Abstract

In Multimodal Discourse Kress and Van Leeuwen began to doubt “whether grammars of distinct modes are quite so uncontentiously ‘there’ as our own efforts […] suggests” (2001: 124). In the present paper, my point of departure is this doubt as I try to find where, if anywhere, these grammars may be found.

I propose the notion of ‘mode-instantiation’ as a theoretical way of describing what goes on in the combination of different meaning-making potentials. The model of mode instantiation is an attempt to describe how modes are instantiated from a ‘pre-mode potential’ into multimodal configurations in the local text. This results in an understanding of modes as a transitory abstraction made from the combined semiotic systems in a given communicational situation. Between the pre-modal potential and the instantiated mode, I describe hyper-mode patterns of typical choices of instantiation as ‘mode-registers’. Subsequently, modes are (re)defined as a locally instantiated assembly of semiotic systems, chosen from the pre-mode potential to obtain optimal meaning-making in a given communicative context.

1. Introduction

The recent work of scholars like Kress & Van Leeuwen (2001) and Baldry & Thibault (2006) on multimodal theory has abandoned seeing “mono-modes” as anything more than a mere theoretical abstraction. The tendency to understand the complex interplay of semiotic systems as poly-semiotic co-distribution of modes, was abandoned at the same time. As Lemke (1998) pointed out, the modes involved in a communicational situation are not simply added, they are multiplied. Hence, the meaning-making of the combined modes is something more and very different from discrete modes put together. This thought is in line with much film and media theory, and the logic behind it can be traced back to Eisenstein’s theories of ‘vertical montage’ (1959: 74). Eisenstein points out about film, that the different visual and audible choices are joined in a combined meaning that is more than simple addition. The meaning is something else and more – is synthesised from the combined choices made within what he calls the vertical montage.
2. The development of social semiotic multimodal theory

Over the last 15 years, social semiotic multimodal theory has evolved from a departure in Systemic Functional Linguistics. The development of the multimodal view of communication has gone through different stages from monomodality, over polymodality to the recent multimodal understanding of how meaning is instantiated in communication. This is elaborated in the following section.

2.1 The monomodal view

The figure above visually demonstrates what Kress & Van Leeuwen call the traditional monomodal view of communication that has been prominent in many linguistic approaches (2001: 4). In this view, choices of distinct elements from a meaning potential are instantiated in the utterance to realise a certain meaning. According to Kress and Van Leeuwen this view is the foundation on which linguistics has been conducted for many years – examining and theorising only one isolated mode (verbal language) at a time (Kress & Van Leeuwen 2001). However, due to the emergence of multimodal theory the view of communication has undergone gradual changes in recent years. The focus has increased to ultimately encompass all modes involved in any communicational situation.
As the founder of Systemic Functional Linguistics Halliday sees the possibility of applying the social semiotic view to all forms of communication. For instance, he states that verbal language is one of many semiotic systems used in communication, and that “We can use our grammatics to interpret these various non-linguistic semiotic systems, asking to what degree of specificity they are like natural language” (Halliday & Matthiessen 1997: 43). Hence, the move towards multimodal theory has evolved gradually from this conviction that many of the same semiotic functions are at play in any kind of meaning-making in any type of communication. This development was initiated by pioneering studies of other mono-modes examining their grammatical likeness to verbal language.

2.2 The polymodal view

Scholars like O’Toole (1994) and Kress and Van Leeuwen (1996) applied Halliday’s systemic functional view of communication to different types of pictorial communication, and later on Van Leeuwen did the same thing with sound (1999). Out of this grew a conviction that not only do the different modes share basic communicational functions, there also always seems to be more than one mode in play simultaneously in communication. The polymodal (polysemiotic) view stresses that choices from several different parallel semiotic systems are expressed simultaneously in communication.

Figure 2: A polymodal view of communication
The figure above demonstrates this basic view of the simultaneous expression of several simultaneously instantiated semiotic systems, and these systems are understood as parallel modes with separate semiotic systems. From this perspective, the analysis of multimodal texts is a matter of simple addition of observations made about each mode. The early polymodal theories made a distinction between verbal and visual communication (O’Toole 1994, Kress & Van Leeuwen 1996). This distinction is problematic since verbal and visual communication are in fact not modes, but rather choices of on the one hand semiotic subsystem (the verbal) and on the other sensory perception (the visual). From a multimodal perspective, verbal language is not actually a mode, rather, it is two very different modes – one is auditory spoken verbal language and the other is visual written verbal language. Even though a range of their semiotic subsystems may have similar functions, they are very differently instantiated. However, the monomodally oriented inquiries into multimodal texts made by the pioneering polymodal research brought about some very interesting insights into the interplay of different semiotic resources in communication. The multimodal approach evolved gradually from this.

2.3 The multimodal view

Because of important work done by scholars like Baldry and Thibault (2006), the multimodal theory of communication has gradually evolved into a view much like the one illustrated by Figure 3 below. This is a multimodal theory in which instantiated elements in the communication still derive from monomodal systems with their own predetermined meaning potential. However, this view is in contrast to the polymodal view. The monomodal potential is now seen as a mere theoretical abstraction because the systems are always intermeshed in the meaning-making. Hence, the different modes involved in any given communicational situation are never found in their pure form, but exist only in combination with other modes. In this view of multimodality, the different semiotic systems overlap and combine in various ways. For instance, spoken verbal language always occurs in a certain tone of voice, with a certain loudness and presence, and often in close interplay with e.g. gesture, posture and facial expressions. Thus, the different modes are ‘pre-meshed’ in their basic form, and their separation is only a descriptive measure that helps manage the theoretical complexity of the many systems. This is based on the belief that thorough description of the grammatical potential is made possible through separating communication into monomodes.
In the following, I will try to take the full consequence of the insights from social semiotic multimodal theory and as a result take the multimodal approach a small step further.

3. The multimodal mode-instantiation model

In its essence, I find the idea of pre-existing monomodal systems which often emerges in multimodal theory and analysis rather problematic. The idea is a reminiscence of the monomodal and polymodal views of communication. In my opinion, utilising the notion of monomodes is not an expedient way of reducing complexity as it does not correspond with the way in which multimodal meaning-making works. Meaning does not stem from choices made within different predefined modes, but from choosing combinations of autonomous systems.

When considering the fact that the traditional modes are always intertwined, I find it more productive to abandon any thoughts of pre-existing modes with more or less fixed meanings, however abstracted these systems may be. Instead, I assume the theoretical existence of a vast overall pre-mode hyper-potential, consisting of all possible semiotic choices in any communication situation.
3.1 Mode instantiation

If there are no monomodes, then multimodal theory should not theorise on them. Instead, I propose the notion of *mode instantiation* where every single communicational utterance instantiates its own unique local mode. This locally instantiated mode is not created by combining certain parts from different monomodes – it is in fact created by choosing and combining little chunks of meaning from more basic semiotic systems. The basic meaning potential is not a priori constrained to any single mode but simply, or rather complexly, consists of every single possible choice for meaning-making in any possible semiotic mode. Therefore this pre-mode potential, or ‘master-potential’, is the basis of all meaning-making. In choosing certain combinations of semiotic systems from the master-potential, the communicator produces the locally instantiated mode. Consisting of every single potential choice in any given communication, the master-potential is not structured in any way. The structuring comes with the instantiation. Based on the cultural context and communication situation, the communicator chooses from the master-potential which mode-configuration will be most apt for realising the intended meaning. Hence, conducting semiotic innovation not only means combining known potentials in new ways but also actually implementing new semiotic potential and thereby expanding the master-potential.

![Figure 4: The Multimodal Mode Instantiation Model](image-url)
Metaphorically speaking, the master potential can be conceived of as a “cloud of semiotic nuclear particles”. The cloud consists of loose semiotic particles that are little chunks of yet unarticulated meaning ready to be instantiated in combination with others. The particles will unite with other particles to form the communicative atoms we call semiotic (sub)systems. A number of atoms unite to join the larger structures of matter which I call instantiated modes. One single particle does not have any fixed meaning in itself; the meaning is created when the systems are united into (multimodal) utterances in context. The concept of multimodal mode instantiation is represented in the multimodal mode instantiation model above. This view functionally acknowledges the complexity of the vast multimodal semiotic meaning potential without theoretical reduction into monomodes. The potential of system-making in the instantiation can be seen as part of the pre-mode potential, but the system-making also seems rather closely related to the mode-registers as we shall see below.

An example of a simple sub-system from the semiotic master-potential could be the gestural moving one hand vertically down with open vertical palm like a slow karate chop. In itself, this movement has very little meaning, but in combination with verbal language and several other systems like facial expressions, stance and proxemics it may take on various very specific metafunctional meanings. The vertical hand movement can be used structurally (textually) to emphasise certain syllables or words, whereby it functions as a choice under the salience system. Structurally, the gesture may also be part of striking a rhythm for the combined multimodal utterance. This is similar to the grid structure in the visual text layout where spacing or framing is used to create a distinct visual pattern across the page to which the different clusters are related (Lupton 2004). Interpersonally, the simple hand gesture can utilise a sensation of ‘straightness’ as opposed to ‘roundness’ to the subject discussed, or it may function as an augmentation of an argument. Ideationally the vertical gesture can be a representation of a participant like placing a border or fence literally or figuratively.

Another example is the semiotic nuclear particles of hue, saturation and brightness which are some of the important distinctive features of the system of colour (Kress & Van Leeuwen 2002). The particles in themselves seem without meaning, but there are some inherent values even in the most basic choices. Colour can have interpersonal modal, textual structural, ideational, perspectival, identificational, and symbolic value depending on the local use. For instance, red is warmer than blue, dark is heavier than bright and strong saturation is more loaded with energy than pale colours (Itten 1961). However, the colour will not get its final meaning before it is implemented into its context of other semiotic systems in the local
mode, and into the context of situation. Red does not mean warm in a traffic signal and it does not mean stop on a flag. The flashing blue light of an ambulance does not signal cold and a blue tint expressing depth in a painting is not meant as a warning. In movies an interpersonal modality scale of basic tint of colour runs from blue to red, expressing a cline from the colder and maybe less emotional to the warmer and more intense (Zettl 1990: 67). Different colour systems potentially function at very different metafunctional levels as well.

3.2 Mode registers
When broadening the perspective to encompass cultural context, we discover patterns of typical ways of expressing meaning. Certain typical instantiation combinations of particles and systems emerge based on their aptness for accomplishing the optimal desired meaning-making in a given communicational situation. In Figure 4 above, I name these mode-registers because they are analytical abstractions based on contextual experience and perhaps precedence of tradition. The mode-registers are at the intermediate level between master-potential and the local mode in the model. Similarly to the traditional notion of modes, the mode-registers are theoretical abstractions based on cultural background knowledge. They are derived from several instances of similar instantiations in similar communication situations as typical ways of choosing particles, systems and their combination. Therefore, any mode-register has a tendency to shape our conception of the choices within the local mode. We find the mode more or less fitting to the mode-register expectations. Less fitting modes seem to have the tension of being more oriented towards other mode-registers, and as a result, they are often perceived as marked choices. In the newspaper article, for instance, we easily recognise the overall combined mode-register choices with different systems of image, layout, written verbal language, colour and so on. Deviations from the typical choices may be perceived as dissatisfactory communication, as creative discursive import, or maybe as a change of mode-register.

The mode-registers are what we relate to when communicating within a given discourse. They function as a theoretical way to structure the vast complexity of different locally instantiated modes. Due to their abstracted nature, mode-registers are not fixed categories, but function more like contours of structures seen through the haze of multiple encounters of communicational usage.

It may not seem clear exactly how this differs from the traditional definition of ‘mode’ (Kress & Van Leeuwen 2001; Baldry & Thibault 2006), as the phenomenon which I
refer to as ‘mode-register’ appears equivalent to what has hitherto been conceived of as ‘mode’. I would argue, however, that the perspective I present here is significantly different. In contrast to the former definition of mode, it stresses the notion that the mode-registers have no pre-existence to the actual instantiations, but are derived from multiple communicative situations created out of patterns of typical use.

In a TV-commercial the choice of medium preselects certain systems and modes and precludes others. Obviously, different olfactory and gustatory systems as well as tactility are not part of the semiosis of TV. On the other hand, obligatory systems are for instance salience, information structure (e.g. given-new), modality as choices in colour, light and detail, and different ideational systems within visual transitivity. There are vast numbers of possibilities of representation of systems – besides spoken or written verbal text as exclusive linguistic systems – such as intonation, voice, facial expression, head-movement, stance, gesture, body position, speech, clothing, gaze and so on. In some instances, text graphics are applied with systems of font, stroke, colour, shaping and others. Architecture, interior design, clothes design and furniture design are other possible modes. Most of these systems should in fact be seen as ideational meaning since they are represented within the moving image of the TV-commercial. The voice of a person is not delivered by vocal chords but by a loudspeaker representing the voice. The clothes design is not experienced as combinations of visual, tactile and audible sensations, but only represented visually in a certain manner and perhaps with augmented fabric sounds. Architecture is not experienced spatially, but is for instance represented through choice of angle, perspective and maybe sound reverb. This phenomenon of one mode being represented through another could be called ‘mode-representation’. Since it is often very difficult to distinguish between primary systems and systems from represented modes this interesting area needs much further clarification in the future.

We can describe the systems of each mode-register, the lexis and the syntagmatic structures in the lexico-grammar, but it is important to stress that the semantic meaning of the lexico-grammatical realisations are by no means fixed. Any utterance, be that verbal, visual or any other, is uttered in a certain context that defines the meaning. We read the context through the text, but also the text through the context. Some mode-registers share parts of their subsystems, sometimes in similar functional ways with similar semantic meaning, and sometimes with different functional consequences. As seen in the example above, colour may have different meaning in different mode-registers, but it may also have the same meaning. For instance, a bluish colour tint in part of the frame is a means of representing depth both in
photographic images and in paintings. The semiotic resources does not have pre-fixed meanings, but are dependent on the context of situation, on the context of the culture it is understood within, and on the choice of mode-register. However, in a given situational context within a given cultural context, precedence constrains the range of meanings possible in a given semiotic potential.

Systems may also be used differently across media. For instance, the typical use of the system of gaze is not the same on film and TV. In movies, a character’s direct eye contact with the viewer is rare. As a result, the typical image act is ‘Offer’ (Kress & Van Leeuwen 2006: 116). When a character in a movie makes eye contact with the camera, making the image act ‘Demand’ (Kress & Van Leeuwen 2006: 116), this is a highly marked choice where the narration comes to a halt. In news programmes, it is the other way around. When the news anchor looks into the wrong camera, away from the viewer making the image act of Offer, this marked choice brings the news-flow to a halt.

4. Multimodal writing
The selected media constrains the possible choices to certain parts of the master-potential. Choices are based on a communicational wish to get a message across, and therefore we may choose what seems to optimally convey the intended meaning. However, the decision is not made between different media or different modes (in the traditional sense), but between combinations of semiotic subsystems that does the job. In those choices, we may relate to a certain mode-register and consider if we want to be close to the typical instantiation or be more innovative at the risk of interfering with the understanding.

When lecturing, as an example of verbal face-to-face communication, I have a choice of using gesture or not. Not instantiating systems of gesture at all is, of course, a communicative choice in itself, which may convey self certainty, self control, strength, clear focus or seriousness, or it may in some cases convey an image of an overly self-controlled, uptight or uninvolved person. This all depends on the situation and on the combination of systems. I may choose to use gesture systems as mainly structural elements, emphasising important parts of my speech (salience to certain sounds through gesture), striking a beat to show the progression of the speech, or visibly using my fingers for counting to illustrate the structural parts of the speech. Systems of gesture may be used to convey interpersonal meaning e.g. in showing enthusiasm through strong or soft movements, conveying the modality of the spoken, making ‘air quotes’ to add sarcasm or pretending to put a finger in the
throat to add the meaning of strong dislike. I may hold my hands in a specific way to show my uncertainty about what I am saying or I may punch my palm to show certainty. Gesture can also be used ideationally to represent elements that I speak about, saying “here we have this opinion, here we have another, and between the two is the new view”, pointing to different areas in the space in front of me. Systems can be utilised to convey meaning in all metafunctions.

When marketing a product for a major company or brand there are many possible choices. The marketing strategy is based on the meaning we wish to convey to the receiver, and as a result, we choose the system combination best suited for that purpose. The modes occur from the chosen combination of systems and they are implemented through the optimal medium. The choice of medium has semiotic importance in itself. A luxury product will usually not be marketed on fliers of poor paper quality, and the marketing of an inexpensive lower quality product will not benefit from choosing an exclusive fashion magazine as medium. A choice of direct mail usually entails the choice of paper and envelope, and within the constraints of these media, we have various systemic choices. Should there be imagery? Should it be photographic, drawn or painted? Should there be written verbal language? Which typographical choices should be made? What about colour and paper quality? Hence, we base a multimodal communication strategy on considerations about the people we wish to communicate to, what we wish to communicate and how we do that to the best effect. The latter includes a number of basic choices: which semiotic systems to implement in the local mode, through which media, in relation to which mode-register, and in what communicational situation.

5. The multimodal grammars are there
This paper emphasises the fact that multimodal complexity should not be reduced by utilising theoretically abstracted monomodes, because meaning is not created in combinations of monomodes but in combinations of systems. Therefore, the description of multimodal grammar must acknowledge this fact and the theory must handle the combined systems as they actually function. This is a challenge because of the complexity of the systems, and even more so because of the way that the same systems may be implemented in various places in the metafunctional structure.

The multimodal mode-instantiation model describes the instantiated versus the potential in the light of typical choices. At the bottom, we find the locally instantiated mode at
the level of our final instantiated choices. At the top, we have the vast pre-modal master-potential, from which we make our selections. In the middle we find the mode-registers to which we relate our choices, because they are patterns of typical ways of instantiating parts of the master potential in modes in certain communicative situations.

The grammars are there, but they exist in two ways: 1) in the grammar of the locally instantiated mode, as a highly specialised system of meaning-making in a certain communicational context, and 2) in what I call mode-registers, which is closer to the commonly understood systemic functional conception, where the systems of the grammar of a given mode-register are derived from the way it is used in many situational contexts within a given culture. However, the number of mode-registers is larger and they are more complex than in the former definition of modes.

### 5.1 The multimodal paradox

The multimodal approach stresses the fact that communication can never be monomodal. As a result, it seems inadequate to apply theoretical and analytical focus to just one single monomode. The consequence of the non-existence of mono-modes, one could argue, is that the notion of multimodality actually becomes a paradox in itself since multimodality requires the existence of exclusive modes. In other words, to speak of multimodal communication is to say that there is monomodal communication, and consequently, to speak of multimodality is to say that there is monomodality. The inherent paradox in multimodal theory may be a reminiscence of the oppositional position of social semiotic multimodal theory to its linguistic origin in Systemic Functional Linguistics. As a result, much effort has been put into arguing that non-verbal modes are equal to the verbal. The term ‘multi-semiotic’ may appropriately convey the fact that it is combinations of very different semiotic (sub)systems that combine into the local modes, but the same paradox emerges here as this would require an existence of mono-semiotic communication, which is another impossibility.

As the multimodal paradigm gradually celebrates its success worldwide, the term multimodality may eventually be abandoned. In the future, we may return to speaking of semiotics instead, but this time around from a different perspective, i.e. from that of seeing the grammar of communication as a diverse combination of systems different from what used to be called monomodes. What lies in the future development of multimodal theory is the description of the distinguishing features of different semiotic systems and their typical combination into grammars.
6. Final Remarks

Multimodal analysis may seem insuperable because qualified analysis of any type of communication entails a detailed knowledge of every single (traditional) mode in the instantiation, and ultimately every part of the great global pre-mode master-potential. It is of course impossible to know the entire master-potential because of the vast complexity of different forms of communication and because this potential is constantly modified through semiotic innovation (Van Leeuwen 2005: 26). In other words, we cannot know the grammar of all possible modes. However, there is no doubt that, from a social-semiotic point of view, multimodal analysis should be grounded in grammatical close analysis of a kind. So, what we can do, since most of us are only specialised in one or a few of the traditional modes, is to use this specialised knowledge to add analytical depth to the multimodal analysis. We can make an informed analysis of multimodal communication from a basic social semiotic multimodal approach, and anchor our work in highly specified lexico-grammatical analysis of one or a few of the modes. Specialists in linguistics may apply this methodological base to the linguistic aspects of a more general multimodal analysis and specialists in visual communication may utilise their specialist knowledge about the visual as the anchoring of their analysis.

Ultimately, perhaps the most qualified approach to multimodal analysis is cooperative analysis where several analysts unite different key competences in very strong multi-semiotic analysis. However, to make this approach feasible, there must be a common methodological point of departure – a common ground that makes it possible to say a great deal about the many modes involved in communication without being an expert on every single one of them. This common ground could very well be social semiotic multimodal theory.

Morten Boeriis, Ph.D.-student
Institute of Language and Communication
University of Southern Denmark
boeriis@language.sdu.dk
References


