This is the published version of a paper published in *Critical Discourse Studies*.

Doing critical discourse studies with multimodality: from metafunctions to materiality

Per Ledin & David Machin

Published online: 29 Apr 2018.

Submit your article to this journal 🖉

Article views: 146

View related articles 🔍

View Crossmark data

Routledge
Taylor & Francis Group

Doing critical discourse studies with multimodality: from metafunctions to materiality

Per Ledin [a] and David Machin[b]

aDepartment Culture and Education, Södertörn University, Huddinge, Sweden; bDepartment Media and Communication, Örebro University, Örebro, Sweden

ABSTRACT

In Critical Discourse Studies (CDS) and in other linguistics oriented scholarly journals we now see more research which draws upon multimodality as part of carrying out analyses of how texts make meaning, in order to draw out the ideologies which they carry. However, much of multimodality is itself based closely on one theory of language called Systemic Functional Linguistics (SFL). And despite calls from some scholars there has been no real interrogation of the concepts and models drawn from this theory as regards how suitable they are both for analyzing different forms of communication and for answering concrete research questions of the nature asked in CDS. In this paper we assess the core principles, taken from SFL into multimodality. Using examples we consider which are more or less suitable for the kinds of work we do in CDS. We make a case that SFL has a narrow notion of 'texts' and a weak notion of context. We show how we can address such problems to deal with what we call the 'materiality' of multimodal communication.

Introduction

Since two key publications in the 1990s (Kress & Van Leeuwen, 1996; O'Toole, 1994) multimodality has been a growing field which has developed into a handful of overlapping and distinctive sub-fields (Jewitt, Bezemer, & O'Halloran, 2016). At its core lies the drive to produce more detailed and predictive forms of analysis of all types of communication. Over the past decade it has become much more usual to find scholarly articles in fields such as critical discourse studies, sociolinguistics and pragmatics, using multimodality are part of their analysis which includes things like images, textbooks, videos, monuments, spaces, etc.

Such critical work draws from the broader field of multimodality in different ways. But what underpins much of it are a number of core principles from the systemic functional linguistics (SFL) of Halliday (1978) and (1985), specifically as described by Halliday and Matthiesen (2014), which themselves form the basis of the approaches taken by Kress and Van Leeuwen (1996) and by O'Toole (1994). In one sense this SFL based multimodality, we

would argue, has become used as a kind of grand theory of all forms of communication, used to look at things like web pages and printed pages (Bauldry & Thibault, 2006), packaging (Wagner, 2015), film (Bauldry & Thibault, 2006; Tseng, 2013), music (Van Leeuwen, 1999) and art (O'Toole, 1994). These are hugely different semiotic phenomena yet approached through a common set of principles and analytical models derived from the work of Halliday's (1978, 1985) SFL.

So far, however, there has been little reflection on whether this is entirely unproblematic. And specifically, which is our concern in this paper, there has been a lack of consideration as to whether such models are appropriate for the work done in CDS. CDS is, broadly, a critical investigation of diverse social phenomena, with language as its core focus, aiming at challenging what is usually taken for granted, in order to draw out buried discourses which support the interests of specific ideologies and dominant groups in society (e.g. Wodak & Meyer, 2016). Language here is seen as a social practice and the focus is on its contexts of use (Flowerdew & Richardson, 2018). Crucially, as Van Dijk (2013) points out, this is a problem-oriented approach which must rely on methods that are 'able to give a satisfactory (reliable, relevant, etc.) answer to the questions of a research project'. Whether indeed the principles of SFL which undedrpin and guide multimodal analysis are suitable for such a critical approach is a discussion which needs to take place. It has now, in fact, been argued amongst multimodal scholars that it is now time to pause, reflect upon, criticize and consolidate, the kinds of concepts used (Bateman, Wildfeuer, & Hiippala, 2017; Ledin & Machin, 2018a).

In this paper we reflect on several of the core concepts and procedures for analysis taken from SFL which have become widely accepted as the basis for multimodal discourse analysis. We argue that these create problems which surface in scholarly work where it is not so clear whether what gets done is actually a process of analysis or purely one of labelling things through set of pre-established categories. In such work it can become unclear whether the payoff is justified from the complexity of analysis. And, we argue, relatively blunt concepts and 'systems' of semiotic features can in fact act as a filter for how data is approached, irrespective of what form of tools might better suit analysis. We make some suggestions as to how we can overcome these problems. SFL offers some highly valuable tools and concepts, but we need to carefully consider how we apply these, and we need a model of analysis that can deal with the material, and social practice based, nature of the forms of communication that we want to critically analysis.

Multimodality as a grand theory

The contemporary SFL based multimodality approach is clarified in a number of recent key texts which present overviews of the field (Djonov & Zhao, 2018; Jewitt et al., 2016). It is common to find this approach across scholarly work which draws on multimodality and formed the basis for earlier key texts in multimodality (Kress & Van Leeuwen, 1996). Here, drawing specifically on Halliday and Matthiesen (2014), the argument is made that analysis should start from Halliday's metafunctions in order to establish the underlying system/grammar of any instance of communication. This would apply to photographs, food packaging, a webpage, or a film.

In simple terms these metafunctions have been used in SFL to account for the way that language is organized to fulfil three basic functions: 1. the need to communicate ideas and

experiences (the ideational metafunction); 2. to form social relationships and identities (the interpersonal metafunction); and 3. to create coherence (the textual metafunction). An assumption is made in this form of grand-theory multimodality that all forms of communication are structured on the basis of these types of meanings. We then identify the system of choices which are used to fulfil the metafunctions, whether in a photograph, food package, or a film. For example, for a photograph this might mean how the interpersonal function is fulfilled through things like 'proximity', 'angle of interaction' and 'gaze'.

This methodological approach is spelled out by Jewitt et al. (2016, p. 49):

(1) Developing metafunctionally organized systems
(2) Analysing the text according to the choices that are selected
(3) Interpreting combinations of choices according to register and genre.

Djonov and Zhao (2018) state, in a similar fashion, that two tenets from SFL underlie (much) multimodal research. First that 'every act of communication simultaneously constructs three broad types of meaning, or "metafunctions"' and second that 'the meaning potential of semiotic modes can be modelled as systems of interrelated choices, paradigmatically, where each has a distinctive structural realization' (ibid., p. 4). The unit of analysis is the clause in the SFL approach to language. In this grand-theory model of multimodality the equivalent of clauses must be found for different forms of communication. Texts or semiotic artefacts are then 'annotated' according to (semantic) system choices attached to the metafunctions.

In what follows we examine these stages and the theoretical assumptions in SFL which lead to these assumptions. We consider how directly these are of use for accounting for other forms of communication and for answering research questions.

The notion of context

Our first point addresses the way that grand-theory multimodality accepts a notion of context from SFL which becomes problematic once it is applied beyond spoken language to different types of communication.

SITUATION: Feature of the context	(realised by)	TEXT: Functional component of semantic system
Field of discourse (what is going on)		Experiential meanings (transitivity, naming, etc.)
Tenor of discourse (who are taking part)		Interpersonal meanings (mood, modality, person, etc.)
Mode of discourse (role assigned to language)		Textual meanings (theme, information, cohesive relations)

Figure 1. SFL's view of the text as a realization of a context of situation (from Halliday & Hasan, 1989, p. 26).

SFL defines text as 'any instance of living language that is playing some part in a context of situation' (Halliday & Hasan, 1989, p. 31). It is seen as a realization of a context of situation, as these authors illustrate in Figure 1. We see that the text is represented in the right hand column and the situation and context in the left hand column.

SFL is about devising a grammar. So we assume that any clause or instance of language, the text, is imprinted with the metafunctions. We see this in the right-hand column. For each metafunction we are given the part of language, or system which realizes these functions. We see that the ideational/experiential meanings are realized by things like transitivity (verb processes) and by how things are named. The interpersonal meanings are realized by moods and modality (or example for indicating certainty such as 'I will' or 'I may'). The textual function is realized by the grammar which allows it to fit together in an information structure. In the left-hand column we see how from this we can make certain assumptions about the context. So the metafunctions and systems which realize them link to contextual features such as 'field' (what is going on) which tends to be realized by the ideational/experiential metafunction, 'tenor' (the relations between participants) by the interpersonal metafunction and 'mode' (the channel of communication – spoken, written, etc.) by the textual. These broad notions of field, tenor and mode are unique to SFL and have, in fact, been crticised for giving a rather shallow and even arbitrary description of context, for example not accounting for the properties of participants, nor conflicting interests and power relations (Van Dijk, 2008, Chap. 2).

For spoken language this general model appears to work. This is because the 'context of situation' where language is used becomes an activity. Context is, therefore, something that unfolds together with language and is infused in language in concrete situations. For written texts or for other forms of communication (such as a photograph or a food package) this notion of context becomes problematic. When Halliday and Hasan (1989) relate written texts to the context of situation in their accounts, they simply project a context from the text, for example a broadcast talk by a bishop (p. 24f.) or a legal document (p. 12f.), using the contextual features of field, tenor and mode. In other words what is going on, who is involved and the channel of communication are projected from the text itself.

Holmberg (2012) has been critical of the way that SFL deals with context showing how this becomes problematic as soon as we apply the notion to written texts. He gives an example of journalistic essays written by students in Swedish upper secondary school. First, Holmberg argues that the essays 'do not seem to have context at all in the sense that Halliday elaborates in his analysis of causal interaction' (ibid., p. 67) – the writing process itself hardly constitutes such a context. Second, even though these (written) texts in a sense has 'context in text', or that a context can be projected in terms of field, tenor and mode as in Figure 1, such a projection is hard to generalize in a systematic way.

There is thefore an untenable assumption in SFL that context is recoverable from the text. This kind of projection becomes problematic when carrying out a critical form of problem-driven analysis. For example, if we are dealing with a kind of news-photograph, how can we project context from the actual image itself, by looking at how the metafunctions are realized through the visual 'syntax'?

Machin (2016) begins to point to the problem here where he shows how much of the meaning of a photograph cannot be established at the level of the contents of the image itself. The image in Figure 2, depicting of a girl during the war in Sarajevo in the 1990s, can

Figure 2. Tom Stoddart photograph titled 'Bosnia, Sarajevo, Girl (4–5) standing near US soldiers in street' (with permission from Getty Images).

help us think a little more about what is context here. In order to understand the meaning of such a photograph we must know something about the 'canons of use' of such forms of photojournalism.

Canons of use accounts for a different level of instances of communication than the semiotic resources (e.g. the actual contents of a photograph). Canons of use are the traditions of use of such instances along with the kinds of semiotic resources that tend to be employed in them. So a canon of use for the photograph are things like photojournalism, art galleries, advertising or personal photographs. Such canons of use come with establish meanings and patterns. And they are used as a means of communication in typical social practices, which for photographs can be journalism, family life, art, advertising, etc. And they are also infused into patterns of production and use.

In the case of photojournalism, the photograph is used in a way that documents reality, it bears witness to a moment, though it is a common criticism that photographs are far from such neutral representations (Tagg, 1988). And photojournalism has established typical frames for representing different kinds of typical news events, such as conflicts (Cottle, 2009). In the photograph of the girl in Sarajevo the photographer (interviewed by the authors) clearly uses such well-trodden frames to represent a highly complex situation. Here this relates to childhood, where children and mothers are often used in war and disaster photojournalism to index break-down in society, human vulnerability and tragedy (Bouvier, 2014). Such images are also designed for specific markets, with a specific grasp of the kinds of world views held by different reader and consumer groups. Obviously, as Bateman et al. (2017, p. 84) point out when discussing context, 'social knowledges' are unevenly distributed. Access to technologies and social practices, as producers or consumers, will vary in a culture and be connected to power relations.

The same photographer was also aware that different kinds of images, different canons of use, were tied to different social practices such as advertising and art. So images, as is described in the field of visual studies, must be accounted for as regards how they relate to different kinds of cultural gazes and industries of presentation (Tagg, 1988). Photographs will therefore have canons of use associated with art and advertising, etc., each of which are produced through established institutional practices and cultures and which viewers have been tutored to understand and to see as simply neutral, as part of the nature of their material world. This material world is structured, interacted with and communicated about using such canons of use which are part of social practices – of reading and understanding the news or going to an art gallery.

It follows that context cannot easily be projected from a text, or only to the extent that we have or acquire competence in the canons of use and social practices at hand. It is not directly present in the way that are say are the actual choices made by the photographer as regards things like framing and angle. So finding out the context, the canons of use and social practices involved, is crucial for a critical multimodal analysis. Clearly in the case of the photograph in Figure 2, the photograph is meaningful because of these.

As we saw, context is not mentioned at all when Jewitt et al. (2016) and Djonov and Zhao (2018) explain the overall SFL multimodal approach. Here all context and meaning can be discovered by looking at how the metafunctions are realized through syntax. Such a form of analysis therefore simply cannot help us to understand how a text is used as a material form of expression. It can only lead to a kind of interpretation of semantics.

Texts as meaningful wholes

In the SFL model of multimodality a problem is also created by the very notion of what comprises a text, since all forms of communication and objects are approached as realization of the metafuntions. This has consequences when we deal with communication such as photographs or food-packaging which are created and experienced as material wholes, missing how such materials communicate and why they are deployed in the first place.

Berge (2012) has been critical of SFL's notion of text arguing that it lacks any notion of text as a meaningful whole or as a semiotic and material artefact and that this 'wholeness' must be the point of departure for the analysis. SFL starts from the systemic resources used for making meaning. It does not attend to texts/semiotic artefacts as wholes as the point of departure for analysis. Berge himself (2012) makes the case through the distinction between utterances and texts. Utterances are unique in a context of situation. But texts are culturally shaped through cultural traditions and norms as we saw for the example of the photograph of the girl in Figure 2.

Treating texts as material wholes entails making a distinction between the configuration/outer form and the inner structure/design (Berge, 2012, p. 82). In other words a text (or semiotic material) is shaped as a whole, which gives it a contextual configuration, where some bonds within which communication unfolds are coded. And it has an inner design which involves the kinds of semtioc choices used which form that whole. For example, Facebook is a semiotic material designed as a whole that affords certain types of communication. It is designed as a page, which codes the contextual configuration, its outer form, the bonds within which communication can unfold. The top section

states who is the owner of the page and contains a photograph. There is a left column that specifies more about the page owner and which displays photos of friends, and a right column with adverts. The middle section is designed for posts ordered in reverse chronology, which points to a culture where 'the latest' is valued. Here a communicative infrastructure is set up enabling some, but not other types, of social interaction, where, for example, boxes for posts and comments are prefabricated, as is the button 'like', affording an immediate and positive response to posts and making comparisons possible in the sense of 'most likes wins'.

At the level of inner structure or design, we would find the choices of colour, typography, iconography, etc., used, or in language the cohesion between clauses or sentences (cf. Halliday & Hasan, 1976, which accounts for this inner structure of texts, but not outer form, their contextual configuration). Communicating by means of semiotic materials involves forming a whole that can be employed in a certain context. For example, a child wanting to create a drawing for grandmother would take an A4 sheet as the bonds in which communication unfolds, an outer form, and then write-draw a 'message'. A child writing in Sweden where we live will often start with 'Hej' (hello) and end with 'Slut' (the end) – once again the bonds, the contextual configuration, inside them the actual communication unfolds. Again at the inner level we find the choices in colours and form used in the drawing and relations between words and clauses in the letter.

Metacommunication/outer form can signal different things (Bateson, 1977). This can relate to the general bonds in which communication occurs, but also to the kinds of social and interpersonal relationships involved and the stance or point taken by a semiotic material. In other words outer form can be more or less elaborated. These kinds of outer forms create a kind of template or road-map, to be filled with our everyday texts. Such a line of thought cannot be captured by SFL. We argue that a notion of text must allow us to understand the nature of a semiotic artifact and its outer form.

Ledin and Machin (2018b) suggest that what we have is, at at the micro level, a text, which is a semiotic material shaped as a whole that we encounter in an actual situation, like a Facebook page or a shampoo bottle. This actual text is part of culture, of a tradition of shared beliefs and behaviours, different social groupings, also including artefacts and technologies, such as writing systems, computers, cameras or grooming products. This is the macro level, and here we find general resources for making meaning in relation to cultural traditions. In between these at the meso level we have canons of use, the potentials for making meaning that semiotic resources/materials have in relation to types of context. So a photographer employs the general cultural resources of his time – like a digital camera and software – to make meaning according to existing 'canons of use', the typical ways of using photography in our culture and, for example, makes some routine and often shared semiotic choices in order to communicate war and human suffering. On this meso level we have choices modelled in relation to canons of use, which, as we have already pointed out, is where we can model choices in a way that is contextually relevant. Put simply materiality and communication become interdependent (Bateman et al., 2017, Chap. 3) where materials can support and reshape communicative practices.

As regards analyzing a 'text' such as the news photograph in Figure 2, we take this to be a semiotic material shaped as a whole. It is of course a unique text on the micro level, but our interest would be to bring our the meso level and the 'canons of use' that it draws upon and how it is infused into social practices. Then, within its outer form/wholeness,

at the inner level, we would point to the semiotic resources are deployed (as regards what is in the image or in a certain design), in the interest of certain actors. This approach is very different from devising the grammar of a semiotic mode, from using the metafunctions to establish the paradigmatic system (formalized in networks for all sub-systems) and then showing that in some context the meaning choices in the networks are drawn upon in a 'text'.

System and choices

The notions of system and choices underlie social semiotics. The term 'system' in particular provided one of the foundations for critical linguistics (e.g. Fowler et al., 1979). Here system comes out as potential (Halliday & Matthiesen, 2014, pp. 20–24). The grammar or 'system' that SFL proposes maps, as we have noted, meanings in a paradigmatic way in system networks. Meanings are construed as choices, as something for users to tap into in different contexts. We can contrast this to a formal or traditional grammar, which would start from forms and syntagmatic relations and formalize rules for phrases and clauses. Such formal linguistic structures would be very hard to apply to for example photographs or newspaper design. But the SFL framework, based on choices of meanings, can and has been transferred to such domains. And when Fowler et al. (1979) set the agenda for critical linguistics the important step was to show how such choices became loaded with ideology.

This idea of choices of meaning works very well and to different degrees for other forms of communication. In the design of shampoo bottles in Figure 3 we clearly see things like choices between width and height, roundness and angularity and as regards colour for the women's and men's products. In the photograph in Figure 2 we can say that the photographer has made certain semiotic choices such as to use the girl, a certain kind of framing, that it is monochrome, that it includes the window frame and bullet holes in the wall.

But whether these choices can also be represented as system networks is a different matter. And we would argue that it is entirely problematic to start with a system network which was designed to be 'one-size-fits-all', one which is used, as in the grand theory multimodality, to look for how a text realizes the types of meanings given by the metafunctions. As Jewitt et al. (2016) suggest such an approach would involve looking for the system, such as processes, participants, circumstances, which realizes the ideational or experiential metafunction. It would then look at things like angle of interaction and gaze to show how these realize the interpersonal function. All of these are often formalized and represented as system networks of choices as is found in the work of Halliday (1985), where for example networks of choices in language will be given for how each metafunction is realized in language.

Kress and Van Leeuwen (1996) take an approach where linguistic processes realized in clauses, which is a sub-system belonging to the ideational metafunction, can be taken as the basis of visual grammar. While both these authors have since shifted away from such an assumption, this model remains highly influential. In Kress and Van Leeuwen's model processes in visual communication are formalized in a way akin to language, so that we get 'actions', 'mental processes', 'verbal processes' and 'reaction' being part of a system network or taxonomy which is six levels deep (p. 74). And, what in linguistics is called 'transitivity' (cf. Figure 1), how roles and actors are dependent on the processes realized in

Figure 3. A shampoo from Garnier Fructis and a shower gel from Axe.

verbs, are included, so that visual processes can be 'transactional' or not (so are people carrying out 'active' processes or not). Since SFL is about semantics, the realization of these processes and actors, how they actually are expressed and materialized, must be specified in a separate diagram (pp. 74–75).

So we see, as both Jewitt et al. (2016) and Djonov and Zhao (2018) explain, how in multimodality system networks are developed for the metafunctions, which amounts to constructing a grammar. Then the meanings are formalized in the networks – in this case the ideational metafunction and (visual) processes – are applied to the multimodal texts chosen. As for Figure 2 we would say that the girl realizes a 'non-transactional reaction', whose realization is defined as this: 'An eyeline vector emanates from a participant, the Reacter, but does not point at another participant' (Kress & Van Leeuwen, 1996, p. 74). Continuing such an analysis we could look at processes in more photographs and also bring in language and scrutinize different processes coded by verbs. We might also say, using Kress and Van Leeuwen's (1996) original model, that the way someone in a photograph is pointing realizes an action and agency through the vector made by the form of their arm, or in Figure 2 perhaps by the gun.

We would argue that this kind of analysis – one which looks for processes and actors in a photograph – while also missing context, the nature of canons of use which are part of social practices – makes the error of missing the affordances of photographs. All semiotic

materials come with affordnaces (Gibson, 1979), with different possibilities to support action and make meaning for different actors, and as humans we develop technologies to shape and design materials ansd use them in canons of use. At the level of affordance photographs do not represent time but captured moments. They can only *index* time and movement and agency. If, rather than coming at the photograph from linguistic processes and syntax, we first describe what we see it comes out slightly differently. We might take look at the photograph and bodily and facial expression and pose as starting points. We might observe that the girl takes a restricted pose and stands still with the hands behind her back. Her eyes are open with a steady gaze and the mouth shut, with the lips slightly going down. This expression indexes, in our interpretation, that she is a bit melancholic. But her pose does not does not suggest fear nor cowering. She stands like this, sort of at ease, even though the soldiers are there.

So this is about indexing – photos do not code processes unfolding over time, as language does, which has to do with affordances, with photos being 'frozen moments', 'reality interrupted' (Sontag, 2004). So as we interpret the indexing here, the girl is thinking, feeling and sensing at the same time, which is exactly what photographs – but not so easily language in a single clause – often represent. And this is all part of a tradition of a specific canon of use of photography. Such images are created to provide just this kind of access to moments of individual emotion.

Within SFL scholarship itself there have been criticisms of the very process of modelling the kinds of networks of choices which are produced uncritically in multimodality. A number of scholars point out that we need to ask in the first place what the choices stand for, or how they are motivated. In SFL, and in grand-theory multimodality, network choices make up an inventory of a sub-system related to a metafunction. But in this model it is not always clear how they are communicatively motivated, how they satisfy different higher order communicative strategies deployed in different contexts (Bache, 2013). In other words such networks tend to be void of context. For example, in an analysis of the registers of radio news and sports bulletins, Berry (2013) demonstrates that the notion of choice should be tied to an account of the relationship between choice and context in this specific case. This very much aligns with the position we have made above where choices must be tied to specific contexts and forms of expressions.

How semantics becomes context: stratification

One reason for the emphasis on networks in SFL is that it is a meta-semantic theory. The system models content on the first place from which meaning is projected. The theory has a strict content-to-expression directionality where the system networks formalize meanings. We saw this in regard to how context is projected from the content of texts like photographs – since texts are imprinted with the need to fulfil the three metafunctions in any setting. No modelling of expression-to-content directionality exists, which, we would argue, is often a fruitful entry point to analyse multimodal communication as we saw in the case of the photograph where there are clearly traditions of forms of expression which are a starting point to grasping why certain choices in content have been made.

The reason for this content to expression directionality can be found in the SFL notion of 'stratification' (Halliday & Matthiesen, 2014, pp. 24–27). Looking at this helps us to understand how it is that grand-theory multimodality risks losing the link between

content and expression as we have suggested above. And this also helps us to see how concepts of the metafunctions used to describe content in the first place, are then projected outwards as an entry point for analysis, mistaking, we suggest, semantics for context.

Stratification deals with the way that language relates to the world and experience. It is usually explained with the help of figures as seen in Figure 4. Halliday and Matthiesen (2014, Chap. 1) explain that language is used to make sense of what we experience and to manage interactions with others. Grammar here must be able to interface with the goings on in the world which require our understandings and with the social processes in which we are involved. Simultaneously it must also organize these to transform them in to words. These are the basis of the three metafunctions. They explain that this has two steps. First experiences and interpersonal relationships are transformed into
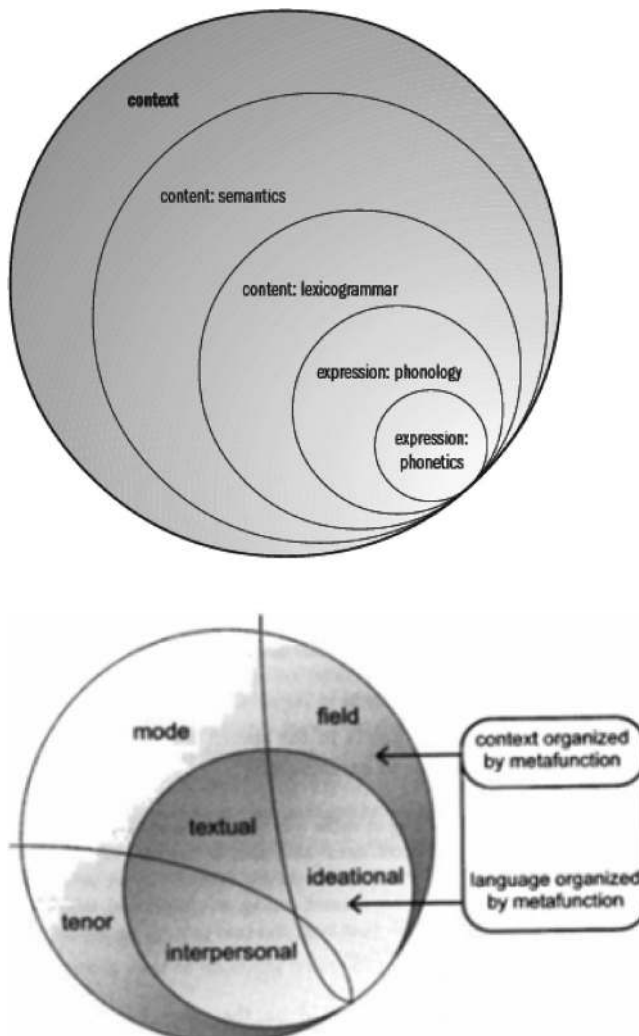
Figure 4. SFL diagrams for stratification. The upper diagram is taken from Halliday and Matthiesen (2014, p. 26) and the lower diagram is from Martin and Rose (2007, p. 297).

meaning (the stratum of semantics). Then the meaning is transformed into wording (the next semantic stratum of lexicogrammar). This, they say, describes the process from the point of view of the speaker. It would be reversed for the listener. Eventually this comes out – is realized – as sounds, as phonology and phonetics, which is the expression. These stratified stages are represented as decreasing circles which sit within the larger circle of 'context'.

Since it is content that is stratified, as we see in the lower diagram, SFL relies on content-to-expression directionality when mapping signs or text. The system networks seek to document and formalize meanings and are supplemented with 'realizations' that point to expressions of these meanings – and this is where expression or materiality comes in. There is strict content-to-expression directionality, and no expression-to-content directionality (cf. Taverniers, 2011). As we argued above, expression, for example in the case of the photograph in Figure 2, is already built in at the level of canons of use and the ideologies and meanings that these carry. And, as noted by the likes of Berry (2013), such networks are not necessarily applicable beyond specific contexts.

Obviously, we would argue, there are choices to be made relating to expression. For example, when designing the shampoo bottles in Figure 3 there would be choices as regards curvature or angularity, whether the plastic of the bottle should be thin or thin, whether the typeface light or with weight. But expression and materiality is just how meanings are realized on the 'lowest level'. This does not account for canons of use and affordances.

If we take the view of Hjelmslev (1961) content and expression always are fused (and only separated by us for purely analytical reasons). In the case of the photograph above, this kind of documentary/art photograph, and the affordance of the photograph to provide a window to reality, along with discourses about representing conflicts and suffering, are already infused into how reality is organized for us. This the same for all semiotic materials. The expressions are in this sense the same as the content, and an analysis might just as well start from expression as from content.

It is important to bear in mind that when carrying out an analysis of graphics or fonts it may indeed be useful to create inventories with expression forms such as 'angularity' and 'curvature'. Such expression forms are also relevant for the shower products in Figure 3, where the female bottle is more slender, whereas the male, Axe bottle is thicker and stands steady. But this must certainly not comprise any grand theory or 'grammar', and such choices must still be placed within the contexts of specific uses.

In SFL based multimodality the metafunctions are assumed to be applicable to any kind of text as we saw in Jewitt et al. (2016). This means that we can approach things like food packaging starting with metafunctions (Wagner, 2015). So an orange juice carton has the texture and shape of an orange so it represents the idea of an orange. It is interpersonal as it creates a relation between the product and the buyer. The textual metafunction tells us how things like the label on a product and the texture combine to create a whole.

But, as we have already indicated, in this approach context tends to become the pre-given semantics that SFL build upon (cf. Ledin & Machin, 2018a; Van Leeuwen, 1999). It misses how packages, or any semiotic artefact, are meaningful to us as material wholes. The green Garnier Fructis shampoo bottle in Figure 3 is a semiotic material which we experience as a whole, and belonging to a 'canons of use', in this case a tradition of construing femininity in women's grooming products. And then, within this outer form/

wholeness, a range of semiotic resources are deployed in a certain design in the interest of certain actors and also in specific sociopolitical contexts – here consumer capitalism and marketing.

So, at the micro level is the material as a whole which we encounter in actual situations. At the macro level we have culture and different semiotic materials/resources with potentials for making meaning, In between these at the meso level we have canons of use, the potentials for making meaning that semiotic resources/materials have in relation to social practices, which relate to consumerism, commodification and associated discourses, like nature as 'pure', notions of 'healthiness', etc. This world of semiotic materials and semiotic choices with their affordances, as Hjelmslev (1961) pointed out, are the very same as our consciousness. And of course in the case of the shampoo bottles it is consumer capitalism which plays a major role in the formation of this materiality. Grand-theory multimodality de-couples texts from the very materiality-human consciousness of which they are a part.

Instantiation or canons of use

One highly important insight in SFL for the problem-driven empirical research done in CDS, as explained by Halliday and Matthiesen (2014, pp. 27–30), is the notion of 'instantiation' (cf. Holmberg, 2012). This has to do with the relation between the system of choices and actual cases of texts. In SFL this is referred to as a 'cline' where system and text are two poles of the cline. So at one end of the cline we have the overall potential and at the other a particular instance of use. In between these two poles there are what are called 'intermediate patterns' (ibid., p. 28). These intermediate patterns are often called 'registers' (Halliday, 1978) and related to different, historically evolved situation types. So ways of thinking and doing have been established as conventional ways of communicating in different types of contexts or situations. Linking this to what we have argued in earlier sections we can say that semiotic materials shaped as wholes can be usefully analyzed as instances of historically evolved canons of use, which resemble this idea of registers. But we must not view these as realisations of a context of situation (micro level), nor of a de-contextualized system of grammar (macro level).

The SFL idea of 'instantiation' is that the system and its potentials (that in principle are endless) have sub-potentials, or 'registers', which accounts for the way that texts then come out as instances of such sub-potentials (a quite different view then seeing texts as a realization of a context of situation). This appears a much more useful starting point for analysis of texts beyond spoken language where we take a food package, a film-clip, web page to be 'an instance' of 'a type', of an established canon of use, also of course associated with certain discourses and social practices.

Put another way we are not interested in making a grammar/system (the macro level), but start from a culturally and historically established sub-potential, a canon of use, that is associated with or deployed in certain contexts (the meso level). An actual material text (the micro level) is seen exactly as an instance of this sub-potential/canon of use. Such a way of thinking about analyzing multimodal communication in fact has some parallels in Fairclough's (1992, p. 71) notion of discourse. He argued that discourses are part of social practices. In other words, the discourses we find in texts are interwoven with how we do things in society. In between the level of text and social practice he placed his notion of 'discursive practices' which relates to the more sociological matters of things

like production, distribution and consumption of texts. For Fairclough (ibid), and later emphasized by other critical discourse scholars (Richardson, 2007), it was only by placing texts as part of social practices and understanding them as part of discursive practices that *critical* analysis could take place. Here we suggest that instances of multimodal texts should be seen as semiotic materials located in social practices and as part of canons of use.

Conclusion

The purpose for writing this paper was to point out that many of the concepts and procedures that have become taken for granted in much of multimodal analysis require further interrogation. The more specific aim here being the extent to which such concepts and procedures provide a useful basis for forms of critical discourse analysis which seek to ask specific research questions about the contents of texts and then carry out relevant forms of analysis to draw out the more buried details in those texts that allow us to answer those questions (e.g. Flowerdew & Richardson, 2018; Van Dijk, 2013; Wodak & Meyer, 2016). What we have shown is that contemporary multimodal theory appears to present a grand theory which can account for all forms of communication, however different these may be. What we show is that this kind of model, based on assumptions in SFL is highly descriptive in the sense that it sorts all phenomena into predetermined networks to show the systemic character of a web page, food-package or film.

At a basic level such a form of analysis runs the risk of becoming entirely self-referential. A system is presented and then applied. What is labelled then appears as part of the system. Such a grand theory, an assumption that all forms of communication have a grammar, risks to corrupting things, blocking out context and ironically enough underpinning a monomodal perspective. In short, such an approach is of little use for answering concrete and critical research questions. And we see the symptoms of attempts to do so where in research papers the process of labelling appears indistinct from the process of analysis, where it is not clear how the labels and networks are producing actual insights, where it is not clear that the payoff is justified from the complexity of analysis and where relatively blunt concepts and lists or 'systems' of semiotic features in fact act as a filter for how data is approached irrespective of what form of tools might better suit analysis. And where the starting point of the analysis, or the identified problem, is based in the need to account for multimodal communication and not based in existing scholarly work which points to existing knowledge and gaps in knowledge.

As regards the SFL based grand theory of multimodality the notion of context and of text, we have shown, are simply not suitable for carrying out problem-driven critical research. We cannot use a model that seeks context from within the text. Nor can we remove texts from their complexity at the macro and meso level, which links them into social practices and how the world and forms of social relations are already to some extent mapped out for us. And for CDS one crucial aim is to discover just this, pointing to the power relations which these legitimize and maintain, whether this is a particular discourse of war, or the way that food packaging allow consumer capitalism to assimilate notions of addressing environmental issues or being healthy. We showed that the notion of choice is important but thay these need to be shown to be contextually relevant and shaped always by the macro level.

As a basic model of analysis for doing multimodal analysis in a way that is aligned with the principles of CDS canons of use must be an important starting point. This is part of showing whose interests are at stake. One way to do this is to consult relevant literature which points to such uses – for example as regards research on photojournalism. We can collect examples of texts which we see as semiotic materials shaped as wholes and explain how they relate to canons of use and discourses. These are instances of canons of use, laden with meanings that not only come from 'within' the semiotic artefact, but are taken for granted and deeply ideological which should be revealed through analysis. We can establish the contextually relevant choices in a certain canon of use. This can be done both with content-to-expression and expression-to-content directionality. We do not wish to make grammar of everything but are dealing with something embodied and materialized that shapes how we think and act in this world. But a detailed analysis must be done, a model of analysis, a method, must be there, just that it must be tied to actual contexts. The texts which form the data can then be analysed, related to the relevant literature, canons of use and discourses. And of utmost importance semiotic aspects only need to be analyzed which are necessary for answering the research question (Ledin & Machin, 2018b). As Bateman et al. (2017, p. 215) stress, any multimodal analysis must start with 'fixing the analytical focus' according to a research question that is to be pursued, which involves making decisions on which kinds of semiotic materials and contexts are to be explored.

As a final comment, SFL and the forms of multimodality built around it, have been hugely inspiring and transformed the landscape of visual communication analysis. The concepts we introduce in this paper and the sequence for carrying our analysis do not reject these by any means, but, use them in slightly different ways, hopefully learning from how we see them used in scholarly work. But we think, at least, we have indicated that these concepts and assumptions must be reflected upon, perhaps approached in new ways, placed alongside and engage with other kinds of theories and models in order to make them more robust.

Disclosure statement

Notes on contributors

Per Ledin, Department Culture and Education at Södertörn University, Sweden. He has published widely in different areas of discourse studies, including writing development, multimodality and critical linguistics. His recent publications include papers on the assessment of writing tests, the semiotics of lists and tables and the language of New Public Management. His recent book is Doing Visual Analysis (2018).

David Machin, Department Media and Communication at Örebro University Sweden. His publications include: The Language of Crime and Deviance (2012), The Language of War Monuments (2013), Visual Journalism (2015) and Doing Visual Analysis (2018). He is co-editor of the international peer reviewed journals Journal of Language and Politics and Social Semiotics.

ORCID

Per Ledin http://orcid.org/0000-0002-3176-7226

References

Bache, C. (2013). Grammatical choice and communicative motivation: A radical systemic approach. In L. Fontaine, T. Bartlett, & G. O'Grady (Eds.), *Systemic functional linguistics: Exploring choice* (pp. 72–94). Cambridge: Cambridge University Press.

Bateman, J. A., Wildfeuer, J., & Hiippala, T. (2017). *Multimodality: Foundations, research and analysis: A problem-oriented introduction*. Berlin: De Gruyter Mouton.

Bateson, G. (1977). *Steps to an ecology of mind*. New York: Ballantine Books.

Bauldry, A., & Thibault, P. J. (2006). *Multimodal transcription and text analysis: A multimedia toolkit and coursebook*. London: Equinox.

Berge, K. L. (2012). Om forsjellene mellom systemisk-funksjonell lingvistikk og tekstvitenksap [On the differences between systemic-functional linguistics and text research]. In S. Matre, R. Solheim, & D. K. Sjøhelle (Eds.), *Teorier om tekst i møte med skolens lese- og skrivepraksiser* [Text theories meet reading and writing practices of schools] (pp. 72–90). Oslo: Universitetsforlaget.

Berry, M. (2013). Towards a study of the differences between formal written English and informal spoken English. In L. Fontaine, T. Bartlett, & G. O'Grady (Eds.), *Systemic functional linguistics: Exploring choice* (pp. 365–383). Cambridge: Cambridge University Press.

Bouvier, G. (2014). British press photographs and the misrepresentation of the 2011 'uprising' in Libya: A content analysis. In D. Machin (Ed.), *Visual communication* (pp. 281–299). Berlin: De Gruyter.

Cottle, S. (2009). Global crises and world news ecology. In S. Allan (Ed.), *The Routledge companion to news and journalism studies* (pp. 473–484). London: Routledge.

Djonov, E., & Zhao, S. (2018). Social semiotics: A theorist and a theory in retrospect and prospect. In S. Zhao, E. Djonov, A. Björkvall, & M. Boeriis (Eds.), *Advancing multimodal and critical discourse studies. Interdisciplinary research inspired by Theo Van Leeuwen's social semiotics* (pp. 1–18). New York: Routledge.

Fairclough, N. (1992). *Discourse and social change*. Cambridge: Polity Press.

Flowerdew, J., & Richardson, J. E. (2018). Introduction. In J. R. Flowerdew & J. E. Richardson (Eds.), *The Routledge handbook of critical discourse studies* (pp. 1–10). London: Routledge.

Fowler, R., Hodge, B., Kress, G., & Trew, T. (1979). *Language and Control*. London: Routledge.

Halliday, M. A. K. (1978). *Language as social semiotics: The social interpretation of language and meaning*. London: Edward Arnold.

Halliday, M. A. K. (1985). *An introduction to functional grammar* (First ed.). London: Edward Arnold.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Halliday, M. A. K., & Hasan, R. (1989). *Language, context, and text: Aspects of language in a social-semiotic perspective*. Oxford: Oxford University Press.

Halliday, M. A. K., & Matthiesen, C. (2014). *An introduction to functional grammar* (Fourth ed.). London: Routledge.

Hjelmslev, L. (1961). *Prolegomena to a theory of language*. Madison: University of Wisconsin Press.

Holmberg, P. (2012). Kontext som aktitivet, situationstyp och praktik: en kritisk anays av kontextxbegreppet i systemisk-funktionell teori [Context as activity, situation type and pactive: a critical analysis of the notion of context in systemic functional theory]. *Språk och stil*, *22*(1), 67–86.

Jewitt, C., Bezemer, J., & O'Halloran, K. (2016). *Introducing multimodality*. Milton Park: Routledge.

Kress, G., & Van Leeuwen, T. (1996). *Reading images: The grammar of visual design*. London: Routledge.

Ledin, P., & Machin, D. (2018a). Multimodal critical discourse analysis. In J. Flowerdew & J. E. Richardson (Eds.), *Routledge handbook of critical discourse studies* (pp. 60–76). London: Routledge.

Ledin, P., & Machin, D. (2018b). *Doing visual analysis. From theory to practice*. London: Sage.

Machin, D. (2016). The need for a social and affordance-driven multimodal critical discourse studies. *Discourse & Society*, *27*(3), 322–334.

Martin, J. R., & Rose, D. (2007). *Working with discourse: Mening beyond the clause*. London: Continuum.

O'Toole, M. (1994). *The language of displayed art*. London: Leicester University Press.

Richardson, J. E. (2007). *Analysing newspapers: An approach from critical discourse analysis*. Basingstoke: Palgrave Macmillan.

Sontag, S. (2004). *Regarding the Pain of Others*. New York: Picador.

Tagg, J. (1988). *The burden of representation: Essays on photographies and histories*. Basingstoke: Macmillan.

Taverniers, M. (2011). The syntax-semantics interface in systemic functional grammar: Halliday's interpretation of the Hjelmslevian model of stratification. *Journal of Pragmatics*, *43*(4), 1100–1126.

Tseng, C.-I. (2013). *Cohesion in film: Tracking film elements*. Houndmills, Basingstoke: Palgrave Macmillan.

Van Dijk, T. (2008). *Discourse and context: A sociocognitive approach*. Cambridge: Cambridge University Press.

Van Dijk, T. (2013). *CDA is NOT a method of critical discourse analysis*. In EDISO Debate – Association de Estudios Sobre Discurso y Sociedad. Retrieved from www.edisoportal.org/debate/115-cda-not-method-critical-discourse-analysis

Van Leeuwen, T. (1999). *Speech, music, sound*. Basingstoke: MacMillan.

Wagner, K. (2015). Reading packages: Social semiotics on the shelf. *Visual Communication*, *14*(2), 193–220.

Wodak, R., & Meyer, M. (2016). Critical discourse studies: History, Angenda, theory and methodology. In R. Wodak & M. Meyer (Eds.), *Methods of critical discourse studies* (pp. 1–22). Los Angeles: Sage.