SEMIOSIS</b> →<br><b>Discourse Semantics</b>                               |                                       |   |              |
|                        | Discourse   | Inter-Visual Relations<br>Work        | Inter-statemental<br>Relations                |              |
|                        | ← <b>INTER-SEMIOSIS</b> →<br><b>Grammar</b>   |                                       |   |              |
|                        | Clause complex<br>Clause<br>Word Group<br>Word  | Episode<br>Figure<br>Part             | Statements<br>Clause<br>Expression<br>Element |              |
|                        | ← <b>INTER-SEMIOSIS</b> →<br><b>Materiality</b>                                       |                                       |   |              |
|                        | Graphology, Typography and Graphics   |                                       |   |              |
| ↓<br><b>DISPLAY</b>    |   |                                       |   |              |

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

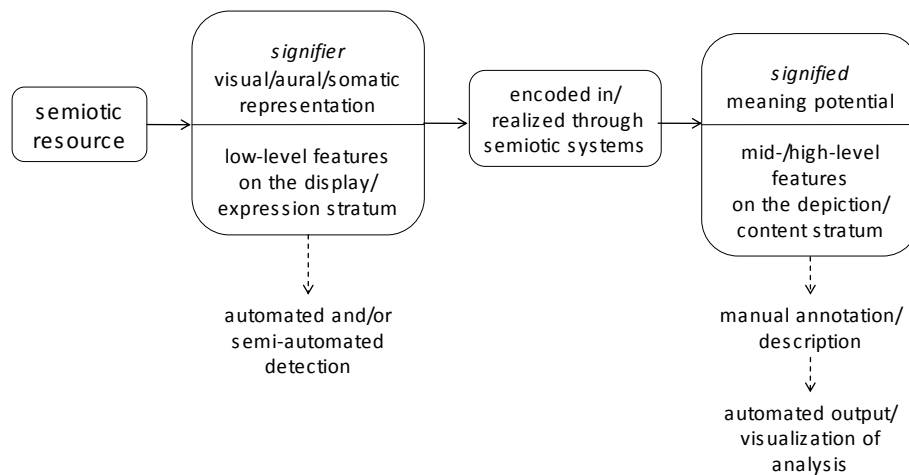
### 3.2 Fusing the Digital Interface with Social Semiotic Theory

Another issue that emerges is one that can also be traced back to more traditional semiotic studies: that of how to relate the different aspects of the multimodal communicative process to one another and within a holistic perspective of meaning as a unified act. These are issues not just of theory but also of practice in terms of software development. The configuration of a software tool for academic (as well as non-academic) purposes is mutually dependent upon abstract theoretical issues as well as how these are applied in the design of the digital interface, as well as the storage of analytical choices – such as annotations in different semiotic systems, resources and modes – in the relational database.

One of the strengths of multimodal social semiotic theories is that they are holistic systems which focus on semiotic resources and their inherent potential for meaning-making. Developing an integrative digital interface nevertheless requires us to re-focus our attention on the structural properties of semiotic systems *per se*. In order to understand how a social semiotic framework can (and needs to) be adapted in designing the digital interface, and perhaps more importantly, to recognize the limits of it, we need to consider the underlying basics of semiotic principles.

According to Turner (1994, p. 121), semiotics sees social meanings as the product of the relationships constructed between ‘signs’. The ‘*sign*’, he says, 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. According to Eggins (1994, p. 15), we have a sign when a meaning (content) is arbitrarily realized through a representation (expression). Eggins sees signs (in other words, semiotic resources) as the fusion or synthesis of content (meaning) and expression (the realization or encoding of that meaning). Designing the digital interface, however, requires us to ‘unpack’ semiotic resources, separating expression from content (see Figure 2 for elucidation).

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



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

In the digital interface, the *Signifiers*, i.e., 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).

Consequently, in the digital interface, the signified meaning potentials realized through semiotic systems can thus be equated to the mid- and high level features on the depiction/content stratum, as they will invariably involve an interpretative element that finds its expression in a higher-order form of user-annotated or computer-assisted output or translation. 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

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

for moving away from pure annotation-based analysis, essentially freeing the analyst to attend to the higher-level aspects of interpretation. Furthermore, computer-assisted semi-automated analysis (e.g.; one-click systemic annotation) 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.

### **3.4 One template/framework interface? Or many?**

Another consideration in the design of a digital interface, apart from the question of how and whether the different approaches to the study of multimodality can indeed be harmonized within a single framework for interface design, is, of course, the all-important 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).

As van Leeuwen (2005, p. 4) points out, “social semiotics resources are signifiers, observable actions and objects that have been drawn into the domain of social communication and that have a theoretical semiotic potential constituted by all their past uses and all their potential uses and an actual semiotic potential constituted by those past uses that are known to and considered relevant by the users of the resource, and by such potential uses a might be uncovered by the users on the basis of their specific needs and interests”. He is adamant that “[s]uch uses take place in a social context, and this context 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). Van Leeuwen (2005) draws our attention to the complexity of meanings available in society at large, in which a single, monolithic repository of semiotic resources will never be able to account for all the possible meanings that a given expression or realization will have in a given context. He nevertheless believes 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”, including the “meanings that have not yet been recognized, that lie, as it were, latent in the object, waiting to be discovered” (van Leeuwen,

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

2005, pp. 4-5). Accordingly, this entails building inventories that are not made with an immediate, urgent purpose in mind (see van Leeuwen, 2005, p. 6).

Consequently, for the software developer, this may mean building not one but many generic framework interfaces for several well-researched and documented popular media genres, such as film and advertising, for example, on the basis of meaning potentials that have already been established in society for a particular phenomenon or event, and leaving the ‘inventorizing’ of meanings that have yet to be discovered in a particular area of practice or specialized research field to the potential future users of our digital interface. Flexibility is thus a guiding principle in our software development agenda; but again, this within the context of usability and accessibility: the functional flexibility must not come at the cost of increased complexity.

### **Conclusion**

As we are confronted with these questions and challenges in the design of digital interfaces and functionalities, we ask ourselves how these tensions identified above are indeed resolvable in a way that both theory and application can harmonize. Although our project is still in the initial stages of development, we have already found that what might appear at first to be a simple task – that is, developing a digital interface for the analysis of multimodal data – in fact represents an exercise in modeling theory, or rather multiple theories. The layout of the interface, configuration of elements; arrangement of ‘annotation strips’ and frames for multiple presentation and analyses of a variety of interacting semiotic systems and phenomena; the decisions as to what categories to include and how to arrange them in such a way as to facilitate particular types or styles (traditions) of analysis; the presentation/visualization (and auralization) of data and of analysis: all of these issues and many more are essentially grounded in the theoretical frameworks and models that one adopts as the basic ‘template’ and motivating guide in developing multimodal tools and technologies for the study of multimodal phenomena. Multimodal interactive digital technologies thus provide for the researcher the same advantages that they provide for the wider community: an opportunity to represent and manipulate multi-modal and multi-semiotic phenomena in texts (and theoretical models of such) in such a way as to increase our (academo-)cultural semiotic potential, and understanding and appreciation of that potential.

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

### 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.
- 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.
- 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.
- Eggins, S. (1994). *An Introduction to Systemic Functional Linguistics*. New York: Continuum.
- 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*. RoutledgeFalmer.
- 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.

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

- 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, David (2007). *Introduction to Multimodal Analysis*. London: Hodder Arnold; New York: Oxford University Press.
- 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.
- 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). *Accumulating Mathematical Knowledge across Semiotic Resources and Modes*. Plenary Paper presented at Disciplinarity Knowledge and Language Symposium, University of Sydney, 8-10 December 2008.
- 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://courses.nus.edu.sg/Course/ellkoh/pubs.html>
- O'Halloran, K. L. (2009b). 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*. London: Equinox.

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

- O'Halloran, K. L. (2009c, forthcoming). 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.
- 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 Guijjaro (Eds.), *The World Told and The World Shown: Multisemiotic Issues*. Palgrave Macmillan: Houndmills, Basingstoke, Hampshire.
- 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 Mmultimediality 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*. Houndsmills, Basingstoke, Hampshire and London: Macmillan Press Ltd.
- 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.