The Visual Basis of Linguistic Meaning and its Implications for Critical Discourse Analysis: Integrating Cognitive Linguistic and Multimodal Methods

Christopher Hart

(Lancaster University)

Abstract

Two important challenges currently facing CDA concern (i) the nature of language processing and (ii) the relation between linguistic and multimodal approaches. In this paper I seek to address both issues by advancing an integrated cognitive and multimodal approach to CDA to account for the communication of ideology in linguistic discourse. This approach is predicated on an argument from Cognitive Linguistics which suggests that understanding language involves the construction of multimodal mental representations, the properties of which can be approached within frameworks of multimodal social semiotics. Specifically, the paper shows how spatial organisation and orientation feature in our linguistic understanding of certain grammatical constructions and, consequently, what evaluative functions those constructions covertly confer. Traditionally, the direction of influence between linguistic and multimodal forms of discourse analysis is unidirectional with the former informing the latter but not the other way around. This paper represents a reversal of this orthodoxy.

1. Introduction

There has recently been a multimodal turn in discourse analysis and critical discourse analysis with researchers addressing the way knowledge and values are constructed through images as well as language usages (Machin, 2013). This multimodal turn has developed in response to the increasingly visual nature of news and other communication such that much of the information that contributes to contemporary discourses is received through the visual channel. This 'visualisation' of discourse requires researchers in discourse analysis to consider visual forms of communication not only in their own right but because, on the view of language to be presented here, to do so will provide better insight into the meanings communicated via language. If understanding language involves the construction of multimodal mental models which encode information derived from visual, including mediatised visual, experience, then the increased visual information we receive will play a significant part in shaping linguistically communicated meaning.

Several approaches to multimodal discourse analysis have been developed (e.g. Machin and Mayr, 2012; Richardson and Wodak, 2009; El Refaie, 2003). Amongst the earliest and most influential, however, have been those based in grammatics (e.g. Kress and van Leeuwen, 1996; O'Halloran, 2004). Here, as Machin (2013: 348) states, researchers saw a need "to develop, in the fashion of linguistics, tools that could better allow us to describe how these non-linguistic forms were being used to communicate ideas, attitudes and identities". The dominant strand of linguistics informing multimodal studies was Halliday's Systemic Functional Linguistics (SFL). Different positions have been taken, however, on the relation between grammars – as systems of semiotic resources – in linguistic and non-linguistic modes of communication.

While linguistics has had a profound impact on the development of multimodality as a field, then, the same cannot be said for the other way round. That is, multimodality has, hitherto, had little or no bearing on linguistics, including applied linguistics in the form of Critical Discourse Analysis (cf. Arnheim, 2004). One recent development, however, suggests we may need, precisely, to turn to multimodality if we want to fully account for the way that meanings are communicated via language. Cognitive Linguistics and Cognitive Linguistic CDA (CL-CDA) argue that meaning-making in language involves the construction of fully modal rather than *a*modal mental representations. These mental representations, which are based in previous visual, including visuo-semiotic experience, encode information of exactly the kind studied in multimodal discourse analysis. It follows that insights from multimodality can, indeed should, be brought to bear in (critical) studies of language. In this paper, I discuss, in light of the view from Cognitive Linguistics, some of the ways in which CDA might be enriched by insights from multimodality. The paper is purely theoretical and I use non-attested examples throughout to illustrate the points being made. However, I hope to show some of the ways in which a cognitive approach to CDA incorporating insights from multimodality might be operationalised.

2. On Language and Image

Studies in CDA and systemic functional approaches to multimodal discourse analysis (SF-MDA) share a view of grammar as a semiotic resource organised into a number of functional systems each specifying particular parameters of realisation. Grammars exist, then, as integrated systems of meaning potentials which are instantiated in the 'choices' presented by texts and actualised in the reader's uptake of the knowledge and value positions expressed in those choices. Different stances are taken on the level of overlap between linguistic and visual grammars. Two principle positions may be identified. From one position, meanings are seen as belonging to culture rather than to specific semiotic modalities. This is the view, for example, of Kress and van Leeuwen (1996). Language usages and images, on this account, are both seen as instantiations of a single underlying semiotic capacity. Of course, Kress and van Leeuwen recognise that some meaning potentials are more readily expressed in one modality than the other but maintain that many of the same systems function in both verbal and visual modes. From this perspective, researchers in SF-MDA have sought to show how grammatical systems identified in the study of language similarly function in visual design. For example, Machin (2007) shows how systems of TRANSITIVITY and MODALITY, as well as the grammar of SOCIAL ACTOR REPRESENTATION first described in van Leeuwen (1996), are exploited in visually articulated discourses. From a second position, images are still seen as instantiations of a systemic functional grammar, designed to perform ideational, interpersonal and textual metafunctions of communication, but the categories that make up the grammar are seen as particular to visual communication. This is the view, for example, of O'Halloran (2008), who states that "the systems for visual imagery are not the same as the language systems, which is an obvious point given the differences between the two semiotic resources" (p. 451). Amongst those systems assumed to be specific to the visual modality are those pertaining to FORM (COLOUR, SHAPE, SIZE) and PERSPECTIVE (DEEP SPACE, POINT OF VIEW) as described by Lim (2004).

Some researchers have warned against comparing linguistic and visual communication where to do so may be to impose a kind of linguistic 'imperialism' (e.g. Machin, 2009; Stöckl, 2004). According to

Stöckl (2004: 18), for example, the danger is that "we tend to somehow look at one mode in terms of another" and that "due to language's dominance, we seem to be asking which linguistic properties images have". Stöckl's point is that this might lead us to overlook important design features of images which lie beyond the scope of language. However, an equally pertinent point is that it precludes us from asking which imagistic properties might language have.

A major motivation for the separation of semiotic systems in O'Halloran's work comes from her (and SFL's) view of language processing as procedural compared to images whose parts are perceived following Gestalt laws of organisation (2008: 447-448; 2011: 124). According to this view, sentences are processed sequentially as an accumulation of parts and progressive dependencies. This view is at odds, however, with the view advanced in CL-CDA where, following Langacker (2008, 2013), grammatical 'constructions' are seen as whole units or 'chunks' paired at the conceptual level with image-schematic representations inside a system of *symbolic assemblies*. Image schemas, crucially, are experienced as gestalts and possess various visuo-semiotic properties. When language is viewed in this way, then the cognitive processes involved in linguistic and visual understanding may not be all that different. Indeed, Cognitive Linguistics explicitly rejects the idea that the systems and processes that support language are unique to language, which is instead seen as reliant on the same systems and processes that are involved in non-linguistic domains of cognition, including visual perception as well as memory and action (Croft and Cruse, 2004). In the following section, I outline the basic principles of CL-CDA in more detail before going on to describe some of the ways in which this approach might benefit from consultation with multimodal approaches to discourse analysis.

3. CL-CDA

A key claim of CL-CDA is that meaning in language is not encoded in amodal propositional structures such as described in Kintsch and van Dijk (1978). Rather, language usages are seen as cues for the co-construction of a dynamic conceptualisation which constitutes an intersubjective mental experience of the verbally represented phenomena. It is this process of conceptualisation, then, that provides meaning in and gives coherence to discourse. Conceptualisation as a process is effected through a range of construal operations (see Croft and Cruse, 2004; Hart, 2014a) but, crucially, involves the activation of imagery (Bergen, 2012; Langacker, 2008, 2013; Talmy, 2000). Meaning in linguistic discourse, in other words, involves multimodal rather than amodal mental representations. Two levels of representation are integrated within the conceptualisation: an image-schematic representation and an enriched simulation.

Image-schematic representation

Image schemas are pre-linguistic conceptual structures which derive from our experiences in and observations of the physical world (Johnson, 1987; Mandler, 2004). They are not images per se but, rather, abstract skeletal structures distilled from repeated patterns of experience. Image schemas arise in basic domains like ACTION, FORCE, SPACE and MOTION to constitute 'folk theories' of the way the world works and provide heuristic guides for interpretation and action. They encode information pertaining, for example, to topology, sequence and causation. These conceptual structures are co-opted in language to form the meaningful basis of both lexical and grammatical units inside a system of *symbolic assemblies* (Langacker, 1987, 1991, 2008). On this account, then, no principle distinction

is made between grammar and the lexicon. Grammatical constructions are stored as discrete units (in the 'constructicon') which are in and of themselves meaningful by virtue of the conceptual structures they conventionally invoke. In discourse, the grammatical constructions presented by texts invoke their conventional conceptual counterparts to constitute our most basic understanding of the target scene, providing information relating to event-type and -structure. Crucially, from a critical perspective, alternative grammatical constructions are available to impose on the target scene alternative understandings or *construals* which may be reflective of wider patterns of belief and value, discourses (see Hart, 2014a). By way of example, consider the contrast between (1) and (2):

(1)   [Protesters TR] [entered MOTION] [the building LM] around 3pm

(2)   [Protesters TR] [smashed their way MOTION] [into PATH] [the building LM] around 3pm

Both constructions encode a MOTION event in which a 'trajector' (TR) moves along a path of motion into a 'landmark' (LM). The difference between (1) and (2) lies in the manner of motion which in (1) has no impact on the LM but in (2) results in a change in state to the LM. The contrasting image schemas invoked by (1) and (2) to construe the same situation are modelled in Figure 1a and 1b respectively.

(a)   (b)
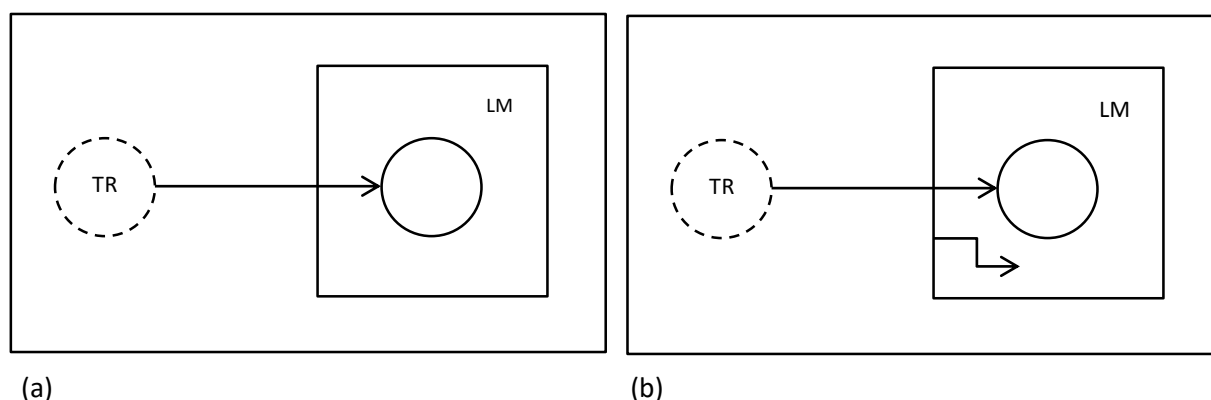
Figure 1. Canonical vs. Impact on LM MOTION schema

TR = Trajector, LM = Landmark. The dashed circle represents that starting point of TR within the event-structure. The vector represents its path of motion into LM. The stepped arrow in (b) represents the change in state to LM as a result of the event.

Another crucial argument of Cognitive Grammar is that situations are subject to further construal as the conceptual content apprehended in their understanding is experienced from a particular perspective. This further level of construal arises as many grammatical constructions include within their semantic frames a point of view specification (Langacker, 2008: 75). This is most easy to see in deictic expressions such as (3) compared to (4). In (3), the point of view from which the event is construed is from inside the LM. In (4), by contrast, the event is construed from outside the LM. It should be noted, however, that spatial point of view is not restricted to deictic expressions but is encoded in other types of grammatical construction (see Section 4).

(3)   [Protesters ~TR~] [came ~MOTION~] [into ~PATH~] [the building ~LM~]

(4)   [Protesters ~TR~] [went ~MOTION~] [into ~PATH~] [the building ~LM~]

Simulation

At a second level of representation, the abstract image-schematic structures invoked by grammatical constructions are 'fleshed out' in a richer simulation of the target scene (see Bergen, 2012 for an overview of research in this area). The basic idea here is that language functions as "a set of cues to the comprehender to construct an experiential (perception plus action) simulation of the described situation" (Zwaan, 2004: 36). By simulation it is meant an "experience resembling perceptual or motor experience occurring in the absence of the relevant external stimuli, in the case of perceptual experience; or without actual execution of motor actions, in the case of motor imagery" (Bergen, 2007: 735). Meaning at this level, then, involves imagining the target scene unfolding spatio-temporally in our mind's eye. At this level of meaning, information not specified in the utterance but derived from relevant past experience, including information relating to the colour, shape, size, orientation of actors, entities, and effectors in the scene, as well as length, rate and relative direction of motion, is integrated into the conceptualisation (Stanfield & Zwaan, 2001; Glenberg & Kaschak, 2002; Kaschak et al., 2005; Matlock, 2004; Zwaan et al., 2002). It is also at this level that systems of affect are integrated. The way this works, according to Bergen (2007: 278), is as follows:

> Through exposure to language in context, language users learn to pair chunks of language like *kick, Mary*, or *John* with perceptual, motor, social, and affective experiences. In subsequent instances of language use, when the original perceptual, motor, social and affective stimuli are not contextually present, the experience of them is re-created through the activation of neural structures responsible for experiencing them in the first place.

Evidence that understanding language involves such a simulation comes from a variety of psycholinguistic experimental and neurofunctional imaging studies which show that the same neural processes involved in perceiving, performing, or being on the receiving end of an action are activated when the same action is designated in the course of discourse (see Bergen, 2012 for an overview; see also Barsalou, 1999; Glenberg & Robertson, 2000; Bergen et al., 2004; Bergen & Chang, 2005).

The simulation, furthermore, is run from the perspective of an 'immersed experiencer' (Zwaan, 2004). That is, with *simulated situatedness*. What kind of experience role the conceptualiser adopts, i.e. whether they imagine themselves as perceiving, performing, or being on the receiving end of the action designated depends, at least in part, on the point of view determined at the schematic level by the selected grammatical construction (Bergen et al., 2004; Bergen and Chang, 2005; Zwaan, 2004).[1]

The connection with multimodality should by now be becoming clear. The meanings invoked by language encode multimodal information of precisely the kind whose potential social semiotic functions have been studied in multimodal discourse analysis. In the following section, I elaborate on this connection to show, under three headings, how insights from multimodal discourse analysis might usefully be incorporated in language-focussed CDA.

4. Connections with Multimodality

Image as Text

What the Cognitive Linguistic account outlined above suggests is that the meanings invoked by particular language usages – text – are multimodal in nature. The conceptualisations constructed in linguistic discourse, in other words, possess many of the same properties as images, properties whose ideological potentials have been extensively studied in multimodal discourse analysis. Multimodal discourse analysis therefore presents a useful paradigm through which to analyse the nuances of meaning communicated via language as well as image.

Two visuo-spatial parameters which have been widely studied in multimodality and which, on the account presented above, feature in linguistically engendered meaning too, relate to topology and point of view (Kress and van Leeuwen, 1996; van Leeuwen, 2008; Lim, 2004). Image schemas, for example, encode topological relations between elements of the target scene relative to the conceptualiser's own situatedness or point of view within the cognitive representation. In this way, the system of PERSPECTIVE can be seen to function in both linguistic and visual, including visuo-semiotic, understanding.[2] Similarly, at the level of simulation, information relating to systems of FORM, such as COLOUR, which has likewise been widely studied in multimodality (Koller, 2008; Kress and van Leeuwen, 1996; Machin, 2007; van Leeuwen, 2010), is integrated into the conceptualisation.

Let us illustrate how insights from multimodality can be incorporated in CDA, then, taking PERSPECTIVE as a case in point.[3] Consider the following four constructions which make up part of the conventional repertoire for reporting on political protests (Hart, 2013a/b, 2014a/b).

(5)   [Protesters AGENT] [attacked ACTION^A] [police officers PATIENT] [in Parliament Square LOCATION]

(6)   [Police officers PATIENT] [were attacked ACTION^A] by [protesters AGENT] [in Parliament Square LOCATION]

(7)   [Police officers AGENT^1] [clashed ACTION^R] with [protesters AGENT^2] [in Parliament Square LOCATION]

(8)   [Protesters AGENT^2] [clashed ACTION^R] with [police officers AGENT^1] [in Parliament Square LOCATION]

All four grammatical constructions can be used to describe the same or similar material situations. A number of differences can be discerned, however, in the conceptualisations they invoke. The conceptualisations are modelled in Figure 2(a)-(d). At the schematic level, conceptualisation takes place inside nested mental spaces – temporary conceptual pockets opened up in discourse for purposes of local understanding (Fauconnier, 1997). At least two mental spaces are involved: an 'event space' in which event-structure is determined and a 'base space' in which point of view is determined. The point of view in the base space does not alter the content of the event space but does affect how it is seen and evaluated. As Kress and van Leeuwen (1996: 146) state: "the addition of perspective adds nothing to the representational meaning [of images] but it does add attitudinal meaning".

The first major difference between (5) – (8), then, relates to the image schema recruited to construe the structure of the event inside the event space. In (5) and (6) the regular transactive or asymmetrical constructions [Sbj ACTION^A Obj] and [Sbj ACTION^A *by* Obj] both invoke a one-sided ACTION schema in

which an AGENT (A) acts upon a PATIENT (P). The direction of energy transfer in the event is unidirectional. By contrast, in both (7) and (8) the reciprocal construction [Sbj ACTION^R *with* Obj] invokes a two-sided ACTION schema in which both participants are equally agentive. The direction of energy transfer, in other words, is bidirectional.[4] Ideologically, of course, the difference serves to attribute agency in and therefore responsibility for the violence in different ways thus realising different (de)legitimating strategies. The second major difference between (5) - (8) consists in the point of view from which the conceptual content in the event space is construed. It is here that the connection with multimodality becomes most clear. The argument I am advancing is that regular transactive constructions encode a point of view in the base space which is either up-stream of the energy flow from behind the AGENT or downstream of the energy flow from behind the PATIENT depending on voice. The active voice as in (5) invites a construal from up-stream of the energy flow as in Figure 2(a) while the passive voice as in (6) invites a construal from down-stream in the energy flow as in Figure 2(b). The reciprocal construction, by contrast, as well as encoding an alternative event-structure, invites a construal of that content from an alternative point of view at right angles with rather than in line with the direction of energy flow. Which of the two available positions is adopted here is motivated by an iconic correspondence between the linear organisation of the clause and the left-right organisation of conceptual content (cf. Perniss et al. 2010). Thus, assuming we arbitrarily assign 'the police' as AGENT^1 and 'protesters' as AGENT^2, (7) invites as construal as in Figure 2(c) while (8) invites a construal as in Figure 2(d).

Now, if these hypotheses are correct, and the grammatical constructions involved invoke a spatialized mental representation of the referential situation, then it follows, since there is no principled reason to suppose that such semiotic features should function differently in linguistic versus visual modes of communication, that the spatial relations encoded can be analysed within frameworks of multimodality. That is (i) conceptualisations can be characterised in the same terms as images[5] and (ii) the ideological/(de)legitimating potentials of spatial (as well as other multimodal) properties can be interpreted with reference to previous findings from multimodal discourse analyses.

The alternative construals encoded by grammatical constructions (5)-(8) establish different spatial and orientational relations between the conceptualiser and participants in the scene under consideration. In multimodal discourse analysis, these different spatial and orientational parameters are attributed different semiotic functions (e.g. Kress and van Leeuwen 1996). Most fundamentally, for example, (5) and (6) present a frontal angle while (7) and (8) present an oblique angle. And functionally, as Kress and van Leeuwen (1996: 136) state, "the difference between the oblique and the frontal angle is the difference between detachment and involvement". Thus, we can characterise the conceptualisation invoked by (5) and (6) as encoding an 'involved' perspective and the conceptualisation invoked by (7) and (8) as encoding an 'observer's' perspective.

More specifically, from the involved perspectives of (5) and (6), participants are arranged, relative to the point of view of the conceptualiser, according to spatial values of proximal versus distal as well as orientation values ego-aligned versus ego-opposed. In (5), the AGENT is proximal and ego-aligned while the PATIENT is distal and ego-opposed. In (6), the PATIENT is proximal and ego-aligned while the AGENT is distal and ego-opposed. According to multimodality studies, the illusion of depth or visual distance in an image serves to signify social distance (Kress and van Leeuwen, 1996: 124; Machin and Mayr, 2012: 96). Thus, where the active voice in (5) establishes a proximal relation with the AGENT and a distal relation with the PATIENT, it may be said to realise an affiliation strategy with respect to the AGENT

and an othering strategy in respect of the PATIENT while the passive voice in (6), which establishes a proximal relation with the PATIENT and a distal relation with the AGENT, invites the viewer to identify with the PATIENT and otherises the AGENT.

(a)

(b)

(c)

(d)
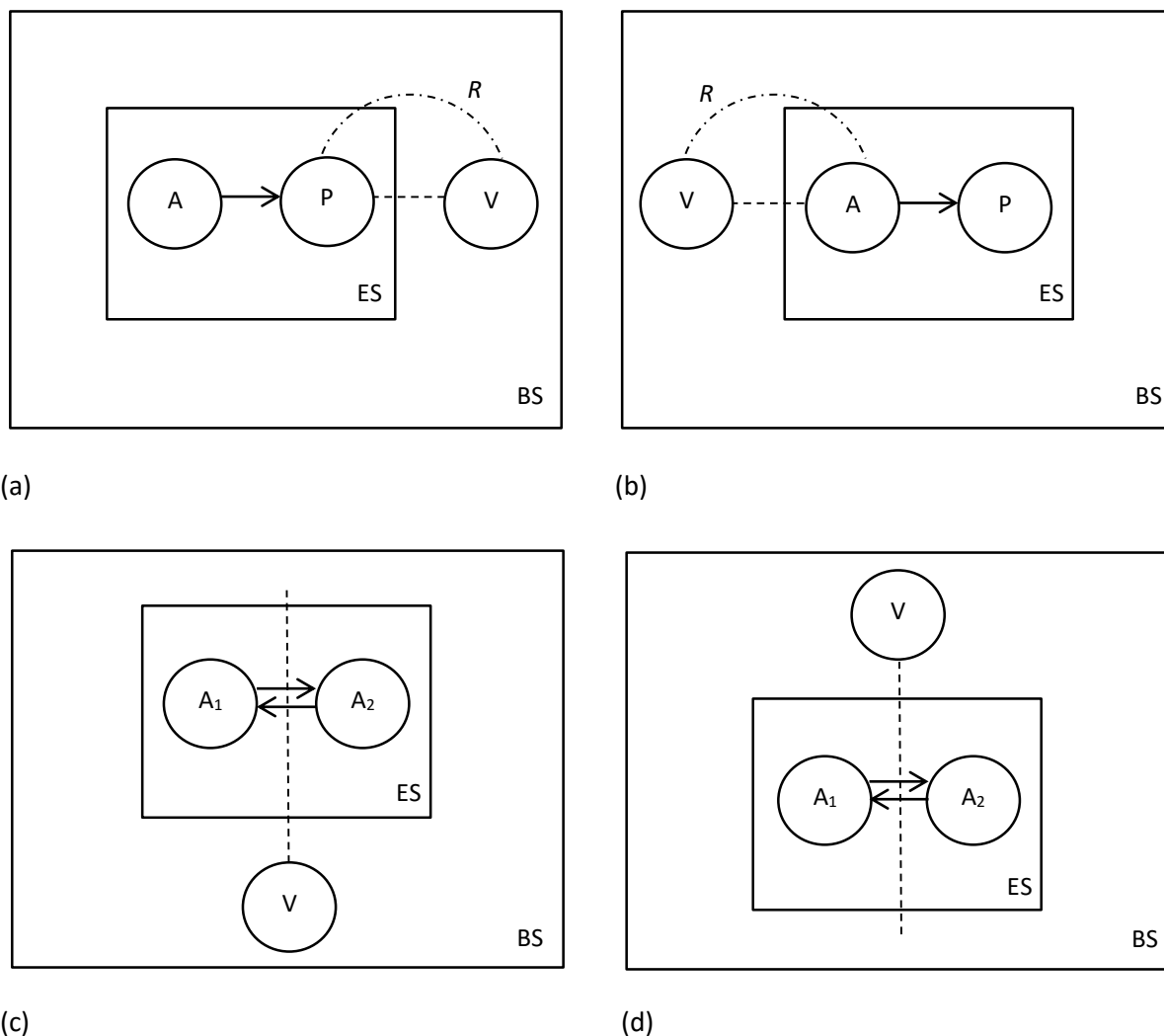
Figure 2. Event-structure and point of view in
regular transactive versus reciprocal constructions

ES = Event space, BS = Base space. A = AGENT, P = PATIENT and V = the viewer or conceptualiser. The solid vectors represent the flow of energy in the event structure. The dashed line represents the viewer's orientation by sightline. The broken line (R) represents a role connector between V and A or P.

Moreover, at the level of simulation, due to the orientational values encoded, the viewer is encouraged not only to see the event from the same perspective as the AGENT or PATIENT but to experience the event as if they were in their shoes.[6] That is, in Kress and van Leeuwen's terms, the viewer comes to temporarily inhabit the world of the image as a participant in the event depicted. In the active voice of (5), then, the viewer is invited into the role of the AGENT to imagine themselves

performing the action. Thus, since no one likes to think of their own actions as being beyond reproach, the active voice is more likely to be used to report legitimated actions of the in-group as in (9), where the event is construed as a MOTION event rather than an ACTION event (see Hart, 2013a/b). Construing the event as a MOTION event, as modelled in Figure 1a, construes it as non-transactive and thereby glosses over any material effect on a second participant.

(9)   [The police TRAJECTOR] [moved MOTION] [into PATH] [the area LANDMARK] around 3pm

By contrast, in the passive voice of (6), the viewer is invited into the role of PATIENT to imagine themselves being on the receiving end of the action. The effect of such a construal, when the simulation is run, is to evoke a sense of personal threat. As Machin and Mayr state:

> Where we see people from behind, this can often serve to offer us their point of view, their perspective on the world. So if in an image we were to see the back of a person and then beyond them another person pointing a gun at them, we would take the perspective of the first person who faces the threat. (Machin and Mayr, 2012: 99)

It is therefore expected, against orthodox interpretations of the ideological functions of the passive voice (Fowler et al., 1979; Kress and Hodge, 1993), that passive constructions will be favoured for delegitimated, e.g. violent, ACTION events in which the perceived out-group are agentive (see Hart, 2013a/b).

From the observer perspectives of (7) and (8), participants are arranged in profile according to spatial values left versus right. According to Kress and van Leeuwen (1996), spatial left signifies 'given' information while spatial right signifies 'new' information. By 'given' it is meant 'known as part of the culture' while 'new' is understood as 'problematic' or 'contested' (Kress and van Leeuwen, 1996: 180-181). Following Kress and van Leeuwen's characterisation, we can therefore interpret the construal encoded in (7) as problematizing the behaviour of the protesters and the construal encoded in (8) as problematizing the behaviour of the police.

By connecting with multimodality, the CL-CDA can reveal ideological/(de)legitimating properties of language usages not previously recognised. Indeed, the critical impetus of CL-CDA may be seen to come precisely from the incorporation of insights from multimodal social semiotics. However, questions arise here concerning how it is we know that different visuo-spatial variables carry different ideological connotations, where the mapping between spatial and orientational values and ideological values seems to be based largely in the subjective readings of the analyst. As Kress and van Leeuwen (1998: 218) concede:

> The major challenge to our approach is the epistemological status of our claim … [H]ow can we know that left and right, and top and bottom, have the values we attribute to them, or more fundamentally, have any value at all?

To address this issue, two sources of evidence are available, both of which suggest alternative semiotic functions for spatial left and right. The first line of evidence is lexical. Polysemy and metaphor in discourse both reflect conceptual connections, in the form of conceptual metaphors, conventionally shared by language users (Lakoff, 1987; Lakoff and Johnson, 1980). Such conceptual metaphors emerge from co-activations in embodied experience. The conceptual metaphors gleaned from patterns of lexical realisation may therefore help establish an experiential basis for the associations

claimed between spatial and ideological values in MDA (see Feng, 2011; Feng and Espindola, 2013). One candidate conceptual metaphor is SOCIAL RELATIONS ARE DISTANCE which is evidenced by expressions like 'close friends' or 'remote cousins'. This conceptual metaphor supports the interpretation of proximal/distal values in images as signifying social distance. Similarly, the semiotic functions of ego-aligned/ego-opposed orientations is supported by a conceptual metaphor STANCE IS POSITION/ORIENTATION IN SPACE (see Hart, 2014a).[7] Our ideological points of view as well as our physical points of view can be 'in line with' or 'opposed to' one another. I can 'stand behind' you both spatially and socially. The etymology of 'confrontation' has its roots in expressions of spatial relation. This conceptual metaphor, however, does not seem to bear out the given/new interpretation of left versus right where, for example, to be 'in the right' is not to be considered problematic but morally correct.

The second line of evidence is experimental. Casasanto (2009), for example, has shown that spatial right is generally associated with positive valence while spatial left is generally associated with negative valence. Of course, the ideological values of spatial properties are unlikely to be universal and may vary according to culture, context and individuals. Experiments such as this, though, suggest ways in which the ideological values of spatial properties in language and image could be tested within the perimeters of particular social and political discourses.   And all of this, of course, is predicated on the spatiality of linguistic meaning in the first place, an hypothesis which still requires empirical validation.
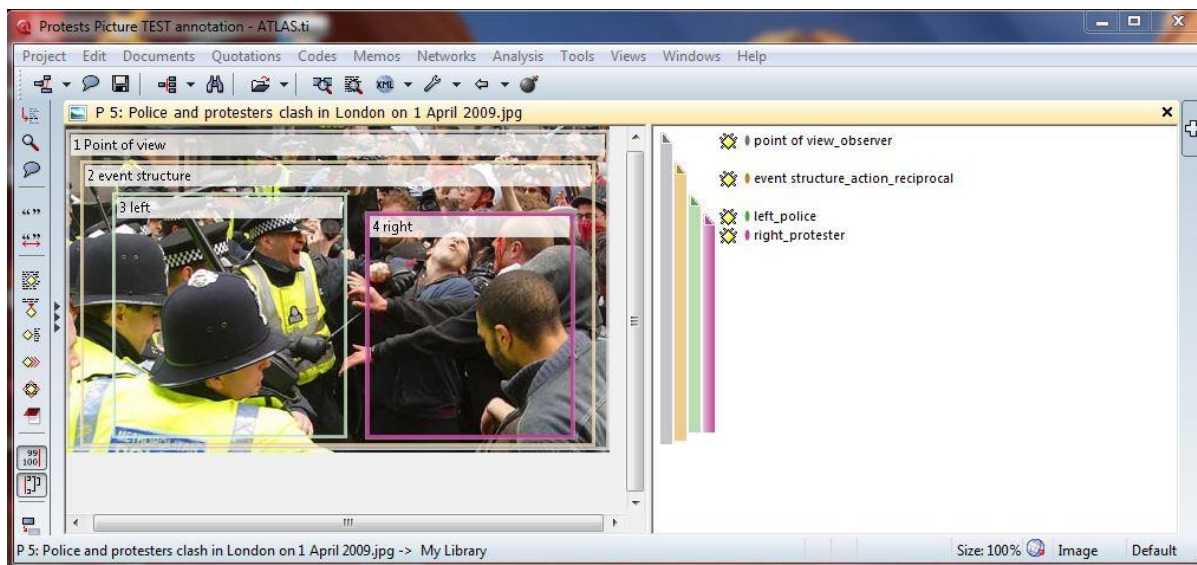
Image as Co-text

The hypothesis advanced above makes a number of predictions concerning behavioural relations between language and image. One prediction is that the multimodal features encoded in language usages and images co-textually linked should, on the whole, tend to match. That is, in other words, there should be a tendency for relevant language usages and images in the same text to encode, e.g., the same spatial point of view. One site where this might most easily be put to the test is in news photographs and their captions.   To achieve this would require a software package in which images could be annotated using a semantic code applicable to both language and image. See Figure 3 for an illustrative example.   The image is of the 2009 G20 protests in London published in the *Guardian*, a paper which was very critical of police behaviour during the summit.

Images can then be compared with co-textually linked sentences, which have been tagged using the semantic code, to see if they match. The image in Figure 3, for example, was accompanied by the caption in (10).

(10)   [Police $_{AGENT}^{1}$] and [protesters $_{AGENT}^{2}$] [clash $_{ACTION}^{R}$] in London on 1 April 2009

Moreover, quantitative analyses may be run over any images saved with specific tags or combinations of tags (collocations) in much the same way as a concordance program used in Corpus Linguistics. This allows one to observe general convergences of language and image across a corpus rather than specific instances of congruity. It also allows for cross-corpus comparisons (e.g. between two different newspapers).

Figure 3. Semantic annotation of image

Image as Context

One further way in which multimodal discourse analysis might be able to contribute to language-based CDA emerges from the claim made in simulation semantics that the imagery invoked by particular language usages is, in part at least, a function of the visuo-spatial and interactional experiences which have accompanied similar language usages in the past (Bergen, 2007). In cases of social and political discourse, of course, one rarely has any first-hand experience of the events being described. Rather, the contexts in which one has been exposed to similar language usages are mediatised and our accompanying experience visuo-semiotic rather than primary visual. The simulated experience, that is, the meaning, evoked by language usages will therefore reflect the semiotic experiences we have had in encountering images of the same or similar events. In other words, it follows that the precise meanings of language usages in any given context are dependent on patterns in visual articulations of the same discourse. In interpreting language, this means that those semantic features not determined by grammatical properties of the utterance but provided by background experience, such as the size, shape, colour of actors, their clothing etc., will be a probabilistic function of the information found in relevant visuo-semiotic experience. To study the nuances of meanings communicated via language, therefore, requires a corpus-based analysis of the images that make up the visuo-semiotic context for the utterance(s) in question. The annotation need not then be restricted to event-structure, point of view and spatial relations but could also usefully include, e.g., categories defined within van Leeuwen's (1996) Social Actor Network as well as categories for body posture, gesture, gaze, facial expressions, clothing and effectors, backdrop and so on.

6. Conclusion

What I have aimed to do in this paper, in accord with the aims of this special issue, is to set out an agenda for future developments in CDA and specifically CL-CDA. The principle claim has been that the

mental representations invoked by language usages encode multimodal information. Insights from multimodal discourse analysis might therefore be usefully integrated within CL-CDA in order to illuminate the nuances of meanings communicated via language. That is to say, where studies in multimodality have been able to shed light on the way images contribute to the social construction of knowledge and values, given the arguments made here concerning the visual basis of linguistic meaning, findings from multimodal studies may equally be able to illuminate some of the most subtle ways in which language contributes to the social construction of knowledge and values. This is especially the case given the increasingly visual practices of news communication. Although I have not presented any attested data, I hope to have illustrated some of the ways in which such an integrated approach might be operationalised. I have illustrated this with a focus on one multimodal feature in particular, POINT OF VIEW. However, I could equally have looked at others. Ultimately, what I hope to have achieved in this paper, then, is to have created a space for a future program of research employing cognitive and multimodal methods to investigate, across a range of features, connections between language, image and ideology.

Notes

[1] It is not just literal language usages which involve simulation. For example, Gibbs (2006) shows that metaphor, a powerful device in the communication of ideology (Charteris-Black 2004), involves a simulation of events which are technically impossible but are imagined as if they were real (see also Fauconnier and Turner 2002). Consider, for example, well-attested metaphors like IMMIGRATION IS A FLOOD. A simulation account of metaphor comprehension provides a cognitive and experiential explanation for the rhetorical effectiveness of metaphor in discourse where the experience provided by the source domain in the simulation may be especially powerful in conjuring imagery which mobilises systems of affect. A metaphor such as IMMIGRATION IS A FLOOD, for example, may result in genuine feelings of drowning (see also van Dijk 1984, 1987).

[2] 'System' in CL-CDA, it should be noted, does not refer to a functional network system designed specifically for communication but to more general cognitive systems which operate across cognitive domains, including language and perception (see Hart 2014a for detailed discussion).

[3] For reasons of space, I concentrate solely on the horizontal or anchorage plane of perspective (see Hart 2015 for a more detailed discussion, including of angle and distance planes)

[4] It is important to recognise that the examples in (5) – (8) are particular instantiations of abstract grammatical constructions. The same construction may be instantiated in utterances involving other semantically similar verbs. For example, the transactive constructions in (5) and (6) may be realised in utterances containing verbs like *hit, strike, punch* etc. while the same reciprocal construction realised in (7) and (8) may be instantiated in utterances containing verb phrases such as *trade punches, exchange blows* etc.

[5] It is therefore not infelicitous to refer to the reader as a 'viewer'.

[6] At the schematic level, this is achieved via a role 'connector' (R) as in Figure 2(a) and 2(b) (Fauconnier 1997).

[7] Interpretations of spatial positions or configurations in the vertical rather than the horizontal dimension, such as the ideal/real interpretation of top/bottom (Kress and van Leeuwen 1996), may similarly be supported by conceptual metaphors, in this case GOOD IS UP/BAD IS DOWN.

References

Arnheim R (2004) *Visual Thinking,* 2nd edn. California: University of California Press.

Barsalou LW (1999) Perceptual motor systems. *Behavioural and Brain Sciences* 22: 577-609.

Bergen B (2007)   Experimental methods for simulation semantics.   In: Gonzalez-Marquez M, Mittelberg I, Coulson S and Spivey MJ (eds.), *Methods in Cognitive Linguistics*. Amsterdam: John Benjamins, pp. 277-301.

Bergen B (2012) *Louder than Words: The New of Science of How the Mind Makes Meaning*. New York: Basic Books.

Bergen B and Chang N (2005)   Embodied construction grammar in simulation-based language understanding.   In: Östman J.-O and Fried M (eds.), *Construction Grammars: Cognitive Grounding and Theoretical Extensions.* Amsterdam: John Benjamins, pp. 147-190.

Bergen B, Chang N and Narayan S (2004) Simulated action in an embodied construction grammar. In: *Proceedings of 26th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum, pp. 108-113.

Casasanto D. (2009)   Embodiment of abstract concepts: good and bad in right- and left-handers. *Journal of Experimental Psychology* 138 (3): 351-367.

Charteris-Black C (2004) *Corpus Approaches to Critical Metaphor Analysis*. Basingstoke: Palgrave.

Croft W and Cruse DA (2004) *Cognitive Linguistics*. Cambridge: Cambridge University Press.

El Refaie E (2003)   Understanding visual metaphor: The example of newspaper cartoons. *Visual Communication* 2 (1): 75-96.

Fauconnier G (1997) *Mappings in Thought and Language*. Cambridge: Cambridge University Press.

Fauconnier G and Turner M (2002) *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.

Feng D (2011) Visual space and ideology: A critical cognitive analysis of spatial orientations in advertising. In: O'Haloran KL and Smith BA (eds.), *Multimodal Studies: Exploring Issues and Domains*. London: Routledge, pp. 55-75.

Feng D and Espindola E (2013) Integrating functional and cognitive approaches to multimodal discourse analysis. *Ilha do Desterro* 64: 85-110.

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

Gibbs RW (2006) Metaphor interpretation as embodied simulation. *Mind & Language* 21 (3): 434-458.

Glenberg A and Kaschak M (2002) Grounding language in action. *Psychonomic Bulletin & Review* 9: 558-565.

Glenberg A and Robertson D (2000) Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language* 43: 379-401.

Hart C (2013a) Event-construal in press reports of violence in political protests: A Cognitive Linguistic Approach to CDA**.** *Journal of Language and Politics* 12 (3): 400-423.

Hart C (2013b) Constructing contexts through grammar: Cognitive models and conceptualisation in British Newspaper reports of political protests. In: Flowerdew J (ed.), *Discourse and Contexts*. London: Continuum, pp. 159-184.

Hart C (2014a) *Discourse, Grammar and Ideology: Functional and Cognitive Perspectives*. London: Bloomsbury.

Hart C (2014b) Construal operations in online press reports of political protests. In: Hart C and Cap P (eds.), *Contemporary Critical Discourse Studies*. London: Bloomsbury, pp. 167-188.

Hart C (2015). Viewpoint in linguistic discourse: Space and evaluation in news reports of political protests. *Critical Discourse Studies* 12 (3): 238-260*.*

Hodge R. and Kress G (1993) *Language as Ideology,* 2nd edn. London: Routledge.

Johnson M (1987) *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago: University of Chicago Press.

Kaschak M, Madden CJ, Therriault DJ, Yaxley RH, Aveyard M, Blanchard AA and Zwaan RA (2005) Perception of motion affects language processing. *Cognition* 94 (3): B79-B89.

Kress G and van Leeuwen T (1996) *Reading Images: The Grammar of Visual Design,* 2nd edn*.* London: Routledge.

Kintsch W and van Dijk TA (1978) Toward a model of text comprehension and production. *Psychological Review* 85 (5): 363-394.

Koller V (2008) "More than just a colour": Pink as a gender and sexuality marker in visual communication. *Visual Communication* 7 (4): 395-423.

Kress G and van Leeuwen T (1998) Front pages: (The critical) analysis of newspaper layout. In: Bell A and Garrett P (eds.), *Approaches to Media Discourse*. Oxford: Blackwell, pp. 186-219.

Lakoff G (1987) *Women, Fire and Dangerous Things: What Categories Reveal about the Mind.* Chicago: University of Chicago Press.

Lakoff G and Johnson M (1980) *Metaphors we Live by.* Chicago, IL: University of Chicago Press.

Langacker RW (1987) *Foundations of Cognitive Grammar, vol. I: Theoretical prerequisites.* Stanford: Stanford University Press.

Langacker RW (1991) *Foundations of Cognitive Grammar, vol. II: Descriptive application.* Stanford: Stanford University Press.

Langacker RW (2008) *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.

Langacker RW (2013) Essentials of Cognitive Grammar. Oxford: Oxford University Press.

Lim FV (2004) Developing an integrative multisemiotic model. In: O'Halloran KL (ed.), *Multimodal Discourse Analysis*. London: Continuum, pp. 220-246.

Machin D (2007) *An Introduction to Multimodal Analysis*. London: Bloomsbury.

Machin D (2009) Multimodality and theories of the visual. In: Jewitt C (ed.), *Handbook of Multimodal Analysis*. London: Routledge, pp. 181-190.

Machin D (2013) What is multimodal critical discourse studies? *Critical Discourse Studies* 10 (4): 347-355.

Machin D and Mayr A (2012) *Critical Discourse Analysis: A Multimodal Approach*. London: Sage.

Mandler JM (2004) *The Foundations of Mind: Origins of Conceptual Thought*. Oxford: Oxford University Press.

Matlock T (2004) Fictive motion as cognitive simulation. *Memory & Cognition* 32: 1389-1400.

O'Halloran KL (ed.) (2004) *Multimodal Discourse Analysis*. London: Continuum.

O'Halloran KL (2008) Systemic Functional-Multimodal Discourse Analysis (SF-MDA): Constructing Ideational Meaning using Language and Visual Imagery. *Visual Communication* 7 (4): 443-475.

O'Halloran KL (2011) Multimodal discourse analysis. In: Hyland K and Paltridge B (eds.), *Companion to Discourse*. London: Continuum, pp. 120-137.

Perniss P, Thompson R and Vigliocco G (2010) Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology* 1: 227.

Richardson J and Wodak R (2009) The impact of visual racism: Visual arguments in political leaflets of Austrian and British far-right parties. *Contraversia:* 6(2): 45-77.

Stanfield RA and Zwaan RA (2001) The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science* 12: 153-156.

Stöckl H (2004) In between modes: Language and image in printed media. In: Vntola E, Charles C and Kaltenbacher M (eds.), *Perspectives on Multimodality*. Amsterdam: John Benjamins, pp. 9-30.

Talmy L (2000) *Toward a Cognitive Semantics.* Cambridge, MA.: MIT Press.

Van Dijk TA (1984) *Prejudice in Discourse: An Analysis of Ethnic Prejudice in Cognition and Conversation.* Amsterdam: John Benjamins.

Van Dijk TA (1987) *Communicating Racism: Ethnic Prejudice in Thought and Talk.* London: Sage.

Van Leeuwen T (1996) The representation of social actors. In: Caldas-Coulthard CR and Coulthard M (eds.), *Texts and Practices: Readings in Critical Discourse Analysis*. London: Routledge, pp. 32-70.

Van Leeuwen T (2008) *Discourse and Practice: New Tools for Critical Discourse Analysis*. Oxford: Oxford University Press.

Van Leeuwen T (2010) *The Language of Colour: An Introduction*. London: Routledge.

Zwaan RA (2004) The immersed experiencer: Toward an embodied theory of language comprehension. In: Ross BH (ed.), *The Psychology of Learning and Motivation, Volume 44*. New York: Academic Press, pp. 35-62.

Zwaan RA, Stanfield RA and Yaxley RH (2002) Do language comprehenders routinely represent the shapes of objects? *Psychological Science* 13: 168-171.