

# Qualitative Research

<http://qrj.sagepub.com/>

---

## **Multimodal ethnography**

Bella Dicks, Bambo Soyinka and Amanda Coffey

*Qualitative Research* 2006 6: 77

DOI: 10.1177/1468794106058876

The online version of this article can be found at:

<http://qrj.sagepub.com/content/6/1/77>

---

Published by:



<http://www.sagepublications.com>

**Additional services and information for *Qualitative Research* can be found at:**

**Email Alerts:** <http://qrj.sagepub.com/cgi/alerts>

**Subscriptions:** <http://qrj.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations:** <http://qrj.sagepub.com/content/6/1/77.refs.html>

>> [Version of Record](#) - Jan 24, 2006

[What is This?](#)

---

## Multimodal ethnography

BELLA DICKS, BAMBO SOYINKA AND AMANDA COFFEY  
Cardiff University

**ABSTRACT** Ethnographers, like other researchers, currently have a broad range of media at their disposal for conducting fieldwork, for aiding analysis and, most challengingly, for representing their completed work. These include digital media such as photographs, video film, audio-recordings, graphics and others besides. Through the computer 'writing space', these media can be integrated together, alongside more conventional written interpretation, into hypermedia environments. However, integration poses a number of potential problems, which this article addresses through a discussion of the semiotics of multimedia. In particular, it argues that different media can be seen to 'afford' different kinds of meaning. The integration of different media, therefore, has potentially significant implications for ethnography. Rather than seeing these media forms as discrete, we suggest an approach to ethnographic work which sees meaning as emerging from the fusion of differently mediated forms into new, 'multi-semiotic' modes. We therefore recognize the need to go beyond the current interest in visual methods, and instead to develop ways of understanding what kinds of meanings are produced in multimodal ethnographic work.

**KEYWORDS:** *hypermedia, multimedia, multimodality, qualitative data analysis, qualitative research methods, representation*

### Introduction

Ethnography is now situated within a world saturated by multimedia technologies. And ethnographers are increasingly utilising a range of communicative resources in their work – including recorded sound, still and moving images, as well as speech and writing. These are often used together within the same

research project. However, our own experiments suggest that mixing media is a complex project, one which is as yet relatively untheorized. A key challenge facing the researcher is how to understand the various semiotic resources (or *modes*) that diverse *media* involve (a distinction to which we return below). This article discusses some of the issues involved in multi-semiotic ethnography, albeit in a preliminary way. Our starting point is that a multi-modal ethnography is not simply a mosaic. Instead (and adapting the arguments of Kress, 1998), we can see it as a new multi-semiotic form in which meaning is produced through the inter-relationships between and among different media and modes. As yet, we lack the conceptual framework to codify how these complex inter-relationships work to produce particular kinds of meaning.<sup>1</sup> What we wish to offer here is a preliminary exploration of how the potential of multimedia in ethnographic fieldwork might begin to be understood.

#### MULTIMEDIA IN THE FIELD VERSUS IN RESEARCH RECORDS

Our first step is to distinguish clearly between data 'in' the field and the means of recording (or, more properly, *representing*) those data for subsequent analysis. One of the problems is that multimedia is often considered a methodological issue only as a feature of the data-records that ethnographers construct (e.g. the photographic images or video footage that they take), rather than as an inherent feature of the worlds they study. But multimedia within the field is not the same thing as multimedia within data-records. Data are the represented world as we know and experience it, rather than the 'world in itself' (Bauer et al., 2000). In other words, data are what we are able to *perceive* in the field. Further, we perceive them through all of our senses, including sight, hearing, touch, smell and even taste. Clearly, data in the field are by their very nature composed of diverse media (they are likely to include sounds, objects, visual designs, people's actions and bodies, etc.). Data, then, are necessarily composed of a diverse and shifting range of media.

If data are intrinsically multimedia, the media that appear in our data-records are necessarily more restricted. For data analysis purposes, we transform our observations of the phenomenal world into a separate set of materials that reduce it to permanent recordings (through media technologies such as fieldnotes, camera images, etc.). The media available to do this – from pen to video camera – are much more restricted than those occurring in the field. Video footage, for example, limits the information recorded to that amenable to audio-capture and camera-work. Fieldnotes limit it to writing. Yet it is here, in the data-records, that methodological issues of multimedia tend to be discussed and 'analysis' conducted. As Emmison and Smith (2000) point out in a discussion focused on the visual realm, social science has traditionally seen the visual only in two-dimensional terms – predominantly as camera images (still or moving). These have been treated as if they were themselves 'data' – and consequently have been under-valued as compared to written records. They argue:

Stated in its bluntest form our reservations about an image-based social science rest on the view that photographs have been misunderstood as constituting forms of data in their own right when in fact they should be considered in the first instance as *means of preserving, storing or representing information*. In this sense photographs should be seen as analogous to code-sheets, the responses to interview schedules, ethnographic fieldnotes, tape recordings of verbal interaction or any one of the numerous ways in which the social researchers seek to capture data for subsequent analysis. (Emmison and Smith, 2000: 2; our italics)

Emmison and Smith argue that the visual should be studied in all elements of the research setting itself – including three-dimensional objects, environments, structures, design and so forth. Visual data are not ‘what the camera can record but . . . what the eye can see’ (Emmison and Smith, 2000: 4). Following this, we propose seeing the media produced by field researchers, whether these are images, sound or written records, not as themselves ‘data’ but as ways of *representing* multimedia field data. In this sense, it is necessary to expand what we think of as media. All kinds of media inevitably characterize the field, and we need to be sensitive to the meanings produced by their various forms. Furthermore, in making records or representations of that field, ethnographers need to think about how the field’s ‘multimediality’ has been reduced, or re-produced, through the recording medium chosen. In order to consider these two levels of multimediality – that of the field, and that of the representational media chosen, we proceed by unpacking a little further what the term ‘multimedia’ means. We do this by introducing a distinction between *modes* and *media*. Accordingly, in what follows, we first draw attention to the various kinds of meaning-making (semiotic *modes*) that allow communication to occur in the worlds we study as ethnographers. We then consider what happens to this multimodality when we try to record it and represent it in the form of recorded observations and materials (or data-records). Our intention is to raise a series of questions and suggested areas for future development, rather than trying to construct a definitive series of methodological recommendations.

## OUR PROJECT

This article draws on work-in-progress currently being undertaken by members of the Cardiff research team as part of an ESRC-funded project on digital ethnography. The project builds on previous ESRC-funded work on hypermedia ethnography undertaken by members of the same team (Dicks and Mason, 1998; Dicks et al., 2005; Mason and Dicks, 2001). The current project has undertaken an ethnographic research project from scratch and ‘done’ all phases of the research process in digital multimedia and, specifically, hypermedia, form. We have also, however, incorporated the tried and tested methods of conventional ethnography: namely, observation and the writing of fieldnotes and memos. Our final ethnography is being prepared in hypermedia form on DVD. The project is both methodological and substantive.

Methodologically, we are trying to explore the potential of digital hypermedia for all phases of ethnographic research, and it is the methodological aspects that are the subject of this article.

Our substantive research site is an interactive science discovery centre in Wales. The centre offers a number of different visitor attractions under the same roof, of which here we will discuss only the two major ones (the exhibits hall and the science theatre). Figures 1–5 and Figure 6 show a selection of field photographs of each venue.

First, there is an exhibits hall (Figure 1A–C), in which are situated multiple gadgets, machines, computers, 3-D puzzles, toys and devices all designed to be easily operated by children so as to help them appreciate selected scientific principles. We call these by the generic term ‘exhibit’. Second, there is a science theatre (Figure 6), in which performances of scientific themes are presented on stage by live presenters to visiting groups of schoolchildren. Our project seeks to understand what kinds of scientific knowledge are being produced in these two venues, and how these are being communicated to visitors, particularly schoolchildren. For example, how does scientific knowledge – produced elsewhere – become *reproduced* in the centre in the guise of interactive material exhibits, dedicated to the production of entertainment as well as education? How do children interact with these exhibits and to what extent do they learn ‘science’ from them? Through interviews and observations (recorded digitally, in various ways) we have sought to understand and document the complex production processes through which the exhibits and the performances are



FIGURES 1A–C. *Photographs of the exhibits hall in the science discovery centre*

FIGURE 1A. *View of the exhibit hall from the first floor*



FIGURE 1B. *View of the exhibit hall from gallery*



FIGURE 1C. *View of the exhibit hall, children playing with blocks*

designed and constructed. We have also been able to observe visitor interactions in order to understand how science in these two novel forms is 'consumed'. Our analysis thus focuses on linking together the spheres of production and consumption in the communication of science.

#### MEDIA AND MODES

In order to understand how the science centre produces meaning (including scientific meanings), we need to understand how its various elements work together to produce a communicative environment. For example, how do the three-dimensional material structures visible in the pictures earlier shape the ways in which science is communicated in this space? The photographs you see here succeed in communicating a number of things about the objects, through their size, shape, position and so forth (including their colour, though here they are necessarily reproduced in black-and-white). But these are only their visual dimensions. In order to understand how they produce meaning, we need to remember that the different physical materials used (plastics, wood and glossy paint, for instance) contribute to this, as do the ways in which embodied visitors physically interact with the three-dimensional *solidity* of the objects. Understanding these dimensions involves undertaking analysis *in the field*, and cannot be undertaken purely from visual records such as photographs. Moreover, we suggest that it involves coming to terms with the distinctive ways in which different media communicate, and this is the essence of our argument here. We suggest that the concept of *multimodality*, as theorized by linguists, is a useful one to aid in this project (see Iedema, 2003; Kress and Van Leeuwen, 2001, for discussion). It is based on the distinction in linguistics between modes and media. *Modes* are the abstract, non-material *resources* of meaning-making (obvious ones include writing, speech and images; less obvious ones include gesture, facial expression, texture, size and shape, even colour). Media, on the other hand, are the specific material forms in which modes are realized, including tools and materials (Kress and Van Leeuwen, 2001: 22). Modes cannot be directly observed, for they are abstract resources; they are rule-governed, codified sets of meaning-resources, involving the idea of formal 'grammars' – for example, the 'grammar' of film and that of writing (though 'grammar' is not always the right term; visual images, for example, are more lexically ordered – i.e. as a loose collection of icons that can be arranged in an indefinite number of ways [Kress and Van Leeuwen, 2001: 113]).

Why is this distinction useful for ethnographers? What we actually observe in the field are the various *media* in which these modes are produced – marks on the page, movements of the body, sounds of voices, pictures on the wall. If these are to be seen as communicative – i.e. capable of affecting the user in particular ways – it is useful to see how each mobilizes a particular set of meaning-resources. For example, expressing something through speech as opposed to writing affects what is said. Science represented in printed images communicates graphically; in embodied performance through a range of



physical media: in each case, what is communicated changes. The concept of multimodality draws attention to how modes co-occur with other modes (graphical, sound-based, gestural, etc.), and how, accordingly, analysis needs to reflect this (Iedema, 2003). It might be objected that the mode/medium distinction is misleading, in that it suggests that meaning is somehow 'attached' to material structures rather than actually inhering in them. And it is certainly true that what ethnographers observe in the field are meaningful bodies, environments and structures (not a technical assemblage of colours, shapes, positions, textures, etc.). This relates to Geertz's point (1973) about the difference between naturalist and interpretivist perspectives: twitches, winks, fake-winks and parody-winks may 'objectively' be the same physical movement, but are all communicatively different. In distinguishing between modes and media, we are not suggesting that meaning is somehow separate from form. Our argument, following theorists of multimodality, is that the media themselves provide part of the signifying power of every observed entity. If one observes a person winking at someone else, the meaning conveyed is not just *affected* by the fact that the speaker has used a gesture rather than a spoken word; it is at least partially *constituted* by that choice, for it belongs to a language of gestures in which we are all competent. Similarly, the physical texture of an exhibit helps to endow it with meaning (as we shall see later) not just in terms of the object at hand, but because we have internalized a 'language of textures' that tells us that a certain kind of physical sensation is to be equated with a certain set of meanings. If we touch something and it feels 'wrong' – for example, if food looks delicious but feels slimy, or cushions look soft but feel hard – then the meanings received are confused and we react accordingly. Designers today know this well; hence their careful attention to the feel, sound, taste and smell of designed goods as well as to their look.

For example, we might want to consider how any field-setting space is organized via the entire ensemble of material objects and structures within it – their positions, colours, shapes and so forth. The aim is to understand how *this* material organization (as opposed to other possible ones) contributes to the production of ethnographic meaning in the research setting. How, for example, does the lay-out of a classroom's furniture, the colours of its walls and the shapes of the desks and toys within it work to anchor particular meanings of childhood, learning and pedagogy? A teacher, for example, may try to convey the scientific meaning of 'physical forces' to schoolchildren through a range of media: by showing them a text-book; a chalk-board drawing; or pushing a door open. In each case, the modes deployed are different, and so the children's perceptions differ too. The distinction between modes and media thus draws attention to the different resources or 'languages' that allow people and environments to communicate. Rather than focusing solely on observable media, it encourages appreciation of the different (or, indeed, similar) kinds of meaning that different media afford. This focus on 'affordances' allows the ethnographer to appreciate how meaning is made



'multi-semiotically', across a variety of media acting in multiple combinations with each other (Kress and Van Leeuwen, 2001: 67).

It follows that in order to understand the meanings of any environment, we need to understand how its various semiotic modes and media work together to produce a particular ensemble of meaning-effects (and these need to be understood, crucially, through including the 'consumption' side of the communicative process: i.e. the ways in which actual users/participants interact with and interpret them). To aid interpretation of these meanings, they somehow need to be representable in a permanent form so that we can analyse them at leisure. But as soon as we use recording technologies (cameras, fieldnotes, video), we are working with a much reduced range of media and modes than those occurring in the field. Let's turn to our own research project for an illustration.

### *Observing multimodality in the field*

In the exhibits hall at the science centre, we can identify three principal means by which the exhibits communicate meaning:

1. *Action/reaction sequences.* The exhibits area is structured around the principle of human/machine interaction. Each exhibit communicates by responding in a specified way (e.g. pumping out a gust of wind) when a certain action is carried out by the user (e.g. pressing a button). Rather than in the reading of written texts (as in traditional models of classroom-based science), here communication occurs in the performance of actions. By setting in play a highly controlled and precise interchange of human action/machine reaction, the exhibits supposedly reveal a principle of science. For example, at the back of the hall is a huge granite ball (the Kugel, see Figure 2), glistening with water, sitting on a plinth. It is



FIGURE 2. *The Kugel Ball exhibit*

obviously massively heavy. When someone pushes the ball with their finger (action), they find to their amazement that they can make it revolve (reaction). What is supposedly communicated here is an appreciation of the incredible forces exerted by the intense pressure of the water-bed it is sitting on. Notably, however, the written mode is not dispensed with altogether; the instructions panel explains that it is not sitting directly on the plinth at all, but on a thin bed of water under intense pressure. The writing 'anchors' the meaning of the action/reaction sequence.

2. *Material semiotics*. Exhibits also communicate meanings through their physical materiality (see Figure 3). This includes their *colour*, *texture*, *shape*, *position*, *opaqueness/light*, *weight*. Each exhibit is encased in a single, wooden module that is hard and solid (*texture* and *shape*). Further, each exhibit is positioned at a certain height, and at a certain distance from other exhibits (*physical position*). And each exhibit casing is machine spray-painted to give a highly glossy finish (*light*) in bright, primary colours of yellow, blue, green and red (*colour*). All these different modes communicate meaning in their own terms. For example, they produce a space which is deliberately reminiscent of children's play areas. In contributing to the overall meaningfulness of the exhibits space, they help to shape the ways in which it is 'read' and experienced by visitors (i.e. as a fun, safe, playful space).



FIGURE 3. *The Wind Blower exhibit*

3. *Interactivity.* Another sphere of multi-modality opens up when we consider the interactions of actual visitors with the exhibits. Indeed, none of the meaning-potential highlighted so far becomes realized until it is picked up and responded to by actual users. We observed a class of 6–7-year-olds as they moved around the exhibit hall. We also initiated spontaneous conversations with them, designed to find out how they were using the exhibits and what the exhibits were communicating to them. Here, we have the human modes of *speech* (the children's, the helpers', the teachers' and the ethnographers' speech), *gesture*, *facial expression*, *bodily movement*, etc. By analysing these various modes in combination, rather than attending to the children's spoken language alone, we can gain an insight into how users are interacting bodily with the exhibits space (Figure 4). We can also try and map the *constraints* operating in that space, which work to shape users' responses to it in particular directions.

All the above modes (colour, texture, speech, position, light, gesture and so forth) are intrinsic to the lived experience of the discovery centre. But what do they tell us about the wider cultural meanings upon which it is drawing? One of the strengths of Kress and Van Leeuwen's approach is that they do not isolate study of the sign from its embeddedness in social contexts. In other words, they see meaning as being produced socially – through the levels of discourse, production, design *and* consumption, as well as 'text'. For example, they argue that a mode accomplishes three things: it 'allows discourses to be



FIGURE 4. *The Benouli Blower exhibit*

formulated in particular ways'; it 'constitutes a particular kind of interaction' and 'can be realised in a range of different media' (2001: 22). We suggest that the action-reaction sequences (described earlier in 1) can be seen as a mode in that they communicate science as a particular kind of 'thing' – as human-initiated mechanical movement, in this case. This reflects the kind of science that is foregrounded in the centre, which is primarily (though not exclusively<sup>2</sup>) mechanical, practical and applied – rather than theoretical or abstract. The centre's emphasis is always on this human-initiated physicality as the learning experience itself (the written caption translates it into more abstract principles, but the 'lesson' is contained in the action-reaction sequence itself). This can be seen as fully consonant with the discourse of task-centred, fun, action-oriented pedagogy favoured by proponents of 'active learning' (reflected in the mantra 'hands-on/minds-on' often repeated by staff at the centre). We could also suggest that the action/reaction mode 'constitutes a particular kind of interaction' (Kress and Van Leeuwen, 2001: 22), which (on the whole though not exclusively) takes place between a single user and a machine. This replicates an individualist, skills-centred and competitive approach to learning, in that each task involves the rewarding or frustrating of individual performance. While a user interacts with the exhibit, other children gather and watch as an 'audience'. Children are thereby positioned as the active performers of designated tasks, rather than the passive recipients of expert knowledge as in older models of classroom and text-book learning (Kress, 1998). The point is that in considering how different modes are employed by different media in the communication of science we can see how *this* particular environment (e.g. a discovery centre) produces science differently to *that* one (e.g. a classroom), and how these differences may be related to historical shifts in discourses of pedagogy and learning. The analysis presented here is necessarily a truncated one, as we wish to move on and consider other aspects of multimodality. Nevertheless, we hope it has served to illustrate how attending to the specificity of any one medium (i.e. the communicative modes it deploys) aids in the interpretation of the social setting under study.

### *Recording multimodality*

In order to give the reader a flavour of the research setting, we have had to rely on photographic images and the written word, which struggle to convey the full multimodality observable within the study setting. To grasp the non-visual modes of texture, solidity and weight, it was necessary to spend time in the hall, moving around the exhibits bodily, interacting with them and experiencing the physical flow that the space encourages. Hence, neither our photographs (as *still images*) nor our written fieldnotes (as *graphical writing*) could reproduce the multi-modal, living, material, kinetic environment that is the science centre itself. In this section, we examine these records as forms of representation in their own right. They each afford a distinctive kind of semiotic potential, but

clearly there are significant cross-overs as well. To illustrate the kinds of comparison we have in mind, we can reexamine the photographs reproduced in Figures 1–4.

#### EXAMPLE 1: PHOTOGRAPHS VERSUS WRITING

Photographs allow us to see modes that are visual: colour, shape, size, position, light. What they do not show us are modes that operate through the other senses – of touch, smell, hearing and taste – such as bodily movement, texture, three-dimensional shape, sounds. If we compare this characteristic of images to writing, certain differences are apparent. For example, consider the following written description of the exhibits hall (Figure 5).

It should be clear from perusing the written description in Figure 5 that a number of modes are in operation. Or rather, one could say that writing employs primarily one mode, verbal language,<sup>3</sup> but that this mode is a particular one in that it allows other modes to be described in their *absence*. So in using language, the modes of colour, sound, shapes, textures as well as actions and movement can be easily conveyed through linguistic description. But in comparison with the photograph of the exhibits hall (Figure 1a) the written description below *appears* more subjective, pointing out just two exhibits for special attention. It affords a particular perspective – the perspective of a narrator who is ‘in’ the scene being described selecting out particular elements for our attention. It also offers the evaluation and description of emotional responses (‘amazingly’, ‘excitedly’). The photographs, too, disclose a perspective – one determined by the position of the camera. Though the camera gaze

When you enter [the science centre], you see a huge, well-lit hall in front of you with high ceilings and a gallery above. It is a white space with big white pillars and vast windows, but there's also lots of colour, noise and movement. On entering, immediately ahead of you is a plastic ball, suspended seemingly by magic in front of a bright yellow solid pyramid. Beyond are more yellows and reds and greens and blues. All around are a variety of different brightly coloured devices or machines, housed in brightly coloured casings and consoles. Many of them move, produce sounds, create visual effects, and so forth when activated by a user. There are lots of children moving around excitedly, flitting from exhibit to exhibit, sometimes bumping into each other. By the side of each exhibit are written instructions, showing you how to activate the machine. When you do, things happen. For example, at the back of the hall is a huge granite ball, glistening with water, sitting on a plinth. It is obviously massively heavy. But, amazingly, when you touch it, it revolves around. You read the instructions, and find out that it is not sitting directly on the plinth at all, but on a thin bed of water under intense pressure. All around you are the sounds of children shouting, adults talking and there is movement everywhere.

FIGURE 5. *Written description of the Exhibit Hall*

appears to be neutral, at the same time it is obviously located in a fixed position. By contrast, the writing jumps back and forth between 'views', affording an impression of both movement and engaged interaction. The pictures appear more static and distanced since the viewpoint is fixed. They effortlessly depict an entire spatial scene and grant the viewer access to certain modes of physicality (the visible ones of position, colour, shape, opaqueness and size, but not texture). The writing, on the other hand, grants a depiction of experience that is more overtly situated and subjective – less neutral and free-floating in its connotations. We are not suggesting here that writing is more selective or subjective than the photographic image: both are the results of the conscious and unconscious adoption of subjective perspectives. What we are suggesting is that the kind of information conveyed by each is qualitatively different. We hope to illustrate this further in what follows.

#### EXAMPLE 2: VIDEO VERSUS WRITING

Now, let's compare writing with video footage. Is video, as often assumed, obviously superior to fieldnotes in its ability to represent the real more fully? In the science theatre, we made a video recording of a live performance of one of the centre's regular shows performed to the group of schoolchildren we studied. This show was intended as a fun-educational demonstration of certain examples of physical forces – namely, pushes and pulls, and was designed to reflect the learning objectives of the national science curriculum used in UK primary schools. In this part of the fieldwork, we were trying to find out how the children were reacting to what was being shown on stage.

Figure 6 shows two stills from video footage we edited into a split-screen sequence. It splices together the views afforded by two cameras, one directed at the stage and one at the children's faces. Though reproduced here by necessity in static form, the edited footage allows us to switch effortlessly and simultaneously between the two perspectives (the stage and the auditorium). In doing so, it suggests strongly that there is a causal relationship between what is seen and heard on stage and what we see registering on the children's faces. This recording also shows up lots of different modes in operation in the performance (although only those modes which are visible or audible and thus amenable to being recorded by video camera). These include colour, position, the interactions of the two performers, the position and shape, size, colour, movement, etc. of the artefacts on the stage. It also, of course, includes sound – both spoken and non-verbal.

Now, let's examine a fieldnote made of the same event (Figure 7).

Far fewer modes are in evidence in the fieldnotes than in the video footage. They employ narrative and syntax, which themselves can describe other modes in their absence. But the fieldnote writer has not, in fact, attempted to describe all the colours, the positions, the movements, etc. that we can see on the video-tape. Colour and shape, for example, do not figure at all. By contrast, these are automatically represented by the video camera. Instead, the





FIGURE 6. Stills from edited video recording of the 'pushes and pulls' show in the science theatre

field-note writer has chosen to record other kinds of observation. For example, she has directed our attention to certain events rather than others (Peter's pushing the monkey along; the boy stopping the monkey; the children clasp their knees and rocking back to front).

But do the fieldnotes give us something that the video does not? The writing uses lots of comparisons ('Peter is louder than Megan and projects his voice more forcefully'). These give immediate clues as to the writer's interests, and what she considers to be of significance in relation to the research questions she has in mind. It also shows a clear sense of a narrator's particular perspective ('Peter seems to ad-lib. My impression is . . .'). And they afford a clear sense of *interpretation*. These notes are not simply observations; they are already interpretations, carried by propositions and arguments. For example: 'Peter uses his body to emphasise/exaggerate aspects of the show.' They are far from a neutral 'recording' of what is happening. There is *more* interpretation in the written fieldnotes than in the edited video recording. The latter is far from a



This 'performance' of the 'pushes and pulls' show is more animated than the show I watched yesterday. Peter is louder than Megan was and projects his voice more forcefully. He exaggerates comments and questions and uses a much wider vocal range. He refers to the monkey as the 'star of the show', and asks the audience if they want to meet him. Peter seems to ad-lib a little. In one of the early set pieces when one boy has come down to the front to push the monkey along, he also stops the monkey. Peter picks up on this, saying we can also use a push (backwards) to stop things. He asks the audience to remember this as they will come back to it later in the show. Peter also uses his body to emphasise/exaggerate aspects of the show – such as hand to mouth, moving around the stage more, using the space more expansively. A number of the children from the school are clapping their knees and rocking from back to front during the show. My impression is that this class are not raising their hands/volunteering/shouting out quite as much as the other two classes that are in the theatre at the same time.

FIGURE 7. *Fieldnote made of a performance of the 'pushes and pulls' show in the science theatre*

merely 'transparent' or neutral reflection (as camera-work never can be), but its meaning is more 'open': it lacks the hierarchical tying down of meaning that the fieldnotes present.

#### THE AFFORDANCES OF DIFFERENT MODES

The above suggests that writing and video can afford quite different kinds of meaning (although it is also the case that common semiotic principles occur across diverse modes). For example, consider the image/writing contrast. Writing enables the abstract relations between ideas (such as their sameness or difference, their relative significance, or their hierarchical ordering) to be represented through syntax. It is also inherently selective, with the writer choosing what to highlight and what to leave out. Images, by contrast, while still the outcome of selections, afford the exposure to view of whole vistas and scenes and the concrete, spatial positioning of things in relation to each other. There is much less hierarchical ordering in the image. This is also the case with sound. In terms of the speech/writing contrast, speech is temporally determined and has no visual component; writing, on the other hand, like images, is spatially determined and is embodied graphically. The spoken word is truly linear and sequential – it unfolds over time and leaves no trace as it is uttered, while writing is a simultaneous, permanent graphical representation that can be repeatedly perused and skimmed in different sequences.

For such reasons, the spoken word is often said to be suited to the description of events ordered into sequences, and to be oriented towards unfolding narrative functions like story-telling. As each spoken word 'disappears' after its utterance, it is easy to generate the effects of suspense and the stringing-out of enigmas. The written word, on the other hand, like other visual forms, is said

to lend itself to the depiction of an arrangement of elements and their relation to each other. It is, therefore, usually considered better suited to hierarchically ordered representations involving argumentation pursued through arrangements of ordered clauses and rhetorics, as in conventional ethnographic authoring (Atkinson, 1990). Writing can be read and re-read; for this reason, it is not strictly true to say that writing is linear, for its reception – i.e. reading – is very often non-linear (see Lemke, 2002). The permanent, sequential arrangement of ordered clauses on the page makes it possible to pursue, at leisure, hierarchically ordered, logical connections between one element and the next (Bolter, 1991). The spoken word, however, is always linear, and does not lend itself to complex, memory-taxing conceptual linkages and relationships. Images, meanwhile, provide the big picture – showing the relations between different aspects of the field. It has been argued by Hastrup (1992), for instance, that ethnographic photography and film give us maps (a cultural tableau that can show us the bigger picture – the concrete differences and characteristics of the field) while written accounts give us itineraries (a guided tour around the social space, which contextualizes the map through an account of lived experience).

Hence, multimodal representation implies the need for careful consideration of the particular kinds of meaning afforded by different media. Voices, for example, vary considerably in tone and quality – giving us immediate clues as to the speaker's sex, age, social status, mood and so forth. Intonation and rhythm, too, convey important dimensions of meaning. Accordingly, when an informant's voice is heard (as opposed to being read as a transcript-extract), there is an important sense in which it stands for itself – as opposed to being incorporated, in its transcribed form, by the modes of writing. Writing, potentially, may for these reasons begin to lose its long-established monopoly in ethnography and start to be used for more mode-specific functions – for explaining sounds and images, for example, or for pointing to them (see Kress, 1998). These possibilities have significant, perhaps even worrying, implications for academic authoring (see Dicks and Mason, 2003; Dicks et al., 2005).

Paying attention to the overlaps among, and distinctiveness of, different modes alerts us to the ways in which different media can be employed for representing multi-modality. We have shown earlier that recording the exhibits hall through the medium of photography as opposed to written or spoken language affords a qualitatively different perspective. We have also suggested that the edited video footage of the science theatre produced a very different representation to the written fieldnotes. However, attending to such differences also shows up some similarities. There are, for example, certain similarities between the meanings conveyed by the fieldnotes and the video. The fieldnotes, like the edited video, enable us to switch back and forwards between the two perspectives of stage and audience. They also frame the event into significant sections, just as the video does. The video points out what is significant by the device of close-ups (e.g. of the children's faces) and the direction of the lens, whereas

writing accomplishes this through syntax (such as the use of prepositions and the ordering of narrative).

In terms of perspective, video always *seems* more objective than writing. However, although the fieldnotes afford a sense of a subjective perspective (i.e. of a narrator positioned within, not outside, the action) in a way that is not afforded by the edited video, this is more an effect of recording style than something inherent in the image-mode. For example, we could have chosen to record the performance by using a single, hand-held video camera and not having recourse to editing at all. This would have granted the sense of a single narrator's perspective, not dissimilar to that which is discernable in the written fieldnotes. Such strategies have indeed been recommended by proponents of visual ethnography as a naturalistic, 'true to life' and scientific form of representation (see, for example, Bazin, 1967; Heider, 1976).

## Conclusion

We have not tried to do more here than raise a number of issues for ethnographers to consider when thinking about multimodality and multimedia. In particular, we have highlighted the importance of attending to the implications of multimodality at two 'stages' of the ethnographic research process: first, the ethnographer's observations of the field and, second, his/her recording of those observations. In terms of the first, we have suggested that the media used in any research setting themselves communicate in particular ways, through utilizing particular modes. In this sense, as McLuhan put it, the medium is (at least part of) the message (McLuhan and Fiore, 1967). We have thereby noted the distinctive semiotic affordances of different media, of which there are certainly other kinds of instance than the ones discussed here. For example, one could hypothesize that children learning about the topic of physical forces through an interactive exhibit might pick up quite different meanings than if they were trying to make sense of it through a classroom textbook. It could also be the case that quite contradictory meanings emerge through the different media, and this is something that future research might fruitfully explore. However, it is also true to say that very different media may employ *similar* communicative resources – for example, a science textbook may use bright, primary colours in similar ways to how they are used in discovery centres (or magazines, for that matter). By applying the principle of multimodality, the field (however constituted) emerges as a complex orchestration of media drawing upon a variety of modes.

In terms of the second – i.e. the production of data-records – we have stressed that, when observations of multi-media data are recorded into differently mediated representations, meaning is necessarily transformed. This is not a novel insight in itself. But what we want to underline here is the complexity of the various transformations produced when these recordings are made in a range of *different* media (typically including, nowadays, audio-visual

recordings). Rather than assuming that multimedia automatically gives us multimeaning, satisfactorily reflecting the multimodality of the field, we might consider *how* different modes are transformed when translated into different media. What semiotic modes do we lose when we use the camera? What meaning potential does speech afford that is closed off by images or writing? By contrast, what common semiotic affordances are produced by both film and writing?

Most significantly, perhaps, and this is a question with which we are still struggling, what happens to the various meaning resources of the different modes when they are combined together via the computer screen? Particularly important here is the capacity for *hyperlinking* in electronic, computer-mediated representation (i.e. the joining together of elements through clickable links on-screen). We have not attempted in this article to introduce the many complex issues involved in the question of hyperlinking in ethnography (issues which we have addressed elsewhere – see Dicks and Mason, 2003; Mason and Dicks, 2001). When we *combine* different modes through different media, and link these together in various ways, what kinds of new, multi-semiotic meaning are produced? Hyperlinking means that multimodality becomes even more complex. In hyperlinking, we are no longer talking simply about the juxtaposition of image, text and sound, but the creation of multiple interconnections and pathways (or traversals) among them, both potential and explicit (Lemke, 2002). This is the new meaning-potential afforded by hypertext + multimodality, and it is what Lemke (2002: 300) calls ‘hypermodality’: ‘the new interactions of word-, image- and sound-based meanings . . . linked in complex networks or webs’.

What ethnographers are starting to consider is how this new potential for generating meaning can be related back to the multimodal (hypermodal?) social world they wish to describe. In particular, as well as tracing the various affordances of different modes and media (discussed in this article) in both the field and their representation of it, ethnographers are also confronting the question of the meaning-potential of linking. How does a piece of video film change when linked to a piece of written text? And what kind of reading or interpretation is produced by that linkage when the reader can pursue an almost infinite number of traversals and linkages of his/her own? These are questions we are currently pursuing, and will, we hope, be the topic of future papers.

#### ACKNOWLEDGEMENT

The project ‘Ethnography for the digital age’, on which the article reports, was funded by the Economic and Social Research Council (2002–2005) under the Research Methods Programme.

## NOTES

1. Kress, Van Leeuwen, Jewitt, Lemke and others' work on multimodality represents a developing theoretical approach that is beginning to shed more light on these processes.
2. The centre does display more theoretical kinds of scientific knowledge – such as astronomy in its small planetarium, and (at the time of our fieldwork) genetics in a space called the Hub, dedicated to temporary exhibitions. However, the exhibits hall is clearly demarcated as the major centre of visitor activity and the prime visitor zone, and this space is dominated by mechanical and other engineering-oriented applications of science.
3. In fact, writing employs other modes too in its graphical dimensions – such as print colour, directionality, size, etc.

## REFERENCES

- Atkinson, P. (1990) *The Ethnographic Imagination: Textual Constructions of Reality*. London: Routledge.
- Bazin, A. (1967) 'The Evolution of the Language of Cinema', in *What is Cinema?* Vol. I. Berkeley, CA: University of California Press.
- Bauer, M.W. and Gaskell, G. (2000) *Qualitative Researching with Text, Image and Sound: A Practical Handbook*. London: Sage.
- Bolter, J.D. (1991) *Writing Space: The Computer, Hypertext, and the History of Writing*. Hillsdale, NJ: Lawrence Erlbaum.
- Dicks, B. and Mason, B. (1998) 'Hypermedia and Ethnography: Reflections on the Construction of a Research Approach', *Sociological Research Online* 3(3), <http://www.socresonline.org~socresonline/3/3/3.html>
- Dicks, B. and Mason, B. (2003) 'Ethnography, Academia and Hyperauthoring', in O.O. Oviedo, J. Barber and J.R. Walker (eds) *Texts and Technology*. New York: Hampton Press.
- Dicks, B., Mason, B., Coffey, A. and Atkinson, P. (2005) *Hypermedia Ethnography*. London: Sage.
- Emmison, M. and Smith, P. (2000) *Researching the Visual: Images, Objects, Contexts and Interactions in Social and Cultural Inquiry*. London: Sage.
- Geertz, C. (1973) 'Thick Description: Toward an Interpretative Theory of Culture', in *The Interpretation of Cultures*, pp. 1–30. New York.
- Hastrup, K. (1992) 'Anthropological Visions: Some Notes in Visual and Textual Authority', in P.I. Crawford and D. Turton (eds) *Film as Ethnography*. Manchester: Manchester University Press.
- Heider, K. (1976) *Ethnographic Film*. Austin, TX: University of Texas Press.
- Iedema, R. (2003) 'Multimodality, Resemiotization: Extending the Analysis of Discourse as Multi-semiotic Practice', *Visual Communication* 2(1): 29–57.
- Kress, G. (1998) 'Visual and Verbal Modes of Representation in Electronically Mediated Communication: The Potentials of New Forms of Text', in I. Snyder (ed.) *Page to Screen: Taking Literacy into the Electronic Era*, pp. 53–79. London: Routledge.
- Kress, G. and Van Leeuwen, T. (2001) *Multi-modal Discourse*. London: Arnold.
- Lemke, J.L. (2002) 'Travels in Hypermodality', *Visual Communication* 1(3): 299–325.
- Mason, B. and Dicks, B. (2001) 'Going Beyond the Code: The Production of Hypermedia Ethnography', *Social Science Computer Review* 19(4): 445–57.
- McLuhan, M. and Fiore, Q. (1967) *The Medium is the Massage*. London: Allen Lane, The Penguin Press.

BELLA DICKS and AMANDA COFFEY are both Senior Lecturers in Sociology at Cardiff School of Social Sciences and worked on the ESRC-funded project, 'Ethnography for the Digital Age'. They are now, together with Bruce Mason and Matt Williams, working on another ESRC-funded project, 'Methodological issues in Qualitative Data Sharing and Archiving under the Qualitative Demonstrator Scheme' (RES-346-25-3010).

*Address:* [email: [dicksb@cf.ac.uk](mailto:dicksb@cf.ac.uk)] [email: [Coffey@cardiff.ac.uk](mailto:Coffey@cardiff.ac.uk)]

BAMBO SOYINKA is a teacher, filmmaker and researcher and was Research Associate working on the project. She is now completing a PhD on Time, Film and Social Theory at Cardiff School of Social Sciences.

*Address:* [email: [soyinkab@cf.ac.uk](mailto:soyinkab@cf.ac.uk)]