

# Multimodal Methods for Researching Digital Technologies

Carey Jewitt

This chapter provides an introduction to the field of multimodality and discusses its potential application for researching digital data and environments. It begins by outlining what multimodality is, its theoretical origins in social semiotics and its underlying assumptions. A number of concepts central to multimodality are introduced: these include mode, semiotic resource, materiality, modal affordance, multimodal ensemble and meaning functions. The scope and potential of multimodality for researching digital technologies are then discussed. The chapter sets out an illustrative example of multimodal research. It concludes with a discussion of the limitations and challenges of a multimodal approach for researching digital technologies.

## WHAT IS MULTIMODALITY?

Multimodality is an interdisciplinary approach drawn from social semiotics that understands communication and representation as more than language and attends systematically to

the social interpretation of a range of forms of making meaning. It provides concepts, methods and a framework for the collection and analysis of visual, aural, embodied and spatial aspects of interaction and environments (Jewitt, 2013; Kress, 2010). While other modes of communication, such as gesture, have been recognized and studied extensively (e.g. McNeil, 1992), multimodality investigates the interaction between communicational means and challenges the prior predominance of spoken and written language in research (Scollon and Scollon, 2009). Speech and writing continue to be understood as significant but are seen as parts of a multimodal ensemble. Multimodality emphasizes situated action – that is, the importance of the social context and the resources available for meaning making, with attention to people's situated choice of resources, rather than emphasizing the system of available resources. Thus it opens up possibilities for recognizing, analyzing and theorizing the different ways in which people make meaning and how those meanings are interrelated.

Multimodality provides resources to support a complex fine-grained analysis of artifacts and interactions in which meaning is understood as being realized in the iterative connection between the meaning potential of a material semiotic artifact, the meaning potential of the social and cultural environment it is encountered in and the resources, intentions and knowledge that people bring to that encounter. That is, it strives to connect the material semiotic resources available to people with what they mean to signify in social contexts. Changes to these resources and how they are configured are therefore understood as significant for communication. Digital technologies are of particular interest to multimodality because they make a wide range of modes available, often in new inter-semiotic relationships with one another, and unsettle and remake genres in ways that reshape practices and interaction. Digital technologies are thus a key site for multimodal investigation.

Underlying this approach is the idea that language, and other systems or modes of communication (e.g. gesture, gaze), is shaped through the things that it has been used to accomplish socially in everyday instantiations, not because of a fixed set of rules and structures. This view of language as a situated resource encompasses the principle that modes of communication offer historically specific and socially/culturally shared options (or 'semiotic resources') for communicating. With this emphasis, a key question for multimodality is how people make meaning in context to achieve specific aims.

Three interconnected theoretical assumptions underpin multimodality. These are briefly introduced and discussed below.

The first assumption underlying multimodality is that, while language is widely taken to be the most significant mode of communication, speech or writing are a part of a multimodal ensemble. Multimodality 'steps away from the notion that language always plays the central role in interaction, without denying that it often does' (Norris, 2004: 3) and proceeds on the assumption

that all modes have the potential to contribute equally to meaning. From a multimodal perspective, language is therefore only ever one mode nestled among a multimodal ensemble of modes. While others have analyzed 'non-verbal' modes, multimodality differs in that language is not its starting point, nor does it provide a prototypical model of all modes of communication. The starting point is that all modes which are a part of a multimodal ensemble – a representation and/or an interaction – need to be studied with a view to the underlying choices available to communicators, the meaning potentials of resources and the purposes for which they are chosen.

The second assumption central to multimodal research is that all modes have, like language, been shaped through their cultural, historical and social uses to realize social functions as required by different communities. Therefore each mode is understood as having different meaning potentials or semiotic resources and to realize different kinds of communicative work. Multimodality takes all communicational acts to be constituted of and through the social. This also draws attention to the ways in which communication is constrained and produced in relation to social context and points to how modes come into spaces in particular ways.

This connects with the third assumption underpinning multimodality – that people orchestrate meaning through their selection and configuration of modes. Thus the interaction between modes is significant for meaning making. Multimodal communication is not in and of itself new, communication has always been multimodal, however, digital media has led to an increased interest in the multimodal character of communication as it foregrounds the need to consider the particular characteristics of modes, multimodal configurations, and their semiotic function in contemporary discourse worlds (Ventola et al., 2004). The meanings in any mode are always interwoven with the meanings made with those of other modes cooperating in

the communicative ensemble. The interaction between modes is itself a part of the production of meaning.

## A BRIEF BACKGROUND

Multimodality was developed in the early 2000s (see Kress and van Leeuwen, 2001; Kress et al., 2001, 2005; van Leeuwen, 2005; Jewitt, 2009). It originated from linguistic ideas of communication, in particular the work of Michael Halliday on language as a social semiotic system. Halliday's work shifted attention from language as a static linguistic system to language as a social system – how language is shaped by the ways that people use it and the social functions that the resources of language are put to in particular settings. In *Language as Social Semiotic* (1978) Halliday sets out a theory of language built on a social functional perspective of meaning and a framework for understanding language as a system of options and meaning potentials: in summary the idea of meaning as choice.

Hodge and Kress in *Social Semiotics* (1998), and later Kress and van Leeuwen in *Reading Images* (2006), expanded attention from language to other semiotic systems (or modes), laying the groundwork for extending and adapting social semiotics across a range of modes and opening the door for multimodality. Kress and van Leeuwen extended principles developed in relation to language to the visual. They examined visual texts to identify a range of semiotic resources, meaning potentials, available choices and the organizing principles underpinning their configuration to visually communicate ideologies and discourses. Multimodality has taken ideas from linguistics that are theoretically transportable to other modes, such as turn taking, coherence, composition, and it has explored the currency of these in relation to the particularities of other modes. In doing so it has extended and adapted Halliday's conception of meaning across a range of modes by taking the specific resources and

organizing principles of spoken and written language as a starting point and extending their essence to other modes in ways recognize and the resources of gesture, gaze, and image differ in significant ways. As multimodality has developed it has also looked beyond linguistics for resources to assist with analysis and to further explore the situated character of meaning making, including sociolinguistics, film theory, art history and iconography and musicology.

Multimodality foregrounds the modal choices people make and the social effect of these choices on meaning. There is therefore a strong emphasis on the notion of context within social semiotic multimodal analysis. The context shapes the resources available for meaning making and how these are selected and designed. Signs, modes and meaning making are treated as relatively fluid, dynamic and open systems, intimately connected to the social context of use. From this perspective analytical interest in the modal system (its resources and principles) is strongly located in (and regulated through) the social and cultural. When making signs, people bring together and connect the available form that is most apt to convey the meaning they want to express at a given moment.

Kress introduced a strong emphasis on the social character of meaning and developed the concept of the motivated sign (Kress, 1997). This served to foreground the agency of the sign maker and the process of sign making. In *Before Writing* (Kress, 1997) he offers a detailed account of the materiality and processes of young children's engagement with texts and how they interpret, transform and redesign the semiotic resources and signs available to them – what has been described as chains of semiosis. From this perspective, signs (e.g. talk, gesture and textual artifacts) are analyzed as material residues of a sign maker's interests. The analytical focus is on understanding their interpretative and design patterns and the broader discourses, histories and social factors which shape that. In a sense, then, the text is seen as a window on to its maker. Viewing signs as motivated and

constantly being remade draws attention to the interests and intentions that motivate a person's choice of one semiotic resource over another (Kress, 1993). This 'interest' connects a person's choice of one resource over another with the social context of sign production – returning to the importance of meaning as choice within social semiotic theories of communication. The modal resources that are available to the person are an integral part of that context – hence the importance of multimodality to understanding the process of meaning making.

Multimodality can, at least in part, be understood as a response to the demands to look beyond language in a rapidly changing social and technological landscape. It is curious to understand how the use of digital technologies extends the range of resources for communication, reshapes the relationship between resources such as image and writing, and has the potential to significantly reconfigure notions of spatiality and embodiment as well as genre conventions, all of which can lead to adapted and some new types of texts and interactions.

## KEY CONCEPTS

This section outlines in more detail six concepts introduced above that are key for multimodality: mode, semiotic resource, materiality, modal affordance, multimodal ensembles and meaning functions.

### *Mode*

This term refers to a set of socially and culturally shaped resources for making meaning: a 'channel' of representation or communication (Kress and van Leeuwen, 2001). One definition of a mode is that it has to comprise a set of elements/resources and organizing principles/norms that realize well-acknowledged regularities within any one community. That is something which can only be recognized as a mode when it is a known/usable system of communication within a community. The

ability for the 'grammar' of the modal system to be broken is seen as a 'test' that it exists. Another 'test' for whether or not a set of resources can count as a mode is if it is possible for it to articulate all three of Halliday's (1978) meaning functions – that is, can a set of resources be used to articulate 'content' matter (ideational meaning), construct social relations (interpersonal meaning) and create coherence (textual meaning). Accepted examples of modes include writing, image, moving image, sound, speech, gesture, gaze and posture in embodied interactions. What constitutes a mode is the subject of debate. For instance, van Leeuwen (1999) has explored when sound and music can be thought of as modes, while Bezemer and Kress (2008) have discussed whether color and layout can be considered as modes. As these examples suggest, modes are created through social processes, fluid and subject to change – not autonomous and fixed. For example, the meanings of words and gestures change over time. Modes are also particular to a community/culture where there is a shared understanding of their semiotic, rather than universal, characteristics.

### *Semiotic Resource*

This term is used to refer to a means of meaning making that is simultaneously a material, social and cultural resource. In other words a semiotic resource can be thought of as the connection between representational resources and what people do with them:

Semiotic resources are the actions, materials and artifacts we use for communicative purposes, whether produced physiologically – for example, with our vocal apparatus, the muscles we use to make facial expressions and gestures – or technologically – for example, with pen and ink, or computer hardware and software – together with the ways in which these resources can be organized. Semiotic resources have a meaning potential, based on their past uses, and a set of affordances based on their possible uses, and these will be actualized in concrete social contexts where their use is subject to some form of semiotic regime. (van Leeuwen, 2005: 285)

This definition highlights the historical development of connections between form and meaning, aligned with Bakhtin's notion of intertextuality. Kress (2010) emphasizes that these resources are constantly transformed. This theoretical stance presents people as agentive sign-makers who shape and combine semiotic resources to reflect their interests.

### **Materiality**

Materiality refers to how modes are taken to be the product of the work of social agents shaping material, physical 'stuff' into cultural semiotic resources. This materiality has important semiotic potentials in itself: sound has different affordances from written inscription, while gesture offers different material potentials from colour, and so on. All modes, on the basis both of their materiality and of the work that societies have done with the material (e.g. working sound to become speech or music) offer specific potentials and constraints for making meaning. The materiality of modes also connects with the body and its senses, that in turn place the physical and sensory at the heart of meaning.

### **Modal Affordance**

The term modal affordance is contested and continuously debated within multimodal research. It originated from the psychologist James Gibson's (1979) work on perception and agent-situation interaction, which defined affordances as the 'action possibilities' latent in an environment, and in which the potential uses of any object arise from its perceivable properties in relation to how it is perceived by an actor's capabilities and interests. Donald Norman later took up this term in relation to the design of artifacts, with an emphasis on both the material and social dimensions of materiality (1990).

Adapted by Kress (e.g. 2010), the term 'modal affordance' refers to the potentialities and constraints of different modes – what it

is possible to express and represent or communicate easily with the resources of a mode and what is less straightforward or even impossible – and this is subject to constant social work. From this perspective, the term 'affordance' is not a matter of perception, but, rather, is a complex concept connected to both the material *and* the cultural, social and historical use of a mode. Modal affordance is shaped by how a mode has been used, what it has been repeatedly used to mean and do and the social conventions that inform its use in context. As indicated by van Leeuwen's definition of semiotic resource, where a mode originates, its history of cultural work, its provenance, shapes the meaning potential of a semiotic resource. These affordances contribute to the different communicational and representational potentials or modal logics of modes (although it is important to note these are open to change and disruption). The affordances of the sounds of speech, for instance, usually happen across time and this sequence in time shapes what can be done with (speech) sounds. The logic of sequence in time is difficult to avoid for speech: one sound is uttered after another, one word after another, one syntactic and textual element after another. This sequence becomes an affordance or meaning potential: it produces the possibilities for putting things first or last or somewhere else in a sequence. The mode of speech is therefore strongly governed by the logic of time. Like all governing principles they do not hold in all contexts and are realized through the complex interactions of the social as material and vice versa – in this sense the material constitutes the social and vice versa. Modal affordance suggests all modes are partial in making meaning, so that the designed selection of modes, into multimodal ensembles, allows this partiality to be managed.

### **Multimodal Ensembles**

Representations or interactions that consist of more than one mode can be referred to as



a multimodal ensemble. The term draws attention to the agency of the sign maker – who pulls together the ensemble within the social and material constraints of a specific context of meaning making. Multimodal ensembles can therefore be seen as a material outcome or trace of the social context, available modes and modal affordances, the technology available and the agency of an individual. When several modes are involved in a communicative event (e.g. a text, a website, a spoken interchange) all of the modes combine to represent a message's meaning (e.g. Kress et al., 2001, 2005). The meaning of any message is, however, distributed across all of these modes and not necessarily evenly. The different aspects of meaning are carried in different ways by each of the modes in the ensemble. Any one mode in that ensemble is carrying a part of the message only: each mode is therefore partial in relation to the whole of the meaning, and speech and writing are no exception (Jewitt and Kress, 2003). Multimodal research attends to the interplay between modes to look at the specific work of each mode and how each mode interacts with and contributes to the others in the multimodal ensemble. This raises analytical questions, such as which modes have been included or excluded, the function of each mode, how meanings have been distributed across modes and what the communicative effect of a different choice would be. At times the meaning realized by two modes can be 'aligned', at other times they may be complementary and at other times each mode may be used to refer to distinct aspects of meaning and be contradictory or in tension. Lemke noted (2002: 303) 'No [written] text is an image. No text or visual representation means in all and only the same ways that text can mean. It is this essential incommensurability that enables genuine new meanings to be made from the combinations of modalities'. Modal affordance in the context of multimodal ensembles raises the question of what image is 'best' for and what words, as well as what other modes and their arrangements are 'best' for in a particular context. The

relationships between modes as they are orchestrated in interactions (and texts) may themselves realize meanings through particular modal combinations, different weightings of modes (Martinec and Salway, 2005) or modal density in an ensemble (Norris, 2009). The structure of hyperlinks, for example, realizes connections and disconnections between elements that may contribute to the expansion of meaning relations between elements. The question of what to attend to, what to 'make meaningful' is a significant aspect of the work of making meaning and is foregrounded by a multimodal focus. Further, as meaning makers decide on modal 'best fit' and how to combine modes for a particular purpose, analysis of the moment-by-moment processes of constructing multimodal ensembles can enable the analyst to unpack how meanings are brought together.

### *Meaning Functions*

As noted earlier, multimodality is built on a functional theory of meaning, an idea of meaning as social action realized through people's situated modal choices and the way they combine and organize these resources into multimodal ensembles. It distinguishes between three different but interconnected categories of meaning choices (also called meta-functions) that are simultaneously made when people communicate.

1. Choices related to how people realize content meanings (known as Ideational meaning) – that is, the resources people choose to represent the world and their experience of it, for example, what is depicted about processes, relations, events, participants and circumstances;
2. Choices related to how people articulate Interpersonal meanings – that is, the resources people choose to represent the social relations between themselves and those they are communicating with – either directly via interaction or via a text or artifact. For example, the visual or spatial depiction of elements as near and far, direct or oblique, are resources used to orientate viewers or inter-actors to a text or one another;

3. Choices concerned with textual or organizational meaning – for example, the choice of resources such as space, layout, pace and rhythm for realizing the cohesion, composition and structure of a text or interaction.

Multimodality applies these meaning functions to all modes to better understand their meaning potential – ‘what can be meant’ or ‘what can be done’ with a particular set of semiotic resources – and to explore how these three interconnected kinds of meaning potentials are actualized through the grammar and elements of their different modal systems.

A key point to draw attention to here is that the concepts outlined in this section can be applied across any kind of representation or interaction – be it a printed or digital text (Jewitt, 2002), a classroom with or without technology (Jewitt et al., 2011) or a complex interaction in a digitally mediated environment such as a surgical operating theatre (Bezemer et al., 2011). Thus, a researcher can employ multimodality to investigate the modal meaning potentials of a resource (e.g. mobile application, tangible environment) as well as how people make use of these resources in interaction.

### THE POTENTIAL OF MULTIMODALITY FOR RESEARCHING DIGITAL TECHNOLOGIES

This section gives a sense of the scope and potential of multimodality for researching digital technologies: how it has been used to date, the kinds of questions it can be used to address and what research insights it can provide to inform the evaluation of technology design and use. The following four potentials of multimodal research are discussed in this section.

1. The systematic description of modes and their semiotic resources.
2. Multimodal investigation of interpretation and interaction with specific digital environments.

3. Identification and development of new digital semiotic resources and new uses of existing resources in digital environments.
4. Contribution to research methods for the collection and analysis of digital data and environments within social research.

### *The Systematic Description of Modes and their Semiotic Resources*

A multimodal approach can be used to create an inventory of the meaning potentials available to people when using a technology in a particular context. This may be done through a systematic description of the modes and their semiotic resources, materiality and modal affordances and the organizing principles of a device and/or application. Building on the notion of meaning as choice and the concept of the meta-functions, some multimodal researchers use a style of diagramming called system networks to map the meaning potentials of a mode. This is a diagrammatic taxonomy of the systematic, semiotic options that are possible within a semiotic or lexico-grammatical system. This maps the potential of modal resources to articulate content, interpersonal and textual or organizational meanings – in an artifact or interaction. The options should preferably be of the either/or type. As described by Kress and van Leeuwen (2006), for instance, a visual image may either be a ‘demand for information’ (a kind of visual question) or an ‘offer of information’ (a kind of statement) – it cannot be both. A ‘demand for information’, in turn may be either ‘polar’ (yes/no question), or open, and so on. When analyzing other modes than language, some semiotic relations are better described as scaled along a continuum – for example the semiotic dimensions of color have been mapped as a set of continuum scales concerning hue, brightness, luminosity and so on (Kress and van Leeuwen, 2002). System networks provide an analytical tool for mapping the range of semiotic resources and options made available by a

mode in a given context. In this way system networks provide a way to push the formal analysis of a mode (or a semiotic resource) to a logical limit.

To date system networks have been used to describe the semiotic options available within a range of modes including language (Halliday and Matthiessen, 2004), visual communication and color (Kress and van Leeuwen, 2002, 2006), action (Martinec, 2000) and sound, voice and music (van Leeuwen, 1999), as well as three-dimensional objects (e.g. tables, Bjorkvall, 2009). Networks have been used to explore multimodal genres and multimodal ensembles including online newspapers (Knox, 2007; Caple and Knox, 2012), film and media texts (Bateman, 2008) and interactive media texts (White, 2012).

In the case of digital texts, mediated interaction and environments, multimodal inventories can be of use in both understanding the meaning-making potentials and constraints that different technologies place on representation, communication and interaction, and how users of those technologies notice and take up those resources in different ways. This can inform both the redesign of technological artifacts and environments as well as their introduction into a set of practices, for example for learning or work.

### ***Multimodal Investigation of Interaction in Specific Digital Environments***

Multimodal researchers have also used system networks to focus on how modal resources are taken up and used in a specific context. They map and compare people's choice of mode, semiotic resources in specific contexts and some examine how these modal choices are shaped by the materiality and affordances of a mode and the research subjects' knowledge and experience. Some multimodal researchers, particularly those who are focused on meaning making as a process and are thus perhaps less concerned with mapping the

resources of the mode itself, use system networks as a much looser heuristic tool to explore meanings. Multimodal studies investigate how these resources are used in specific contexts and how people talk about them, justify them and critique them in order to understand how semiotic resources are used to articulate discourses across a variety of contexts and media, such as school, workplaces, online environments, textbooks and advertisements.

The import of the body and spatiality in the contemporary digital landscape is evident in emergent bodily interaction-based technologies (Price et al., 2009). Much work has been done on the classroom as a multimodal environment of learning and the role of position, posture, gesture and gaze has been shown to be key to learning and teaching in the production of school English and science (e.g. Kress et al., 2001, 2005). Multimodal attention to how bodily modes and space feature in interaction – their semiotic resources and affordance has potential for researching digital technologies. For instance, Wii games serve to reconfigure the relationships between players' physical (and therefore social relationship) bodies, now with digital sensory feedback via wristbands and body straps, virtual avatars, and the screen in ways that require physical digital mapping in interesting ways for what it means to collaborate and 'play together'. Multimodality provides a set of resources to describe and interrogate these remappings – for example to get at the interaction across the 'physical' and the 'virtual' body. This type of digital remapping and extending of the physical is paramount in a range of digitally remediating contexts. The question of how screens and digital technologies mediate the role of the body is also relevant for understanding online multimodal interaction. Jones in his analysis of how people construct and consume multimodal displays of their selves in social networking environments examines 'how the different digital technologies available for producing and consuming displays affects the kinds of



relationships that are possible between users of these sites and the kinds of social actions that these displays allow them to take' (Jones, 2009: 82). A focus on mode, semiotic resources, materiality and modal affordance provides a descriptive language for examining interaction in these complex sites. For instance, multimodal research in a surgical operating theatre shows the interactional impact of digital technologies being inserted into older established social environments (Bezemer et al., 2011). Surgeons undertaking keyhole surgery work in screen-based digital environments that, like the Wii, reorientate their gaze, body posture, team configurations and require them to engage in physical–digital mapping. A multimodal approach also asks if the use of blended physical–digital tools of applications like those discussed here generate new forms of interaction and enable new action, as well as physical, perceptual and bodily experiences.

Multimodality has been applied to a range of multimodal digital genres to explore questions of digital identities and literacy, notably in the field of education (Alvermann, 2002; Jewitt and Kress, 2003; Marsh, 2006). It has also been used to analyze the orchestration of music, filmic shots and editing features in video productions, digital animation and games (e.g. Walton, 2004; Burn, 2009), as well as online environments, (Jones, 2009) and more recently interactions with mobile and Geographic Information System (GIS) technologies (Hollett and Leander, this volume).

The relationships across and between modes in multimodal texts and interaction are a central area of multimodal research, and multimodal research often investigates the relationship between a given context and the configuration of modes in a text or situated interactions – both to better understand the modal resources in use and to address substantive questions. The textual or organizational meta-function has been a focus of this work, for instance understanding how multimodal cohesion (van Leeuwen, 2005) is realized (or not) through the integration of different semiotic resources in multimodal

texts and communicative events via rhythm, composition, information linking and modal density or intensity (Norris, 2004).

The ways in which contemporary digital texts are organized via textual features such as digital layering and hyperlinking and the impact of this on how people navigate multimodal digital texts has also been examined (Lemke, 2002; Zammit, 2007). This work is potentially useful when thinking about the take up of designed resources (e.g. Jewitt, 2008). There is a large body of multimodal research that explores the dynamics of the interaction between image and language. This includes the early work of Kress and van Leeuwen (2006) on the visual articulation of meaning, Lemke's (1998) work on the role of image and writing in science textbooks, work by Martinec and Salaway (2005) rethinking Barthes's classification of image–text relations, and Bezemer and Kress's (2008) development of a framework for the analysis of image, writing, typography, color and layout in school textbooks. Focusing on multimodal texts, for instance, Kress and Bezemer investigated the learning gains and losses of different multimodal ensembles of learning resources in science, mathematics and English from the 1930s to the first decade of the twenty-first century, including digitally represented and online learning resources. They provide a multimodal account of the changes to the design of learning resources and their epistemological and social/pedagogic significance. They conclude that image and layout are increasingly meshed in the construction of content and color so that layout and typography can increasingly be seen as communicative modes. With a focus on multimodal interaction Jewitt, in her book *Technology, Literacy and Learning* (2008), explores the fundamental connection between a range of modal resources (including color, image, sound, movement and gesture and gaze), digital technologies, knowledge, literacy and learning. In this and other work she shows how teacher and student engagement with the modal resources made available by technologies reshapes practices such

as reading and writing and the particular ways in which students and teachers interact in school science and English classes and looks at the impact of this on learning. These studies show how digital technologies stretch, foreground and in some cases remake modes, semiotic resources and their configurations in contemporary materiality and modal affordances, as well as the intersemiotic relations possible in multimodal ensembles.

### ***Identification and Development of New Digital Semiotic Resources and New Uses of Existing Resources***

In addition to creating inventories of modes and semiotic resources, and analyzing how these are used in a range of specific contexts, multimodality contributes to the discovery and development of new semiotic resources and new ways of using existing semiotic resources.

Studying the semiotic potential of a given semiotic resource is studying how that resource has been, is, and can be used for purposes of communication, it is drawing up an inventory of past and present and maybe also future resources and their uses. By such inventories are never complete, because they tend to be made for nature specific purposes. (van Leeuwen, 2005: 17)

The discovery and development of new modal resources is linked to social change and society's need for new semiotic resources and new ways of using existing semiotic resources as the communicational landscape changes. Two factors central to this are the potentials of digital technology and the importing of semiotic resources in a global society. Digital synthesizers and other digital technologies, for example, have reshaped the possibilities of the 'human' voice to create new semiotic resources and contexts for the use of 'human' voices – in digital artifacts, public announcements, music, and so on (van Leeuwen, 2005). This digital reshaping of voice has in turn impacted the non-digital use of voice – for example, providing different

tonal or rhythmic uses of the non-digital voice not previously imagined. Modal semiotic resources common to print-based texts, such as textual linking, layering, layout and the organization of time, are also foregrounded and reconfigured in significant ways by digital technologies. Knox (2007), for example, has explored how online newspapers have reshaped newspaper layout, genres and the relationship of image, writing and video, and has mapped the 'wash-back' influence from online to print-based newspapers, as well as reading pathways (Knox, 2007; Caple and Knox, 2012). Adami (2009) has examined the multimodal patterns of coherence and turn-taking on the social networking site YouTube. Multimodal tools also have the potential to identify and describe the reconfigurations of space, time and embodiment that digital technologies (e.g. mobile and GIS) make available and address questions about how these technologies influence people's interaction and experiences.

Multimodality moves beyond intuitive ideas about what a technology can do, to provide detailed analysis of the way semiotic resources of digital technologies work, what they can and cannot do. It enables the construction of explicit understandings of a form of communication and thus makes it possible for these to be discussed, taught and evaluated. Multimodality can also help to design and implement new uses for semiotic innovations.

### ***Contribution to Research Methods***

Researchers increasingly need to look beyond language to better understand how people communicate and interact in digital environments. This places new demands on research methods with respect to digital texts and environments where conventional concepts and analytical tools may need rethinking. Multimodality makes a significant contribution to existing research methods for the collection and analysis of data and environments within social research. It provides methods for the collection and analysis of

types of visual digital data including screen capture data and eye-tracking data (e.g. see Holsanova, 2012) and researcher-generated and naturally occurring digital video data (e.g. Kress et al., 2001, 2005; Norris, 2004; Bezemer and Jewitt, 2010). The use of digital video technology and a multimodal focus pose what has become a key challenge for social research, namely how to transcribe or *re-present* multimodal data (e.g. movement in time and space). Increasingly, the topic of transcription is subject to innovation and experimentation in multimodal approaches. This might range from the inclusion of line drawings and stills from video footage to the use of software such as Comic Life and Transana (e.g. Baldry and Tibault, 2005; Plowman and Stephen, 2008; Flewitt et al., 2009; Bezemer and Mavers, 2011). As already discussed, multimodality provides tools for mapping and analyzing the visual, embodied and spatial features of interaction with digital technologies as well as the analysis of music, film, digital animation, games, adverts and other new media (e.g. Knox, 2007; Burn, 2009; Jones, 2009; Adami, 2009; Caple and Knox, 2012).

Having outlined the scope and potential of a multimodal approach for researching digital technologies in general terms, the following section illustrates its application.

### ILLUSTRATIVE CASE STUDY

This short case study concerns the learning of mathematical concepts in a digital programming game environment and is focused on the interaction of two students (aged seven years) with the resources of Playground, an object-orientated programming tool (Jewitt and Adamson, 2003). The excerpt discussed here focuses on how the students' emergent conception of 'bounce' was shaped through their selection and use of the modal resources available to them (the full case study is reported in Jewitt, 2008).

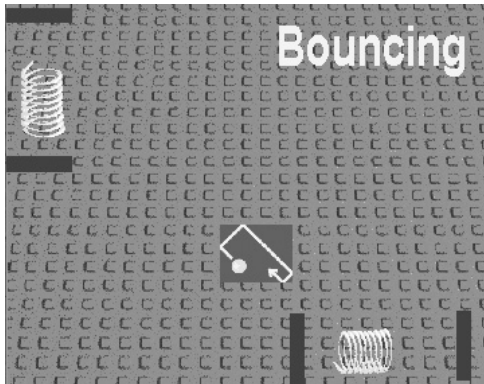
In the students' original design using paper and pen, the game concerned a small creature

being chased by an alien that fired bombs to catch it. The movement of their characters (a creature and an alien) and bounce of the bullets were realized using modes and semiotic resources drawn from a static image, writing and cartoon-visual genres (e.g. a time-lapse drawing and wiggly lines to signify vibration and the sound of an explosion).

Programming the game in Playground offered the students additional modes and semiotic resources for their design, notably ready-made visual elements and backgrounds, color, movement and sound and the removal of the written mode. Detailed analysis of the students' game as a product as well as video data of the process of production shows that these modal resources demanded different kinds of representational commitment, design decisions and thinking on the part of the students. The move from the page to screen underpinned changes in ideational, interpersonal and textual meaning, resulting, for instance, in increasing the stakes for the little creature: now it will be killed instead of being caught, suggesting a shift in the students' understanding of the affordance (social rules and expectations) of genre from board game to adventure/action game on the screen. The students' digital redesign of the multimodal frame of the game redefined the game narrative and the necessity to consider the movement of the elements. In addition, they needed to specify the spatial and dynamic relationships between the elements in the game.

In the students' written description of the game, bounce is represented as 'the bombs go sideways by arrows and then if [the bomb] touches the bars it goes different ways'. That is, bounce is represented as a matter of movement and change of direction when something is touched. The semiotic resources and affordance of writing as a mode do not require the students to make explicit the 'cause' of this change in movement – the player, the bomb or the bars.

The digital environment of Playground represents the idea of bounce in three modes and each provides different semiotic resources for the students' construction of the entity

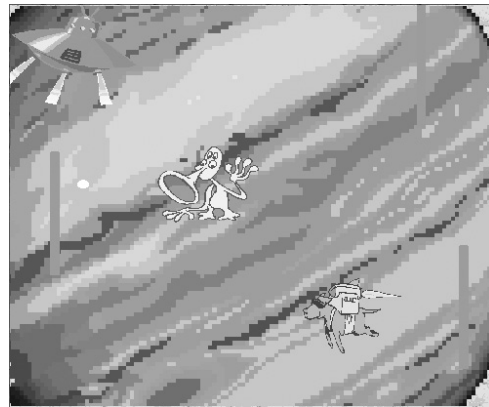


**Figure 17.1** The program for 'bounce'.

'bounce'. Figure 17.1 shows the program for 'bounce'. It uses the mode of writing – the word 'Bouncing' – to name and classify the movement in everyday terms. It uses the mode of still image – two images of a spring and an image of a ball – to specify particular potentials of bounce as a mechanical and regularly ordered entity rather than an organic, unpredictable bouncing (e.g. a rabbit). Third, it uses the mode of animated movement – three repeated animated sequences, one of a spring moving up and down between two bars, another of a spring moving sideways between two bars and a third sequence of a ball moving at angles within a square. The animated sequences work to give meaning to the entity 'bounce' in the context of the Playground program.

These modal resources work together as a multimodal ensemble to associate the (ideational) meaning of 'bouncing' within the mathematical paradigm of the system. This introduction of movement as a design resource raised a key question for the students in their design, 'What is it that produces bounce?' and 'What is it that bounces?'

Initially, the students in their game design (shown in Figure 17.2) programmed the sticks to bounce (that is, they added the behaviour of bouncing to the sticks), placed them on the game and then played the game. The sticks bounced off. It was the visual experience of playing the game that led the



**Figure 17.2** The students' game design in Playground.

students to realize their mistake and how to rectify it.

Through their engagement with the Playground environment the students worked out their ambiguities about agency – ambiguities that the affordances of writing and static image in their own paper design masked.

The students used gaze and gesture as a resource to address these questions and the process of programming bounce in their game. They students created different kinds of spaces on the screen through their gesture and gaze with/at the screen itself and their interaction with and organization of the elements displayed on the screen. These spaces marked distinctions between the different kinds of practices that the students were engaged with. In their creation and use of these spaces the students set up a rhythm and distinction between game planning, game design and construction and game playing. The students gestured 'on' the screen to produce a plan of the game: an 'imagined-space', overlaying the screen in which they gesturally placed elements and imagined their movement, and used gesture and gaze to connect their imagined (idealized) game with the resources of the application as it ran the program. The temporary and ephemeral character of gesture and gaze as modes enabled their plans of the game to remain fluid and ambiguous.

The role of gesture was central in their unfolding programming of the bouncing behaviour in three ways.

1. Gestures gave a way into understanding how the students are thinking about the concept 'bounce'. Initially the two students' talk and gesture is strongly coordinated and suggestive of a shared vision of how they imagine the bullet moving (from the alien to the left stick, then to the top-right stick). When the students stop acting in unison, however, two alternative versions of the movement of the bullet emerge (Figure 17.3). Student 1 traces the bullet moving in a *vertical* line down to the bottom-right stick. She then traces it in a *horizontal* line to the dog, wiggles the pen to indicate somewhere in that area. She is working with the entity 'bounce' as a generalized concept of movement, as going from one place to another. Student 2 works with the entity 'bounce' as a more specialized kind of movement. She indicates that a bullet would not move in a perpendicular line from the top-right to the bottom-right stick (as gestured by student 1). Holding her finger on the top-right stick she then gesturally traces an 'imagined' stick to the right of the alien before slowly trailing her finger off the edge of the screen. This 'gestural overlay' adds another stick to the visual design of the game, which in turn enables her to imagine the movement of the bullet bouncing from the top-right stick to the bottom-right stick, then off past the dog.
2. Examining the students' use of gesture in this way helped to identify areas of difficulty. The two students' accounts both end with a faltering tone

of voice and lexical (e.g. 'whatever', 'ends somewhere') and gestural vagueness of wiggles and trailing off. These gestures are material signs of uncertainty of how the movement of a bullet would come to an end if it did not hit the dog. Would the ball keep bouncing or would it go off screen? This is itself an uncertainty of what is *producing the bounce*, is it the ball or the something that is hit by the ball.

3. The students' use of gesture can be analyzed to explore their hypothesis. Student 2 used a gestural overlay to 'estimate' where the ball would bounce, which in turn led to the amendment of the game – student 2's suggestion that they need to place some horizontal sticks on the planet.

The *invisibility*, the visual absence, of the bullets at this stage of the design is what proves to be problematic for the students. They prioritized the meaning of the visual within the multimodal ensemble of the game and, modally, at that point in the game-design, the students were working visually and not multimodally. The students were looking at the game to decide where to 'attach' the bounce. The 'sticks' (bars) were visible on screen but the bullets are 'within the alien' and are only visible when the game is being played. In this visual mode of working the system does not make the bullets available as something that the students could specify as the object that the 'I bounce' refers to. In short, when working visually, the notion of agency depends on visual presence. In sum, what was made visible on the

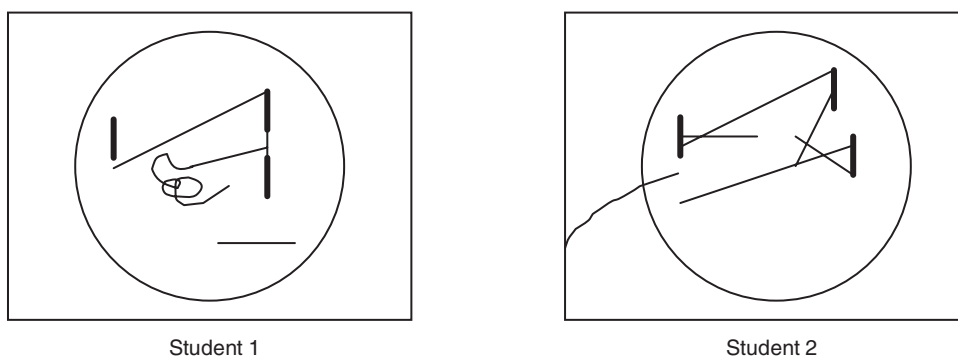


Figure 17.3 The students' gestures with the screen.



screen proved to be particularly important in the students' design process. The students appeared to associate visual presence with agency: 'If it couldn't be seen it couldn't be acting' seemed to stand behind the students' programming process.

This example shows how the availability of multimodal resources changes the representations that students are working with, as well as the work of interpreting them, particularly what it is that the students need to attend to and what they need to specify. Finally, it highlights the potential of examining multimodal interaction and the range of representational resources available on screen to understand technology-mediated learning.

## LIMITATIONS AND CHALLENGES

Although multimodal research has much to offer, it also has several limitations. A criticism sometimes made of multimodality is that it can seem rather impressionistic in its analysis. How do you know that this gesture means this or this image means that? In part this is an issue of the linguistic heritage of multimodality – that is, how do you get from linguistics to all modes. In part it is the view of semiotic resources as contextual, fluid and flexible – which makes the task of building 'stable analytical inventories' of multimodal semiotic resources complex. It is perhaps useful to note that this problem exists for speech or writing. The principles for establishing the 'security' of a meaning or a category are the same for multimodality as for linguistics and other disciplines. It is resolved by linking the meanings people make (whatever the mode) to context and social function. Increasingly, multimodal research looks across a range of data (combining textual/video analysis with interviews, for example) and towards participant involvement to explore analytical meanings as one response to this potential problem.

Linked with the above problem of interpretation is the criticism that multimodality

is a kind of 'linguistic imperialism' that imports and imposes linguistic terms on everything. These critics overlook the fact that much of the work on multimodality has its origins in social semiotic theory of communication and the social component of this perspective sets it apart from narrower concerns with syntactic structures, language and mind and language universals that have long dominated the discipline. This view of communication can be applied (in different ways) to all modes.

Multimodal analysis is an intensive research process, both in relation to time and labor. Multimodal research can be applied to take a detailed look at 'big' issues and questions through specific instances. Nonetheless, the scale of multimodal research *can* restrict the potential of multimodality to comment beyond the specific to the general. The development of multimodal corpora may help to overcome some of these limitations, as might the potential to combine multimodal analysis with quantitative analysis in innovative ways.

## CONCLUSION

This chapter has provided an introduction to the field of multimodality. It has discussed what multimodality is, sketched its theoretical origins and presented its underlying assumptions. Throughout the chapter the key concepts central to this approach have been introduced, discussed and illustrated through their application within the literature and in the case study example presented above. In this way the chapter has set out the scope and potential of multimodality for researching digital technologies with reference to four significant areas: (1) the systematic description of modes to research meaning making in complex digitally mediated environments and the evaluation and design of multimodal digital artifacts, interactions and experiences; (2) the investigation of interpretation and interaction in specific digital environments; (3) the identification

and development of new digital semiotic resources and new uses; and (4) a contribution to research methods. Finally, the chapter points to some of the limitations and challenges of a multimodal approach for digital technologies.

## REFERENCES

- Adami, E. (2009) 'We/YouTube': exploring sign-making in video-interaction', *Visual Communication*, 8(4): 379–399.
- Alvermann, D. (ed.) (2002) *Adolescents and Literacies in a Digital World*. New York: Peter Lang.
- Baldry, A. and Tibault, P. (2005) *Multimodal Transcription and Text Analysis*. Harlow: Equinox.
- Bateman, J. (2008) *Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents*. Harlow: Palgrave Macmillan.
- Bezemer, J. and Jewitt, C. (2010) 'Multimodal analysis', in L. Litosseliti (ed.), *Research Methods in Linguistics*. London: Continuum. pp. 180–197.
- Bezemer, J. and Kress, G. (2008) 'Writing in multimodal texts: a social semiotic account of designs for learning', *Written Communication*, 25(2): 166–195.
- Bezemer, J. and Mavers, D. (2011) 'Multimodal transcription as academic practice: a social semiotic perspective', *International Journal of Social Research Methodology*, 14(3): 191–207.
- Bezemer, J., Murtagh, G., Cope, A., Kress, G. and Kneebone, R. (2011) "'Scissors, please": the practical accomplishment of surgical work in the operating theatre', *Symbolic Interaction*, 34(3): 398–414.
- Bjorkvall, A. (2009) 'Practical function and meaning: a case study of IKEA tables', in C. Jewitt (ed.), *The Routledge Handbook of Multimodal Analysis*. Abingdon, Oxon: Routledge.
- Burn, A. (2009) *Making New Media: Creative Production and Digital Literacies*. New York: Peter Lang.
- Caple, H. and Knox, J. (2012) 'Online news galleries, photojournalism and the photo essay', *Visual Communication*, 11(2): 207–236.
- Flewitt, R., Hampel, R., Hauck, M. and Lancaster, L. (2009) 'What are multimodal data and transcription?', in C. Jewitt (ed.), *Routledge Handbook of Multimodal Analysis*. Abingdon, Oxon: Routledge. pp. 40–53.
- Gibson, J. (1979) *The Ecological Approach to Visual Perception*. Hillsdale, NJ: Lawrence Erlbaum.
- Halliday, M.A.K. (1978) *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.
- Halliday, M.A.K. and C. (2004), *An Introduction to Functional Grammar*. London: Hodder Education
- Hodge, R. and Kress, G. (1988) *Social Semiotics*. Cambridge: Polity Press.
- Holsanova, J. (2012) 'Methodologies for multimodal research', *Visual Communication* (Special Issue), 11(3).
- Jewitt, C. (2002) 'The move from page to screen: the multimodal reshaping of school English', *Visual Communication*, 1(2): 171–196.
- Jewitt, C. (2008) *Technology, Literacy and Learning: A Multimodal Approach*. Abingdon, Oxon: Routledge.
- Jewitt, C. (ed.) (2013) *The Routledge Handbook of Multimodal Analysis* (2<sup>nd</sup> edition). Abingdon, Oxon: Routledge.
- Jewitt, C. and Adamson, R. (2003) 'The multimodal construction of rule in computer programming applications', *Education, Communication and Information*, 3(3): 361–382.
- Jewitt, C., Bezemer, J. and Kress, G. (2011) 'Annotation in school English: a social semiotic historical account', *National Society for the Study of Education Yearbook*, 110(1): 129–152.
- Jewitt, C. and Kress, G. (eds) (2003) *Multimodal Literacy*. New York: Peter Lang.
- Jones, R. (2009) 'Technology and sites of display', in C. Jewitt (ed.), *Routledge Handbook of Multimodal Analysis*. Abingdon, Oxon: Routledge. pp. 114–126.
- Knox, J. (2007) 'Visual-verbal communication on online newspaper home pages', *Visual Communication*, 6(1): 19–36.
- Kress, G. (1993) 'Against arbitrariness: The social production of the sign as a foundational issue in critical discourse analysis', *Discourse and Society*, 4(2): 169–191.
- Kress, G. (1997) *Before Writing: Rethinking Paths to Literacy*. London: Routledge.
- Kress, G. (2010) *Multimodality: A Social Semiotic Approach to Contemporary Communication*. Abingdon, Oxon: Routledge.
- Kress, G., Jewitt, C., Bourne, J., Franks, A., Hardcastle, J., Jones, K. and Reid, E. (2005) *English in Urban Classrooms: A Multimodal Perspective on Teaching and Learning*. Abingdon, Oxon: Routledge.
- Kress, G., Jewitt, C., Ogborn, J. and Tsatsarelis, C. (2001) *Multimodal Teaching and Learning*. London: Continuum.
- Kress, G.R. and van Leeuwen, T. (2001) *Multimodal Discourse: The Modes and Media of Contemporary Communication*. London: Edward Arnold.
- Kress, G. and van Leeuwen, T. (2002) 'Colour as a semiotic mode: Notes for a grammar of colour', *Visual Communication*, 1(3): 343–368.

- Kress, G. and van Leeuwen, T. (2006) *Reading Images: The Grammar of Visual Design*. (2nd revised edition). Abingdon, Oxon: Routledge.
- Lemke, J. (1998) 'Metamedia literacy: Transforming meanings and media', in D. Reinking, M. McKenna, L. Labbo and R. Kieffer (eds), *Handbook of Literacy and Technology: Transformations in a Post-Typographic World*. Hillsdale, NJ: Erlbaum. pp. 283–302.
- Lemke, J. (2002) 'Travels in hypermodality', *Visual Communication*, 1(3): 299–325.
- Marsh, J. (2006) 'Global, local/public, private: Young children's engagement in digital literacy practices in the home', in K. Pahl and J. Rowsell (eds), *Travel Notes from the New Literacy Studies*. Clevedon, North Somerset: Multilingual Matters. pp. 19–39.
- Martinez, R. (2000) 'Types of processes in action', *Semiotica*, 130 (3/4): 243–268.
- Martinez, R. and Salway, A. (2005) 'A system for image–text relations in new (and old) media', *Visual Communication*, 4(3): 337–371.
- McNeil, D. (1992) *Hand and Mind: What Gestures Reveal about Thought*, Chicago, IL: University of Chicago Press.
- Norman, D. (1990) *The Design of Everyday Things*. New York: Doubleday.
- Norris, S. (2004) *Analyzing Multimodal Interaction*. Abingdon, Oxon: RoutledgeFalmer.
- Norris, S. (2009) 'Modal density, modal configurations: Multimodal actions', in C. Jewitt (ed.) *Routledge Handbook of Multimodal Analysis*. Abingdon, Oxon: Routledge. pp. 78–90.
- Plowman, L. and Stephen, C. (2008) 'The big picture? Video and the representation of interaction', *British Educational Research Journal*, 34(4): 541–565.
- Price, S., Roussos, G., Falcão, T.P. and Sheridan, J.G. (2009) 'Technology and embodiment: relationships and implications for knowledge, creativity and communication', [www.beyondcurrenthorizons.org.uk](http://www.beyondcurrenthorizons.org.uk) (accessed 24 August 2012).
- Scollon, R. and Scollon, S. (2009) 'Multimodality and language: a retrospective and prospective view', in C. Jewitt (ed.), *The Routledge Handbook of Multimodal Analysis*. Abingdon, Oxon: Routledge. pp. 170–180.
- Van Leeuwen, T. (1999) *Speech, Music, Sound*. London: Macmillan.
- Van Leeuwen, T. (2005) *Introducing Social Semiotics*. Abingdon, Oxon: Routledge.
- Ventola, E., Charles, C. and Kaltenbacher, M. (2004) *Perspectives on Multimodality*. Amsterdam: John Benjamins Publishing.
- Walton, M. (2004) 'Behind the screen: the language of web design', in I. Snyder and C. Beavis (eds), *Rewriting Literacy in the Network Society*. Hampton, Australia: New Dimensions.
- White, P. (2012) 'Reception as social action: the case of marketing', in S. Norris (ed.), *Multimodality in Practice* (Chapter 11). Abingdon, Oxon: Routledge.
- Zammit, K. (2007) 'The construction of student pathways during information-seeking sessions using hypermedia programs: A social semiotic perspective'. Unpublished PhD thesis, University of Western Sydney.