Christopher Hart* and Javier Marmol Queralto

What can cognitive linguistics tell us about language-image relations? A multidimensional approach to intersemiotic convergence in multimodal texts

https://doi.org/10.1515/cog-2021-0039
Received March 11, 2021; accepted September 18, 2021; published online October 5, 2021

Abstract: In contrast to symbol-manipulation approaches, Cognitive Linguistics offers a modal rather than an amodal account of meaning in language. From this perspective, the meanings attached to linguistic expressions, in the form of conceptualisations, have various properties in common with visual forms of representation. This makes Cognitive Linguistics a potentially useful framework for identifying and analysing language-image relations in multimodal texts. In this paper, we investigate language-image relations with a specific focus on *intersemiotic convergence*. Analogous with research on gesture, we extend the notion of *co-text images* and argue that images and language usages which are proximal to one another in a multimodal text can be expected to exhibit the same or consistent construals of the target scene. We outline some of the dimensions of conceptualisation along which intersemiotic convergence may be enacted in texts, including event-structure, viewpoint, distribution of attention and metaphor. We take as illustrative data photographs and their captions in online news texts covering a range of topics including immigration, political protests, and inter-state conflict. Our analysis suggests the utility of Cognitive Linguistics in allowing new potential sites of intersemiotic convergence to be identified and in proffering an account of language-image relations that is based in language cognition.

Keywords: cognitive linguistics; intersemiotic convergence; language-image relations; multimodality

***Corresponding author: Christopher Hart**, Linguistics and English Language, Lancaster University, Lancaster, UK, E-mail: c.hart@lancaster.ac.uk
Javier Marmol Queralto, Linguistics and English Language, Lancaster University, Lancaster, UK, E-mail: j.marmolqueralto@lancaster.ac.uk

1 Introduction

Within linguistics, many paradigms have undergone a multimodal turn to view language as only one part of a much broader communicative complex and to thus include within their analytical purviews other non-linguistic modes. The two most widely recognised approaches here are multimodal interaction analysis (e.g., Norris 2004) and social semiotics (e.g., Kress and van Leeuwen 2006). Multimodal interaction analysis is concerned with situated communicative interaction. It has its roots in conversation analysis and interactional sociolinguistics but extends its remit beyond language to take account of the role played by other 'embodied' modes like gesture, gaze, facial expression, body posture and proxemics. Social semiotics, by contrast, extends principles of Systemic Functional Linguistics (SFL) to provide a 'grammar of visual design' intended to account for meaning in 'textual' artefacts like images and sculptures.

Cognitive Linguistics has similarly seen a multimodal turn in recent years (e.g., Pinar Sanz 2015). This has largely fallen into two strands. In one strand, researchers are interested in the role of gesture and its relation to language and conceptualisation in situated usage events where many of the dimensions of construal postulated in Cognitive Linguistics are shown to receive potential expression in gesture also (see Cienki 2013 for an overview). In more textual forms of analysis, it is metaphor, as one particular dimension of construal, which has received by far and away the most attention, where Forceville's (1998, 2002, 2006, 2008) model of multimodal metaphor has been used to interrogate an impressively wide range of text-types, including advertisements, political cartoons, comics, films and musical scores (Forceville and Urios-Aparisi 2009). More recently, expressions of viewpoint across different modes and multimodal text-types have also been subject to investigation (e.g., Dancygier and Vandelanotte 2017; Vandelanotte and Dancygier 2017).[1]

What all approaches to multimodality have in common is a view of meaning as being greater than the sum of its parts. That is, meaning in any communicative act is not just a product of the individual modes that contribute to it but of the interplay

1 In Cognitive Semiotics (Louhema et al. 2019; Zlatev 2015) – a new transdisciplinary field for the study of meaning that combines semiotics with cognitive science and linguistics, with a focus on cognitive linguistics – the term "polysemiotic" is used in preference to "multimodal" to describe communication that relies on a combination of semiotic systems. This is partly motivated by a perceived ambiguity in the way that the terms "mode" versus "modality" are sometimes used. We stick to the term "mode" and understand it as any system of signs available for the communication of meaning. This stands in contrast to "modality" which refers to the 'channel' through which communication is delivered. Thus, speech and writing are different communicative modalities which rely on different combinations of semiotic modes.

between them. Particularly in the case of textual research, a key endeavour here has been to explore the way that language and image interact with one another to create a sense of intersemiotic coherence. As Royce (2007: 63) puts it, researchers want to know "what features make multimodal text visually-verbally coherent". In other words, what gives a multimodal text *texture?* Various attempts have been made to address this question in terms of language-image (L-I) relations.

Much of the research investigating L-I relations has come from the perspective of social semiotics where L-I relations have been modelled on the basis of various types of relation defined within the architecture of SFL (e.g., Liu and O'Halloran 2009; Martinec and Salway 2005; Royce 1998, 2007) as well as Rhetorical Structure Theory (e.g., Taboada and Habel 2013). L-I relations have only recently been addressed from the perspective of Cognitive Linguistics (e.g., Dancygier and Vandelanotte 2017). In this paper, we offer a Cognitive Linguistic treatment of one particular L-I relation in the form of what Lui and O'Halloran call *intersemiotic parallelism* but which we will refer to as *intersemiotic convergence* (cf. also *intersemiotic repetition* [Royce 1998, 2007]). Liu and O'Halloran (2009: 372) define intersemiotic parallelism a "a cohesive relation that interconnects both language and images when the two semiotic components share a similar form". From an SFL perspective, this phenomenon is realised through the transitivity configurations presented by both language and, following Kress and van Leeuwen (2006), images (Liu and O'Halloran 2009; Royce 1998, 2007). Importantly, from a Cognitive Linguistics perspective, the shared form that characterises this echoic relation does not reside in the linguistic and visual structures of the text per se but between images and the mental imagery, in the form of conceptualisations, which both language usages and images instantiate. This view helps to address Forceville's (1999: 170) concern that SFL approaches "compare visual structures too much with surface language instead of with the mental processes of which both surface language and images are the perceptible manifestations". It also affords a multi-dimensional perspective. Since various dimensions of imagery contribute to the meaning of a linguistic expression simultaneously, e.g., in image-schematic structuring, viewpoint and attentional distribution, language and image may converge in multiple respects. That is, there are multiple sites where language and image may potentially overlay one another, giving rise to different degrees of intersemiotic parallelism. Owing to the different affordances of language and image, as well as the register conventions and genre constraints operating over any text, however, it is unlikely that language and image will converge in every possible respect. We therefore see any reduplication between the two modes as being multi-dimensional and scalar rather than absolute and, hence, prefer the term *intersemiotic convergence* which seems to more accurately capture this idea.

2 Multimodality in cognitive linguistics

Cognitive Linguistics makes several defining assumptions about language that make it particularly accommodating of multimodal analysis and which make it amenable to investigating L-I relations specifically. Following Croft and Cruse (2004), these assumptions are (i) that language is not an autonomous cognitive faculty; (ii) that linguistic structure is usage-based; and (iii) that grammar is conceptualisation. From these positions, several important corollaries arise. For example, it follows from the first assumption that the cognitive processes involved in language are not unique to language but are manifestations of more general cognitive processes found to function in other non-linguistic domains of cognition like memory, attention, perception, imagination, reason and motor execution. It further follows that the meanings evoked by linguistic expressions are conceptual in nature and that the conceptual processes which provide meaning to linguistic expressions will have analogues in other areas of cognitive experience, including vision and action. This opens up the possibility for psychologically real parallels to be drawn in our understanding of language, images and bodily movements.

The second assumption is that linguistic structure emerges, via processes of abstraction, from usage events whereby recurrent form-meaning pairings become conventionalised inside a system of symbolic units or constructions (Goldberg 1995; Langacker 1987, 1991, 2008). A usage event is defined as "a comprehensive conceptualisation, comprising an expression's full contextual understanding, paired with an elaborate vocalisation, in all its phonetic detail" (Langacker 2001: 144). In Cognitive Grammar, a symbolic unit thus consists of two poles: a phonological pole and a semantic pole. The semantic pole consists of semantic structure in the form of conceptualisations while the phonological pole consists of representations whose "essential feature is being overtly manifested, hence able to fulfil a symbolising role" (Langacker 2008: 15).

While many people in linguistics would have no problem viewing words in these terms, the radical claim of Cognitive Grammar is that all units of language are characterizable in this way and that the difference between lexical and grammatical constructions lies only in the degree of abstractness or schematicity which they encode in their semantic structure. Similarly, from this perspective, metaphorical expressions are principally no different from other forms of linguistic expression in so far as they conventionally index semantic structure in the form of conceptual metaphors. A second important claim of Cognitive Grammar is that the representations included under the rubric of phonological structures "include not only sounds but also gestures and orthographic representations" (Langacker 2008: 15). This has led researchers to develop the idea that the symbolic units or constructions

that are constitutive of language have multimodal potential and may incorporate other semiotic forms as part of their phonological pole (e.g., Kok and Cienki 2016; Steen and Turner 2013; Zima 2017). In other words, any form of expression, whether visual, manual, or auditory, that features alongside language in a usage event has the potential to become part of a multimodal construction, as represented in Figure 1 (where the forms that make up the 'phonological' pole of a construction belong to different semiotic modes). Zima and Bergs (2017) therefore argue that the usage-based model is "particularly well-equipped to unite the natural interest of linguists in the units that define language systems with the multimodality of language use". Following Goldberg's (2006: 5) criteria for linguistic constructionhood, Zima and Bergs (2017) outline two criteria for a multimodal form-meaning pairing to achieve constructional status: (1) that the non-verbal feature is used *recurrently* with a given verbal structure and its meaning contribution is "not strictly predictable" from its form; or (2) that the two forms co-occur with "sufficient frequency". Thus, as Kok and Cienki (2016: 70) state:

> whether or not elements of expression qualify as linguistic does not depend on the modality through which they are expressed. Rather the grammatical potential of co-verbal behaviours is to be assessed according to their degree of entrenchment as symbolic structures in an individual's mind and the degree of conventionalisation of those symbolic structures within a given community.

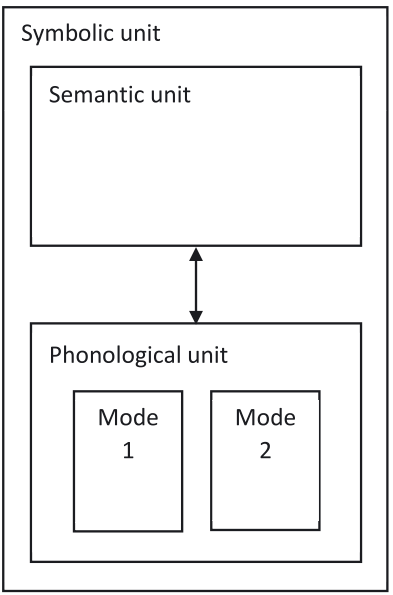

Figure 1: Multimodal construction.

An interesting line of research recently opened up considers the sociocultural level at which 'community' is defined and thus the level at which constructions may be identified. While most of the research in Construction Grammar has addressed more general constructions found at the level of a given language, constructions may also be particular to a given discourse or genre (see Antonopoulou and Nikiforidou 2011; Groom 2019). In other words, a specific discourse or genre may have its own distinct repertoire of conventionalised form-meaning pairings. And this includes multimodal form and meaning pairings. While constructions, by definition, exist at different levels of schematicity, constructions particular to, and characteristic of, a given discourse or genre will tend to be more specific in both their forms and functions.

The third assumption, which follows from the previous two, encapsulates the idea that the linguistic expression used on any occasion encodes a particular *construal* of the situation it describes by virtue of the conceptualisation it conventionally evokes. As Langacker (1991: 295) states, every linguistic structure "embodies conventional images and thus imposes a certain construal on the situation it codes". Indeed, Langacker (1991: 295) argues, it is "precisely because of their conceptual import – the contrasting images they impose – that alternate grammatical devices are commonly available to code the same situation".

The dimensions of construal along which conceptualisations may vary are described across frameworks in Cognitive Linguistics. Due to the non-autonomy of language, in many cases these have correlates in visuospatial experience (Langacker 2008; Talmy 2000). Construal operations postulated in Cognitive Linguistics include schematisation, viewpoint, distribution of attention, fictive motion, and metaphor.

What all of this means for multimodality is that the conceptualisations evoked by language share semiotic features, e.g., in spatial arrangement, perspective and salience, with visual and manual forms of representation, including images (Hart 2016). This is reflected in the extensive use that Cognitive Linguistics makes of diagrammatic notation. The diagrams found in Cognitive Linguistics, however, are not just ad hoc impressions. They are intended to capture the modal nature of meaning and the specific visuospatial properties that seem to account for semantic distinctions made by alternate forms of linguistic expression in a way that, while falling short of a mathematical formalism, is nevertheless systematic (Langacker 2008: 11). This approach to meaning receives considerable support from Simulation Semantics, which has shown experimentally that visuospatial properties of the kind posited in Cognitive Linguistic analyses do indeed form part of the meanings of linguistic expressions and that representations encoding this information get activated in online linguistic comprehension (see Bergen 2012 and Matlock and Winter 2015 for overviews). What this means for intersemiotic

convergence in particular is that, in any usage event, language and visually apprehended modes such as gesture and images may be seen to coincide in several conceptual dimensions.

Research into gesture has shown that co-speech gestures – i.e., those which are co-timed with a linguistic expression in a usage event – frequently reflect, in their shape, size, axis and direction of movement, dimensions of construal encoded by the linguistic expressions they accompany. For example, where conceptual metaphors are expressed verbally, aspects of source domain imagery are observed to receive gestural representation also (Cienki 1998; Cienki and Müller 2008). Thus, when speakers talk about 'small' versus 'large' numbers, relying on a QUANTITY IS SIZE metaphor, the metaphoric construal of magnitude is reflected, correspondingly, in the size of their co-speech gestures (Woodin et al. 2020). Similarly, in the case of aspect, the temporal unboundedness of event-conceptualisations encoded by progressive verb forms is reflected in gestures of greater duration or involving repetition (Duncan 2002; Hinnell 2018; Parrill et al. 2013). A further conceptual dimension in which language and gesture may coincide is viewpoint. For example, McNeill (1992) shows how speakers, when retelling a story, assume the perspective of either a character within the story or of an observer external to it and that this viewpoint may get reflected in both the linguistic and the gestural forms of narration/re-enactment used. In cases such as these, where gesture provides no additional information, the relation between the two modes is sometimes described in terms of *redundancy* (Abner et al. 2015). However, this may equally be characterised as intersemiotic convergence where the meanings conveyed in each mode actively reflect and reinforce the meanings conveyed by the other. Conversely, language and gesture may exist in a *supplementary* relation where information in one mode complements information given in the other by providing a particular specification, framing or perspective. For example, metaphors may be expressed in gesture where they are not co-expressed verbally (Cienki 1998). Guilbeault (2017) has shown that, in the case of viewpoint, different modes may simultaneously express *competing* perspectives. We may see this as an instance of intersemiotic divergence rather than convergence.

While multimodality in Cognitive Linguistics has primarily been taken to refer to the relationship between language and gesture (Dancygier and Vandelanotte 2017: 567), it is recognised that multimodality focussed on language and image in texts "urgently needs more detailed analysis in cognitive linguistics circles" (Dancygier and Vandelanotte 2017: 567). Analogous with co-speech gestures, co-text images in multimodal texts may interact in different ways with adjacent linguistic expressions as part of a multimodal process of meaning construction. One area where L-I relations in multimodal texts have been addressed is metaphor. For example, in his model of multimodal metaphor, Forceville (2006, 2008) has shown

how metaphors in multimodal texts like advertisements depend on an interaction between language and image such that one provides the *source* and the other provides the *target* for metaphorical mapping. It has also been shown that the same underlying conceptual metaphor may receive representation in both linguistic and visual texts. For example, El Refaie (2003) shows how the metaphors NATION IS A BUILDING and IMMIGRATION IS MOVING WATER, evidenced verbally in news reports, are also realised visually in the genre of editorial cartoons. Similarly, Koller (2005) found that illustrations in business magazines make use of the same metaphors, such as those based on a WAR or FIGHTING frame, as verbal expressions belonging to the same discourse domain. Catalano and Musolff (2019) analysed verbal and visual metaphor and metonymy in media representations of migration in the U.S and found migrants to be similarly de-humanised in both modes. A less commented on feature, however, is the multimodal reduplication of metaphors within the same text (cf. El Refaie 2015). That is, where a metaphor is realised simultaneously in both language and image to constitute a site of intersemiotic convergence.

A more recent dimension of construal that has been investigated as it occurs in visual and multimodal texts is viewpoint. For example, Dancygier and Vandelanotte (2017) show how the image schema of BARRIER may be instantiated visually (e.g., in images of walls) as well as verbally and that this image schema comes with a range of potential viewpoints that get exploited in texts in order to evoke different experiences. Thus, from one viewpoint, a barrier may be construed as an obstacle while from another viewpoint the same barrier may be construed as an object offering protection. Borkent (2017) analyses multimodal viewpoint construction in comics and shows how the asynchronicity of viewpoint cues in this genre may be exploited to create a sense of tension as the reader flits alternatively between character and narrator viewpoints. In this case, the dissonant experience may be said to arise from intersemiotic divergence in the dimension of viewpoint.

Vandelanotte and Dancygier (2017) also emphasise the role of viewpoint in their analysis of internet memes but focus on memes as multimodal constructions, as understood in various forms of construction grammar (e.g., Goldberg 1995; Langacker 1987). They analyse the different relations that images may enter into with language and thus the different contributions they may make to the overall meaning of the meme. For example, images may serve to fill in constructional roles that are left unexpressed linguistically. Conversely, images may serve to supply a constructional frame on which the meaning of the meme is contingent. In which case, language and image exist in a supplementary relation.

In what follows, we focus specifically on intersemiotic convergence, as it is realised across several dimensions of construal, and as it occurs in another multimodal text-type, namely news photographs and their captions. From a Cognitive Linguistic perspective, we understand intersemiotic convergence to refer to the co-instantiation of one or more dimensions of construal in a multimodal semiotic unit.

3 Data and scope

In the remainder of this paper, we explore some of the dimensions of construal which may be simultaneously enacted by language and image in multimodal texts to constitute sites of intersemiotic convergence. We focus our analysis on online news texts covering a variety of topics/events including immigration, political protests and inter-state conflict. News texts represent a useful source for multimodal research in Cognitive Linguistics. As Steen and Turner (2013: 13) state in relation to the audio-visual correlates of language in news texts:

> Cognitive linguists routinely study basic mental operations and phenomena that are not exclusive to language but that are deployed in language and leave their mark on its structure … Since the news deploys other modalities than speech and text, it is an obvious project to look for the ways in which these basic mental operations and phenomena are deployed in those other modalities. (Steen and Turner 2013: 13)

Analogous with the term *co-speech gesture*, we focus specifically on what might be termed *co-text images*. Co-text images are images which, owing to their proximity with particular language usages in the organisation of a text, are likely to be viewed together with those language usages as part of a single semiotic unit or syntagm. In news texts, the prototypical example of this, and where we concentrate the majority of our analyses, is news photographs and their captions, which, surprisingly, have not received detailed treatments in multimodality studies (Bateman 2014: 251) (cf. Bednarek and Caple 2012: 126–127; Hart 2017).[2] Although no quantitative analysis is presented, again analogous with Cognitive Linguistic research into gesture, we note the possibility for recurrent L-I combinations to hold or attain constructional status with linguistic and visual forms each being represented at the phonological pole of a multimodal symbolic unit (Steen and Turner 2013). We also note that a tendency for language and image to converge in the dimensions postulated may be taken as evidence for the hypothesis that linguistic expressions encode visuospatial properties as part of their semantic values. What we hope to do in this paper is to lay the foundations for the kind of cross-modal coding scheme, based in Cognitive Linguistics, that could be used to address questions such as these in a future large-scale corpus-based study (cf. Carter and Adolphs 2008).

2 L-I relations between news photographs and their captions have been addressed in analyses of news texts reporting migration (Crespo Fernandez and Martínez Lirola, 2012; Martínez Lirola and Zammit, 2017). These works, however, are focused on the ideological implications of L-I relations rather than on their cognitive basis.

Specifically, in what follows, we treat four dimensions of construal: schematisation, viewpoint, windowing of attention and metaphor. These are by no means the only dimensions of construal that are likely to be reflected multimodally but since they have been the subject of extensive research in Cognitive Linguistic studies of co-speech gesture (see Cienki 2013 for an overview) they provide a natural starting point for investigating L-I relations in multimodal texts. It is not the case that on any given occasion language and image will necessarily overlay one another in every aspect of construal. As Zima and Bergs (2017) note in relation to co-speech gestures, "gesture-speech combinations fall on a continuum with respect to the semantic overlap between the two modalities. They range from co-expressive to cases in which very disparate meanings are expressed in [each mode]". Indeed, it is perfectly possible, as Steen and Turner (2013: 17–18) point out, for language and image to exist in a divergent rather than convergent relationship whereby information provided in one mode may either be supplementary to information provided in the other or else may contradict it in order to create some discordant effect for purposes such as humour, satire, ambivalence etc. For example, the L-I combination in (1) displays divergence in the dimension of schematisation where the event, as construed in the verbal mode, is one of static motion (Talmy 2000) in which migrants, indexed by 'wait', simply occupy a location in space while, conversely, in the visual mode, the image instantiates a force-dynamic construal (Talmy 2000) in which immigrants are confined to a given location by a second interactant in the form of a fence.

(1)

Migrants wait nearby the entrance of Hungarian transit zone near Roszke village on May 31.
Telegraph.co.uk, 13 June 2016

In the case of news texts, however, and particularly in the case of news photographs and their captions, language and image are expected to more often be intersemiotically convergent. From a rhetorical standpoint, in this discursive context, intersemiotic convergence serves to tell the same story, from the same perspective. In other words, it serves to maintain a consistent narrative, with the version of events presented in each semiotic mode corroborating the version given in the other. In our examples, language and image may therefore be observed to overlap in several dimensions of construal. Certainly it is the case that, in any one example, multiple dimensions of construal are operating concomitantly with one another. For example, schematisation necessarily involves a viewpoint (Dancygier and Vandelanotte 2017; Langacker 2008). Similarly, attentional selection and distribution is a necessary feature of all conceptions (Talmy 2000; Langacker 2008). However, for purposes of exposition, we isolate different dimensions of construal and structure our analysis along these lines. Where particularly pertinent to the example, we acknowledge the contribution of multiple conceptual dimensions to its overall meaning.

4 Dimensions of construal as sites of intersemiotic convergence

4.1 Schematisation

Arguably the most fundamental dimension of conceptualisation lies in image-schematic representations of event-structure (Langacker 2008; Talmy 2000). Image schemas stand as *archetypal conceptions* representing basic patterns of experience (Langacker 2008: 355). Their role in structuring knowledge within semantic domains like action, force, space and motion, where they serve as folk theories of the way the world works, has been widely studied in Cognitive Linguistics (Hampe 2005; Johnson 1987; Oakley 2010). A key claim of Cognitive Linguistics is that such schemas "work their way up in to our system of meaning" (Johnson 1987: 42) to constitute the semantic basis of linguistic – lexical and grammatical – forms. One pervasive pattern of experience, for example, is an interaction involving the transference of energy, through forceful physical contact, from one participant to another. This type of interaction is represented by an *action chain* schema which, Langacker (2002, 2008) argues, forms the conceptual basis of the prototypical transitive clause. Talmy (2000) has similarly argued that

image schemas representing various types of force and motion event are encoded in the meanings of both open and closed class linguistic elements. In discourse, the linguistic form selected to describe an event is not determined by any objective properties of the referential situation but rather invites a particular conceptualisation of it. Image-schematic structuring through language is thus a matter of construal. The same or ostensibly the same type of material situation may be schematised differently through alternate linguistic formulations. For example, in the context of media discourse on political protests, Hart (2013) has shown that violent interactions between police and protesters can be schematised as action, force or motion events depending on the ideological perspective of the text-producer and that within these domains alternate schemas are available to further construe the same situation in different, ideologically vested ways. Crucially, in terms of L-I relations and intersemiotic convergence, the same image-schematic patterning may underpin co-text images. Take, as a first example, the image-caption combination in (2). In this usage event, language and image converge in instantiating the three-participant action chain represented in Figure 2. Different types of action and event are associated with different *archetypal roles* (Langacker 2008: 356). A three-participant action chain represents a transfer of energy from an AGENT to a PATIENT via an INSTRUMENT. In (1), this schema is instantiated in the ditransitive construction where the agent is encoded as subject, the instrument as direct object and the patient as indirect object. The same event-structure is seen in the image where all three participants are represented and the interaction between them is suggested by the dynamicity of the image. For Kress and van Leeuwen (2006), this dynamicity is what gives the image its 'narrative' structure and is a product of vectors formed by visually depicted elements. In (2), such a vector is formed by the outstretched limb of the refugee. The continued trajectory of this 'effector' implies the bottle's direction of travel toward the police and thus the sequence of energy flow between participants. The vector formed in the image may therefore correspond with the arrows representing energy transference in Figure 2. From a perspective more akin with Simulation Semantics, it has been shown that static photographs of human actions where there is implied motion activate motor areas of the brain (Kim and Blake 2007; Kourtzi and Kanwisher 2000; Proverbiaet al. 2009). This suggests that in understanding images viewers 'complete the picture' by performing dynamic simulations which unfold along the lines laid down by corresponding image schemas.

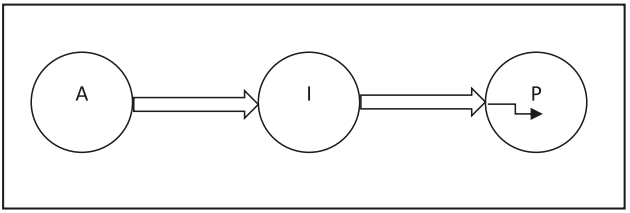

Figure 2: Three-participant action chain.

(2)

A refugee throws a bottle toward Hungarian police at the "Horgos 2" border crossing into Hungary, near Horgos, Serbia.
Telegraph.co.uk, 22 September 2015

In (2), the archetypal conception instantiated in both language and image is a one-tailed action schema in which the transfer of energy flows unidirectionally from an agent as the energy *source* to a patient as the energy *sink*. However, another archetypal conception represents a bi-directional exchange of energy between two participants who are both agentive and thus simultaneously act as energy source and energy sink in the interaction. This two-tailed action schema, represented in Figure 3, forms the meaningful basis of reciprocal verbs. In discourse, it serves to construe a physical interaction as two-sided and thus assigns mutual blame and responsibility for the event (Hart 2018). In example (3), this schema is instantiated linguistically by the reciprocal verb 'clash' but also visually by the co-text image. In contrast to the image in (2), in which only one participant is depicted in an 'action shot' while the second participant is more passive, in (3) both sets of participants are shown engaged in the mutual transfer of force.

There is also a clear contrast in viewpoint between examples (2) and (3) with the viewpoint in (2), realised both verbally and visually, being from that of the agent of the action and the viewpoint in (3), again realised both verbally and visually, being from that of an observer (see Section 4.2).

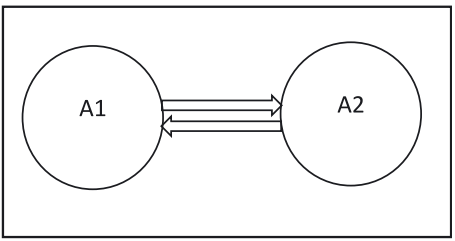

Figure 3: Two-tailed action schema.

(3)

Police and protesters clash in Oxford Circus.
Telegraph.co.uk, 27 March 2011

In the domain of motion, Talmy (2000) identifies several different types of motion event. However, a basic distinction that can be made is between motion events which are force-dynamically neutral and those that involve a force-dynamic component. In a force-dynamically neutral event, the impetus for motion begins with the agent and their ability to move is not hindered in any way. Motion events that involve a force-dynamic component include caused motion and impeded motion. In impeded motion, the agent's ability to move freely is constrained by the presence of some 'barrier' which they are able to circumvent or penetrate in order to complete the intended translocation.[3] These two types of impeded motion event are represented by image schemas such as modelled in Figure 4.[4] The arrows in Figure 4a and 4b represent a path of motion rather than a transfer of energy and the stepped arrow in Figure 4b represents the change in state to the barrier brought about in the course of realising the

3 Talmy (2000) uses the terms 'agonist' and 'antagonist' to refer to these two types of force-interacting entity.

4 These diagrams are based on Johnson (1987). See Talmy (2000) for an alternative form of diagrammatic notation.

event. Many linguistic expressions, both open and closed class, include a force-dynamic conceptualisation as part of their meaning (Talmy 2000). This includes the *try* + *infinitive* construction which focuses on an effort to overcome an obstacle (without making known the outcome of this effort) (Talmy 2000: 436–437).[5]

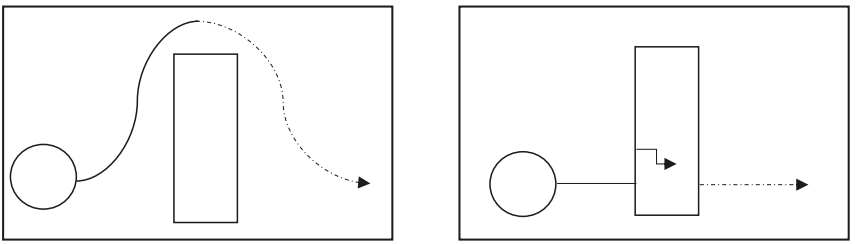

Figure 4: Impeded motion schemas.

In news texts reporting immigration, we find L-I combinations such as (4) where an impeded motion schema is instantiated linguistically by forms like 'trying to reach' and 'attempting to get to' but also visually by the image of a border fence, which immigrants are shown climbing through or over.

(4)

Migrants in Calais attempting to get to Britain.
Express.co.uk 15 August 2015

5 This is in contrast to, say, *manage* which also encodes knowledge of the outcome (Talmy 2000: 436–437).

These L-I combinations may thus be said to converge in schematising immigration in force-dynamic terms. They are candidates for a discourse-level construction where, in the visual mode, the BARRIER element of the impeded motion schema is conventionally instantiated by images of fences – a recurrent visual trope in immigration discourse (see Dancygier and Vandelonotte 2017 for discussion of visual instantiations of the BARRIER schema). However, there is a difference between the two modes in levels of specificity. While in the language, the form or nature of the impediment is not specified and the manner by which it is overcome (e.g., circumvented vs. penetrated) is not expressed, this information is contained within the co-text image. Such L-I combinations may therefore be described as exhibiting a hyponymic relation whereby the image instantiates a more specific type of impeded motion schema. In this sense, while convergent at a basic level of schematisation, the image in (4) may also be said to supplement information expressed in its verbal co-text. The schema instantiated by the image in (4) is the one represented in Figure 4b. Of course, intersemiotic convergence is not limited to images and their captions but may also be found between images and conceptualisations evoked by other regional and functional parts of the text (Bednarek and Caple 2012: 96–100). In the case of (4), for example, the headline of the article instantiated the same schema as the image: "Migrants trying to 'break into' Britain".

Dancygier and Vandelanotte (2017) argue that image schemas like the BARRIER schema include as part of their meaning viewpoint affordances whose realisation in usage events yields different experiential results with clear emotional consequences. The viewpoint in (4), instantiated multimodally, construes the barrier from the perspective of 'us' on the other side of the barrier to the migrants.

Another aspect of schematisation concerns not the event-structure itself but the participants within it. Talmy (2000) suggests that distinctions within the linguistic category of number, i.e., singular versus plural (as well as aspectual distinctions like semelfactive versus iterative), are accounted for conceptually in terms of plexity of structure (see Figure 5).

For Talmy (2000: 48), plexity is "a quantity's state of articulation into equivalent elements". While singular nouns specify a uniplex referent, plural nouns

Figure 5: Multiplex versus uniplex.

specify a multiplex referent. The construal evoked by examples like (3) and (4) is therefore one in which the referents are treated as multiplex. However, certain nominal forms, including collective nouns and noun phrases, construe multiplex referents as uniplex. In opposite relation to what Talmy describes as *multiplexing*, this may be said to represent a cognitive operation of *uniplexing*. It is found in example (5) where the collective NP 'column of migrants' construes the group of immigrants being referred to as a uniplex structure of a specific oblong shape. In images, plexity is realised in the dispersion of, or degree of agglomeration of, visually depicted elements. The higher the degree of agglomeration (and therefore lower dispersion), the more uniplex the structure. In (5), language and image are therefore intersemiotically convergent in the dimensions of plexity and shape where the co-text image displays a high degree of agglomeration such that the individual migrants come to form a uniplex structure that is similarly oblong in shape. Rhetorically, this does several things which are worth commenting on. In both modes, immigrants are aggregated or de-individuated so that their own personal stories are not recognised and they can all be viewed and treated in the same way. It is also worth noting that the image in (5) is unbounded – other than by the frame of the photo. A structure that is unbounded is conceived as "continuing on indefinitely with no necessary characteristic of finiteness intrinsic to it" (Talmy 2000: 50). In this context, the unboundedness of the image functions rhetorically to suggest a 'column' of significant magnitude.

(5)

The huge column of migrants passes through fields in Rigonce, Slovenia, after having been held at the Croatia border for several days.
MailOnline, 25 October 2015

Plexity is related to viewpoint where the further away one is from a given scene, the more uniplex that scene becomes (see Hart 2015). Hence, the camera angle presented by (5) is that of an aerial shot. We turn to viewpoint in the proceeding section.

4.2 Viewpoint

If schematisation represents one dimension of construal, in so far as it involves the apprehension of particular conceptual content in order to conceptualise the structural properties of the referential situation, further dimensions of construal concern the way in which that conceptual content is viewed. As Langacker (2008: 55) states:

> An expression's meaning is not just the conceptual content it evokes – equally important is how that content is construed. As part of its conventionalised semantic value, every symbolic structure construes its content in a certain fashion … In viewing a scene, what we actually see depends on how closely we examine it, what we choose to look at, which elements we pay most attention to, and were we view it from.

Other dimensions of construal, then, are based in cognitive systems of perspective and attention (Croft and Cruse 2004; Langacker 2008; Talmy 2000). In terms of perspective, a key claim of Cognitive Grammar is that "many expressions undeniably invoke a vantage point as part of their meaning" (Langacker 2008: 75). Zwaan (2004) argues that in comprehending an utterance, hearers are immersed experiencers who simulate the scene described from a particular linguistically cued perspective. On the basis of these claims, Hart (2015) proposes an embodied 'grammar' of viewpoint modelled in three dimensions – anchor, angle and distance – and argues that the meanings of alternate linguistic expressions can be characterised, in part, as a shift in one or other of these aspects and that linguistic forms which include a viewpoint specification as part of their meanings therefore have analogues in images, which, by the hypothesis, share the same perspective. According to this model, collective nouns such as found in (5) combine vertical angle with maximal distance to engender a construal analogous to the aerial shot found in the image of (5). On the anchor plane, where a viewpoint shift equates to *panning*, transitive versus reciprocal verb constructions are analysed as not only indexing alternative image schemas but also as including contrasting viewpoints as part of their semantic values. On this hypothesis, transitive constructions encode a viewpoint which construes the event sagittally from the perspective of the agent or the patient depending on voice while reciprocal constructions encode a viewpoint which construes the event transversally from a perspective that is orthogonal to that of participants within it and which is equidistant between them (Hart 2015). The contrasting *viewing arrangements* (Langacker 2008) that result are represented in Figure 6.

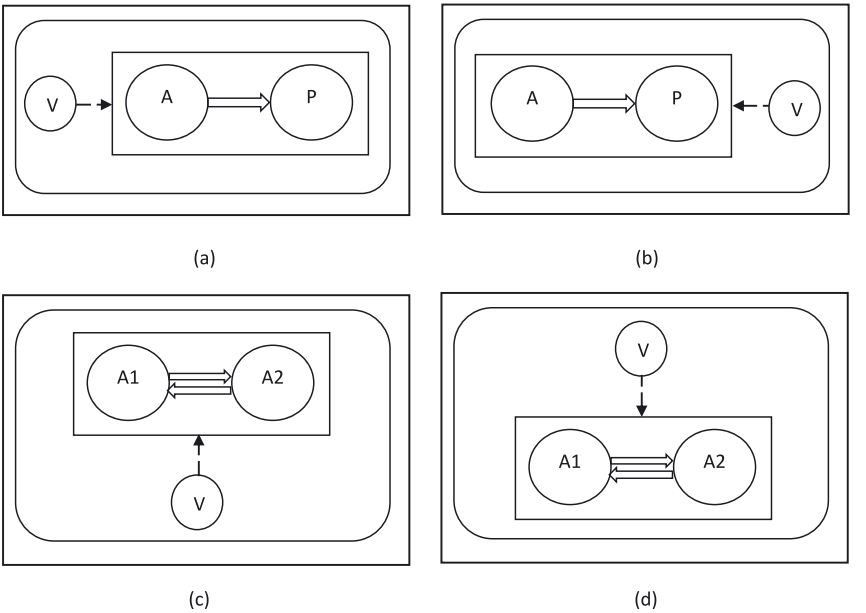

Figure 6: Viewing arrangements in the anchor plane.

This hypothesis is borne out experimentally. In a sentence-image matching task, Hart (2019) gave subjects transitive and reciprocal verb constructions and asked them to indicate which schematic image, presented in four orientations, best represented the situation described in each of the sentence types. The results showed a significant level of agreement between subjects where, for transitive constructions, subjects converged on sagittal images while, for reciprocal constructions, they converged on transversal images. The study therefore demonstrates that viewpoint is indeed a meaningful aspect of at least these linguistic expressions. Accordingly, viewpoint constitutes a potential site of intersemiotic convergence in multimodal texts which, in the context of news texts, we should expect to be realised. This is the case, for example, in (2) and (3). In (2), the ditransitive construction occurs with a co-text image in which the vector representing the action unfolds along the sagittal axis observed from the perspective of the agent. By contrast, in (3), the reciprocal construction occurs with a co-text image whose viewpoint locates the vectors representing the action along the transversal axis. Thus, in (2), language and image converge in instantiating the

viewing arrangement in Figure 6a while in (3) they converge in instantiating the viewing arrangement in Figure 6c or 6d.[6]

A further finding of Hart (2019) was that information structure in reciprocal constructions is associated iconically with contrasting left-right arrangements on the transversal axis and thus with opposite viewpoints. Thus, if we assign police to Agent 1 and protesters to Agent 2, the viewing arrangement instantiated by the L-I combination in (3) is that of Figure 6c. By contrast, in (6) below, the L-I combination presents a multimodal instantiation of the viewing arrangement in Figure 6d.[7] On this analysis, then, the L-I combinations presented in (2), (3) and (6) converge not only in their basic schematic patterning but also in viewpoint.

(6)

Protesters waving placards clash with police at a pro-migration demonstration in central London.
MailOnline, 19 March 2016

The motivations and functions of different viewing arrangements are discussed in detail by Hart (2015). But further support for the claim that viewpoint is a meaningful feature of these linguistic expressions comes from the consistent way that language usages instantiating these constructions and images which, by the hypothesis, share a common viewpoint behave in eliciting textual effects. Hart (2018,

6 The *X CLASH with Y + transversal image* combination seems a particularly strong candidate for multimodal constructional status at the level of language more generally. It is also reflected in gesture where, intuitively, the clap-like gesture that would accompany spoken instances occurs on the transversal axis.

7 Images will not always necessarily correspond to cardinal viewpoints but, as in (5), will normally approximate one or other value.

2019) tested the effects of transitive versus reciprocal verbs and images where the only relevant variable is viewpoint on blame assignment and perception of aggression. Reciprocal verbs and transversal images both invite more equal distribution of blame than their opposites in transitive verbs and sagittal images. Within reciprocal constructions and transversal images, participants are judged as more aggressive when they appear left in the organisation of the clause and the image. This is consistent with Casasanto's finding that, for the majority of people (i.e., right-handers), leftward space is associated with negative valence. It may also be explained by an association between leftward space and agency in both language and image that is, in turn, based in the left-right conceptualisation of the timeline along which events unfold.

The area of language most recognised for encoding viewpoint is, of course, deixis. The viewpoint encoded by deictic expressions represents the situatedness of interlocutors in space and time. Thus, in spatial deixis, *coming* is used to describe movement toward a destination where at least one interlocutor is located while *going* is used to describe movement toward a destination where neither interlocutor is located. In the context of migration discourse, where destination countries are construed as containers (Charteris-Black 2006; Chilton 2004), deictic expressions are typically of the 'coming' variety and may thus be described as encoding a perspective from inside of the container. Although more background knowledge and pragmatic inferencing is required to determine the viewpoint, this vantage point interior perspective may also be encoded in co-text images. Example (7) is an interesting example in which the conceptualisation evoked by the lead paragraph is instantiated visually across two consecutive images in the text (each with their own captions). When read narratively as a visual 'subject' + 'predicate', the two images fulfil the dynamic script-like structure of the conceptualisation evoked by the linguistic expression in the lead paragraph, consisting of a motion schema + viewpoint interior perspective as represented in Figure 7. The first image depicts the agent or FIGURE (Talmy 2000) undergoing the motion ('children as young as six years old'). The rest of the script is fulfilled in the second image. Motion itself and the manner of motion are instantiated in the image of a train, which implies mode of transport. And although the train's direction of travel and thus PATH is not made explicit, the image depicts the GROUND ('the channel tunnel') from a perspectival location inside of the UK, which coincides with the presumed deictic point of reference encoded by 'coming' in the linguistic co-text. When read in conjunction with the linguistic co-text, the train's simulated direction of travel is therefore likely to be emerging out of the tunnel and into the UK. In this sense, while intersemiotically convergent in some aspects of meaning, language and image also exist in a supplementary relation.

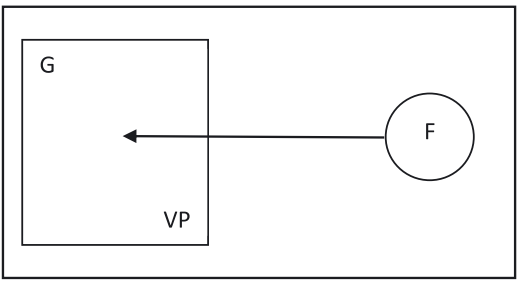

Figure 7: Motion schema exterior to interior + viewpoint interior perspective.

(7)

Children as young as six years old are coming through the Channel Tunnel without parents or guardians and claiming asylum in Britain, it was claimed today.
MailOnline, 13 April 2016

4.3 Windowing of attention

Viewpoint is linked to attention in so far as it defines which aspects of a scene are in the foreground and which are in the background (Chilton 2014; Talmy 2000). L-I combinations therefore often converge in both viewpoint and attentional configuration. One aspect of attentional configuration resides in what Talmy (2000) calls *windowing* which has its reflex in *gapping*.[8] According to Talmy (2000: 257), language places a portion of a wider image schema or *event-frame*, representing a coherent referential situation, into the foreground of attention by virtue of explicit mention. Windowing of attention has clear analogues in visual forms of representation where images, defined by the scope of their viewing frame, necessarily only capture a part of the events they document, representing a particular snapshot in time and space. Intersemiotic convergence in the dimension of attention

8 'Windowing of attention' (Talmy 2000) and 'profiling' (Langacker 2008) capture the same phenomenon. We prefer the term 'windowing of attention' because of its resonance with the notion of a viewing frame in multimodality.

may therefore occur where co-text images capture the same portion of an event that gets windowed in the language usages they accompany. To illustrate this, let us consider the example in (8). The viewpoint in both the language and the image is from the perspective of the Palestinian protesters. The event they are 'witnessing' is an instance of what Talmy (2000: 265) calls an *open path* event. An open path event is an event involving an object physically in motion in the course of a period of time that is "conceptualised as an entire unity thus having a beginning and an end, and whose beginning point and ending point are at different locations in space" (Talmy 2000: 265). Path windowing occurs when language directs attention over different facets of the conceptually complete path. Talmy identifies three forms of windowing that may be imposed on different regions of the path: initial, medial and final. This is represented in Figure 8. In (8), the object in motion is tear gas canisters. Such an event-type involves a launch site and a landing site. The beginning and end points of the path, however, do not receive linguistic representation. Instead, the nominalised form 'falling tear gas canisters' is an example of medial path windowing which leaves the beginning and end points of the event-frame attentionally backgrounded. This is represented in Figure 7b. Likewise, in the image we see only the tear gas canisters as they are moving through the air and do not see from where they emanated or where they end up. The multimodal representation is thus convergent in distribution of attention as well as perpsective. A contrasting image, involving final path windowing, would be one capturing the moment of impact. Of course, as Talmy (2000: 266) points out, given sufficient context, we can mentally trace the whole path to reconstruct or complete it, but it is only the medial path portion that is foregrounded for attention in both the language and the image. Thus, in (8), language and image converge in instantiating the conceptualisation represented in Figure 8b.

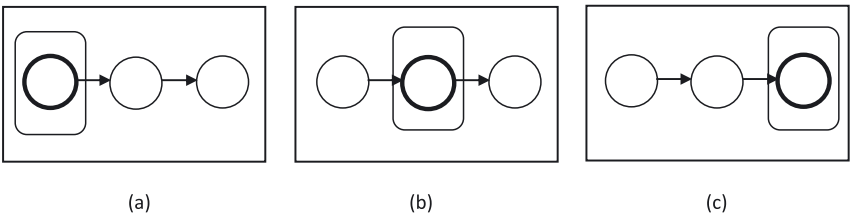

Figure 8: Path windowing.

(8)

Palestinian protesters look up at falling tear gas canisters near the border with Israel in the southern Gaza Strip on Tuesday.
Wall Street Journal, 14 May 2018

We also find intersemiotic convergence in path windowing in the context of immigration discourse. Immigration, similarly, is an instance of an open path event in which migrants depart a country of origin and end up at a destination country. L-I combinations can converge in directing attention over particular aspects of this process. Compare examples (9) and (10). In (9), the linguistic expression ('crossing the Channel') and the co-text image both involve medial path windowing. In the image we are not shown the beginning or the endpoint of the journey (though the language suggests the intended destination). By contrast, in (10), both the language and the image involve final path windowing. Final path windowing is an inherent semantic feature of the verb *arrive*. In the image, final path windowing occurs where migrants are shown exiting a boat at a shoreline representing the terminus of their journey. The L-I combination in (10) is thus intersemiotically convergent in instantiating the conceptualisation represented in Figure 8c. Of course, in the image, the shoreline could be any shoreline but it is specified in the co-text as being that of the UK. Conversely, while there is nothing to indicate viewpoint in the linguistic expression, the image specifies a viewpoint interior perspective, from the land side of the shoreline, which, when interpreted together, lends the linguistic expression a deictic quality. In this sense, while intersemiotically convergent in windowing of attention, language and image in (10) are supplementary in respect of viewpoint.

(9)

A group of migrants crossing The Channel in a small boat headed in the direction of Dover, Kent, on 10 August.
Independent.co.uk, 10 August 2020

(10)

The number of child migrants arriving to the UK to claim asylum has rocketed.
Immigrationnews.co.uk, 23 June 2020

4.4 Metaphor

The final dimension of construal we will address is metaphor. Metaphor is a central topic in Cognitive Linguistics and is the construal operation that has been most widely investigated in multimodal research. Metaphor, as a conceptual process that is instantiated in and evoked by metaphorical expressions, involves the apprehension of a source frame to provide a template for sense-making inside a target frame (Lakoff and Johnson 1980). Based on Lakoff and Johnson's (1980: 153) claim that metaphor is "primarily a matter of thought and action and only derivatively a matter of language", Forceville (2006: 381) argues that metaphors should be expected to "occur non-verbally and multimodally as well as verbally". And indeed, many metaphors that find linguistic expression in a given discourse are also found to receive visual forms of representation elsewhere within the same discourse (e.g., El Refaie 2003; Fridolfsson 2008; Koller 2005, 2009). In multimodal metaphor theory, the focus of attention has been on articulations of metaphors where source and target frames are represented separately in different semiotic modes which work together in realising the metaphor (Forceville 2006, 2008). Comparatively less attention has been given to the co-articulation of metaphors across semiotic modes within multimodal texts. That is, to metaphor as a potential site of intersemiotic convergence (cf. O'Halloran 1999). As gesture research has shown, however, metaphors may be expressed simultaneously in more than one mode (Cienki and Müller 2008). Of course, given the affordances of different modes, even when expressing the same metaphor, there are likely to be differences between them. For example, in degree of specificity or in the extent to which the metaphorical interpretation is forced. In relation to the latter, we can distinguish at least two types of intersemiotic convergence in metaphor. In the first type, verbal and visual modes are fully convergent in so far as the basic underlying metaphor is recoverable from each mode independently of the other. In the second type, images are consistent with the metaphorical framing presented verbally but their *potential metaphoric reading* is unlikely to be realised in the absence of a verbal co-text metaphor. For example, images of migrants contained by fences such as found in (1) have a potential metaphorical reading in which migrants are construed as caged animals. However, such a reading is unlikely without a verbal instantiation of the metaphor being co-present. Language may therefore serve to highlight or downplay the potential metaphoricity of images. In other words, language may have a *metaphor anchoring effect* (cf. Barthes 1977). Where images have a potential metaphoric reading that is consistent with a metaphoric framing

presented verbally, we refer to them as *frame-consistent images*.[9] It is important to note that these categories are not absolute or discrete and that the experience of images as metaphorical or not will depend on individual subjectivities. Let us consider some examples to illustrate.

(11)


THE TIMES

SUBSCRIBE LOG IN

VIDEO

Paris becomes a war zone as police clash with protesters

Demonstrations against high taxes and falling living standards erupted into the worst violence yet as officers battled rioters.
By Peter Conradi, in Paris

Teargas was fired as rioters in masks and helmets hijacked what had been a peaceful protest yesterday
ALAIN JOCARD/AFP/GETTY

The Sunday Times, December 2 2018, 12:01am

Dozens of cars were set on fire and shops and restaurants looted as a protest by more than 10,000 members of the gilets jaunes — yellow vests — movement turned the centre of the French capital into a war zone last night.

The Sunday Times, 2 December 2018

In the multimodal text given as example (11), reporting on the *gilets juanes* (yellow vests) protests in Paris, the metaphor PROTEST IS WAR is strung throughout verbal portions of the text, realised directly by repeated descriptions of Paris as a 'war zone' as well as indirectly by the description of police 'battling' protesters. The same metaphor is also expressed in the image but with a greater degree of specificity. While the WAR frame evoked in the verbal portions of the text is a generic WAR frame, the frame evoked by the image represents a specific historic event, namely the Second French Revolution of July 1830. The metaphoricity of

9 Images can also be consistent with verbally presented metaphorical frames without having any potential metaphoric reading where visually depicted elements instantiate particular aspects of the source frame. For example, linguistic expressions of a metaphor IMMIGRATION IS FLOODING may be accompanied by images containing water such as those found in (9) and (10).

the image is achieved via the intertextual reference it makes to Eugène Delacroix's famous painting *Liberty Leading the People* which, produced to commemorate the events of July 1830, shows Marianne, a national symbol of the French Republic, personifying the Goddess of Liberty. The intertextually referenced image provides an *access point* to the frame it instantiates which is then brought to bear in the interpretation of the current image. Of course, where intertextuality is a vehicle for metaphor, the metaphorical interpretation of the image depends on the reader having the requisite background knowledge to recover the intertextual reference (Werner 2004).

In (11), the metaphoric reading of the image is possible based on the image alone. As an instance where a potential metaphoric construal is dependent on linguistic co-text, reconsider example (5). Since 'column of X' is the designation for a group of soldiers and, by metaphorical extension, a group of ants, the caption in (5) may be analysed as expressing one of two metaphors: IMMIGRANTS ARE SOLDIERS or IMMIGRANTS ARE INSECTS. Both are well documented metaphors in media discourses of immigration (e.g., Hart 2010; Santa Ana 1999). The image in (5) is potentially consistent with the imagery of both of these metaphors. While in (11) it is the *content* of the image that resembles another specific image, in (5) the image bears a *structural resemblance*[10] to typical images of invading armies and insects which may thus be interdiscursively rather than intertextually brought to bear in interpreting the current image. The metaphoric construal of both language and image in (5) is therefore likely to be determined by metaphorical expressions in other prominent regions of the text which perform a frame-setting function. Here we find that the headline of the text in which (5) is embedded contains a militarising metaphor:

(12) On the march to western Europe: Shocking pictures show thousands of determined men, women and children trudging across the Balkans as politicians warn EU could collapse in weeks. (*MailOnline*, 25 October 2015)

Thus, the L-I combinations in both (11) and (5) present instances of intersemiotic convergence in the dimension of metaphor with WAR providing the source frame in each example. The images in each case, though, do not fit neatly within Forceville's (2008) classification of pictorial metaphor as *contextual* or *hybrid*. Rather, both instances represent a third type of pictorial metaphor (*holistic*) where it is the image as a whole that is reminiscent of another iconic image or type of image that belongs to a different context (Hart 2017).

10 As revealed by a Google image search for 'column of soldiers' or 'column of ants'.

5 Conclusions

Across theories in Cognitive Linguistics, various dimensions of imagery are posited as providing meaning to linguistic expressions as part of the conceptualisations they conventionally evoke. Many of these dimensions of conceptualisation have a basis in visuospatial experience. In this paper, we have sought to demonstrate the utility of Cognitive Linguistics in general as a framework for investigating intersemiotic relations between language and image in multimodal texts, focusing on intersemiotic convergence in particular. We have explored the quite natural hypothesis, which emerges from the modal account of meaning given in Cognitive Linguistic analyses, that the conceptualisations evoked by linguistic expressions have semiotic features in common with visual forms of representation and that consequently the dimensions of conceptualisation proposed in Cognitive Linguistics exist as potential sites of intersemiotic convergence between language usages and co-text images in multimodal texts. In the context of news discourse, we have explored intersemiotic convergence in four aspects of conceptualisation: schematisation, viewpoint, distribution of attention and metaphor. Although in our analyses we isolated these for purposes of exposition, the account we offer is multidimensional with language and image having the potential to coincide in several dimensions of construal simultaneously.

What we hope to have achieved is a programmatic paper which (a) responds to calls for Cognitive Linguistics to address issues of multimodality beyond the language/gesture interface and (b) invites further research into L-I relations within Cognitive Linguistics. We have focussed on the context of news discourse and so our findings are necessarily genre-specific. Similarly, we have focussed solely on intersemiotic convergence. However, there are myriad ways that the conceptualisations instantiated by language and image in other text-types may diverge from one another to achieve various kinds of textual effect. This is an area that needs investigation. Indeed, the textual effects, for example, on memory or event-perception, of different degrees of intersemiotic convergence is something that requires experimental investigation. We have also ignored temporal and aspectual dimensions of meaning. Although in principle these could be addressed in static texts, they are perhaps more amenable to analysis in dynamic texts such as TV news stories. Indeed, moving texts of the kind found in TV corpora offer a further exciting data type for multimodal Cognitive Linguistic research (Steen and Turner 2013). Finally, although no quantitative analysis has been presented, a crucial question concerns the extent to which certain L-I combinations found in specific usage events hold multimodal constructional status within a given language or discourse. Moreover, regular co-occurrence of language usages and

images which, on the analyses presented, are congruent would constitute a form of evidence for those analyses. We hope to have provided the beginnings of a framework for analysing L-I combinations and instances of semiotic convergence which can be used to address such quantitative questions in future research.

Data availability statement

All data generated or analysed during this study are included in this published article.

References

Abner, Natasha, Kensy Cooperrider & Susan Goldin-Meadow. 2015. Gesture for linguists: A handy primer. *Language and Linguistics Compass* 9(11). 437–451.

Antonopoulou, Elini & Kiki Nikiforidou. 2011. Construction grammar and conventional discourse: A construction-based approach to discoursal incongruity. *Journal of Pragmatics* 43. 2594–2609.

Barthes, Roland. 1977. *Image music text*. London: Fontana Press.

Bateman, John A. 2014. *Text and image: A critical introduction to the visual/verbal divide*. London: Routledge.

Bednarek, Monika & Helen Caple. 2012. *News discourse*. London: Bloomsbury.

Bergen, Benjamin. 2012. *Louder than words: The new of science of how the mind makes meaning*. New York: Basic Books.

Borkent, Mike. 2017. Mediated characters: Multimodal viewpoint construction in comics. *Cognitive Linguistics* 28(3). 539–563.

Carter, Ronald & Svenja Adolphs. 2008. Linking the verbal and the visual: New directions for corpus linguistics. In Andrea Gerbig & Oliver Mason (eds.), *Language, people, numbers: Corpus linguistics and society*, 275–291. Amsterdam: Rodopi.

Catalano, Theresa & Andreas Musolff. 2019. "Taking the shackles off": Metaphor and metonymy of migrant children and border officials in the. *U.S. Metaphorik* 29. 11–46.

Charteris-Black, Jonathan. 2006. Britain as a container: Immigration metaphors in the 2005 election campaign. *Discourse & Society* 17(5). 563–581.

Chilton, Paul. 2004. *Analysing political discourse: Theory and practice*. London: Routledge.

Chilton, Paul. 2014. *Language, space and mind: The conceptual geometry of linguistic meaning*. Cambridge: Cambridge University Press.

Cienki, Alan. 1998. Metaphoric gestures and some of their relations to verbal metaphoric expressions. In Jean-Pierre Koenig (ed.), *Discourse and cognition: Bridging the gap*, 189–204. Stanford, CA: Center for the Study of Language and Information.

Cienki, Alan. 2013. Cognitive Linguistics: Spoken language and gesture as expressions of conceptualization. In Cornelia Müller, Alan J. Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill & Sedinha Teßendorf (eds.), *Body - language - communication, volume 1: An*

international handbook on multimodality in human interaction, 182–201. Berlin: Mouton de Gruyter.
Cienki, Alan & Cornelia Müller (eds.). 2008. *Metaphor and gesture*. Amsterdam: John Benjamins.
Crespo-Fernandez, Eliecer & María Martinez Lirola. 2012. Lexical and visual choices in the representation of immigration in the Spanish press. *Spanish in Context* 9(1). 27–57.
Croft, William & D. Alan Cruse. 2004. *Cognitive linguistics*. Cambridge: Cambridge University Press.
Dancygier, Barbara & Lieven Vandelanotte. 2017. Internet memes as multimodal constructions. *Cognitive Linguistics* 28(3). 565–598.
Duncan, Susan D. 2002. Gesture, verb aspect, and the nature of iconic imagery in natural discourse. *Gesture* 2(2). 183–206.
El Refaie, Elisabeth. 2003. Understanding visual metaphor: The example of newspaper cartoons. *Visual Communication* 2(1). 75–96.
El Refaie, Elisabeth. 2015. Cross-modal resonances in creative multimodal metaphors: Breaking out of conceptual prisons. In María J. Pinar Sanz (ed.), *Multimodality and cognitive linguistics*, 13–26. Amsterdam: John Benjamins.
Forceville, Charles. 1998. *Pictorial metaphor in advertising*. London: Routledge.
Forceville, Charles. 1999. Review: "Educating the eye? Kress and van Leeuwen's *Reading Images: The Grammar of Visual Design (1996)*". *Language and Literature* 8(2). 163–178.
Forceville, Charles. 2002. The identification of target and source in pictorial metaphors. *Journal of Pragmatics* 34(1). 1–14.
Forceville, Charles. 2006. Non-verbal and multimodal metaphor in a cognitivist framework: Agendas for research. In Gitte Kristiansen, Michael Achard, René Dirven & Francisco J. Ruiz de Mendoza Ibáñez (eds.), *Cognitive linguistics: Current applications and future perspectives*, 372–402. Berlin: Mouton de Gruyter.
Forceville, Charles. 2008. Metaphor in pictures and multimodal representations. In Ray W. Gibbs Jr (ed.), *The Cambridge handbook of metaphor and thought*, 462–482. Cambridge: Cambridge University Press.
Forceville, Charles & Eduardo Urios-Aparisi (eds.). 2009. *Multimodal metaphor*. Berlin: Mouton de Gruyter.
Fridolfsson, Charlotte. 2008. Political protest and metaphor. In Terell Carver & Jernej Pikalo (eds.), *Political language and metaphor: Interpreting and changing the world*, 132–148. London: Routledge.
Goldberg, Adele. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
Goldberg, Adele. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
Groom, Nicholas. 2019. Construction grammar and the corpus-based analysis of discourses: The case of the WAY IN WHICH construction. *International Journal of Corpus Linguistics* 24(3). 335–367.
Guilbeault, Douglas. 2017. How politicians express different viewpoints in gesture and speech simultaneously. *Cognitive Linguistics* 28(3). 417–447.
Hampe, Beate (eds.). 2005. *From perception to meaning: Image schemas in cognitive linguistics*. Berlin: Mouton de Gruyter.
Hart, Christopher. 2010. *Critical discourse analysis and cognitive science: New perspectives on immigration discourse*. Basingstoke: Palgrave.

Hart, Christopher. 2013. Constructing contexts through grammar: Cognitive models and conceptualisation in British Newspaper reports of political protests. In John Flowerdew (ed.), *Discourse and contexts*, 159–184. London: Continuum.
Hart, Christopher. 2015. Viewpoint in linguistic discourse: Space and evaluation in news reports of political protests. *Critical Discourse Studies* 12(3). 238–260.
Hart, Christopher. 2016. The visual basis of linguistic meaning and its implications for CDS: Integrating cognitive linguistic and multimodal methods. *Discourse & Society* 27(3). 335–350.
Hart, Christopher. 2017. Metaphor and intertextuality in media framings of the (1984–85) British Miners' Strike: A multimodal analysis. *Discourse & Communication* 11(1). 3–30.
Hart, Christopher. 2018. Event-frames affect blame assignment and perception of aggression: An experimental case study in CDA. *Applied Linguistics* 39(3). 400–421.
Hart, Christopher. 2019. Spatial properties of ACTION verb semantics: Experimental evidence for image schema orientation in transitive versus reciprocal verbs and its implications for ideology. In Christopher Hart (ed.), *Cognitive linguistic approaches to text and discourse: From poetics to politics*, 181–204. Edinburgh: Edinburgh University Press.
Hinnell, Jennifer. 2018. The multimodal marking of aspect: The case of five periphrastic auxiliary constructions in North American English. *Cognitive Linguistics* 29(4). 773–806.
Johnson, Mark. 1987. *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago: University of Chicago Press.
Kim, Chai-Youn & Randolph Blake. 2007. Brain activity accompanying perception of implied motion in abstract paintings. *Spatial Vision* 20. 545–560.
Koller, Veronika. 2005. Designing cognition: Visual metaphor as a design feature in business magazines. *Information Design Journal and Document Design* 13(2). 136–150.
Koller, Veronika. 2009. Brand images: Multimodal metaphor in corporate branding messages. In Charles Forceville & Eduardo Urios-Aparisi (eds.), *Multimodal metaphor*, 45–71. Berlin: Mouton de Gruyter.
Kourtzi, Zoe & Nancy Kanwisher. 2000. Activation in human MT/MST by static images with implied motion. *Journal of Cognitive Neuroscience* 12. 248–255.
Kok, Kasper I. & Alan Cienki. 2016. Cognitive grammar and gesture: Points of convergence, advances and challenges. *Cognitive Linguistics* 27(1). 67–100.
Kress, Gunther & Theo van Leeuwen. 2006. *Reading images: A grammar of visual design*, 2nd edn. London: Routledge.
Lakoff, George & Mark Johnson. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.
Langacker, Ronald W. 1987. *Foundations of cognitive grammar, volume I: Theoretical prerequisites*. Stanford, CA: Stanford University Press.
Langacker, Ronald W. 1991. *Foundations of cognitive grammar, volume II: Descriptive application*. Stanford, CA: Stanford University Press.
Langacker, Ronald W. 2001. Discourse in cognitive grammar. *Cognitive Linguistics* 12(2). 143–188.
Langacker, Ronald W. 2002. *Concept, image, and symbol: The cognitive basis of grammar*, 2nd edn. Berlin: Mouton de Gruyter.
Langacker, Ronald W. 2008. *Cognitive grammar: A basic introduction*. Oxford: Oxford University Press.
Liu, Yu & Kay O'Halloran. 2009. Intersemiotic texture: Analysing cohesive devices between language and images. *Social Semiotics* 19(4). 367–388.

Louhema, Karoliina, Zlatev Jordan, Maria Graziano & Joost van de Weijer. 2019. Translating from monosemiotic to polysemiotic narratives: A study of Finnish speech and gestures. *Sign Systems Studies* 47(3/4). 480–525.

Martinec, Radan & Andrew Salway. 2005. A system for image-text relations in new (and old) media. *Visual Communication* 4(3). 337–371.

Martinez Lirola, Maria & Katina Zammit. 2017. Disempowerment and inspiration: A multimodal discourse analysis of immigrant women in the Spanish and Australian online press. *Critical Approaches to Discourse Analysis across Disciplines* 8(2). 58–79.

Matlock, Teenie & Bodo Winter. 2015. Experimental semantics. In Bernd Heine & Heike Narrog (eds.), *Oxford handbook of linguistic analysis*, 771–790. Oxford: Oxford University Press.

McNeill David. 1992. *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.

Norris, Sigrid. 2004. *Analysing multimodal interaction: A methodological framework*. London: Routledge.

O'Halloran, Kay. 1999. Interdependence, interaction and metaphor in multisemiotic texts. *Social Semiotics* 9(3). 317–354.

Oakley, Todd. 2010. Image schemas. In Dirk Geeraerts & Hubert Cuckyens (eds.), *The Oxford handbook of cognitive linguistics*, 214–235. Oxford: Oxford University Press.

Parrill, Fey, Benjamin K. Bergen & Patricia V. Lichtenstein. 2013. Grammatical aspect, gesture, and conceptualization: Using co-speech gesture to reveal event representations. *Cognitive Linguistics* 24(1). 135–158.

Pinar Sanz, Maria J. (eds.). 2015. *Multimodality and cognitive linguistics*. Amsterdam: John Benjamins.

Proverbio, Alice M., Dederica Riva & Alberto Zani. 2009. Observation of static pictures of dynamic actions enhances the activity of movement-related brain areas. *PLoS One* 4. e5389.

Royce, Terry D. 1998. Synergy on the page: Exploring intersemiotic complementarity in page-based multimodal text. *Japan Association for Systemic Functional Linguistics Occasional Papers* 1(1). 25–49.

Royce, Terry D. 2007. Intersemiotic complementarity: A framework for multimodal discourse analysis. In Terry D. Royce & Wendy L. Bowcher (eds.), *New directions in the analysis of multimodal discourse*, 63–110. London: Lawrence Erlbaum Associates.

Santa Ana, Otto. 1999. 'Like an animal I was treated': Anti-immigrant metaphor in US public discourse. *Discourse & Society* 10(2). 191–224.

Steen, Francis & Mark Turner. 2013. Multimodal construction grammar. In Mike Borkent, Barbera Dancygier & Jennifer Hinnell (eds.), *Language and the creative mind*, 255–274. Stanford, CA: CSLI Publications.

Taboada, Maite & Christopher, Habel. 2013. Rhetorical relations in multimodal documents. *Discourse Studies* 15(1). 65–89.

Talmy, Lenard. 2000. *Toward a cognitive semantics*. Cambridge, MA: MIT Press.

Vandelanotte, Lieven & Barbera Dancygier (eds.). 2017. Special Issue: Multimodal artefacts and the texture of viewpoint. Journal of Pragmatics 122.

Werner, Walt. 2004. On political cartoons and social studies textbooks: Visual analogies, intertextuality and cultural memory. *Canadian Social Studies* 38(2). http://www.educ.ualberta.ca/css/Css_38_2/ARpolitical_cartoons_ss_textbooks.htm.

Woodin, Greg, Bodo Winter, Max Perlman, Janette Littlemore & Teenie Matlock. 2020. 'Tiny numbers' are actually tiny: Evidence from gestures in the TV News Archive. *PLoS One* 15(11). e0242142.

Zima, Elisabeth. 2017. Multimodal constructional resemblance: The case of English circular motion constructions. In Franciso J. Ruiz de Mendoza Ibáñez, Alba L. Oyón & Paula Pérez-Sobrino (eds.), *Constructing families of constructions: Analytical perspectives and theoretical challenges*, 30–337. Amsterdam: John Benjamins.

Zima, Elisabeth & Alexander Bergs. 2017. Multimodality and construction grammar. *Linguistics Vanguard* 3(s1). 20161006.

Zlatev, Jordan. 2015. Cognitive semiotics. In Peter P. Trifonas (ed.), *International handbook of semiotics*, 1043–1067. Dordrecht: Springer.

Zwaan, Rolf A. 2004. The immersed experiencer: Toward an embodied theory of language comprehension. In Brian H. Ross (ed.), *The psychology of learning and motivation: Advances in research and theory*, 35–62. New York: Academic Press.