Multimodality: A guide for linguists
Jeff Bezemer & Carey Jewitt

Manuscript to appear in L. Litosseliti (ed), *Research Methods in Linguistics* (2nd edition). London: Continuum.

Chapter outline

This chapter provides a rationale for a multimodal perspective on meaning, communication and discourse. It draws attention to the range of different modes that people use beyond speech and writing and explores the theoretical and methodological implications of multimodality. The chapter addresses two key questions. First, what is multimodality? Why and how is it relevant to linguistics? Second, how has multimodality been taken up? What questions are addressed, what materials are collected, and what methods are used to analyze these materials? Taking one approach – Social Semiotics – and one area of research – online text making – as an example, the chapter defines and illustrates key concepts and steps in multimodal inquiry. It concludes with consideration of future directions.

1. What is multimodality?

'Multimodality' is a term that is now widely used in the academic world. The number of publication titles featuring the term has grown exponentially since it was first coined in the mid-1990s. Since then, a myriad of conferences, monographs, edited volumes, and other academic discussion forums have been produced that are dedicated to multimodality. Its contributing scholars come from many different disciplines, including linguistics, semiotics, media studies, new literacy studies, education, sociology and psychology.

With the term being used so frequently and widely, it may seem as though these scholars have identified a shared phenomenon of interest, a common object of study. Indeed, we can, in relatively generic terms, describe that phenomenon, or object of interest as something like, 'we *make meaning* in a variety of different ways', or, 'we *communicate* in a variety of different ways'. Yet we must immediately add to that that 'multimodality' (and related concepts, including 'mode'/'modality', '(semiotic) resource') is differently defined. Exactly how the concept is articulated and 'operationalized' varies widely, both across and within the different disciplines and research traditions in which the term is now commonly used. Therefore, it is very difficult and potentially problematic to talk about multimodality without making explicit one's theoretical and methodological stance.

If a 'means for making meaning' is a 'modality', or 'mode', as it is usually called, then we might say that the term 'multimodality' is a recognition of the fact that people use multiple means of meaning making. But that recognition alone does not accurately describe the notion of multimodality. After all, Saussure, writing in the early 20th century, already suggested that 'linguistics' was a 'branch' of a more general science he called 'semiology'. Since then the branches of that imaginary science have continued to specialise in the study of one or a small set of means for making meaning: linguistics on speech and writing, semiotics on image and film,

musicology on music; and new sub-disciplines have emerged: visual sociology, which is concerned with, e.g., photography; visual anthropology, which is concerned with, e.g., dress. These (sub-)disciplines focus on the means of meaning making that fall within their 'remit'; they do not systematically investigate synergies between the modes that fall inside and outside that remit.

Multimodality questions that strict 'division of labour' among the disciplines traditionally focused on meaning making on the grounds that in the world we're trying to account for *means of meaning making* do not operate in isolation, but almost *always appear together*: image with writing, speech with gesture, maths symbolism with writing, and so forth. It is that recognition of the need for studying how different kinds of meaning making are combined into an *integrated, multimodal whole* that scholars called for when they started using the term multimodality in the late 1990s. It was a recognition of the need to move beyond the empirical boundaries of existing disciplines and develop theories and methods that can account for the ways in which we use gesture, inscription, speech and other means together, to produce meanings that cannot be accounted for by any of the existing disciplines. This fact only became more noticeable with the introduction of digital technologies, which enable people to combine means of making meaning that were more difficult or impossible to disseminate before, for the majority of people anyway (moving image being one pertinent example). So that is how the introduction of the notion of 'multimodality' marks a significant turn in theorizing and analysing meaning.

What the early adopters of the term recognized was not only the need to look at the co-occurrence and interplay of different means of making meaning, but also that each 'mode' offers *distinct possibilities and constraints*. It had often been argued (e.g. by Saussure, and Vygotsky) that language has, ultimately, the highest 'reach', that it can serve the widest range of communicative functions, or that it enables the highest, most complex forms of thinking, and is therefore the 'most important'. In multimodality, the working assumption is that there are *differences* between semiotic resources in terms of the possibilities they offer for making meaning, and that it is not the case that one resource has *more or less* potential than the other. Thus multimodality marks a departure from the traditional opposition of 'verbal' and 'non-verbal' communication, which presumes that the verbal is primary, and that all other means of making meaning can be dealt with by one and the same term.

Based on these early references, we can formulate three *key premises* of multimodality:

> 1. Meaning is made with different semiotic resources, each offering distinct potentialities and limitations;
> 2. Meaning making involves the production of multimodal wholes;
> 3. If we want to study meaning, we need to attend to all semiotic resources being used to make a complete whole.

We should add some important footnotes to this. First, not everyone working in multimodality uses the notion of 'meaning making'. Depending on their disciplinary background and focus, they might say that they are interested in 'multimodal *communication*', 'multimodal *discourse*', or 'multimodal *interaction*'. We will use the term 'meaning making' unless we are writing about a specific approach to

multimodality. Nor does everyone working in multimodality use the the term 'mode': some prefer to talk about 'resource', or 'semiotic resource', and generally avoid to draw strong boundaries between different resources, highlighting instead the significance of the multimodal whole ('Gestalt'). Indeed, some scholars whose work we subsume under the heading of 'multimodality' do not use that term themselves, for that very reason; while otherwise committing to the three key premises we just presented.

Second, scholarly interest in the connections between different means of making meaning does pre-date the notion of multimodality. For instance, the study of gesture and its relation to speech, gaze and the built environment has a long history, in linguistic anthropology interactional sociology, and other disciplines (see e.g. Mehan 1980, Goffman 1981, Kendon 2004), while the relation between image and writing has been studied in semiotics (e.g. Barthes 1977/1964). These early contributions have produced important insights in what we now call multimodality. At the same time, we should note that the potential empirical scope of multimodality goes further still. We can see a development from an exclusive interest in language to an interest in language and its relations to other means of making meaning, to an interest in making meaning more generally, without a clear base point: language, or any other mode.

Third, while those using the term 'multimodality' generally aim to develop a framework that accounts for the ways in which people combine distinctly different kinds of meaning making, their *epistemological perspectives* (i.e. their perspective on how we can know the world) are different. As we shall see later on in this chapter, in some approaches to multimodality the assumption is that it is possible and indeed necessary to develop an integrated theoretical and methodological framework for *some* kinds of meaning making, for instance for the study of speech, gesture, gaze and the material environment. In other approaches, the assumption is that it is possible and necessary to develop an encompassing theoretical and methodological framework to account for *all* kinds of meaning making –whether in image or in gesture or in writing, or any other mode. So researchers who adopt the notion of multimodality (or whose work is treated by others as being part of the field of multimodality) still draw different boundaries around what it is in the empirical world that they aim to account for. This is not a matter of ambition, but a matter of epistemology: some argue that the differences between, say, image and speech are too great to handle within one and the same framework; others argue that, notwithstanding the differences, it is still possible, at a more general level, to establish common principles of meaning making.

2. Theories and methods in multimodality

Linguistics has traditionally defined its object of study around two particular modes: speech and writing; and from there some linguists have 'branched out' to explore connections with other modes. Other disciplines, such as psychology, sociology, (social) semiotics, and anthropology, had defined their object of study in more general terms, around notions of 'meaning making'. Within linguistics, as indeed in the other disciplines that contribute to the field of multimodality, there's considerable variation in terms of theoretical and methodological outlook. For example, multimodality has been taken up in Discourse Analysis (Scollon & Scollon 2003), Conversation

Analysis (Streeck et al 2011), Systemic Functional Linguistics (O'Halloran & Lim 2014) and Cognitive Linguistics (Forceville 2006). There's also significant variation in methodology: some studies are in-depth analyses of single cases (e.g. texts or strips of interaction); others draw on large corpora and aim to test hypotheses. In many studies, selected elements of different approaches have been adopted and brought into connection with concepts and methods derived from other disciplines. For instance, eye-tracking technology has been used to 'test' certain concepts proposed in social semiotics (Holsanova 2012). Other work has attempted to bring together concepts from social semiotics with ethnography (Dicks et al. 2011).

One consequence of this take-up of multimodality is that old names for disciplines have become misnomers. For instance, the terms 'Conversation Analysis' (CA), or 'Systemic-Functional Linguistics' (SFL) no longer match the scope of the disciplines they describe. New terms have been suggested to mark the changing scopes of these disciplines, ('Multimodal Discourse Analysis', 'Multimodal CA') but they have not been widely adopted, and are unlikely to 'settle'.

Each of these (linguistic) traditions in which multimodality was taken up also has different terminological preferences coupled with different conceptualisations of what we have described so far as 'means for making meaning'. In some traditions, such as social semiotics, the terms 'mode' and 'semiotic resource' are both used, and definitions have been proposed that make a distinction between the two. In others, such as conversation analysis, '(semiotic) resource' is used, but 'mode' is not, or very rarely; and some attempts at defining '(semiotic) resource' have been made. Yet none of these definitions is (as yet) widely and consistently used beyond those who proposed them.

There is, put simply, much variation in the conceptualization and definition of mode and (semiotic) resource. Gesture and gaze, image and writing seem plausible candidates, but what about colour, or layout? And is photography a separate mode? What about facial expression, and body posture? Are action and movement modes? You will find different answers to these questions not only between different research publications but also within. To avoid potential confusion, it is important to make a deliberate decision on what categories and terms to use when engaging with multimodal research. It will be helpful to formulate some 'working definitions', drawing on the ones already put forward by the approach you adopt. Even though the working definition is unlikely to be entirely satisfactory, it is important to strive for maximum conceptual clarity and consistency.

As well as theoretical variation, there are significant differences in epistemological positions. Some 'multimodalists' look at human artefacts –technologies, tools, objects- only insofar as they are being oriented to in observed interactions. So for instance, Charles Goodwin (2000) looked at the 'Munsell chart', a tool used to determine the colour of soil by the archaeologists participating in the interactions he had videotaped. Yet others have studied artefacts away from specific situated interactions. For instance, Bezemer & Kress (2008) studied textbooks. Through close analysis of selected texts they explored how makers of textbooks –authors, picture editors, graphic designers- use writing, image, layout etc. to make meaning; they did not look at how these texts were actually interpreted and used by teachers or students

in classrooms. Underpinning this difference are quite different ideas about whether and how researchers can 'get at' meanings made.

3. A multimodal perspective on language

Multimodality challenges the idea –still widespread in linguistics and beyond- that 'language' is the single most powerful mode of communication. Here's how.

Communities and contexts of use shape what modes are used and how

It is tricky to make general claims about what 'people' do with language. We need to ask, for *who*, if at all, is language the most resourceful, widely used and important mode of all, and in what *contexts of use*?

Most scholarly work on communication is focused on a very specific community of meaning makers: the hearing community. If we consider signing communities, we can see that gesture can be just as important in a community as speech can be, in that it can be used to serve all social and communicational needs. If we based our account of the significance of language on an inquiry of communication between babies and their parents, or between blind people, or between people who do not share a language, we would get an entirely different picture. We could carry on: we could observe the language use of people with aphasia, or of people on the autism spectrum; in each case, language plays a different, particular role. And of course, even when considering the hearing and speaking majority, we would find significant variation. People spending a lot of time taking pictures, making music, dancing, painting, pottering, whatever – may well ascribe the highest expressive potential to modes other than language. Put simply, language has a different status in different *communities* and in the *repertoires* of different people.

Language also has differing status in different *contexts of use*. Much, if not most, communication happens without the use of speech (or writing). We only need to walk out into the public space to realize just how much is being communicated without the use of speech and writing. There are, of course, signs (in the everyday sense of the word) placed everywhere, often not featuring any written language, which we may or may not attend to and interpret; and there are signs made (in a semiotic sense) by other users of the public space. When we cross a street, we might momentarily communicate with the driver of a car that is approaching. The driver might use facial expression and gestures to communicate that they give way for you to cross the street. But before that happened, you will have anticipated what the driver is likely to do next on the basis of your recognition and interpretation of changes in speed and direction of the car / driver. On these occasions, gesture, facial expression and other modes fulfil all the social needs of a situation.

The point is that language, like any other mode, is used differently in different activities and artefacts. Of course, if you are drawn to those activities in which speech and/or writing appear to play a central role, you are likely to be left thinking that language is, overall, the dominant mode in the lives of everybody, always and everywhere.

Another point is that if you are interested in social practices and you use language as a 'way in' to those practices (as e.g. linguistic ethnographers do, see Creese and Copland, this volume) you need to be aware of what might remain inaccessible to you.

Many 'linguistic' principles are actually general semiotic principles

The claim that language is the most resourceful, important and widely used mode of all is, at the very least, premature. There is a relatively long history of the study of language and much progress has been made in developing means of characterising language in a highly detailed manner. As long as such sophisticated toolkits do not exist for the description of, say, the resources of gesture, colour, dress, or scent, we have no means of 'evidencing' the resourcefulness of such modes. In the meantime, it would be premature to conclude that language has, overall, more meaning potential than other modes.

Indeed, studies in multimodality do suggest that some of the principles and properties traditionally attributed to language can also be found in other modes. For instance, in their study of *image*, Kress and van Leeuwen (2006) showed that image not only has the equivalent of what linguists call lexis, it also has a 'grammar':

> "Just as grammars of language describe how words combine in clauses, sentences and texts, so our visual 'grammar' will describe the way in which depicted elements –people, places and things- combine in visual 'statements' of greater or lesser complexity and extension." (p. 1)

We should point out that the risk of linguistics 'branching out' is that the new territory is described in the terms of the originating discipline. Indeed this is a common critique of linguistic perspectives on multimodality. When expanding the traditional scope it is important to keep a close eye on what is typical of a mode or semiotic resource and what may count as a more general principle of meaning making; and make sure that linguistic categories are not imposed onto other modes. Every time the frame is expanded, old terms and categories need to be revisited and re-evaluated, in the light of the wider range of empirical cases being considered. So we might ask, 'What would the counterpart be of a verb in image?' Only if we then immediately add: 'Maybe image doesn't have anything like the verb. Maybe it has categories unlike anything language has.'

Each mode offers distinct possibilities and limitations

Language may have unique possibilities, but it also has unique limitations, like any other mode. While this may seem obvious, linguists such as Jerrold Katz have, not so long ago, suggested that one unique property of language is its 'principle of effability': "…there is nothing to indicate that there is any type of information that cannot be communicated by the sentences of a natural language" (Katz 1972: 19). Let's take a moment to reflect on the implications of this statement. It means that someone being 'lost for words' points to a limitation on the side of the language user,

always. It is never a reflection of the limitations of the meaning potential of language. Katz's claim also means that any meaning made in any mode can be *transcribed*, without any significant 'gains' or 'losses' in meaning.

Observation of meaning making suggests otherwise. To give one brief example: surgeons have developed a specialist language for describing the human body. And yet, when you listen to what surgeons say when they operate, you will often hear them referring to 'that stuff', or 'that bit over there' (Bezemer et al. 2014). In spite of a history of anatomical study that began in 1600BC, language only provides a fraction of the resources required to communicate the information needed in this situation. The pointing gesture cannot, in fact, be transcribed without losing some of its precision. Indeed, social semioticians have argued that any attempt to translate something into words always involves a kind of 'transformation', or 'transduction' (Kress 2010).

If we want to map the potentialities and limitations of different modes we need to attend to what modes have in common as much as what is distinct about them. When studying meaning making across different modes, we can identify which semiotic principles are shared, and how these principles are realised differently in each mode. This goes back to an ambition formulated by Jakobson, who proposed that semiotics 'deals with those general principles which underlie the structure of all signs whatever and with the character of their utilization within messages, as well as with the specifics of the various sign systems.' (1968: 698)

An example of how this ambition is approached in present day social semiotics is given by Bezemer & Kress (2016):

> "All communities need means for expressing/realizing (the general semiotic feature) *intensity*. In the mode of *speech* that is realized by the intensity of sound – 'loudness'; it is also realized lexically, e.g. as "very". Lexis is available in the mode of *writing* as well; here intensity can also be indicated by visual prominence, as in the use of a bold **font**, or by CAPITALIZING. In the mode of *gesture* intensity might be realized by the *speed of movement* of the hand, or by the *extent* of the movement. In the mode of *colour* it might be done through degrees of saturation." (Bezemer & Kress 2016: 17)

In other words, a multimodal perspective draws attention to the *general* and the *particular* in speech and writing, recognizing both what they have in common, with each other and with other modes, and how they are distinctly different. From a multimodal perspective, therefore, speech, or writing, are not 'more' resourceful, but 'differently' resourceful; they do not have more potential, but different potential for making meaning – just like any other mode.

As you are reading this you might be thinking, "Well, that's all very exciting, but I'd define my interest more narrowly; if there's no or hardly any language involved, it falls outside my area of interest. I'm a linguist, after all." To this we would say that the meaningful units that people produce are almost always multimodal. In some ways the argument is not unlike that put to medical specialists who lose sight of the whole body when dealing with only a part of it. It is the 'body' that constitutes the meaningful whole – organs are merely constituent, interacting parts of it. In the same

way, multimodality encourages us to establish what the contribution is of each to the construction of a meaningful whole – a 'text'.

4. Social semiotics

To illustrate how multimodality has been taken up and how it can be 'operationalised' we now turn to one specific approach to multimodality: Social Semiotics (remember that this is just one of many different approaches to multimodality).

The key features of a social semiotic approach to multimodality may be summarised as follows:

- *Aim:* To recognize the agency of social actors and social/power relations between them.
- *History:* Pioneered by Gunther Kress and Bob Hodge in the early 1980s in Australia, building on critical linguistics, SFL, semiotics and social theory. van Leeuwen brought inspiration from music and film studies.
- *Theory of meaning:* Based on the notion of the motivated sign (Kress 1993), which holds that the relation between 'signifier' and 'signified' is always motivated, and never 'arbitrary' as Saussure suggested.
- *Concept of mode:* Central to social semiotic theory. Indeed most theorizing on what counts as mode comes from social semiotics. A short definition would be 'a set of socially and culturally shaped resources for making meaning' that has distinct 'affordances' (cf. Kress 2010).
- *Empirical focus:* Initially focused on 'artefacts' (especially print media, film and games – both 'professional', e.g. an advert in a magazine; and 'vernacular', e.g. a child's drawing), then also began to account for social interaction recorded on video through field work.
- *Methodology:* Typically detailed analysis of selected small fragments (e.g. one or a small set of drawing(s)), sometimes involving historical comparisons; is often combined with ethnography.
- Typical *research question*: Mavers (2011) looked at a teacher's instructions and the drawings that children made subsequently in a science classroom. As in any social semiotic study, questions she addressed included: How did the sign makers use the modes available to them (in this case drawing and writing) to re-present the world? What did they attend to? What did they highlight? What was gained and lost in the process of 'translating' from one mode to another?

Social semiotics takes the notion of 'sign' as its starting point. Signs are elements in which the 'signified' (a 'meaning') and the 'signifier' (a material 'form') have been brought together. In our social semiotic take, the 'sign' has three characteristics. First, the relation of form and meaning is 'motivated', that is, the relation between the two is not an *arbitrary* one (see e.g. Kress 1993). What that means is that the form is, in some of its aspects (characteristics, features) taken by the maker of the sign as being 'apt' to serve as the means of expressing the meaning at issue. In research by one of the authors of this chapter (Bezemer et al. 2014) a surgeon was observed making a relatively restricted back and forth sideways

movement with an instrument three times in quick succession. This gesture (the short back and forth movement with a 'grasper') 'means differently' to, say, a more extended movement, made more slowly, and made only once or twice. That is, the characteristics of this gesture are an apt means for what the surgeon wishes to communicate to his audience, namely 'a lack of obstruction'. These motivated relations of form and meaning are shaped by and materialize the 'interest' of the surgeon at the moment of making the sign.

The second characteristic is that the sign is always shaped by the environment in which it is made, and its 'place' in that environment. To make their signs, sign-makers choose from a range of modes which are available in their environment (often made available by other people). Each of the modes has distinct and different meaning potentials and therefore offers potentials for signs with different social effects. These potentials are drawn on by the maker of a sign in each specific instance, always in relation to the needs and requirements of that instance.

The third characteristic is that each mode offers certain potentials for making meaning: it has specific *affordances*. Signs are always made in a particular mode. As modes offer different potentials for making meaning, this entails that signs - and their effects - made in one mode differ from signs made in other modes. Sign-makers use existing signifiers in the different modes. In their use of these - always in specific environments - at times these resources are used more closely in line with 'convention' than at other times. For instance, a teacher might pick up a pen to point at a specific area on the whiteboard. The pen was not designed for pointing but, nevertheless, offers that as a potential signifier. The material qualities (and the history of use of this instrument) make it an apt resource for communication in this environment. In some ways this is not all that different to using a very large book as a door-stop: the book's potential in one of its characteristics – as large, heavy, moveable – is recognized for a present and 'un-conventional' purpose.

In other words, what is at issue here is a play, a tension, between stability and instability: resources with recognized potentials – whether as mode (e.g. *writing*) or as signifier (e.g. *word*, or syntactic form such as a sentence) – are constantly newly drawn into use. In periods of relative stability and in stable settings, this happens in relatively predictable (though nevertheless still always novel) ways. In periods of rapid social change, as in the present era of superdiversity and globalisation, and in settings that are in constant flux, uses of modes and of signs do take the forms and meanings envisaged by their makers, though less constrained and less fixed by convention.

The use of modes and the making of signs in and for specific environments, has to be seen in relation to broad social, geographical and temporal/historical conditions. Signs are made out of the awareness of the maker of the sign of a wide range of factors. The sign maker's rhetorical assessment of the physical-social site, and of the conditions which exist at the moment of the making of the sign all appear in how the sign is made. At the base of this perspective lies an insistence on maintaining the known yet often unrecognized distinction between signifier and sign: as has been pointed out (Scollon and Scollon 2003) a 'traffic sign', say, one indicating a speed-limit, is actually not a sign in the semiotic sense while it is lying on the back of the truck on which it is being transported to the site where it will be positioned.

The seeming difficulty of dealing with the 'meaning of colours' in some semiotically consistent fashion is another instance. How can 'red' mean danger or love or wealth or happiness; or 'white' mean death or joy; and so on. The problem ceases to be one when we recognize that 'a colour' is a signifier, not a sign. As *material* (and *signifier* therefore) the colour 'red' offers a range of potentials for meaning-*making*, different ones of which are differently taken up in different cultures and societies.

These two examples can in turn serve as metaphors for the use of specific modes as much as for the making of signs. In a study on changes in textbooks from the 1930s to the year 2004 (Bezemer & Kress 2008), it is clear that the role, the use, of modes in that particular medium has changed significantly over the last decades. For the first fifty years of that period – broadly speaking - writing had a central function – a *functional specialization*, in terms of 'carrying' the major *functional load* in relation to curricular content. Now modes such as image and layout have taken over some of the functional load previously carried by writing alone, and the functional specializations of each of these three modes is undergoing quite significant changes. What looks, seemingly, like merely a quantitative change (more images) is due to (deeper and prior) social and technological changes, which lead to a reshaping of the functions which a mode - writing in this case - will be used to fulfil in specific ensembles of modes in particular settings and media. These social and semiotic changes have profound implications for the conceptualisation and investigation of, e.g. 'reading' and 'writing', and it is these implications that social semiotics aims to address.

5. Example: Exploring text making on Facebook

We will now present an example of a social semiotic analysis, one that was aimed at this question of what it means to 'read' and 'write' in online platforms. The example is from Bezemer & Kress (2017). They addressed the following research question: How do young people make multimodal texts for dissemination on Facebook? The focus of the analysis was on posts made by one 12-year-old Dutch boy, Daan, on Facebook. One of these posts is represented here by Figure 1.
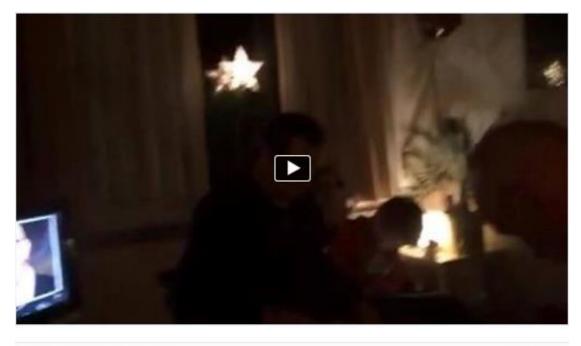
Figure 1: Daan's post on Facebook

Exploring the conditions in which the signs were made

Before going into the textual details some preliminary observations are in order. First, we note that these sign makers do not operate in the same *spatio-temporal frame*: they are in different locations, and their sign-making is differently organised temporally. There may, for instance, have been a 'pause' between Daan's post and a 'gap' in response, of an hour, a whole day, or even more.

Second, communication in this example is not face to face, but is mediated by an *online platform* (run by a major corporation), which provides a set of resources for making meaning. Third, the platform itself gives rise to, 'facilitates' the use of a certain set of *modes: writing*, for instance, and (*moving) image*. These modes enable sign makers to communicate across time and space, and to '*edit'*, that is, to review and remake a sign-complex before it is made available to others for interpretation. Fourth, the sign-making realises a specific social relation: Friend-Friend.

Yet beyond these particularities we can identify common principles of composition and of communication. First, here as in the other cases, sign-makers make sign-complexes with their (imagined) audience in mind; they make signs which they believe are apt for that audience; this shows the sign-maker as rhetor. Second, sign-making can be described in terms of *design*, with sign-makers selecting modes for making signs and sign-complexes which they believe are apt for their rhetorical purposes, given the affordances of the modes chosen. A perspective on the sign-

maker as both rhetor and designer highlights the semiotic work of distributing functions across modes, or, of allocating different 'semiotic tasks' to different modes, much like a composer uses the characteristics of different musical instruments to carry the melody or to provide a particular background.

We then consider the *interest* of the sign maker (Kress 2010). 'Interest', in this sense, arises out of the text maker's social, cultural, affective, material experiences and present position in the world, shaping his or her attention to and engagement with the world. Daan's post demonstrates his *interest* in communicating to an audience of his friends and family about a festive occasion which seems to have appealed to him, as the event is unfolding. He selects features to which, given his *interest*, his attention is drawn: selected elements in this event and the social, cultural and material environment in which the event is taking place. We might say that these, with a heightened sense of the event, are the kinds of things that seem to this 12 year old's imagination particularly worth selecting and showing. He, in turn, wishes to draw the attention of his audience to these heightened, defining elements of this occasion.

In this, Daan makes choices about what to select and how best to represent what he wishes to convey, which also includes that this event has 'an atmosphere'. Only what is 'criterial' or essential in terms of his *interest* is represented; other features are left out. Those which are selected are then subject to Daan's design.

In the post, Daan draws on a potential of this platform for this instance; that is, the possibility of sharing (experiences, messages) with people with whom one is connected on the platform. Here, in the case of 'Facebook', the text is in the genre of *status update*. A common version of the genre 'status update' is the *report*, in which authors describe and/or show what they are doing or what they have just done: using the modes of *writing* and *video*.

Daan's post exemplifies the genre. Daan reported on what he was doing on the evening of New Year's Eve, 2013. On that evening, Daan was at his uncle's, celebrating with his father, his sister and his two younger brothers, and his two cousins. Central to their celebratory gathering was a type of dining the Dutch call 'gourmetten'. It involves gathering around a dinner table with a raclette at the centre. The raclette comes with several little pans, allowing the guests to prepare their own food, grilling charcuterie and vegetables from platters provided. It is a popular form of dining for special occasions, for instance during the festive season.

This makes the post coherent with its *context of use:* it was posted on New Year's Eve, when people are celebrating, engaging in more or less predictable activities. The date of posting appearing above the *post* (which is automatically generated by Facebook) enables readers to reconstruct the temporal frame, if they don't read the post as it has come in.

From exploring Daan's Facebook 'profile', and from talking to him, we know that his Facebook audience (then) consisted of 33 Friends. They included 20 'peers' (13 boys, 7 girls), including classmates, friends from the neighbourhood, and his sister; 3 cousins, all 16+; and 10 adults, including his mother, 6 uncles and 1 aunt, and two female adult friends of the family. Except for 3 of his uncles and the aunt, all 'Friends' lived in the same city; the majority in the same neighbourhood, where he

met them face-to-face on a daily basis. He himself did not (yet) post frequently on Facebook—23 posts in the first year; but he read what his Friends post every day. Most of the time, he accessed Facebook on his iPhone 5, using a Facebook App.

Exploring the sign complex

So what sign-complex did Daan design for this audience, what multimodal ensemble did he design, using the modes available to him? His 'report' consists of signs made in *writing*, and in a *video*, with both writing and video presented simultaneously in a single frame.

The *written* element is in Dutch, and would, roughly, translate to 'Enjoying dining with uncle and father and cousin' (a word-by-word translation would be, 'Nice with uncle and father and cousin dining'). The adverb in first position, 'Leuk' ('nice'), modifies what follows: 'met oom en vader en neef koermeten', suggesting why that which follows is noteworthy / worth 'sharing': the event affected the sign maker's mood. The prepositional phrase in second position, 'met oom and vader and neef', *describes* a selection of the people around him. He describes the participants using terms that indicate their family relation, well suited for an audience that is not familiar with the proper names of his relatives. The verb in last position, *koermeten* (spelled as a non-standard variation of 'gourmetten'='dining'), describes the social event well understood by his (predominantly Dutch) friends on Facebook.

The *video* which Daan made using his iPhone is 12 seconds long. The camera work is shaky. The frame moves from right to left and back, a 'pan' giving a 'panoramic', 180 degrees close-up view of the camera holder's surround from a low/eye-level angle. The video *shows* selected elements of the environment: partial shots of some people in a room; where in the room they are; and what they orient to. In the foreground, one adult is shown standing, orienting to an object on the table; a child tries to get in the frame of the camera. In the background, some people are on a sofa. The TV is on. It is night time. Two lights are visible, including one star-shaped light hanging in front of a window. In the dimmed light the people appear as silhouettes; vision is blurry; and as the camera moves quickly it is all the more difficult to identify people and objects. Fragments of speech are audible: one adult refers to food ('shoarma'); a child calls for mama; and there's sizzling of some kind.

Part of the design of any sign complex is the production of links between its constituent elements. This is about *cohesion*: "a potential for relating one element in the text to another, wherever they are and without any implication that everything in the text has some part in it." (Halliday and Hasan, 1976: 27). For instance, in Daan's sign-complex, the writing names an *activity* ('koermeten', i.e., the grilling of food on a raclette) which is coherent with the actions made visible and audible in the video: the sizzling sound and the spoken reference to 'shoarma' in the video. The writing also introduces *people* - an uncle, father and cousin - some of whom potentially feature in the moving image, which shows an adult and two children.

Cohesion is also produced by the *layout* (i.e. spatial arrangement as mode) of all the elements in the site where they appear, i.e. on the 'news feeds' of Daan's Friends. *News feeds* are vertically organised, a column marked by hairlines on either

side, with the most recent feeds appearing on top. Readers scroll down to read older feeds, which are separated by horizontal hairlines. In Daan's post, the writing appears *above* the film; so if we follow Facebook's top-down structure, we might say that in Daan's post *writing* appears *before* ('earlier than') the video. This ordering allows for the written element to serve as a frame for the video; a frame which is to be read before the video is watched. In Barthes' (1977) terms, writing here 'anchors' the image; it "*directs* the reader through the signifieds of the image" (p.40).

How *writing* and *video*, and the signs made in each, operate as a multimodal ensemble and as a sign-complex can also be explored by asking: What if the video was left out, or what if readers chose not to play the video? What does the video provide that the written sentence does not? We might say that *writing* is used here to produce an 'abstraction': concrete events are 'transcribed' by means of writing into generic categories, selecting some of the constituent elements of the occasion itself and of the video, while leaving out others. In this case, the writing does not describe many of the specifics of the circumstances, such as features of the setting; or indeed characteristics of the participants: what they look like, how they sound, et cetera. Above all, it does not give an 'impression' of the atmosphere as the video does (in Kress & Van Leeuwen's (2006) terms as a 'symbolic suggestive process'), depicting a 'generalized essence', the gist rather than the detail, complementing, filling in the sparse description given in writing: 'leuk' ('nice').

The example shows how modes operate in ensembles to serve *complementary functions*. *Writing* describes the social relations between the sign maker and the people represented in the written part of the overall text; it names the occasion of the gathering; and it provides an appraisal of the situation from the sign maker's point of view. None of this information is provided by the *moving image*, by *speech* or other modes in the video. The video shows some of the more specific actions involved in the event, and some of the characteristics of the participants not mentioned in the written sentence, including visual and vocal features, giving an impression of 'mood' or 'atmosphere'. Without one or the other, the text would not be the same; the signs are interwoven, mutually modifying.

Exploring responses to the sign complex

We can consider how a sign complex can 'prompt' the consecutive making of a sign-complex by the sign-makers who were addressed. It shows responses from some of the 'Friends' who have noticed and interpreted his post: six people responded, all of them 'liking' the post, while one – one of Daan's uncles - 'comments', 'Miss you Daan!!!'

When Friends noticed Daan's post and 're-made' - interpreted - the sign-complex, following their own principles, communication had happened. Whatever else it may be, communication is never a straightforward transmission of messages. Daan's Friends may have chosen not to follow the reading path suggested by the layout of the (visual) elements of the post, and they will have chosen their own degree of commitment for and focus of engagement. The signs made in response to Daan's post give us a glimpse of their engagement. For instance, some signified *affect,* e.g. by pressing 'like'.

6. Conclusions

The example demonstrates that sign makers are immensely semiotically 'resourceful': using resources creatively to serve their interests. A social semiotic framework aims to be generous in its recognition of those resources, and to highlight the full repertoire of semiotic resources that people need to develop in order to participate in seemingly straightforward instances of communication. Daan shows that in order to be successful at 'social networking', writing skills no longer adequately describe the competencies involved in using video, layout, writing (and possibly other resources we may not have attended to) which are made available by a 'platform' for the design of sign-complexes.

The example also illustrates how language becomes part of a bigger whole, namely a 'text' that is made with a number of different modes. If we wanted to analyse what the meaning maker constructed as a meaningful, coherent whole, we would need to treat the entire video as our object of inquiry, not just extract and examine a part (such as the spoken or the written) of that whole. If you want to understand how language is used within that text, you will have to attend to those other modes in the text as well: you can't interpret, or analyse what was said or written in isolation.

The example is an instance of communication on social media. It is the case that the texts we find on the internet are almost always multimodal - in fact, the challenge is: find one that is not! Multimodal texts such as these do of course pre-date the internet era, but digital technologies have given a majority access to the resources needed to produce and disseminate multimodal texts at relatively low cost. The new technologies now draw our attention to what had previously been possible to overlook.

We mentioned at the beginning of this chapter that multimodality is a diverse field. Indeed, it remains to be seen whether multimodality will develop into a distinct field, or whether, and if so how, theories and methods in multimodality will be integrated in disciplines like applied linguistics. Our prediction is that these two trends will continue in parallel. This means that multimodality will feature quite differently on different research agendas.

For communities that have adopted selected ideas about multimodality, the question will be whether and how these can be reconciled with their current theories and methods. In many disciplines, notably linguistics, the term multimodality is often used to evoke old ideas about the role of the 'non-verbal', rather than as a short-hand term for the far-reaching premises we outlined in this chapter.

The community of researchers 'doing' multimodality will continue to explore principles and means of making meaning across different modes, media and social domains. Work on now relatively well-documented modes, such as image, might focus on testing empirically some of the hypotheses that were proposed in the early days of multimodality, using quantitative methods such as multimodal corpus analysis (Bateman et al. 2017). Yet other work will aim to advance understanding of resources that have until now been largely ignored, such as touch (Jewitt 2017); or that have

been investigated in isolation in experimental rather than naturalistic inquiry, such as gaze and facial expression (Korkiakangas, in press). Exploration of different types of activity –for instance, 'walk-and-talk' (Broth & Mondada 2013) and play (Cowan 2014), will continue to advance understanding of meaning making beyond the traditional 'speech event'. For the 'multimodalists', the integration and synthesis of these understandings into encompassing frameworks will remain one of the biggest challenges (Bezemer & Kress 2016).

References

Barthes, R. (1977[1964]) Rhetoric of the image. In Image Music Text. London: Fontana.

Bateman, J. J. Wildfeuer & T. Hippala. (2017). Multimodality. Foundations, Research and Analysis: A Problem-Oriented Introduction. Berlin: Walter de Gruyter.

Bezemer J. & G. Kress. (2017). Young People, Facebook and Pedagogy: Recognizing Contemporary Forms of Multimodal Text Making. In: M. Kontopodis, C. Varvantakis & C. Wulf. Global Youth in Digital Trajectories. London: Routledge.

Bezemer, J. & G. Kress (2008). Writing in Multimodal Texts: a Social Semiotic Account of Designs for Learning. *Written Communication* 25 (2), 166-195.

Bezemer, J. & G. Kress. (2016). *Multimodality, Learning and Communication: A Social Semiotic Frame.* London: Routledge.

Bezemer, J., A. Cope, G. Kress & R. Kneebone (2014). Holding the Scalpel: Achieving Surgical Care in a Learning Environment. *Journal of Contemporary Ethnography* 43 (1), 38- 63.

Broth, M., E. Laurier, & L. Mondada (Eds.), *Studies of video practices: Video at work.* New York: Routledge/Taylor & Francis Group (pp. 1-29).

Cowan, K. (2014). Multimodal transcription of video: examining interaction in Early Years classrooms. *Classroom Discourse, 5 (1)*, 6-21.

Dicks, B., R. Flewitt, L. Lancaster & K. Pahl. (2011). Multimodality and ethnography: working at the intersection. Qualitative Research 11(3), 227-237.

Forceville, C. (2006). Non-verbal and multimodal metaphor in a cognitivist framework: Agendas for research. *Applications of Cognitive Linguistics* 1, 379-402.

Goffman, E. (1981). *Forms of Talk.* Oxford: Blackwell.

Goodwin, C. (2000). Action and Embodiment Within Situated Human Interaction. *Journal of Pragmatics* 32, 1489-1522.

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

Heath, C., P. Luff and J. Hindmarsh (2010). *Video in qualitative research.* London: Sage.

Holsanova, J. (2012). New Methods for Studying Visual Communication and multimodal integration. *Visual Communication* 11, 251-257.

Jakobson, R. 1968. *Selected Writings. Word and Language*. The Hague: Mouton and Co.

Jewitt, C. (2017) Towards a Multimodal Social Semiotic Agenda for Touch, In s. Zhao, E. Djonov, A. Bjorkvall and M. Boerils (Eds.) *Advancing Multimodal and Critical Discourse Studies: Interdisciplinary Research* Routledge: London. p. 79-93.

Jewitt, C., J. Bezemer & K. O'Halloran (2016). Introducing Multimodality. London: Routledge

Katz, J J. (1972). *Semantic Theory*. New York: Harper & Row.

Kendon, A. (2004). Gesture. Visible Action as Utterance. Cambridge: CUP.

Korkiakangas, T. (in press). Communication, gaze and autism: A multimodal interaction perspective. London: Routledge.

Kress, G. (1993). Against arbitrariness: The social production of the sign as a foundational issue in critical discourse analysis. *Discourse and society* 4 (2), 169-191.

Kress, G. (2010). *Multimodality. A Social Semiotic Approach to Contemporary Communication*. London: Routledge.

Kress, G. & T. van Leeuwen (2006). *Reading Images: A Grammar of Visual Design.* London: Routledge.

Mavers, D. (2011): *Children's Drawing and Writing: The Remarkable in the Unremarkable.* London: Routledge.

Mehan, H. (1980). The competent student. *Anthropology & Education Quarterly*, 11, 3, 131-152.

O'Halloran, K. & F. Lim (2014). Systemic Functional Multimodal Discourse Analysis. In: In C.D. Maier & S. Norris (eds). *Texts, Images and Interaction: A Reader in Multimodality.* Mouton de Gruyter (pp. 137-153).

Scollon, R. and S. Scollon (2003). *Discourses in Place. Language in the Material World.* London: Routledge.

Streeck, J., C. Goodwin & C. LeBaron (eds). (2011). *Embodied Interaction. Language and Body in the Material World.* Cambridge: Cambridge University Press.

Van Leeuwen, T. (2005). Introducing Social Semiotics. London: Routledge.

Further Reading

Jewitt, Bezemer & O'Halloran (2016)
Outlines, illustrates and compares a range of different approaches to multimodality, including social semiotics, systemic functional linguistics and conversation analysis; and discusses how to design a study in multimodality.

Bateman, Wildfeuer & Hippala (2017)
Outlines and illustrates different methods for analyzing multimodal materials, including diagrams, films, webpages, social media, and computer games.

Van Leeuwen (2005)
Introduction to social semiotics, outlining semiotic analytical procedures, illustrated (mainly) with paper- and screen-mediated texts and three-dimensional objects.

Heath, Luff & Hindmarsh (2010)
Introduction to a multimodal approach grounded in Conversation Analysis and Ethnomethodology, drawing on a range of different studies by the authors.

Online Resources

http://www.routledgetextbooks.com/textbooks/9780415639262/
Includes resources for self-study related to *Introducing Multimodality.*

https://mode.ioe.ac.uk/resources/
Includes interviews with key figures in multimodality, glossaries, examples of multimodal transcripts and bibliographies.

Discussion Questions

1. How does multimodality challenge previously well-established theories, concepts and terms in linguistics?

2. How might linguists refute the arguments put forward by multimodalists?

3. How would you analyse a how-to video on YouTube (e.g. on how to replace a kitchen tap)?

Glossary of key terms

Mode
Modes are semiotic resources that provide distinct possibilities for human expression, meaning making and communication.

Multimodality
A field of study that aims to explore, theorize and explain human expression, meaning making and communication in all its manifestations.

Sign
The outcome of the semiotic act of making a connection between a signifier ('form') and signified ('meaning').

Social semiotics
The study of sign making and multimodality.