International Journal of Social Research Methodology

ISSN: 1364-5579 (Print) 1464-5300 (Online) Journal homepage: http://www.tandfonline.com/loi/tsrm20

Learning from linguistics: rethinking multimodal enquiry

Kirsten Kohrs

To cite this article: Kirsten Kohrs (2017): Learning from linguistics: rethinking multimodal enquiry, International Journal of Social Research Methodology, DOI: 10.1080/13645579.2017.1321259

To link to this article: http://dx.doi.org/10.1080/13645579.2017.1321259

INTERNATIONAL JOURNAL OF SOCIAL RESEARCH METHODOLOGY, 2017
https://doi.org/10.1080/13645579.2017.1321259

Learning from linguistics: rethinking multimodal enquiry

Kirsten Kohrs

Culture, Media and Creative Industries, King's College London, London, UK

ABSTRACT

Multimodal studies posit that meaning is not only communicated through spoken and written words, but also through other modes such as images, gesture, gaze, proximity etc. The widespread availability of high-quality, miniaturised audio and video recording and storing technology has made multimodal data collection cheap and easy. However, the transcription and analysis of the resulting avalanche of recorded data is complex, time-consuming, labour-intensive and expensive. To date there is no established practice or consensus as to scope, methods, objectives or definitions. In fact, concern has been voiced that the field risks expanding to the point of incoherence, sometimes building theory from intuition and generalising from single case studies. Lessons from the 200-year-old discipline of modern linguistics can provide one way forward for the vibrant emerging field of multimodal studies by introducing methods that generate results and hypotheses which can be critically evaluated and empirically tested.

ARTICLE HISTORY
Received 29 November 2016
Accepted 14 April 2017

KEYWORDS
Corpus linguistics; cross-disciplinary methodology; multimodality; linguistics; semiotics; theory; visual communication

Introduction

The visual culture and media theorist JWT Mitchell (1994) who coined the term 'the pictorial turn' to describe the (re-)orientation of modern society towards the visual, notes that in an age dominated by pictures, paradoxically, we still do not know the exact role of the codes and conventions of non-linguistic symbol systems in meaning-making:

> The simplest way to put this is to say that, in what is often characterized as an age of 'spectacle' (Guy Debord), 'surveillance' (Foucault), and all-pervasive image-making, we still do not know exactly what pictures are, what their relation to language is, how they operate on observers and on the world, how their history is to be understood, and what is to be done with or about them. (Mitchell, 1994, p. 13)

The French intellectual and cultural theorist Roland Barthes (1967) refers to 'mixed systems' regarding the different mediators of meaning-making, such as sound, image, writing and objects working in conjunction. Landmark studies building on ground-breaking early work in semiotics such as de Saussure's (1972/1983) linguistic signs and Barthes's (1957/2009, 1964/1999) concept of 'myth' as a higher-level sign include, for instance, studies of advertising which often expose its ideological dimension (Goffman, 1979; Goldman, 1992/2000; Messaris, 1997; Williamson, 1978/2002). More recently these 'mixed systems' have been investigated under the label 'multimodality' extending the field to texts such as moving/dynamic pictures, sculpture, architecture and so on.

An explosion of interest in multimodal methodologies in the last two decades has resulted in scholarship covering a heterogeneous range of research interests under the heading of multimodality,

CONTACT Kirsten Kohrs Kirsten.Kohrs@kcl.ac.uk

cultural studies, visual studies or visual rhetoric. The types of visual phenomena under investigation continue to expand (Hill, 2009). Focus of enquiry are visual representations as well as 'pretty much anything created by human hands – a building, a toaster, a written document, an article of clothing' (Hill & Helmers, 2004, p. ix).

The relatively young academic field of multimodal studies received a major impetus from increasingly affordable technology which enables the recording, replaying, transcription and analysis of complex multimodal phenomena (Norris, 2013; O'Halloran & Smith, 2013a; O'Halloran, Tan, Smith, & Podlasov, 2011). Unsurprisingly, as a young field multimodal study is characterised by a heterogeneity of perspectives and diversity of approaches (Holsanova, 2012; Müller, 2007; Müller, Kappas, & Olk, 2012; O'Halloran & Smith, 2013a). Indeed, a recent special issue of *Visual Communication* illustrates the multitude of approaches and objects of study of multimodal analysis:

> The methods include content analysis, social semiotic analysis, eye tracking measurements – in combination with think aloud protocols and retrospective interviews – as well as iconology and psychophysiological real time measurements. The respective approaches are exemplified through detailed analyses of a variety of materials, including press photography, art, multimodal health education materials, PowerPoint presentations, internet advertisements and TV media discussions. (Holsanova, 2012, p. 251)

One of the principal pioneers of multimodal approaches, Carey Jewitt, and her collaborators, engage in methodological innovation by focusing on research of the body in the arts and digital environments (Jewitt, Xambo, & Price, 2016), for example. The guest editors of a recent special issue of *Qualitative Inquiry* titled *Hypermodal Inquiry* also challenge traditional notions of scholarship in an intriguing manner in the context of the arts (Kaufmann & Holbrook, 2016).Though some of the contributions of the special edition contain textual components, it is the digital component that is pioneering. Four section introductions consist exclusively of videos between 28 and 58 s posted on YouTube in which a camera slowly moves away from a close-up of an apple's wet stem ('Section Introduction: Bounds of Meaning,' 2016; 'Section Introduction: Disciplining Hypermodality,' 2016; 'Section Introduction: Presence and Absence,' 2016; 'Section Introduction: Textual Explosions,' 2016). Other contributions include engaging with sensory experience by showing a static view of the outside of a building with a soundtrack of unseen students submitting papers (Dean, 2016) or inventing, performing and becoming data while being videoed making marks on paper with charcoal (Waterhouse, Otterstad, & Jensen, 2016). The editors of the special edition thus suggest that '"How does one read work such as this?" "Is it art?" "Is it scholarship?"' are the wrong questions to ask (Kaufmann & Holbrook, 2016, p. 160), hence, seeking to dissolve the demarcation between art and scholarship.

Theories in the field of visual communication offer complementary angles of analysis rather than competing ones. Scholars appear 'less interested in debunking or overthrowing older theories than in developing new methods for shedding more light on the many complex ways in which images mean' (Hill, 2009, p. 1002). While many scholars embrace the thrill of the openness and experimental nature of multimodal approaches, such as Kaufman and Holbrook who note that they 'laughed with excitement at the hypermodal array that filled our inboxes' (Kaufmann & Holbrook, 2016, p. 160), criticism of a lack of focus and agenda of visual studies as a discipline also exists. Thus, it has been suggested that two 'ontological perils' may need to be avoided, namely 'the lack of a specific object of study [and] the expansion of the field to the point of incoherence' (Dikovitskaya, 2005, p. 3).

The results and methods of the 200-year-old discipline of modern linguistics, though perhaps not unique in identifying issues and offering solutions, suggests a range of important lessons for the emerging field of multimodality as the latter experiences significant growth. The next sections will look at the theoretical and practical challenges to multimodal studies and then examine a way forward that modern linguistics can offer. The final section will suggest future directions.

Multimodal analysis: key directions

In the exciting field of multimodal research, the scope, namely the complexity and number of modes and variables to be defined and integrated into a single framework for the numerous media under

investigation, poses many challenges. Two examples of multimodal analysis stand out among recent scholarship through their broad scope as opposed to the myriad of research focussing on a single or a very small number of hand-picked case studies, namely Kress and van Leeuwen's groundbreaking *Reading Images: The Grammar of Visual Design* (1996/2007) and Baldry and Thibault's (2006) *Multimodal Transcription and Text Analysis: A Multimedia Toolkit and Coursebook*. However, neither of these two examples can be considered entirely successful. A brief analysis will explain why not, and a comparison with general linguistics theory and approaches will shed light on issues to be addressed in future research.

Reading Images: The Grammar of Visual Design and its revised second edition (Kress & van Leeuwen, 1996/2007) propose a grammar of, for the most part, linguistic and visual elements in static two-dimensional texts. The authors note the importance of multimodal data to meaning-making such as composition, framing and a range of devices to highlight the salience of various elements of a multimodal text (relative size, focus/sharpness, foregrounding/back-grounding, colour and so on). In spite of the book's great strengths in terms of presenting innovative ideas and concepts, a number of theoretical weaknesses have, however, been identified, such as the categorical manner in which generalisations are considered proven based on a few illustrations, and problematic classifications and interpretations are presented as certainty.

Kress and van Leeuwen's (1996/2007) framework for describing how the different modes work together to create meaning in illustrated texts has been accused of lacking analytic rigour as the authors make broad and sometimes problematic generalisations from small samples (Forceville, 2009; McLoughlin, 2008). For instance, Lakoff and Johnson (1980/2003) established that spatial relationships such as above/below, in front/behind, close/distant, left/right, north/south/east/west, and inside/outside (centre/periphery) are not semantically neutral but are linked to cultural concepts. Kress and van Leeuwen adopt three of these dimensions of visual space, left/right, top/bottom, centre/margin (1996/2007), in their proposal for classification.

Left/right spatial relations often have sequential significance. Reading and writing in European cultures (as opposed to Arabic, Hebrew and Chinese for instance) is from left to right and top to bottom, so a movement from left to right in reading texts is the default stance 'unless attention is diverted by some salient feature' (Chandler, 2007, p. 111). Thus, Kress and van Leeuwen argue a principle of a continuous movement from left to right in reading, but furthermore assert that the left side is the key site for the 'already given', that is information that the reader is already aware of, while the right side is the site for 'new' information that the reader does not yet know (1996/2007, p. 183). In the second edition of *Reading Images* these claims about the semantic significance of the left/right spatial relations are reiterated and extended: 'Looking at what is placed on the left and what is placed on the right in other kinds of visuals has confirmed this generalisation' (2007, pp. 180–181). However, only a single example from an Australian women's magazine is provided as evidence. Moreover, these results could not be replicated in British magazines (McLoughlin, 2008), raising questions as to the generalisability of the findings.

Concerns have also been voiced that spatial concepts such as in the example above, namely, that the left side equals 'already given'/the right side equals 'new', and others (see details below) are taken as facts, 'whereas in reality these concepts are no more than hypotheses requiring further critical evaluation as well as empirical testing' (Forceville, 2009, p. 1462).

In fact, van Leeuwen (2000) himself treats principles derived in one context without further validation as definitive and broadly generalisable in another, for example, regarding the meaning of the top/bottom elements in composition in his own study on multimodal texts by school children. Basing his claim once more on a non-specified and thus non-verifiable sample of 'many different types of images and other visuals' (2000, p. 283), van Leeuwen claims that the authors were able to conclude that

> vertical, top - bottom structures, if used to polarize two different elements (for example, to oppose the past and the present, writing and drawing, dream and reality, etc.), present the top element as the generalized and/or idealized essence of the information, and the bottom element as contrasting with this by being more detailed and specific and/or more 'down to earth,' more oriented towards facts and practicalities. They therefore called the top element the 'ideal' and the bottom element the 'real'. (van Leeuwen, 2000, p. 283)

Based on these claims, van Leeuwen concludes about multimodal texts by school children:

> The texts I am analyzing in this paper polarize writing and drawing in this way, positioning them as different from each other, rather than as integrated and intermingled: writing is presented as the generalized and idealized essence of the information, drawing as 'illustration,' as a more detailed and factual complement (and perhaps also as the element which most 'really' brings out the child's reactions). (2000, p. 283)

Illustration: Example of Child's Description of Science Museum School Visit (van Leeuwen, 2000, p. 297)

This conclusion, however, does not take into consideration, for instance, how the task was presented to the children, namely, that children were first asked to write about their experience and, only as a fifth task at the very end, were they asked if they could draw it:

> This is the task for the children:
> Write about one or two things in Launch Pad (some prompts you might like to use to help them)
> What do you remember best?
> What did it do?
> What did you like most about it?
> Why?
> Can you draw it?
>
> (van Leeuwen, 2000, p. 304)

The possibility that the children may merely have worked through their tasks in the given sequence as a plausible explanation why the written part of the assignments was found at the top of the page of the children's reports is not considered while the author speculates about the 'ideal essence' of written elements.

Kress and van Leeuwen (1996/2007) furthermore posit that texts have a centre/margin composition emphasising the importance of a nucleus and key focus of attention literally at the geometric centre of texts while elements that are literally in the margins are of less or little significance. However, they have to draw on examples from Byzantine and Buddhist art and children's drawings to find data to support this hypothesis. Their thinking also overlooks centuries of a quest for ideal proportions, frequently called the 'golden' mean, ratio, proportion or section in which 'a smaller part relates to a larger part as the larger part relates to the whole' (Ocvirk, Stinson, Wigg, Bone, & Cayton, 2013, p. 76), that is, approximately a rule of two thirds. Today, this principle of composition is recognised by scientists in nature and frequently used in the visual arts, including photography, to achieve balance and harmony (Freeman, 2013; Lewis & Lewis, 2014; Ocvirk et al., 2013). However, acknowledging the 'relative infrequency of centred compositions in contemporary Western representation' themselves,

Kress and van Leeuwen seek to explain away the divergence between the hypothesis and the data: '[it] perhaps signifies that, in the words of the poet, "the centre does not hold" any longer in many sectors of contemporary society' (1996/2007, p. 197), rather than revise their theory as a result of having been disproved by the data.

Thus, even though *Reading Images* can truly be considered a 'courageous attempt to fill a glaring gap' (Forceville, 1999, p. 163), it is highly problematic:

> In several instances, KvL [Kress and van Leeuwen] are carried away by their theoretical and ideological framework, arbitrarily or rigidly applying it to the new picture, and this sometimes yields highly unconvincing results (…) In short, KvL often too easily assume (a) that their examples are representative and (b) that their personal interpretations have intersubjective validity. (Forceville, 1999, pp. 171, 172)

Furthermore, there are many examples of subsequent scholarship that blithely builds on Kress and van Leeuwen's unproven 'general principles' as if they were tested and validated (Bell, 2008; Bell & Milic, 2002; Jewitt & Oyama, 2001/2008).

Like *Reading Images* (Kress & van Leeuwen, 1996/2007), a second key approach to multimodal analysis, Baldry and Thibault's *Multimodal Transcription and Text Analysis: A Multimedia Toolkit and Coursebook* (2006) stands out among recent scholarship as it is notable for its broad scope. However, unlike Kress and van Leeuwen (1996/2007), the authors build up detailed description from a close analysis of specific texts to arrive at generalisations. In their approach analytical detail is crucial and the complexities of the analytical process, namely itemisation and classification, are exacerbated by the challenges of accessing and transcribing different types of multimodal texts. The challenge of presenting the quantity and complexity of detail for publication appears self-evident, particularly in the case of dynamic texts such as audiovisual film or digital hypertexts (O'Halloran & Smith, 2013a). *Multimodal Transcription and Text Analysis* (Baldry & Thibault, 2006) has thus been criticised for its scope since the complexity and quantity of detailed description of numerous modes and their variables are listed rather than integrated into a coherent framework resulting in an overwhelming dominance of description over insight (Forceville, 2007).

A third notable large-scale line of inquiry into multimodal discourse in addition to those cited is an empirical case study based on corpus analysis. Bateman, Delin, and Henschel (2002) rightly argue that to arrive at theory, multimodal meaning making needs to be grounded on a more solid empirical basis; multimodal analysis must become more verifiable and less impressionistic. Plausible or interesting claims based on selective, informal analyses must be underpinned by a systematic analysis of a larger data-set. Thus, Bateman, Delin, and Henschel (2007) showed how the electronic versions of newspaper front pages differ from their supposedly identical printed versions in a corpus-based analysis. Furthermore, they have been developing a computerised annotation system to model genre in document layout to enhance corpus-based multimodal analysis delimiting the scope of the study by excluding the pictorial elements. (Bateman, 2012; Bateman, Delin, & Henschel, 2004; Bateman et al., 2002).

Modern linguistics

Since human language is the most researched communication system, a linguistic perspective suggests itself as a starting point for theorising pictorial and multimodal communication. Naturally, there are advantages and drawbacks. One major approach, Systemic Functional Linguistics (SFL) founded by M.A.K. Halliday in the 1960s, demonstrates strengths in that it is applicable to a variety of different types of text and in that it extends scope from focus on grammaticality to functionality as well as from isolated sentences to discourse in specific contexts (Halliday, 1994/2006; Herriman, 2012; O'Halloran, 2008). However, it presents an 'unwarranted faith in the "systematicity" of non-verbal modes of communication' (Forceville, 2011, p. 39) as it perhaps applies a 'vocabulary' and 'grammar' analogy too rigidly to visuals while risking to overlook important variables in non-verbal modes that are absent in verbal language. Other major linguistic approaches are based on the relatively new field of cognitive linguistics whose conceptual approach contrasts with the structural approach of generative grammar

(Talmy, 2006). These approaches are the Conceptual Metaphor Theory which, however, *nomen est omen*, exclusively focuses on the notion of conceptual metaphor and Relevance Theory which is currently heavily focussed on verbal discourse (2011; Forceville, 1998; Wilson & Sperber, 2002). Hence, I will argue that multimodal analysis has much to learn from the older and more established discipline of general linguistics, even if it is perhaps not unique in identifying issues and proposing solutions.

While the study of visual communication as an academic field is comparatively new (Harris, 2006; Messaris, 2003) modern linguistics has existed since the 1800s. This section will look at some of the approaches and developments in linguistics that may provide direction to the new and exciting field of multimodal analysis starting with a very brief introduction to general linguistics.

General linguistics has been defined as the systematic study of language, describing languages as well as attempting to answer the question as to what language is and formulating theories as to how it works (Aitchison, 2010, 2012; Akmamjian, Demers, Farmer, & Harnish, 2010; Trask & Mayblin, 2012). Language 'inherently involves sharing a code with other people' (Pinker, 1995, p. 242). A shared grammatical code would be futile, if only a single person possessed it. It allows 'speakers to produce any linguistic message' which can be comprehended by the receiver (Pinker, 1995, p. 237).

Grammar includes sounds (phonology), syntax and semantics. Syntax pertains to the form and arrangement of words (Aitchison, 2010, 2012; Pinker, 1995). Meaning (semantics) has two sides in general linguistics. Even though words are also linked to various classes of recognisable objects in the external world, linguistics is primarily focussed on the meanings of lexical items (words) as they are interrelated within the language system (Aitchison, 2010, 2012). A competent speaker of a language can differentiate between sentences that are grammatically (in-)correct(1), nonsensical (1, 3), interpretable (2), have similar meanings (2 and 3 have the same literal meaning) or are ambiguous (4):

(1) The wickwock grunched the mobe.
(2) My brother is a spinster.
(3) My male sibling is an unmarried female.
(4) Visiting great aunts can be a nuisance.

(Aitchison, 2010, pp. 101, 117, 105) [my categorisation]

Syntax and semantics overlap in that the meaning of a word depends not only on its semantic field (meaning grouped according to a specific subjects) but also on its relationship with other words within the structure of the sentence. Verbs, for instance, convey both semantic and syntactic information. For example, in the sentence above, it is clear that the *wickwock* did something to the *mobe*.

Of course, not all aspects of the meaning of language can be captured through the dimensions of linguistic sentence structure as described above. Sentences are part of a larger discourse and interpreted in context (Aitchison, 2010). Since the 1950s linguistics has also intersected with other disciplines producing new fields such as neurolinguistics, psycholinguistics or sociolinguistics (Kienpointner, 2001).

Unlike the emerging field of multimodal studies, general linguistics benefits from a long tradition of critical stance towards its own methodology (Bresnan, 2007; Culicover & Jackendoff, 2010; Gibson & Fedorenko, 2010, 2013) Thus, conventions have been established in general linguistics over time that describe and test results and methodology to ensure validity and replicability. The methods for gathering data in general linguistics that have proven viable over time are threefold: Firstly, judgement can be elicited from a respondent as to whether two sentences mean the same thing or are grammatical; secondly, controlled experiments can test a native speaker's reactions and behaviour; or, thirdly, a body of naturally-occurring utterances, a corpus, can be collected (Gibson & Fedorenko, 2013; Scholz, Pelletier, & Pullum, 2015).

An evaluation of these three approaches must contend with methodological questions. Firstly, to ensure elicitation's reliability and validity, too small a number of respondents must be avoided. Differences in judgement as to whether something is grammatical or acceptable, and differences in scale of acceptability must be taken into account. Furthermore, the influence of preceding context or extraneous variables on outcomes must be assessed. Secondly, experimentation must ensure a sufficient

number of participants and stimuli to ensure the reliability of theoretical claims and generalisations. Finally, corpus-based research needs to ensure both the relevance and reliability of the corpus evidence (Gibson & Fedorenko, 2013; Scholz et al., 2015). A careful selection of a corpus to ensure relevance and reliability, can in principle avoid the methodological weaknesses of elicitation or experimental constructions, namely an insufficiently large sample or number of respondents, investigator bias, and lack of control for external variables. A closer look at corpus research is therefore warranted.

A corpus (Baker, 2010; Hunston, 2006, 2012; Scholz et al., 2015; Stubbs & Halbe, 2012; Trask, 1999) is a collection of written or spoken texts of naturally occurring language, often carefully selected to represent a type of communication or a variety of language for linguistic analysis. Corpora allow systematic analysis, either manually or via dedicated software, to reveal linguistic patterns of association between groups of words or a linguistic feature or a text type, relative frequency of information or other linguistic phenomena. The objective can be to present the normal features of a language in typical proportions or to contrast patterns of different varieties (Bonelli & Sinclair, 2006). For instance, collocations, the tendency of words to occur together with particular other words, can be observed to identify recurring phraseology. Comparisons in frequency of occurrence can be made between different types of texts, registers and historical periods which may be statistically or culturally significant (Hunston, 2006, 2012). While corpus analysis was carried out manually in the past, the increasing availability and sophistication of computer hardware (memory) and software (search and analysis programs) today enables a rapid advance of computational collecting, searching and analysing of corpora (Krishnamurthy, 2006; Scholz et al., 2015).

Corpus studies rely on annotation which is ultimately based on the judgements of native speakers and consistency between annotators is crucial, but they also rely on subjective judgement in collecting and categorising observations (Bresnan, 2007; Gibson & Fedorenko, 2010; Scholz et al., 2015). Manual or semi-automatic annotation of corpora can consist of parts-of-speech tagging, that is labelling words with their syntactic category. However, a corpus can also be annotated with semantic or pragmatic information, for example, for sentiment analysis to determine whether a product review is positive or negative (Scholz et al., 2015). Information can be retrieved from corpora only through annotation, that is only through what is made salient by human intervention. The next section will show how these results and methods from linguistics offer insights for the emerging field of multimodality in overcoming problems and challenges inherent in the richness and complexity of its data and the sheer size of its scope.

Multimodal methodology: theoretical challenges

Rigorous systematic analysis of multimodal texts established the existence of individual visual tropes, such as visual metaphor in advertising (Forceville, 1998), and will no doubt model genre features in document layout in the future (Bateman, 2012; Bateman et al., 2002, 2004). A more recent corpus-based empirical study illustrates how gender is communicated in visual culture (Kohrs & Gill, in press). However, as shown above, gaps in multimodal scholarship still exist, both in terms of a systematic analysis based on a corpus of work and in terms of a theoretical framework regarding how the basic components of meaning-making operate individually and collectively. Opportunity exists for an attempt to build theory towards a more systematic as well as a more comprehensive understanding of how visual codes operate to convey meaning. In the remainder of this paper, I will highlight five dimensions in which a comparison with the methods and results of linguistics can add value to multimodal scholarship.

Empirical robustness and methodological rigour

In order to bestow greater confidence as to the validity of research results, insights from general linguistics demonstrate the value of systematic rigour and corpus-based studies rather than the examination

of single, cherry-picked case studies in order to produce a body of scholarship that can be empirically tested and thus (dis-)proved or improved.

The review of large-scale multimodal studies above has shown that methodological weaknesses can be the result of relying on individual, hand-picked examples and overgeneralisation e.g. Kress and van Leeuwen (1996/2007) or the result of too little restriction in variables and modes creating a large amount of description rather than a coherent framework (Baldry & Thibault, 2006). A corpus-based approach to multimodal theory-building addresses concerns as to these flaws and limitations of data analysis in that only a sufficiently large amount of naturally occurring multimodal communication, namely a corpus, permits the development of a system of annotation that reliably establishes relative frequencies or patterns of association between groups of linguistic features and text types.

Terminology and definitions

Following on from the heterogeneity in visual studies as to scope, objectives, and methods, there is currently also no consensus as to terms and definitions in multimodal studies. Terms such as multimodality, multimodal analysis, semiotic resources, modes, modalities are used rather loosely, have no clear delimitations and often overlap (Kress, 2010; Norris, 2004; O'Halloran, 2008). The experience in general linguistics would suggest that the lack of shared terminology is prone to hinder the development of the field. I would also suggest avoiding unnecessary neologisms wherever possible to aid terminological and thus conceptual clarity. Familiarisation with existing expertise as well as vocabulary/terminology based on excellent robust scholarship in disciplines that focus on specific non-verbal communication, for example, such as art or film studies (Bordwell & Thompson, 2013; Bowen & Thompson, 2013; Lewis & Lewis, 2014; Ocvirk et al., 2013) might enhance transparency, insight and thus theoretical progress.

Data

Crucially, the question of what constitutes data in multimodal analysis is currently not clear. As models and techniques for classifying the basic units of the different modes largely do not exist, the annotation and analysis and validation of theory become problematic in empirical studies (Bateman et al., 2002). In terms of the automated analysis of meaning-making, even software development for the study of spoken or written language faces stumbling blocks, for example in semantics where ambiguity in meaning such as right/wrong and right/left (sense disambiguation) creates difficulties (Krishnamurthy, 2006). In multimodal analysis the challenge to progress from the automated extraction of features to determining meaning is exacerbated since annotations make only particular features of a corpus salient based on a particular chosen theoretical approach (Hunston, 2012). Guidelines and established practice for multiple levels of annotation still need to be established to manage the labour-intensive and expensive process of multimodal analysis (Martin, Paggio, Kuehnlein, Stiefelhagen, & Pianesi, 2008).

Theories and methodological frameworks, moreover, need to be developed (O'Halloran, 2008). The difficulties are obvious. Single dimensions of non-verbal communication such as facial expressions and their meaning (Ekman, 1985, 2003) have been theorised based on extensive empirical evidence, however, a detailed analysis of a political television talk show, for instance, involves verbal language, body language, voice, camera angle, framing and so on showing 'the detail and complexity involved in annotating, analysing, searching and retrieving multimodal semantics patterns within and across complex multimodal phenomena' (O'Halloran, 2008, p. 25).

While audio and video recording technology have made data collection relatively cheap and easy, the transcription of the recorded material is highly complex and time-consuming (Norris, 2002; Siszsons, 2013). Attempts have been made to automate this process using software (Baldry & Thibault, 2006, 2008; Bateman, 2012; Bateman et al., 2002; O'Halloran, Podlasov, Chua, & K.L.E, 2012; O'Halloran & Smith, 2013b; Smith, Tan, Podlasov, & O'Halloran, 2011) and the usefulness and usability of software has been analysed (Rohlfing et al., 2006). It is almost a truism that '[a]ll linguistic data depend

on classifications of meaning and form, which are, at bottom, subjective. In that sense, it is wrong to imagine that we somehow escape from subjectivity when we count and statistically summarize data' (Bresnan, 2007, p. 302). Nonetheless, to further the development of multimodal scholarship, a growing consensus as to what constitutes data based on a corpus of work would allow hypotheses or research outcomes to be compared, proven or disproven.

Scope

The scope of multimodal analysis is also critical. Too much detailed micro-analysis of a single mode can limit understanding of comprehensive multimodal meaning-making while the inclusion of too many different modes risks not giving sufficient attention to how each individual mode works (Bezemer & Jewitt, 2010). Furthermore, theoretical questions arise in data preparation and collection as to how different features of different modes (e.g. gesture and language) can be integrated into a single framework (Adolphs, 2012). Moreover, when attempts are made to identify a multitude of variables in numerous different semiotic modes in various case studies (Baldry & Thibault, 2006), the ensuing complexity risks criticism of yielding more description than insight (Forceville, 2007).

It is worth noting, that a key branch of linguistics has focussed only on the structure of words and sentences for more than half a century to arrive at the integrated framework that exists today. A way forward for multimodal research could be to focus on types of text that are similar in key dimensions as, for example, in genre or the medium, but, in addition, to privilege a corpus of static media (rather than dynamic or interactive media) which is further delimited, such as 'print advertising in 1990s glossy magazines aimed at upper-class women'; 'Splasher [*sic*] horror movies of the 1990s'; 'Websites promoting art museums' (Forceville, 2007, p. 1237). Only such a purposive restriction to a particular type of visual language will make the rigorous systematic analysis of a body of work feasible, and ultimately enable progress towards answering the fascinating broader questions regarding comprehensive frameworks for visual language.

Native speaker competence

To add to the list of challenges, a crucial difficulty in multimodal analysis is the absence of the equivalent of the adult native speaker's linguistic competence. For example, design professionals might evaluate the standard of a body of graphic texts differently (Bateman et al., 2002), that is, more critically than 'ordinary punters' lacking expertise in the field. Many multimodal researchers have been educated in one field of study (linguistics, literature, education, art history etc.) and are therefore limited in their knowledge regarding other disciplines and, thus, potentially unaware of important work done by experts in those other fields (Forceville, 2007). The visual arts (Hudson & Noonan, 2015; Lewis & Lewis, 2014; Ocvirk et al., 2013), film studies (Bordwell & Thompson, 2013; Bowen & Thompson, 2013), photography (duChemin, 2012; Freeman, 2013; Langford, 2000; Präkel, 2010; Wells, 2015), literature (Abbott, 2010; Culler, 2000), rhetoric (Lanham, 1991; Leith, 2011; Lotman, 2006; Sloane, 2001) to name but a few, all have long established traditions of scholarship on which multimodal studies can build.

Conclusion

This paper critically reviewed crucial issues and problems the dynamic emerging approach of multimodal analysis is currently facing and has drawn on the two-hundred-year old discipline of modern linguistics to offer insight and food for thought for viable future directions for this vibrant field.

One of the challenges for this emerging field is the difficult but unquestionably necessary task of finding salient means of categorising its modes and variables. Clearly, the richness of its data makes its study extraordinarily complex. Close critical scrutiny and comparison with general linguistics highlights that the absence of established practices or consensus as to objectives, definition

and methods in multimodal studies hinders progress in terms of building a much needed coherent framework for multimodal analysis. The scope of the research constitutes further challenges due to the overwhelming cultural richness and complexity of visual and multimodal communication. Thus, too much micro-analysis of a single mode limits a comprehensive understanding of how the modes work together. Analysis of too many modes risks limiting the understanding of how each individual mode works. Analysis of too many variables in too many modes in too many media, moreover, tends to yield more description than insight leaving the key challenge of incorporating the various features of different modes into an integrated framework unachieved.

Though linguistics does not uniquely identify issues and offer solutions, of course, lessons from the systematic, comprehensive, tried-and-tested discipline can, however, provide direction to the exiting emerging field of multimodal scholarship. Building theory without a sound empirical basis, that is, relying merely on intuition and generalising from single or a small number of case studies is inadequate. It is particularly unsatisfactory and problematic for scholarship, if these generalisations are furthermore taken at face value and proliferated without substantiation. A corpus approach adapted from linguistics, could bring much needed academic rigour to multimodal studies while limiting the scope of the research to a fairly narrowly defined type of text can overcome the particular challenges that the lack of consensus on objectives, definitions and methods currently pose.

Learning from the science of linguistics indicates that the staggering complexity and richness of data inherent in multimodal texts makes finding patterns a pivotal research objective. In the fast-moving field of multimodal analysis focus must shift from selectively choosing data to illustrate a theory to systematically developing and testing theory on the largest practicable body of data. Consequently, once patterns become apparent, comparative studies based on sound empirical data become feasible. Thus, multimodal studies can evolve into a rigorous academic approach which can generate results in the form of theoretical concepts and hypotheses which can be critically evaluated and empirically tested, and hence (dis-)proven and refined.

Disclosure statement

No potential conflict of interest was reported by the author.

Notes on contributor

Following an extensive and stellar career creating commercial communication, Kirsten Kohrs is now a senior lecturer at the University of Greenwich, London. Her research interests focus around visual language.

References

Abbott, H. P. (2010). *The Cambridge introduction to narrative* (2nd ed.). Cambridge: Cambridge University Press.

Adolphs, S. (2012). Corpora: multimodal. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1188–1190). Oxford: Blackwell.

Aitchison, J. (2010). *Aitchison's linguistics*. London: Hodder Education.

Aitchison, J. (2012). *Linguistics made easy*. London: Hodder Education.

Akmamjian, A., Demers, R. A., Farmer, A. K., & Harnish, R. M. (2010). *Linguistics: An introduction to language and communication* (6th ed.). Cambridge: MIT Press.

Baker, P. (2010). Will Ms ever be as frequent as Mr? A corpus-based comparison of gendered terms across four diachronic corpora of British English. *Gender and Language, 4*, 125–149.

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

Baldry, A., & Thibault, P. J. (2008). Applications of multimodal concordances. *Hermes – Journal of Language and Communication Studies, 41*, 11–41.

Barthes, R. (1957/2009). *Mythologies* (A. Lavers & S. Reynolds, Trans.). London: Vintage.

Barthes, R. (1964/1999). Rhetoric of the image. In J. Evans & S. Hall (Eds.), *Visual culture: The reader* (pp. 33–40). London: Sage.

Barthes, R. (1967). *Elements of semiology* (A. Lavers & C. Smith, Trans.). London: Jonathan Cape.
Bateman, J. (2012). Multimodal corpus-based approaches. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 3983–3991). Wiley.
Bateman, J., Delin, J., & Henschel, R. (2002). Multimodality and empiricism: methodological issues in the study of multimodal meaning-making. *GeM Report, 1*. Retrived from http://www.fb10.unibremen.de/anglistik/langpro/projects/gem/downloads/bateman-delin-henschel-Salzburg.pdf
Bateman, J., Delin, J., & Henschel, R. (2004). Multimodality and empiricism: Preparing for a corpusbased approach to the study of multimodal meaning-making. In E. Ventola, C. Charles, & M. Kaltenbacher (Eds.), *Perspectives on multimodality* (pp. 65–87). Amsterdam: John Benjamins.
Bateman, J., Delin, J., & Henschel, R. (2007). Mapping the multimodal genres of traditional and electronic newspapers. In T. D. Royce & W. L. Bowcher (Eds.), *New directions in the analysis of multimodal discourse* (pp. 147–172). Mahwah, NJ: Lawrence Erlbaum.
Bell, P. (2008). Content analysis of visual images. In T. van Leeuwen & C. Jewitt (Eds.), *Handbook of visual analysis* (pp. 10–34). London: Sage.
Bell, P., & Milic, M. (2002). Goffman's gender advertisements revisited: Combining content analysis with semiotic analysis. *Visual Communication, 1*, 203–222.
Bezemer, J., & Jewitt, C. (2010). Multimodal analysis: Key issues. In L. Litosseliti (Ed.), *Research methods in linguistics* (pp. 180–197). London: Continuum.
Bonelli, E. T., & Sinclair, J. (2006). Corpora. In K. Brown (Ed.), *Encyclopedia of language & linguistics* (2nd ed., pp. 206–220). Oxford: Elsevier.
Bordwell, D., & Thompson, K. (2013). *Film art: An introduction* (10th ed.). New York, NY: McGraw-Hill.
Bowen, C. J., & Thompson, R. (2013). *Grammar of the shot* (3rd ed.). New York, NY: Focal Press.
Bresnan, J. (2007). A few lessons from typology. *Linguistic Typology, 11*, 297–306.
Chandler, D. (2007). *Semiotics: The basics*. London: Routledge.
Culicover, P. W., & Jackendoff, R. (2010). Quantitative methods alone are not enough: Response to Gibson and Fedorenko. *Trends in Cognitive Sciences, 14*, 234–235.
Culler, J. (2000). *Literary theory: A very short introduction*. Oxford: Oxford University Press.
Dean, J. (2016). "Submitting Love?" A sensory sociology of Southbourne. *Qualitative Inquiry, 22*, 162–168.
de Saussure, F. (1972/1983). *Course in general linguistics* (R. Harris, Trans.). London: Duckworth.
duChemin, D. (2012). *Photographically speaking: A deeper look at creating stronger images*. Berkeley, CA: New Riders.
Dikovitskaya, M. (2005). *Visual culture: The study of the visual after the cultural turn*. Cambridge: MIT Press.
Ekman, P. (1985). *Telling lies: Clues to deceit in the marketplace, politics, and marriage*. New York, NY: Norton.
Ekman, P. (2003). *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. New York, NY: Holt.
Forceville, C. (1998). *Pictorial metaphor in advertising*. London: Routledge.
Forceville, C. (1999). Educating the eye? Kress & Van Leeuwen's reading images: The grammar of visual design (1996). *Language and Literature, 8*, 163–178.
Forceville, C. (2007). Multimodal transcription and text analysis: A multimedia toolkit and coursebook: Anthony Baldry, Paul J. Thibault, Equinox, London/Oakville, 2006, 270 pp., $40/£24.99 (paperback), ISBN 1-904768-07-5. *Journal of Pragmatics, 39*, 1235–1238.
Forceville, C. (2009). New directions in the analysis of multimodal discourse. *Journal of Pragmatics, 41*, 1459–1463.
Forceville, C. (2011). Practical cues for helping develop image and multimodal discourse scholarship. In K. Sachs-Hombach & R. Totzke (Eds.), *Bilder – Sehen – Denken. Zum Verhältnis von begrifflich-philosophischen und empirisch-psychologischen Ansätzen in der bildwissenschaftlichen Forschung* [Images – Looking – Thinking. On the relationship between conceptual-philosophical and empirical-psychological approaches in image-scientific research] (pp. 33–51). Cologne: Halem.
Freeman, M. (2013). *Michael Freeman's photo school fundamentals*. New York, NY: Focal Press.
Gibson, E., & Fedorenko, E. (2010). Weak quantitative standards in linguistics research. *Trends in Cognitive Sciences, 14*, 233–234.
Gibson, E., & Fedorenko, E. (2013). The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes, 28*, 88–124.
Goffman, E. (1979). *Gender advertisements*. London: The Macmillan Press.
Goldman, R. (1992/2000). *Reading ads socially*. London: Routledge.
Halliday, M. A. K. (1994/2006). Systemic theory. In K. Brown (Ed.), *Encyclopedia of language & linguistics* (2nd ed., pp. 443–448). Oxford: Elsevier.
Harris, R. (2006). Modern linguistics: 1800 to the present day. In K. Brown (Ed.), *Encyclopedia of language & linguistics* (2nd ed., pp. 203–210). Oxford: Elsevier.
Herriman, J. (2012). Systemic functional linguistics. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 5509–5517). Oxford: Blackwell.
Hill, C. A. (2009). Visual communication theories. In S. W. Littlejohn & K. A. Foss (Eds.), *Encyclopedia of communication theory* (pp. 1002–1005). Thousand Oaks, CA: Sage.

Hill, C. A., & Helmers, M. (Eds.). (2004). *Defining visual rhetorics*. Mahwah, NJ: Lawrence Erlbaum.
Holsanova, J. (2012). New methods for studying visual communication and multimodal integration. *Visual Communication, 11*, 251–257.
Hudson, S., & Noonan, N. (2015). *The art of writing about art* (2nd ed.). Stamford, CT: Cengage Learning.
Hunston, S. (2006). Corpus linguistics. In K. Brown (Ed.), *Encyclopedia of language & linguistics* (2nd ed., pp. 234–248). Oxford: Elsevier.
Hunston, S. (2012). Corpus linguistics: Historical development. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1366–1372). Oxford: Wiley.
Jewitt, C., & Oyama, R. (2001/2008). Visual meaning: A social semiotic approach. In T. van Leeuwen & C. Jewitt (Eds.), *Handbook of visual analysis* (pp. 134–156). London: Sage.
Jewitt, C., Xambo, A., & Price, S. (2016). Exploring methodological innovation in the social sciences: The body in digital environments and the arts. *International Journal of Social Research Methodology, 20*, 105–120.
Kaufmann, J., & Holbrook, T. (2016). Introduction. *Qualitative Inquiry, 22*, 159–160.
Kienpointner, M. (2001). Linguistics. In T. O. Sloane (Ed.), *Encyclopedia of rhetoric* (pp. 426–449). Oxford: Oxford University Press.
Kohrs, K., & Gill, R. (in press). Confident appearing: Revisiting 'Gender Advertisements' in contemporary culture. In J. Baxter & J. Angouri (Eds.), *The Routledge handbook of language, gender and sexuality*. Abingdon: Routledge.
Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Abingdon: Routledge.
Kress, G., & van Leeuwen, T. (1996/2007). *Reading images: The grammar of visual design* (2nd ed.). New York, NY: Routledge.
Krishnamurthy, R. (2006). Corpus lexicography. In K. Brown (Ed.), *Encyclopedia of language & linguistics* (2nd ed., pp. 250–254). Oxford: Elsevier.
Lakoff, G., & Johnson, M. (1980/2003). *Metaphors we live by*. Chicago, IL: The University of Chicago Press.
Langford, M. (2000). *Basic photography* (7th ed.). Oxford: Focal Press.
Lanham, R. A. (1991). *A handlist of rhetorical terms* (2nd ed.). Berkeley: University of California Press.
Leith, S. (2011). *You talkin' to me?: Rhetoric from Aristotle to Obama*. London: Profile Books.
Lewis, R., & Lewis, S. I. (2014). *The power of art* (3rd ed.). Boston, MA: Wadsworth, Cengage Learning.
Lotman, M. (2006). Rhetoric: Semiotic approaches. In K. Brown (Ed.), *Encyclopedia of language & linguistics* (2nd ed., pp. 582–589). Oxford: Elsevier.
Martin, J.-C., Paggio, P., Kuehnlein, P., Stiefelhagen, R., & Pianesi, F. (2008). Introduction to the special issue on multimodal corpora for modeling human multimodal behavior. *Language Resources and Evaluation, 42*, 253–264.
McLoughlin, L. (2008). *The language of magazines*. London: Routledge.
Messaris, P. (1997). *Visual persuasion: The role of images in advertising*. Thousand Oaks, CA: Sage.
Messaris, P. (2003). Visual communication: Theory and research. *Journal of Communication, 53*, 551–556.
Mitchell, W. J. T. (1994). *Picture theory: Essays on verbal and visual representation*. Chicago, IL: University of Chicago Press.
Müller, M. G. (2007). What is visual communication? Past and future of an emerging field of communication research. *Studies in Communication Sciences, 7*, 7–34.
Müller, M. G., Kappas, A., & Olk, B. (2012). Perceiving press photography: A new integrative model, combining iconology with psychophysiological and eye-tracking methods. *Visual Communication, 11*, 307–328.
Norris, S. (2002). The implications of visual research for discourse analysis: Transcription beyond language. *Visual Communication, 1*, 93–117.
Norris, S. (2004). Multimodal discourse analysis: A conceptual framework. In P. LeVine & S. R (Eds.), *Discourse and technology: Multimodal discourse analysis* (pp. 101–115). Washington, DC: Georgetown University Press.
Norris, S. (2013). Multimodal communication: Overview. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 3977–3978. Oxford: Blackwell.
O'Halloran, K. L. (2008). Systemic functional-multimodal discourse analysis (SF-MDA): Constructing ideational meaning using language and visual imagery. *Visual Communication, 7*, 443–475.
O'Halloran, K. L., Podlasov, A., Chua, A., & K.L.E, M. (2012). Interactive software for multimodal analysis. *Visual Communication, 11*, 363–381.
O'Halloran, K. L., & Smith, B. A. (2013a). Multimodal text analysis. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 4124–4129). Oxford: Blackwell.
O'Halloran, K. L., & Smith, B. A. (2013b). Multimodality and technology. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 4089–4094). Oxford: Wiley.
O'Halloran, K. L., Tan, S., Smith, B. A., & Podlasov, A. (2011). Multimodal analysis within an interactive software environment: Critical discourse perspectives. *Critical Discourse Studies, 8*, 109–125.
Ocvirk, O. G., Stinson, R. E., Wigg, P. R., Bone, R. O., & Cayton, D. L. (2013). *Art fundamentals: Theory and practice* (12th ed.). New York, NY: McGraw Hill.
Pinker, S. (1995). *The language instinct: The new science of language and mind*. London: Penguin Books.
Präkel, D. (Ed.). (2010). *The visual dictionary of photography*. Lausanne: AVA.

Rohlfing, K., Loehr, D., Duncan, S., Brown, A., Franklin, A., Kimbarra, I., … Wellinghoff, S. (2006). Comparison of multimodal annotation tools: Workshop report. *Online-Zeitschrift zur Verbalen Interaktion, 7*, 99–123.
Scholz, B. C., Pelletier, F. J., & Pullum, G. K. (2015). Philosophy of linguistics. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2015 ed.). Retreived from https://plato.stanford.edu/archives/spr2015/entries/linguistics/
Section introduction: Bounds of meaning. (2016). *Qualitative inquiry, 22*, 200.
Section introduction: Disciplining hypermodality. (2016). *Qualitative Inquiry, 22*, 191.
Section introduction: Presence and absence. (2016). *Qualitative Inquiry, 22*, 161.
Section introduction: Textual explosions. (2016). *Qualitative Inquiry, 22*, 175.
Siszsons, H. (2013). Transcribing multimodal interaction. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 5894–5901). Oxford: Blackwell.
Sloane, T. O. (Ed.). (2001). *Encyclopedia of rhetoric*. Oxford: Oxford University Press.
Smith, B. A., Tan, S., Podlasov, A., & O'Halloran, K. L. (2011). Analysing multimodality in an interactive digital environment: Software as a meta-semiotic tool. *Social Semiotics, 21*, 359–380.
Stubbs, M., & Halbe, D. (2012). Corpus linguistics: Overview. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1377–1379). Oxford: Blackwell.
Talmy, L. (2006). Cognitive linguistics. In K. Brown (Ed.), *Encyclopedia of language & linguistics* (2nd ed., pp. 542–546). Oxford: Elsevier.
Trask, R. L. (1999). Key concepts in language and linguistics. *Key Concept Series*.
Trask, R. L., & Mayblin, B. (2012). *Introducing linguistics: A graphic guide*. London: Icon Books.
van Leeuwen, T. (2000). It was just like magic – A multimodal analysis of children's writing. *Linguistics and Education, 10*, 273–305.
Waterhouse, A.-H. L., Otterstad, A. M., & Jensen, M. (2016). … anything but synchronized swimming/methodologies … artistic movements in/with unknown inventions. *Qualitative Inquiry, 22*, 201–209.
Wells, L. (Ed.). (2015). *Photography: A critical introduction* (5th ed.). Abingdon: Routledge.
Williamson, J. (1978/2002). *Decoding advertisements: Ideology and meaning in advertisements*. London: Marion Boyars.
Wilson, D., & Sperber, D. (2002). Truthfulness and relevance. *Mind, 111*, 583–632.