

# Using Generative AI to Investigate the Multimodal Representation of Refugees and Migrants

Christopher Hart (Lancaster University)

Email: c.hart@lancaster.ac.uk

## Abstract

This article uses AI-generated images to consider patterns in the way refugees and migrants are visually represented on the internet. AI text-to-image generation is treated as a diagnostic tool enabling the identification and analysis of recurring features in the visual representation of social groups. The theoretical departure point is the notion of multimodal constructions developed in cognitive linguistics. 144 images are generated by prompts instantiating a [refugees/migrants + motion verb] construction. The images are then analysed through the lens of Cognitive Critical Discourse Analysis to establish the visual correlates of these verbal forms and to reveal their ideological and (de)legitimizing functions in reproducing stereotypes, stoking fear and ultimately sustaining power differentials. The results show that the image most likely to be entrenched as part of a multimodal construction is of a large mass of people moving on foot toward the viewer. The ideological implications of this and other patterns detected are discussed as are differences in the visual representations associated with people designated as ‘refugee’ versus ‘migrant’.

## Introduction

Human migration is a globally significant phenomenon. Population dynamics continue to shift as a consequence of conflict, persecution, poverty and climate change among other contributing factors. People displaced by these circumstances are among the most vulnerable people on the planet. Yet, public perceptions of and attitudes toward migration are often hostile, which creates a space for equally hostile government policies. To a large extent, it is mediated experience in representations of refugees and migrants that is responsible for shaping public knowledge and opinion in connection with migration. In an increasingly multimodal world, it is visual as well as verbal representations that are of critical influence, including visual representations generated through Artificial Intelligence (AI). AI-generated images not only reflect our multimodal world and therefore provide a lens through which it can be examined, they increasingly come to populate it and are thus themselves important sites of critical investigation. This study analyses AI-generated images of migration from the perspective of Cognitive Critical Discourse Analysis, focussing specifically on the motion event inherent to the migration process. Despite the social and methodological significance of AI-generated images, the study is the first to apply Critical Discourse Analysis to images of migration produced by AI.

## Background

Images published on and shared over the internet increasingly define our knowledge and understanding of social and political realities. Indeed, according to Stocchetti (2011: 33), images are now “the dominant form of political communication”. Millions of images are viewed and downloaded

daily via search engines like Google (Zhang and Rui 2013). The images returned often reflect and reinforce pre-existing power structures and prejudices (Guilbeault et al. 2024; Noble 2018). Online news media, compared to traditional print formats, provide extensive amounts of visual content. Caple and Knox (2015) found that almost half of English-language newspaper websites around the world include dedicated multimedia sections with the most common form being photo galleries. Images also generate traffic to and influence social media engagement with news articles. For example, links to online news articles that contain images are more likely to be clicked and shared (Collier, Dunaway and Stroud 2021; Li and Xie 2019). Eye-tracking studies show that the images contained in news articles themselves are among the first entry-points to the text and receive a disproportionate amount of attention compared to written material (Bucher and Schumacher 2006; Quinn et al. 2007; Holsanova and Nord 2010; Leckner 2012). As an example of a more general ‘picture superiority effect’ (Hockley 2008), images in news texts are also shown to enhance recall (Graber 1990). Besides such cognitive processing and engagement effects, images achieve important framing effects and can even lead to falsely recalling information not present in news stories (Garry et al. 2007). Images produced as part of digital news content provide a visual narrative of political topics and events and are especially influential in shaping public knowledge and opinion and thus in driving policy directions. For example, the images included as part of a news story affect evaluations of social actors and actions (Arpan et al. 2006) and influence behavioural intentions (Giese et al. 2021; Powell et al. 2015). With respect to immigration, negative portrayals of migrants increase ingroup favouritism and outgroup hostility (Schermer 2012; Schemer and Meltzer 2019; Conzo et al. 2021; Scherman et al. 2022). When linked with crime, a negative framing of immigration in the news influences voting behaviour in favour of anti-immigrant parties (Burscher, van Spanje and De Vreese 2015). Conversely, exposure to more positive coverage, including instances of successful intergroup contact, results in more positive attitudes toward refugees and migrants and decreased support for restrictive border and security policies (Joyce and Harwood 2014; Djourelova 2023).

All of this suggests that the visual representation of refugees and migrants online is highly significant for public perceptions of and attitudes toward migration and that certain patterns of representation, if found, will contribute to perpetuating and sustaining stereotypes, stoking prejudice and fear, and legitimating hostile immigration policies. Critical Discourse Analysis (CDA) has developed a set of tools, based in linguistics but extended to visual and multimodal communication, that enables the identification and quantification of visual elements implicated in discursive constructions of power and inequality (Machin 2013; van Leeuwen 2014).

In connection with migration, research in CDA has uncovered several forms of visual representation which, recurrent in online news, construct refugees and migrants as a dangerous and alien ‘Other’ (Batziou 2011; Bleiker et al. 2013; Martínez Lirola 2016, 2022; Martínez Lirola and Zammit 2017; Wilmott 2017; Farris and Silber Mohamed 2018; Catalano and Musolff 2019; Romano and Dolores Porto 2021). For example, refugees and migrants tend to be depicted in large groups rather than small groups (Martínez Lirola 2016; Wilmott 2017), which results in dehumanising effects as audiences are less likely to ascribe to them human emotions like compassion, guilt and tenderness (Azavedo et al. 2021) as well as increased support for anti-immigration policies (Azavedo et al. 2021; Madrigal and Soroka 2023). Refugees and migrants are also frequently represented in explicitly dehumanising contexts (López 2020; Martínez Lirola 2022), where they appear in ways comparable to wild animals, or in contexts of crime/security, where they become criminalised (Martínez Lirola 2016; Wilmott 2017; Farris and Silber Mohamed 2018; Catalano and Musolff 2019). Images also tend to capture refugees and migrants somewhere along the way in their migratory journey rather than depicting the circumstances they are escaping or showing them settled into life in a new place (Romano and Dolores Porto 2021). Such patterns of visual representation are consistent with verbal forms of representation

where refugees and migrants are described in metaphorical terms as animals/diseases, natural disasters or invading armies (Abid et al. 2017; Charteris-Black 2006; Cisneros 2008; Dolores Porto 2022; El Refaie 2001; Hart 2021; Musolff 2015; Santa Ana 1999, 2002) and these metaphors likewise achieve framing effects in eliciting negative emotions and legitimating anti-immigration policies (Chkhaidze et al. 2021; Jimenez et al. 2021; Marshall and Shapiro 2018; Utych 2018). Gabrielatos and Baker (2008) used corpus linguistic methods to show that the label 'refugee' collocates with numerical signifiers expressing large quantities, though interestingly in their sample the label 'migrant' did not.

Collocation is "the phenomena of certain words frequently occurring next to or near each other" in texts (Baker 2006: 96). As a result of this statistical relationship, the concepts lexicalised by the collocate of a word come to form part of its semantic profile (Stubbs 1995) and may be primed even on occasions when the collocate is not actually present (Stubbs 1996, 2002). Thus, where the words *illegal* and *immigrant* regularly co-occur with one another (Gabrielatos and Baker 2008), the concept ILLEGAL may still be primed by references to *immigrants* without the adjectival modifier. The notion of collocation is extended to relations of intersemiotic co-occurrence by McGlashan (2021) who uses the term *collustration* to refer to "repeated co-occurrences between representations in linguistic and visual semiotic resources across numerous texts" (p. 227).

A similar idea developed in cognitive linguistics is that of *multimodal constructions* (Dancygier and Vandelanotte 2017; Cienki 2017; Hoffman 2021; Kok and Cienki 2016; Lehmann 2024; Schoonjans 2018; Steen and Turner 2013; Uhrig 2022; Ziem 2017; Zima 2017; Zima and Bergs 2017). The difference between collustration and multimodal constructions, however, is that collustration describes a behavioural relationship between verbal and visual elements at the level of texts while multimodal constructions are entrenched structures in the minds of language users. Collustrations can thus be seen as the predictable outcome of multimodal constructions which, following a usage-based account, are at the same time derived from the collustrational behaviours of language and image.

Spoken and written language nearly always occur in multimodal situations. From a usage-based constructionist perspective, it therefore "appears very likely that the mental grammars of speakers will also contain a considerable number of multimodal constructions" (Hoffman 2021: 82). Usage-based traditions like cognitive grammar (Langacker 2008) and construction grammar (Goldberg 2006; 2003), as part of a more general 'multimodal turn' in linguistics (Cohn and Schilperoord 2024; Stöckl 2020), have thus been extended to incorporate within their theoretical architectures non-verbal forms like gesture and image.

Goldberg (2006: 5) defines a linguistic construction as a "learned pairing of form with semantic or discourse functions" where forms include morphemes or words, idioms, and partially lexically filled as well as fully general phrasal patterns. Constructions are stored in the minds of language users as cognitive units which together make up the 'constructicon'. A multimodal construction, accordingly, represents a routinised form-meaning pairing whose form consists of elements belonging to more than one semiotic mode. That is, multimodal constructions consist of intersemiotic forms that conventionally combine in the expression of meaning and which together with that meaning are entrenched as stored cognitive units in the minds of language users. For Ziem (2017), there are two types of multimodal construction. One is "inherently" multimodal comprising multimodal forms which obligatorily co-occur, as in deictic expressions that require a kinetic element in order to determine their referent. The other consists of two or more semiotic forms which co-occur with "sufficient frequency" as to have become conventionalised and thus entrenched, where entrenchment is a correlate of conventionalisation (Lehmann 2024).

Multimodal constructions are therefore distinct from incidental co-occurrences with the question of when a multimodal construct achieves constructional status being a largely quantitative one dependent on frequency of co-occurrence within a speech community. In so far as they “provide insights into the conventionalization of a construction” (Lehmann 2024: 413), corpus frequencies have thus become paramount to claims about multimodal constructions. Indeed, corpus data has been the primary source of evidence for research investigating multimodal constructions. No specific frequency threshold is set, however. This is because entrenchment is a gradual phenomenon and, as Uhrig (2022) points out, we are therefore better thinking in terms of degree of entrenchment than of a binary opposition between constructional and non-constructional status. For example, in their respective corpora, Schoonjans (2018) showed that the German particle *einfach* was accompanied by a headshake in 24% of instances, while Uhrig (2022) showed that constructions involving throwing verbs (e.g. fling, lob) occurred with a ‘throwing’ gesture 54% of the time, and Zima (2017) showed that the construction [all the way from X PREP Y] occurred with a specific gestural form in 80% of cases.

Crucially, constructions may exist in the language at large or may be particular to specific discourses and genres (Antonopoulou and Nikiforidou 2011; Groom 2019). While most of the research addressing multimodal constructions has been focussed on combinations of verbal and kinetic elements in spoken language, there is some research investigating the visual and audio-visual correlates of linguistic forms, including specifically in the context of news communication (Steen and Turner 2013; Hart and Marmol Queralto 2021; Hart 2024). For example, Steen and Turner (2013) used the NewsScape corpus to show that when narrating a news event, the past-tense + proximal deictic construction (as in “was/were + now”) has a tendency to occur alongside a zooming in on the entity whose experiences are being recounted. Hart (2024) also used the NewsScape corpus to investigate the images that regularly co-occur with four constructions describing motion events in TV news coverage of migration. The four constructions crossed referential nominal with grammatical aspect ([refugees/migrants have VERBed] / [refugees/migrants are VERBing]) and focussed on the top six verbs expressing motion in the corpus. Among various patterns observed, the most common form of visual representation found to occur with these constructions depicted large groups of refugees and migrants moving on foot over land somewhere along the migratory journey. An important point to note here is that the imagery that figures in a multimodal construction is not a specific image but is a more abstract visual form representing features common across images. As with collocation, aspects of meaning contributed by the visual component of a multimodal construction become part of the meaning of the verbal component alone and are likely to be evoked even when the visual component is not itself co-instantiated in text.

Investigations into multimodal constructions are based largely on correlational analyses of intersemiotic forms. However, since multimodal data annotation is extremely time-consuming, this means analyses are necessarily based on a relatively limited number of data-points. The argument I wish to make here is that AI provides a methodological shortcut to a dataset that is infinitely vast by comparison and that, by aggregating millions of images across the internet to generate images which typify those most recurrently associated with a given verbal form, AI can be used as an effective tool in establishing the visual components of multimodal constructions.

I do not claim to have a computational understanding of the way AI text-to-image generation works but instead adopt the position of a non-technical, critical user (Putland, Chikodzore-Paterson and Brookes 2023). AI-image generators are given textual prompts by a human user and then, relying on trained machine learning models, produce images that match the prompt. According to Canva, a free-to-use graphic design platform with a built-in text-to-image generator:

to create AI-generated images, the machine learning model scans millions of images across the internet along with the text associated with them. The algorithms spot trends in the images and text and eventually begin to guess which image and text fit together. Once the model can predict what an image should look like from a given text, it creates entirely new images.<sup>1</sup>

AI-generated images will naturally reflect the social identities, values and stereotypes enshrined in the images that the models are trained on and can therefore be critically analysed to reveal the ideologies and prejudices propagated over the internet.

At the same time, those ideologies and prejudices will be reinforced and amplified as the image pool comes to be populated more and more by AI-generated images. AI-image generation is therefore described as a “semiotic technology”, that is, “a technology for meaning-making that is deeply inscribed with certain sets of social norms, values and ideologies” (Westberg and Kvåle 2024: 2). From this perspective, AI-generated images constitute, in their own right, “important, but currently understudied, social texts to attend to in relation to complex social phenomena” (Putland, Chikodzore-Paterson and Brookes 2023: 2). However, since AI text-to-image generation is relatively new technology, research investigating it is still limited, especially within discourse studies. A few recent exceptions are nevertheless to be found. For example, across a broad range of prompts, Banchi et al. (2023) found that AI generates images which reinforce racial and gendered stereotypes, promote whiteness as an ideal, and reproduce American norms. Prompts mentioning social groups (e.g. by race or nationality) produce images “that tie specific groups to negative or taboo associations like malnourishment, poverty, and subordination” (p. 1495). These biases are so robust that they are reproduced even when they are explicitly countered in the prompts. Putland, Chikodzore-Paterson and Brookes (2023) examined AI-images generated in connection with dementia. They found that images recycle previously existing and prominent discourses surrounding the syndrome by maintaining a biomedical framing, presenting narratives focussed on loss and dementia as a ‘living death’, and displaying a distinct lack of diversity characterised by an over-representation of older, light-skinned individuals. Westerberg and Kvåle (2024) analysed AI-generated images representing teenagers. Although the images generated present a diverse range of ethnicities, they do not appear to mix between ethnic provenance. Instead, through juxtaposition of teenagers with contrasting biological attributes (e.g. skin colour, hair quality), the images reinforce exclusive ethnic categories. Other dimensions of diversity are not visually denoted. For example, there is conformity to body normativity as well as with respect to cultural attributes like clothing.

Putland, Chikodzore-Paterson and Brookes (2023: 5) and Westerberg and Kvåle (2024: 2) see the social functions of AI-generated images as continuing from stock images but providing users with a greater ability to generate content for themselves. Machin (2004) addresses the increasing role of image banks like Getty Images in defining the visual language of various media, including digital news media. He shows that the images contained in such banks provide users with a pre-structured world that is organised into ideologically constrained categories that are consistent with the values of consumerism and globalisation at the expense of social security and mobility. AI and image banks now exist in a symbiotic relationship with image banks providing a large proportion of the data that AI models are trained on and image banks having a large stock of AI-generated images.

---

<sup>1</sup> <https://www.canva.com/ai-image-generator>

This study both uses and targets AI-generated images to critically analyse patterns in the visual representation of refugees and migrants and to establish the imagery that figures as part of a specific multimodal construction.

## Method

### Analytical Framework: Cognitive CDA

CDA has successfully extended frameworks developed in linguistics to account for patterns of representation in non-linguistic modes of communication, including images, and their role in reproducing social identities and unequal relations of power (Kress and van Leeuwen 2006; Machin 2013, 2016; Machin and Mayr 2012; van Leeuwen 2008). While much of multimodal CDA has been based on extensions of Systemic Functional Linguistics, multimodal CDA is not characterised by any particular method and is instead conceived as a *field of application* (cf. Jewitt 2009: 2). Any approach to CDA may therefore be extended to multimodal data (e.g. Richardson and Wodak 2009). However, Cognitive CDA (e.g. Hart 2025) is especially well appointed for multimodal analysis since its analytical parameters, stemming from the embodied view of language presented in cognitive linguistics, are based in aspects of visual perception.

Cognitive CDA is an approach to CDA which, drawing on cognitive linguistics, considers the ideological qualities and legitimating potentials of conceptualisations evoked by linguistic and other semiotic forms in contexts of political communication (Hart 2025). A key concept in Cognitive CDA is that of construal, which refers to our “manifest ability to conceive and portray the same situation in alternate ways” (Langacker 2008: 43), and is a fundamental feature of conceptualisation. Cued by the linguistic expressions selected in its description, a referential event is necessarily construed in a particular way, which may be ideologically vested. The parameters of construal along which ideology is enacted in language have analogues in visual perception where “in viewing a scene, what we actually see depends on how closely we examine it, what we choose to look at, which elements we may most attention to, and where we view it from” (Langacker 2008: 55). Given the connection with visual experience, it is obvious that the same dimensions of construal will also inhere in images. One inherent aspect of migration that is especially sensitive to ideological construal in both language and image and which appears to be of particularly high news value is the motion event (Hart 2024, 2025; Hart and Marmol Queralto 2021).

The linguistic representation of motion has received extensive treatment in cognitive linguistics (Pourcel 2010; Slobin 2004, Talmy 1985, 2000; Zlatev, Blomberg and David 2010). The motion event is a conceptual archetype (Langacker 2008) or event-frame (Talmy 2000) made up minimally of four conceptual elements: (i) a Figure, namely the entity or object that undergoes motion; (ii) motion itself, which is defined by motion or the continuation of a stationary location; (iii) a path along which the motion takes place; and (iv) a Ground providing a frame of reference relative to which the motion is characterised. The canonical motion event and the one inherent to migration is translational motion, which contrasts with other types of motion like rotation and involves a change in location to the figure over the period of time under consideration (Talmy 2000/II: 25). In addition to the four core elements, another element that may also receive representation is manner of motion, which refers to the way a motion event is accomplished. These elements are semantic elements that get expressed in surface elements like nouns, verbs and prepositions. Crucially, they may receive expression in visual as well as verbal elements of composition (Hart and Marmol Queralto 2021). Research in Cognitive CDA has

shown that all elements that make up a motion event are subject to construal leading to various ideological and (de)legitimizing effects (Hart 2025; Hart and Marmol Queralto 2021). The present study extends principles of Cognitive CDA to focus specifically on visual representations of the motion event that is inherent to the migration process.

## Data Selection

To generate images, an AI image generator provided by the free-to-use online graphic design platform Canva was used. Canva integrates several different AI image generators that meet different user needs, including Dream Lab, Magic Media, DALL.E by OpenAI and Imagen by Google Cloud. The specific tool selected for this study was Magic Media. Magic Media was selected because it allows users to create images in particular styles including photos and because it produces images that are highly realistic as opposed to the more surreal images that are often produced by other tools.<sup>2</sup> Canva's Magic Media uses as its underlying model Stable Diffusion. Stable Diffusion is a deep learning, latent diffusion model whose open-source code was originally released in 2022 (Barazida 2022) and which is currently one of the most popular models available. While Stable Diffusion can be accessed directly via its own web user interface, Canva was used, in keeping with the perspective of a non-technical critical user, owing to its greater accessibility and more user-friendly environment.

The text prompts used to generate images were a simplified version of the search queries used by Hart (2024). The prompts crossed referential nominal (*refugees* vs *migrants*) with six verbs of motion (*arrive*, *cross*, *enter*, *flee*, *flood* and *pour*) which were the most frequent to occur with the two denominations in the NewsScape Corpus.<sup>3</sup> Since tense and aspect did not impinge on visual representations in Hart's (2024) study, these variables were not included here with prompts given instead in the simple present form. This resulted in a total of twelve noun + verb combinations like "refugees arrive", "migrants arrive", "refugees cross", "migrants cross", etc. The decision to investigate plural noun forms and not singular forms reflects their relative frequency in media discourses.

Canva's Magic Media allows users to select from different styles, including photography, digital art and fine art. Each style includes further sub-categories with photography, for example, including filmic, photo, moody and vibrant and digital art including anime, dreamy and psychedelic. To ensure maximum levels of realism, the style selected was photography: photo. Magic Media also allows users to generate images in three different formats: square, landscape and portrait. To avoid any biases arising from the affordances of different formats, for each prompt an equal number of images in each format was included in the sample. Magic Media generates four images per prompt in the selected style and format. Only one round of image generation was performed for each prompt with all four images generated being included in the sample. This resulted in a final sample comprised of 144 AI-generated images. All data were collected in a single day on 13<sup>th</sup> February 2025.

---

<sup>2</sup> The tool is accessible here: <https://www.canva.com/ai-image-generator/>.

<sup>3</sup> The terms *refugee* and *migrant* were selected to focus the study on movement of people across borders rather than legal aspects of migration like asylum-seeking. The two terms are also closely related where, although they have distinct definitions in law, they are often used interchangeably in media and public discourse. The categories are treated here as socially and discursively constructed, including as a function of the images with which their designations regularly co-occur. Other migration-related categories like *aliens*, *invaders* etc. were not investigated owing to their already inherently pejorative status.

## Data Coding

Previous studies show a range of conceptual parameters according to which the elements that make up the motion event may vary in language and image to ideological effect (Hart 2024, 2025). These include conceptual distinctions pertaining to the Figure such as presence, size, boundedness and dividedness, which are each coded for here. Despite restricting the style of images to photographic, a small number of images generated were more expressionist in style. Images were therefore coded for whether refugees and migrants are corporeally present or represented instead through a more abstract form. Size refers to the quantity of people represented. Here, a quantity is classed as small if it features less than twenty people and as large if it features twenty or more people.

Dividedness refers to a quantity's internal segmentation (Talmy 2000/I: 55). A quantity's state of dividedness is classed as discrete if it is conceived as having breaks or interruptions in its composition (ibid.). It is classed as continuous when otherwise separate elements are melded so that they come to cohere as a perceptual continuum or gestalt (Talmy 2000/I: 56). In images of migration, a discrete image would be one in which refugees and migrants are more widely dispersed with a low degree of agglomeration. A continuous image would be one that shows refugees and migrants more tightly interspersed with a high degree of agglomeration, such as being huddled together with no degree of physical separation between them. Note, however, that individuals do not have to be spatially contiguous in reality to be construed as continuous. The distinction is a visual perceptual one and camera angles and other semiotic properties can work to impose on the Figure a continuous construal. Equally, images showing individuals in spatially contiguous arrangements may still be classified as discrete if the individuals depicted are perceived as such.

Boundedness refers to whether a quantity is "conceived as continuing on indefinitely with no necessary characteristic of finiteness intrinsic to it" (Talmy 2000/I: 50), in which case it is unbounded, or whether it is conceived as having a clearly demarcated extension, in which case it is bounded. In the context of migration, bounded images would include those depicting lines of people the endpoints of which are both clearly visible inside of the viewing frame or crowds of people the entire contour of which falls inside the viewing frame. Unbounded images would include images of individuals or groups of individuals whose limits of extent lie outside of the viewing frame or up to and beyond the vanishing point in the image to create the illusion of continuing indefinitely. Because they often pattern together, with boundedness/discrete and unbounded/continuous coinciding, the categories of boundedness and dividedness are prone to confusion. However, the two categories vary independently (Talmy 2000/I: 55). For example, in images of refugees and migrants, individuals might be shown in a loosely gathered crowd whose total expanse extends beyond the viewing frame, in which case the image would be discrete but unbounded.

Other conceptual distinctions pertain to motion, manner of motion and the Ground (path is subsumed under viewpoint). With respect to motion, events are classified as motion where the image can be read as 'dynamic', implying a change in location to the Figure. This includes images in which refugees and migrants are themselves static but are depicted aboard a vehicle that is read as being in motion. The event is classified as stationary where no motion is detected, such as images of an assembled crowd or people posing for a photograph. Manner refers to the means by which the motion event is accomplished. This includes whether motion is achieved pedally or by means of a transport vehicle such as a bus, boat or train. Ground refers to the situation in which the motion event occurs. It is classified according to geographical features like landscapes (roads, tracks, mountain paths, fields, deserts), rivers, oceans and cityscapes as well as temporary structures connected to the politics of migration (camps, borders, processing centres). In some instances, images are decontextualised so



that no Ground is discernible. A further conceptual distinction relating to the Ground but connected to the overall migratory journey is source-path-goal. Migration involves departing a country of origin (source) and travelling along a specific route (path) to start a new life in a destination country (goal). An image is coded as source when the Ground depicts the circumstances refugees and migrants are leaving behind. An image is coded as path when refugees and migrants are shown anywhere on their migratory journey, including arriving in the destination country and/or being processed in holding centres. An image is coded as goal when it shows refugees and migrants settled into and actively participating in life within the destination country. A final conceptual distinction related to the Ground concerns the presence or not of security elements such as fences, police or military personnel within it. Finally, a necessary feature of images and conceptualisations is viewpoint.

Conceptual distinctions pertaining to viewpoint are defined with respect to three dimensions: path, angle and distance. With respect to path, images are coded as toward, away or across (or unclear/mixed) depending on the direction of motion relative to the viewer. With respect of angle, images are coded as horizontal, diagonal or from above. Distance is similarly coded based on a tripartite distinction between close-up, medium and distal shots. The values for viewpoint variables are necessarily ideals which images do not always perfectly align with. Viewpoint is therefore coded according to the values that images most closely approximate. The full coding scheme used in the analysis is given in Table 1.

Twenty per cent of the data (36 images) was independently coded by a second coder. Inter-coder reliability was overall almost perfect with a mean kappa score of 0.854. Scores for individual variables ranged from substantial to perfect.<sup>4</sup> Unsurprisingly, given the more subjective degree of judgement involved, the lowest score was for the viewpoint variable distance (0.641) while the highest scores (1.00) were achieved for presence, source-path-goal and manner. Since inter-coder reliability was high for all variables, analysis proceeded on the original coding.

Variable	Values
Presence	Corporeal Abstract
Size	Small (<20) Large (20+)
Dividedness	Discrete Continuous
Boundedness	Bounded Unbounded
Motion	Motion Stationary
Manner	On foot Bus Boat Train
Ground	Landscape (roads, tracks, mountain paths, fields, desert) Rivers Sea/coast

<sup>4</sup> The kappa scores for all individual variables were as follows: Presence (1.000); Size (0.800); Dividedness (0.724); Boundedness (0.923); Motion (0.769); Manner (1.000); Ground (0.884); Source-Path-Goal (1.000); Security (0.893); Path (0.769); Angle (0.845); Distance (0.641).

	Camps/borders/processing centres Decontextualised
Source-path-goal	Source Path Goal
Security	Present Absent
Path	Toward Away Across Unclear/mixed
Angle	Horizontal Diagonal From above
Distance	Close-up Medium Distal

**Table 1.** Coding scheme

## Results and Discussion

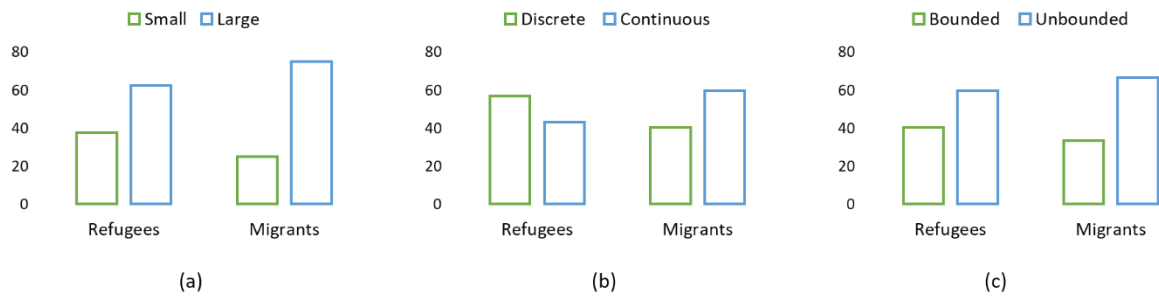
### Figure Properties

*Size:* The Figure is depicted more often as a large group (68.8%) than a small group. An exact binomial test confirms that this difference is statistically significant assuming equal proportions ( $p < .001$ ). As shown in Figure 1a, the result applies to images prompted by both nomination categories ‘refugees’ and ‘migrants’.

*Dividedness:* Overall, images are split evenly between discrete (49%) and continuous (51%) Figures. However, as shown in Figure 1b, images prompted by ‘migrants’ are significantly more likely to depict a continuous Figure (59.7%) compared to images prompted by ‘refugees’ (43.1%) ( $\chi^2 = 4.003$ ,  $p < .05$ ).

*Boundedness:* The Figure is depicted significantly more frequently as unbounded (63.2%) than it is bounded ( $p < .01$ ). As shown in Figure 1c, this applies to images prompted by both nomination categories.

*Presence:* ‘Refugees’ and ‘migrants’ receive corporeal representation in 98% of images ( $n = 141$ ).



**Figure 1.** Bar plots showing Figure properties. (a) = size; (b) = dividedness; (c) = boundness

When Figure properties are considered together, the most striking pattern that emerges in connection with the Figure is the propensity for refugees and migrants to be represented in large, continuous, unbounded forms such as those in Figure 2. This configuration is the dominant pattern accounting for 42.4% ( $n=61$ ) of all images ( $\chi^2=21.886$ ,  $p<.001$ ).<sup>5</sup> It occurs in two main variants with refugees and migrants shown fused together in non-delimited line or crowd formations as in Figure 2a and 2b respectively.



**Figure 2.** Large, continuous, unbounded Figures in line (a) and crowd (b) formations

The depiction of refugees and migrants in large rather than small groups is consistent with previous findings for online news images (Martínez Lirola 2016; Wilmott 2017). Size is ideologically significant where, for example, refugees and migrants are judged as less capable of experiencing human emotions like tenderness, guilt and compassion when they are shown in large group sizes (Azavedo et al. 2021). Large group depictions also lead to decreased perceptions of vulnerability (Bleiker et al. 2013). Conversely, among people who are already high in threat sensitivity, anti-immigration attitudes are mitigated by images of individual migrants but not by large groups of migrants (Madrigal and Soroka 2023). In a phenomenon known as the *identifiable victim effect*, people express greater empathy and willingness to help in situations involving specific individuals than in situations involving a larger group

<sup>5</sup> The second most frequent configuration representing a small, discrete, bounded Figure accounts for 23.6% ( $n=34$ ) of images.

(Jenni and Lowenstein 1997; Lee and Hugh Feeley 2016). The images in Figure 2 thus contrast with individual or small group depictions such as represented in Figure 3, where the Figure is also both discrete and bounded. Where the default image of migrants is of large groups of younger men (Banks 2012; Martínez Lirola and Zammit 2017; Olier and Spadavecchia 2022), small group images are also more likely to include women, children and the elderly or individuals who are identifiable as a family and therefore invite greater degrees of sympathy and support.



(a)



(b)

**Figure 3.** Small, discrete, bounded Figures

Although images prompted by ‘refugees’ and ‘migrants’ both tend to depict large groups, they differ in the dividedness of the Figure, with images prompted by ‘migrants’ more likely to depict a continuous Figure compared to images prompted by ‘refugees’. The contrast can be seen in Figure 4 where, compared to people in Figure 4a, people in Figure 4b are more densely compacted such that they coalesce to form a single defined shape.<sup>6</sup>



(a)



(b)

**Figure 4.** Discrete (a) versus continuous (b) Figures.

<sup>6</sup> Unsurprisingly, there is a correlation between distance and dividedness ( $\chi^2 = 17.2226$ ,  $p < .001$ ) where a close-up viewpoint correlates with discrete Figures and a distant viewpoint correlates with continuous Figures. This results from the loss in granularity experienced as the camera zooms out.

The parameter of dividedness has the function of individualising people in discrete images or construing them as a single mass in continuous images. The distribution between different nominations points to a difference in the way people designated as ‘migrants’ are typically conceived compared to people designated as ‘refugees’. Through discrete representations, those designated as ‘refugees’ are more likely to be recognised as individual beings with their own histories, motives and emotions. By contrast, ‘migrants’ are presented as an homogenous mass that moves and behaves as a single entity. Such a construal is consistent with bovine metaphors which liken migrants to herds of animals (Santa Ana 1999) or metaphors which construe migrants as bodies of water (Charteris-Black 2006; Santa Ana 2002). Indeed, a particular type of continuous Figure is one generated by the verb ‘flood’ in which migrants appear like a sprawling ‘sea’ of people as in Figure 5. Such imagery is strongly associated with the verb ‘flood’ occurring with 75% (n=18) of all instances. Forceville (2008, 2009) has shown that metaphors get expressed visually as well as verbally. Images like those in Figure 5 may be analysed as visual instantiations of an IMMIGRATION IS FLOOD metaphor. There appear to be two distinct means by which this metaphor is rendered. In images like the one in Figure 5a, the metaphor results from the *integration* of water and migrants while for images like the one in Figure 5b, the metaphor results from the *substitution* of water with migrants.<sup>7</sup> The images prompted by ‘flood’ are less realistic and, especially in images like Figure 5a, represent a more creative output of AI. They may nevertheless be indicative of the kind of mental imagery associated with flood metaphors in this context.



(a)



(b)

**Figure 5.** Continuous Figures prompted by ‘flood’

Figures prompted by ‘refugees’ and ‘migrants’ are more frequently unbounded than they are bounded. In bounded Figures, as in the image in Figure 6, the full extent of the Figure is clearly defined such that its entire range is discernible.

---

<sup>7</sup> In Forceville’s (2008) terms, the visual metaphors expressed by the images in Figure 5a and 5b are examples *hybrid* versus *contextual* metaphors respectively.





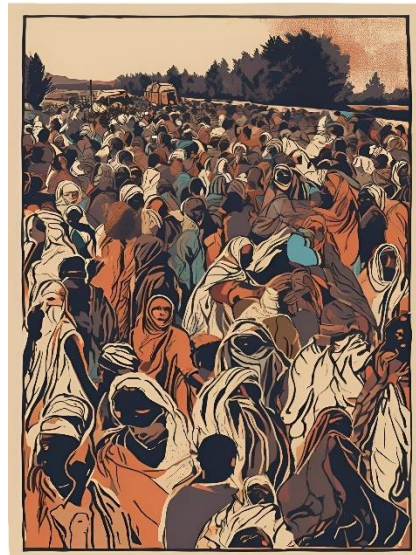
**Figure 6.** Bounded Figure.

Unbounded Figures come in several forms, as illustrated in Figure 7. In one especially common form, shown by the images in Figure 7a, a linear Figure extends indefinitely beyond the horizon or vanishing point in the image so that its endpoint cannot be made out. Another form involves a line or crowd of refugees/migrants that extends beyond the edges of the viewing frame, as shown by the images in Figure 7b. A third form is one in which the Figure occupies the entire frame as shown by the images in Figure 7c. Unboundedness functions to construe migration as a never-ending phenomenon. It thus supports arguments which claim that there exists a limitless and unsustainable number of refugees and migrants who will continue to impose on host countries unless changes to migration policies are made.



**Figure 7.** Unbounded Figures

Previous studies have shown that in human-produced images of migration, refugees and migrants frequently get represented through abstract visual forms such as silhouettes (Hart 2024). Although such *erasure* is less common in the present data, occurring in only three instances, it is nevertheless worth highlighting as a form of visual representation picked up by the AI image generator despite the instruction to return images only in a photographic style. As one example, in Figure 8, refugees are represented in the style of an expressionist picture. Such images strip the Figure of human features so that they are presented as faceless or even bodiless beings. They therefore deny the corporeal experience of refugees and migrants or remove them entirely from the same existential order as the viewer. The image in Figure 8 recalls Edvard Much's 1893 painting *The Scream* and presents a somewhat ghoulish Figure.



**Figure 8.** Non-presence

## **Motion, Manner and Ground**

*Motion:* The Figure is construed more frequently as being in motion (59%) than being stationary (Figure 9a). An exact binomial test confirms that this difference is significant assuming equal proportions ( $p < .05$ ). The pattern applies to images prompted by both nomination categories ‘refugees’ and ‘migrants’.

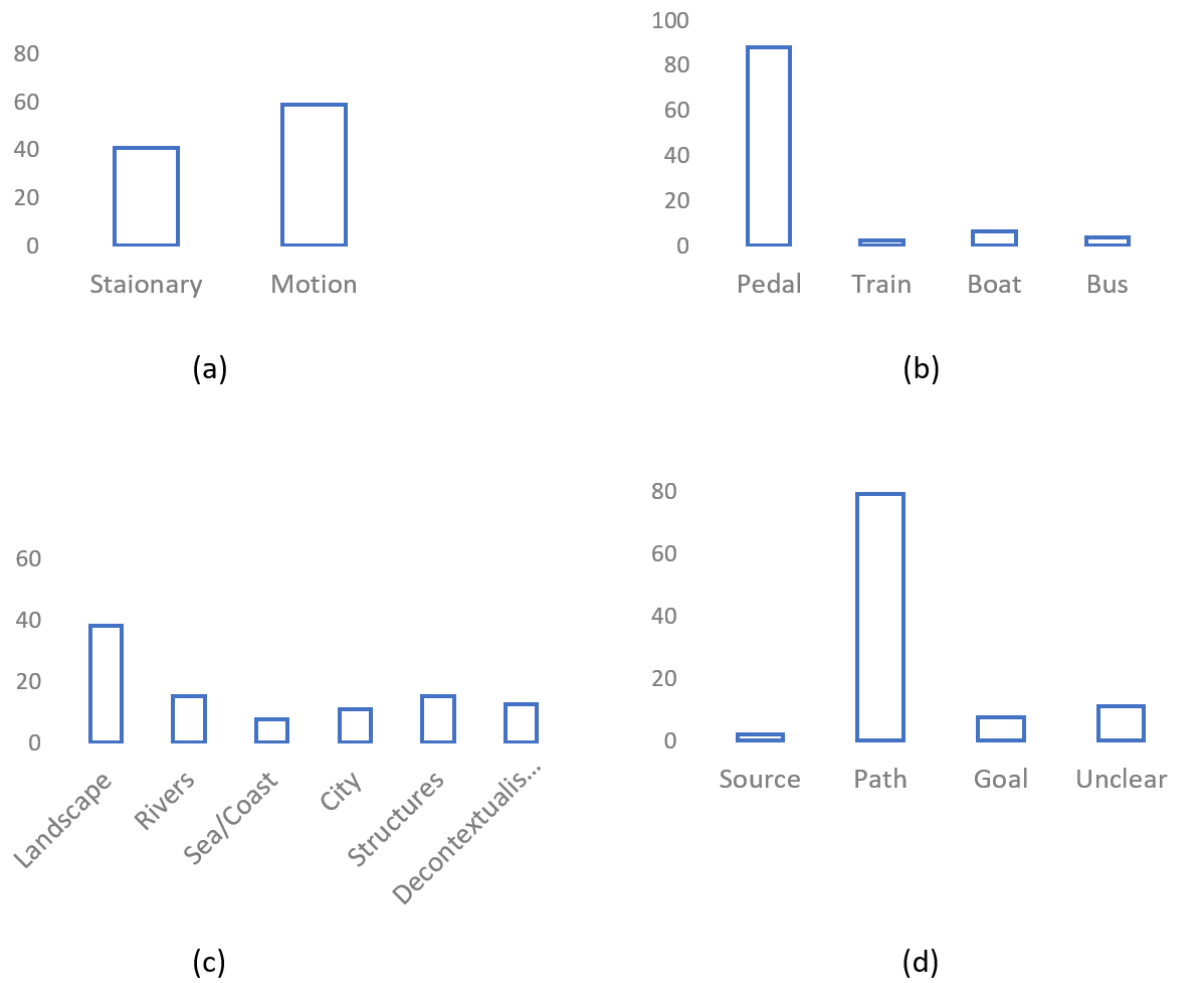
*Manner:* The Figure is construed more frequently in pedal motion (88.2%) than any other manner of motion (Figure 9b). The result is significant assuming equal proportions ( $p < .001$ ). When non-pedal manners of motion are examined separately ( $n = 17$ ), there is a difference in the specific modes of transport associated with the nominations ‘refugees’ versus ‘migrants’. As shown in Figure 10a, ‘migrants’ are most frequently associated with boats (72.7%,  $n = 8$ ) while ‘refugees’ are most frequently associated with buses (83.3%,  $n = 5$ ). A Fisher’s Exact test confirms that this distribution is significant ( $p < .001$ ).

*Ground:* The most common form of Ground represented is landscape, which accounts for 38.2% of images overall (Figure 9c). As shown in Figure 10b, this is the most frequent form of Ground for both nomination categories. However, as Figure 10b also shows, there are significant differences in the other types of Ground with respect to which ‘refugees’ versus ‘migrants’ are depicted ( $\chi^2 = 38.210$ ,  $p < .001$ ). The Ground element prompted by ‘refugees’ is more likely to be a temporary structure connected to the politics of migration while it is exclusively the prompt ‘migrants’ that places the Figure in the context of the sea/coast.

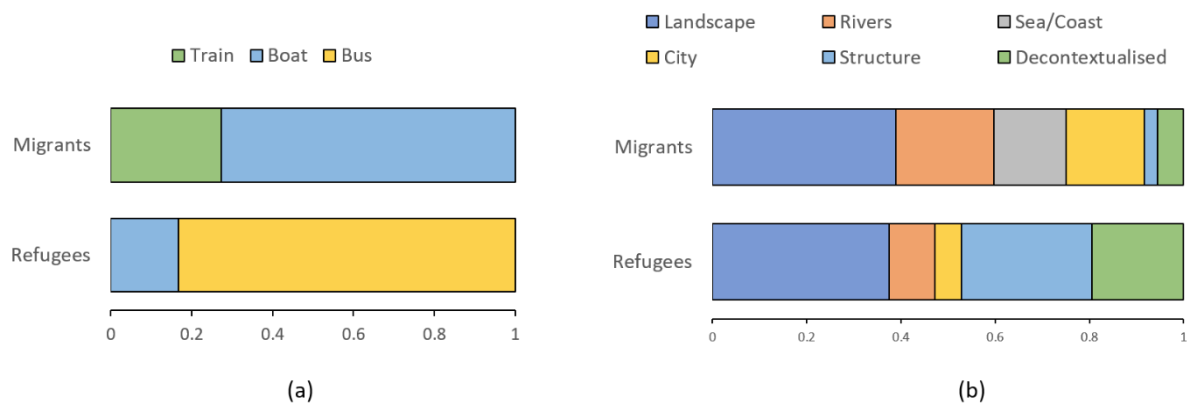
*Source-Path-Goal:* Ground elements where a location is identifiable represent path locations (79.2%) significantly more frequently than source (2.1%) or goal (7.6%) locations ( $p < .001$ ). This pattern applies to images prompted by both nomination categories ‘refugees’ and ‘migrants’.

*Security:* Ground elements contain a security presence in 6.3% ( $n = 9$ ) of images.





**Figure 9.** Bar plots showing proportions of motion (a), manner (b), Ground (c) and source-path-goal (d) categories.



**Figure 10.** Stacked bar plots showing proportions of non-pedal manners (a) and ground types (b) per 'refugees' vs 'migrants'.

From the results above, the image that emerges as the most likely to figure alongside the [refugees/migrants + motion verb] form in a multimodal construction is one in which a large, continuous, unbounded Figure moves on foot through geographical landscapes including roads, tracks, mountains, fields and deserts whilst on route in the migratory process.

The depiction of refugees and migrants in path locations rather than source or goal locations in the AI-generated images is consistent with findings for online news images (Romano and Dolores Porto 2021). By failing to represent refugees and migrants in source locations, images ignore the difficult or tragic circumstances that lead to displacement, such as poverty, war and persecution, and allow the argument that people are migrating out of choice and opportunity rather than need. Likewise, by failing to represent refugees and migrants settled into life in goal locations, images ignore the positive outcomes of migration. Two exceptions to the general pattern, which focus on source and goal locations respectively, are given in Figure 11.



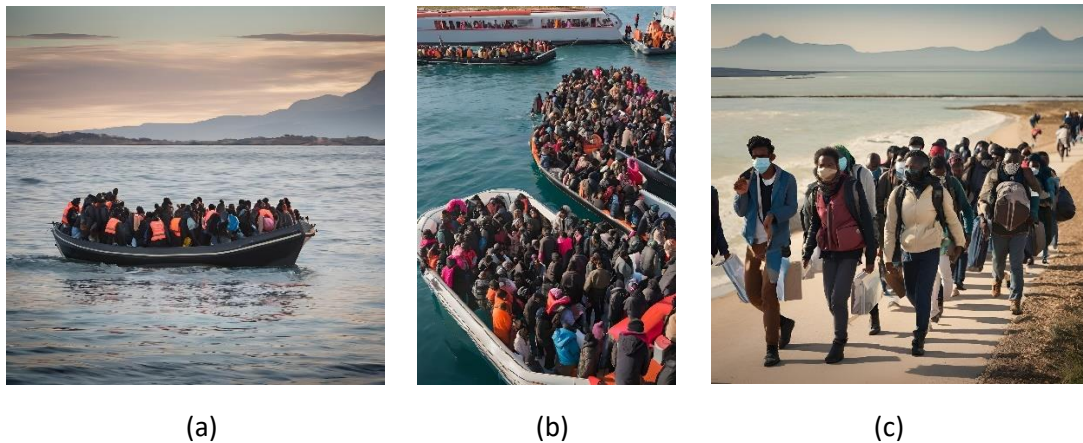
(a)



(b)

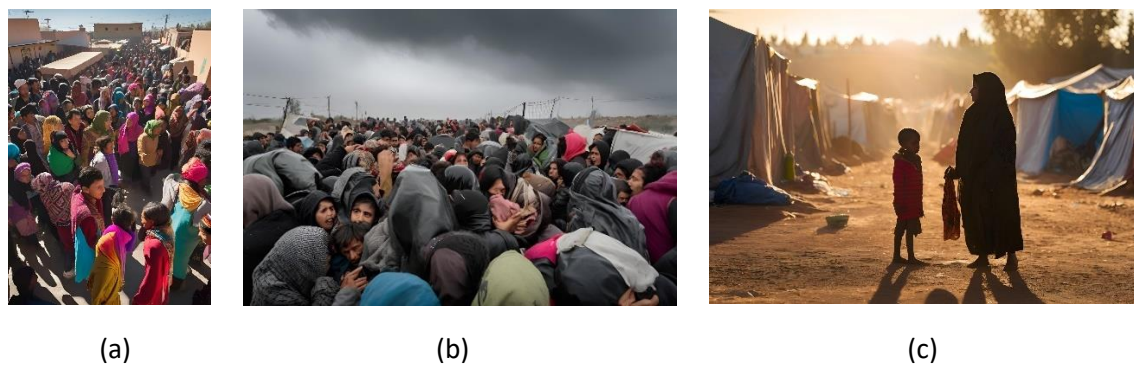
**Figure 11.** Source (a) and goal (b) images

While pedal motion through geographical landscapes is the predominant pattern in images generated by “refugees/migrants + motion verb” prompts, a number of other patterns are also detected and represented in the data with sufficient frequency as to be worthy of discussion. For example, the AI-generated images detect an association between ‘migrants’ and the sea/coast plus boats which is not present for ‘refugees’, suggesting slightly different and ideologically contrasting semantic profiles for the two designations. For images prompted by ‘migrants’, the Figure is shown in the context of the sea/coast in 15.3% of cases ( $n=11$ ) with 54.5% of such cases generated by the prompt ‘migrants arrive’. Interestingly, the AI-generated images appear to be tapping into a specifically European narrative that is preoccupied with the idea of people arriving to countries like Italy, Greece and the UK on small boats. Azavedo et al. (2021) show that images of migrants in sea contexts further amplifies dehumanisation effects and increases perceptions of realistic threat (i.e. threat to the in-group’s economic or political power of physical well-being) compared to land-based images which increase perceptions of symbolic threat (i.e. threat to the in-group’s norms, values and culture).



**Figure 12.** Sea images prompted by ‘migrants arrive’

Refugees and migrants are depicted as stationary in 41% of images. Stationary images include images of people gathered in crowds as in Figure 13a, images of people contained and unable to move as in Figure 13b, and images of people sitting or standing at locations somewhere along their migratory journey as in Figure 13c.



**Figure 13.** Stationary images

A particularly noteworthy kind of stationary image is portrait photographs as in Figure 14. Such images do not so much document the process of migration as they do specific types of people. In doing so, they place refugees and migrants before the camera for inspection in a way that invites curiosity and is comparable to more anthropological forms of photographic documentation, which are used to classify cultures and often construct their subject as an exotic Other or subaltern (Leon-Quijano 2022; Poole 2005). Portrait photos such as those in Figure 14 therefore reinforce a view of refugees and migrants as different from ourselves.





(a)



(b)



(c)

**Figure 14.** Portrait photos

A final point worth discussing is the presence of security features within the Ground as in the images shown in Figure 15. Previous studies have shown that images of migration frequently include a security presence such as border walls, fences, police or military personnel (Catalano and Musolff 2019; Hart 2024; Martínez Lirola 2016, 2022). Again, although such features are relatively infrequent in the current data, they are nevertheless indicative of an association between migration and security that is sufficiently widespread in internet images as to be considered part of the visual discourse of migration. Such images contribute to a securitisation of immigration (Kataba and Jacobs 2023; Lazaridis and Wadia 2015; Vezonik 2018) and thus serve to criminalise displaced people (Martínez Lirola 2016) and further justify the militarisation of borders (Catalano and Musolff 2019).



(a)



(b)



(c)

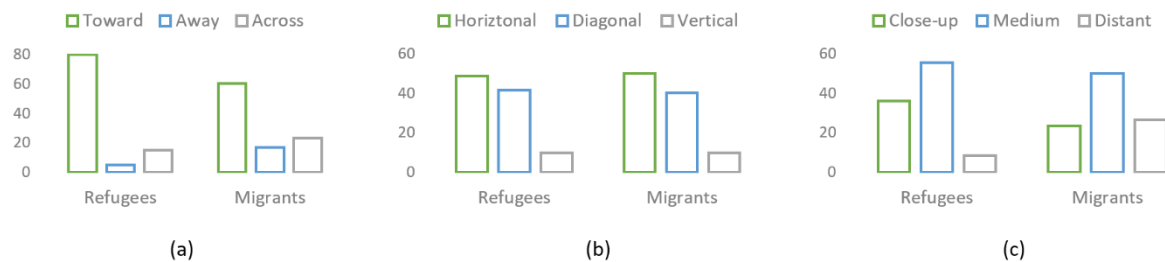
**Figure 15.** Security images

## Viewpoint

*Path:* 25 images displayed either mixed or undiscernible paths. Of the remaining 119 cases, a path toward the viewer is displayed by 69.8% of images, a path away from the viewer is displayed by 10.9% of images, and a path across the view is displayed in 19.3% of images. An exact binomial test confirms that this difference is significant assuming equal proportions ( $p < .001$ ). As shown in Figure 16a, the pattern applies regardless of nomination.

**Angle:** The most common angle, displayed by 49.3% of images, is horizontal with diagonal and vertical angles displayed by 41% and 9.7% of images respectively. The difference is significant assuming equal proportions ( $p < .001$ ). The pattern is consistent across nominations as shown in Figure 16b.

**Distance:** The most common distance prompted by both nominations is a medium shot, which is found in 52.8% of images overall. Close-up shots are found in 29.8% of images while distant shots are found in 17.4% of images. However, as shown in Figure 16c, the prompt ‘refugees’ generates significantly more close-up shots while ‘migrants’ generates more distant shots ( $\chi^2 = 8.4376$ ,  $p < .01$ ).



**Figure 16.** Bar plots showing proportion of viewpoints in three dimensions: (a) path; (b) angle; (c) distance

The functions of viewpoint are examined most extensively by Kress and van Leeuwen (2006) who show how perspectival variables position the viewer with respect to participants in the image and invite the viewer to enter into different kinds of social relation with the subject. Viewpoint is inherently multidimensional with every image presenting simultaneously a perspectival value in path, angle and distance. The functions of any particular viewpoint value are therefore not only sensitive to the social context of the image but depend on other semiotic configurations within it, including viewpoint values in other dimensions.

Of a possible twenty-nine viewpoints combining values for path, angle and distance, the one that is the most common in the data is toward + horizontal + medium-shot, which accounts for a fifth ( $n=29$ , 20.1%) of all images. When the predominant image of a large, continuous, unbounded Figure moving on foot through a landscape is seen from this perspective, as in Figure 17, the effect that arises is a visual form of *proximisation* (cf. Cap 2013) as the mass of people shown approaching the viewer creates a sense of impending threat. However, when the image is of a small group, as in Figure 18, the same viewpoint specification does not produce the same effect with the image instead likely to evoke sympathy. Conversely, when the Figure is a large group but the viewpoint variable of path is away as in Figure 19, the effect is not one of proximisation but of depersonalisation or anonymisation, which discourages feelings of empathy or affiliation toward the Figure.



(a)



(b)



(c)

**Figure 17.** Large-group depiction with viewpoint toward (path) + horizontal (angle) + medium (distance)



**Figure 18.** Small-group depiction with viewpoint toward (path) + horizontal (angle) + medium (distance)



**Figure 19.** Large-group depiction with viewpoint away (path) + horizontal (angle) + medium (distance)

Van Leeuwen (2005: 138) argues that distance too is symbolic where it “indicates the closeness, literally and figuratively, of our relationships”. The AI images identify a contrast in distance between



images associated with the nominations ‘refugees’ versus ‘migrants’. People designated as ‘refugees’ are more likely to be depicted through a close-up shot as in Figure 20a while people designated as ‘migrants’ tend to be depicted through a distant shot as in Figure 20b.



(a)



(b)

**Figure 20.** Distance = close-up (a) versus distant (b)

In the context of migration discourse, close-up shots are the viewpoint “most likely to elicit empathy in viewers” (Wilmott 2017: 74). Long-shots, by contrast, create distance between migrants and the viewer highlighting their “otherness” (ibid.) and “suggest that immigrants’ situation and problems are not ours” (Martínez Lirola 2022: 494). Although both nominations are associated primarily with medium shots, the differential association with close-up versus distal shots again indicates that the two terms, while overlapping, denote ideologically distinct categories which are constructed in part through the types of images they are accompanied by.

While close-up shots as in Figure 20 can evoke sympathy, they can equally invite pity and contribute to an *aestheticisation of suffering* (Chouliaraki 2006), especially when combined with a diagonal viewpoint as in the images in Figure 21. For Chouliaraki (2006: 92), aestheticised suffering “seems to rest on the spectator’s indulgent contemplation of the spectacularity of the scene of human pain”. Such an aestheticization of suffering is evident in the images in Figure 21 in the faces of the mothers and the vulnerability of the babies they are clutching. For Chouliaraki (2006: 82), such images belong to a “regime of pity” which “produces the spectacle of suffering as authentic for spectators”. Recalling the image of Madonna and Child, the image of a mother holding a baby is a particularly iconic symbol of suffering (Chouliaraki 2006: 57).



**Figure 21.** Close-up shots

A difference in degree of angle is the difference between pity and superiority. Images with a diagonal angle account for two fifths of the AI generated images, suggesting that a downward angle, while not quite as frequent as a horizontal angle, is a common feature of internet images of migration. From such an elevated position, as in the images in Figure 22, the viewer is literally and metaphorically "looking down" on the subject where, as van Leeuwen (2008: 139) states, "to look down on someone is to exert imaginary symbolic power over that person". In images like those in Figure 22, refugees and migrants are therefore subjugated or disempowered suggesting the right of more powerful actors to determine their freedom and autonomy. The image in Figure 22b, showing people contained, is even capable of being read in a way that compares refugees and migrants to penned animals.



**Figure 22.** Angle = diagonal.

## Conclusion

AI generated images detect and reflect patterns of representation in the millions of images that occur together with specific text forms across the internet. They are therefore a useful diagnostic tool for investigating multimodal constructions which shortcuts the need to collate and annotate massive multimodal corpora. In the context of social and political discourses, AI images reveal patterns of visual



representation responsible for reinforcing stereotypes and prejudices, stoking fear and hostility, and thus legitimating and sustaining harmful and discriminatory policies and practices.

The image in a multimodal construction is not a specific one but a schematic one. Neither does it necessarily represent an actual image or images. Rather, it is a bundle of features derived probabilistically as a function of features presented across different images. This composite form therefore stands as a prototype which instantiations may deviate from in one or more respect and which is not necessarily realised in every respect even by the majority of images. Analysis of the AI-images in the context of migration suggests that the [refugees/migrants + motion verb] construction has as its counterpart in a multimodal construction the image of a large, continuous, unbounded Figure moving on foot through a landscape toward the viewer. This image contributes to the construction of migration as a substantive and unrelenting 'problem' that directly affects the addressee and to construals of refugees and migrants which ignore their individual characteristics and identities.

Multimodal constructions do not preclude other forms of visual representation from also regularly occurring with verbal forms and these alternative patterns are also constitutive of the visual discourses surrounding a particular issue. Indeed, the AI images analysed here suggest several other ideologically significant forms of representation that also abound on the internet and thus make up part of the multimodal discourse on migration. For example, the AI images suggest that refugees and migrants are often criminalised through securitising Grounds or denied corporeal identities through more abstract Figure forms.

It should be noted that the type of images identified and analysed in the present study are the ones associated with plural noun + motion verb constructions. Singular noun + motion verb constructions are likely to have as their counterpart in a multimodal construction other visual forms, which present an alternative, potentially more positive, discourse. It is plural noun + motion verb constructions, however, which figure more frequently in hegemonic discourses of migration and which are therefore especially worthy of investigation.

The analysis also shows that while the [refugees/migrants + motion verb] construction exists at one level of schematicity, where it is conventionally associated with a particular visual form, other visual forms associated with the construction differ depending on the specific noun or verb in the relevant slot. For example, images prompted by 'refugees' are more likely to be close-up shots and engage in a politics of pity while images prompted by 'migrants' are more likely to be distant shots making migrants a remote concern. The AI-generated images also highlight a (Eurocentric) connection between 'migrants' and the sea/coast which is not present for 'refugees'. Aligning with previous research showing that the terms *refugees* and *migrants* collocate with different verbal forms (Gabrielatos and Baker 2008), this suggests that two designations are associated with overlapping but distinct sets of visual forms, and thus denote ideologically distinct categories constructed through recurring relationships with visual as well as verbal forms. The verb *flood* also seems to yield a unique type of image, though this is less faithful to attested images and is more reflective of the creative potential of AI. This perhaps points to a limitation of the study. Exactly how AI works is, for most discourse analysts and non-computer scientists in general, something of an enigma. The extent to which it is truly representative of all images on the internet or only of a specific sub-set is unclear. A further limitation is that only one AI text-image generating tool was used in the study. It remains unclear whether other tools would detect the same patterns and thus produce the same results. Further investigation is required to understand the extent to which different tools are convergent in the patterns they uncover. Interestingly, however, the patterns of representation detected by the AI-generated images in the present study confirm those found in previous studies involving much smaller, more targeted datasets. There are also ethical concerns in using AI-images as a tool in academic research. By generating new

images and publishing them on the internet, however small the impact may be and even though for purposes of critical exposition, I am contributing back to the image pool the very type of images I am problematising and therefore further biasing the models on which AI text-to-image generation works. Notwithstanding such issues and limitations, I hope in this paper to have demonstrated how AI text-to-image technology may be harnessed in research investigating multimodal constructions and the multimodal construction of society, social identities and social relations.

## References

- Abid, R.Z., Manan, S.A. and Rahman, Z.A.A.A. (2017). 'A flood of Syrians has slowed to a trickle': The use of metaphors in the representation of Syrian refugees in the online media news reports of host and non-host countries. *Discourse & Communication* 11 (2): 121–140.
- Antonopoulou, E. and Nikiforidou, K. (2011). Construction grammar and conventional discourse: A construction-based approach to discursial incongruity. *Journal of Pragmatics* 43: 2594-2609.
- Arpan, L.M., Baker, K., Lee, Y., Jung, T., Lorusso, L. and Smith, J. (2006). News coverage of social protests and the effects of photographs and prior attitudes. *Mass Communication* 9 (1): 1-20.
- Azevedo, R.T., De Beukelaer, S., Jones, I.L., Safra, L. and Tsakiris, M. (2021). When the lens is too wide: The political consequences of the visual dehumanization of refugees. *Humanities and Social Sciences Communications* 8: 115.
- Banks, J. (2012). Unmasking deviance: The visual construction of asylum seekers and refugees in English national newspapers. *Critical Criminology* 20 (3): 293-310.
- Barazida, N. (2022). Stable Diffusion: Best Open Source Version of DALL-E 2. Medium. Towards Data Science (blog). [https://towardsdatascience.com/stable-diffusion-best-open-source-version-of-dall-e-2-ebcdf1cb64bc\(open in a new window\)](https://towardsdatascience.com/stable-diffusion-best-open-source-version-of-dall-e-2-ebcdf1cb64bc(open%20in%20a%20new%20window)).
- Batziau, A. (2011). Framing 'otherness' in press photographs: The case of immigrants in Greece and Spain. *Journal of Media Practice* 12 (1): 41-60.
- Bianchi, F., Kalluri, P., Durmus, E., Ladhak, F., Cheng, M., Nozza, D., Hashimoto, T., Jurafsky, D., Zou, J. and Caliskan, A. (2023). Easily accessible text-to-image generation amplifies demographic stereotypes at large scale. *FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*: 1493-1504.
- Bleiker, R., Campbell, D., Hutchison, E. and Nicholson, X. (2013). The visual dehumanisation of refugees. *Australian Journal of Political Science* 48 (4): 398-416.
- Bucher H-J. and Schumacher P. (2006). The relevance of attention for selecting news content: An eye-tracking study on attention patterns in the reception of print and online media. *Communications* 31 (3): 347–368.
- Burscher, B., van Spanje, J. and De Vreese, C.H. (2015). Owning the issues of crime and immigration: The relation between immigration and crime news and anti-immigrant voting in 11 countries. *Election Studies* 38: 59–69.
- Cap, P. (2013). *Proximization: The Pragmatics of Symbolic Distance Crossing*. John Benjamins.

- Caple, H. and Knox, J.S. (2015). A framework for the multimodal analysis of online news galleries: What makes a “good” picture gallery. *Social Semiotics* 25 (3): 292-321.
- Catalano, T. and Musolff, M. (2019). “Taking the shackles off”: Metaphor and metonymy of migrant children and border officials in the U.S. *Metaphorik.de* 29.
- Charteris-Black, J. (2006). Britain as a container: Immigration metaphors in the 2005 election campaign. *Discourse & Society* 17(6): 563-582.
- Chkhaidze, A., Buyruk, P., and Boroditsky, L. (2021). Linguistic metaphors shape attitudes towards immigration. *Proceedings of the Annual Meeting of the Cognitive Science Society* 43.
- Chouliaraki, L. (2006). *The Spectatorship of Suffering*. Sage.
- Cienki, A. (2017). Utterance Construction Grammar (UCxG) and the variable multimodality of constructions. *Linguistics Vanguard* 3(s1): 20160048.
- Cisneros, D.J. (2008). Contaminated communities: The metaphor of “immigrant as pollutant” in media representations of immigration. *Rhetoric and Public Affairs* 11 (4): 569-601.
- Cohn, N. and Schilperoord, J. (2024). *A Multimodal Language Faculty: A Cognitive Framework for Human communication*. Bloomsbury.
- Collier, J.R., Dunaway, J. and Stroud, N.J. (2021). Pathways to deeper news engagement: Factors influencing click behaviors on news sites. *Journal of Computer-Mediated Communication* 26 (5): 265–283.
- Conzo, P., Fuochi, G., Anfossi, L., Spaccatini, F., and Onesta Mosso, C. (2021). Negative media portrayals of immigrants increase ingroup favoritism and hostile physiological and emotional reactions. *Sci. Rep.* 11:16407.
- Dancygier, B. and Vandelanotte, L. (2017). Internet memes as multimodal constructions. *Cognitive Linguistics* 28 (3): 565-598.
- Djourelouva, M. (2023). Persuasion through slanted language: Evidence from the media coverage of immigration. *American Economic Review* 113 (3): 800-835.
- Dolores Porto, M. (2022). Water metaphors and evaluation of Syrian migration: The flow of refugees in the Spanish press. *Metaphor and Symbol* 37 (3): 252-267.
- El Refaie, E. (2001). Metaphors we discriminate by: Naturalized themes in Austrian newspaper articles about asylum seekers. *Journal of Sociolinguistics* 5 (3): 352-371.
- Farris, E.M. and Silber Mohamed, H. (2018). Picturing immigration: How the media criminalizes immigrants. *Politics, Groups, and Identities* 6 (4): 814-824.
- Forceville, C. (2008). Metaphor in pictures and multimodal representations. In R.W. Gibbs, Jr (ed.), *The Cambridge Handbook of Metaphor and Thought*. Cambridge University Press. pp.462-482.
- Forceville, C. (2009). Non-verbal and multimodal metaphor in a cognitivist framework: Agendas for research. In C. Forceville and Urios-Aparisi (eds.), *Multimodal Metaphor*. Mouton de Gruyter. pp.19-42.

- Gabrieletos, C. and Baker, P. (2008). Fleeing, sneaking, flooding: A corpus analysis of discursive constructions of refugees and asylum seekers in the UK press 1996-2005. *Journal of English Linguistics* 36 (1): 5-38.
- Garry, M., Strange, D., Bernstein, D.M. and Kinzett, T. (2007). Photographs can distort memory for the news. *Applied Cognitive Psychology* 21 (8): 995-1004.
- Geise, S., Panke, D., and Heck, A. (2021). Still images-moving people? How media images of protest issues and movements influence participatory intentions. *Int. J. Press* 26: 92–118.
- Groom, N. (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.
- Goldberg, A.E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences* 7 (5): 219-224.
- Goldberg, A.E. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford University Press.
- Graber, D.A. (1990). Seeing is remembering: How visual contribute to learning from television news. *Journal of Communication* 40 (3): 134-156.
- Guilbeault, D. Delecourt, S., Hull, T., Desikan, B.S., Chu, M. and Nadler, E. (2024). Online images amplify gender bias. *Nature* 626: 1049-1055.
- Hart, C. (2021). Animals vs. armies: Resistance to extreme metaphors in anti-immigration discourse. *Journal of Language and Politics* 20 (2): 226-253.
- Hart, C. (2024). Multimodal meaning making in news communication about immigration: Using the NewsScape corpus to explore co-verbal images in TV news. *Frontiers in Communication* 9: 1451105.
- Hart, C. (2025). *Language, Image, Gesture: The Cognitive Semiotics of Politics*. Cambridge University Press.
- Hart, C. and Marmol Queralto, J. (2021). What can cognitive linguistics tell us about language-image relations? A multidimensional approach to intersemiotic convergence in multimodal texts. *Cognitive Linguistics* 32 (4): 529–562.
- Hockley, W. E. (2008). The picture superiority effect in associative recognition. *Mem. Cognit.* 36: 1351–1359.
- Hoffman, T. (2021). Multimodal construction grammar: From multimodal constructs to multimodal constructions. In W. Xu and J.R. Taylor (eds.), *The Routledge Handbook of Cognitive Linguistics*. Routledge. pp.78-92.
- Holsanova, J. and Nord, A. (2010). Multimodal design: Media structures, media principles and users' meaning-making in printed and digital media. In H-J. Bucher, T. Gloning and K. Lehnen (eds.), *Neue Medie - neue Formate. Ausdifferenzierung und Konvergenz in der Medienkommunikation*. Campus Verlag. pp.81-103.
- Jenni, K.E. and Lowenstein, G. (1997). Explaining the “identifiable victim effect”. *Journal of Risk and Uncertainty* 14: 235-257.

- Jewitt, C. (2009). Introduction: Handbook rationale, scope and structure. In C. Hewitt (ed.), *Routledge Handbook to Multimodal Analysis*. Routledge. pp.1-7.
- Jiminez, T., Arndt, J. and Landau, M.J. (2021). Walls block waves: Using an inundation metaphor of immigration predicts support for a border wall. *Journal of Social and Political Psychology* 9 (1): 159-171.
- Joyce, N. and Harwood, J. (2014). Improving intergroup attitudes through televised vicarious intergroup contact: social cognitive processing of ingroup and outgroup information. *Commun. Res.* 41: 627–643.
- Kataba, M. and Jacobs, A. (2023). The ‘migrant other’ as a security threat: The ‘migration crisis’ and the securitising move of the Polish ruling party in response to the EU relocation scheme. *Journal of Contemporary European Studies* 31 (4): 1223-1239.
- Kok, K. and Cienki, A. (2016). Cognitive grammar and gesture: Points of convergence, advances and challenges. *Cognitive Linguistics* 27(1): 67–100.
- Kress, G. and van Leeuwen, T. (2006). *Reading Images: The Grammar of Visual Design*. 2<sup>nd</sup> end. Routledge.
- Langacker, R.W. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford University Press.
- Lazaridis, G. and Wadia, K. (eds.) (2015). *The Securitisation of Migration in the EU: Debates since 9/11*. Palgrave.
- Leckner, S. (2012). Presentation factors affecting reading behaviour in readers of newspaper media: an eye-tracking perspective. *Visual Communication* 11(2): 163–184.
- Lee, S., and Hugh Feeley, T. (2016). The identifiable victim effect: A meta-analytic review. *Soc. Infl.* 11: 199–215
- Lehmann, C. (2024). What makes a multimodal construction? Evidence for a prosodic mode in spoken English. *Frontiers in Communication* 9: 1338844.
- Leon-Quijano, C. (2022). Why do “good” pictures matter in anthropology? *Cultural Anthropology* 37 (3): 572-598.
- Li, Y. and Xie, Y. (2020). Is a picture worth a thousand words? An empirical study of image content and social media engagement. *Journal of Marketing Research* 57 (1): 1-19.
- López, R.M. (2020). Discursive de/humanizing: A multimodal critical discourse analysis of television news representations of undocumented youth. *Education Policy Analysis Archives* 28 (47):
- Machin, D. (2004). Building the world’s visual language: The increasing global importance of image banks in corporate media. *Visual Communication* 3 (3): 316-336.
- Machin, D. (2013). What is multimodal critical discourse studies? *Critical Discourse Studies* 10 (4): 347-355.
- Machin, D. (2016). The need for a social and affordance-driven multimodal critical discourse analysis. *Discourse & society* 27 (3): 322-334.
- Machin, D. and Mayr, A. (2012). *How To Do Critical Discourse Analysis: A Multimodal Introduction*. Sage.

- Madrigal, G. and Soroka, S. (2023). Migrants, caravans, and the impact of news photos on immigration attitudes. *The International Journal of Press/Politics* 28 (1): 49-69
- Marshall, S.R. and Shapiro, J.R. (2018). When “scurry” vs. “hurry” makes the difference: Vermin metaphors, disgust and anti-immigrant attitudes. *Journal of Social Issues* 74 (4): 774-789.
- Martínez Lirola, M. (2016). Linguistic and visual strategies for portraying immigrants as people deprived of human rights. *Social Semiotics* 27 (1): 21-38.
- Martínez Lirola, M. (2022). Critical analysis of dehumanizing news photographs on immigrants: Examples of the portrayal of non-citizenship.
- Martínez Lirola, M. and Zammit, K. (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.
- McGlashan, M. (2021) Linguistic and visual trends in the representation of two-mum and two-dad couples in children’s picture books. In A.J. Moya-Guijarro and E. Ventola (eds.), *A Multimodal Approach to Challenging Gender Stereotypes in Children’s Picture Books*. Routledge. pp.268-305.
- Musolff, A. (2015). Dehumanizing metaphors in UK immigrant debates in press and online media. *Journal of Language Aggression and Conflict* 3 (1): 41-56.
- Noble, S.U. (2018). *Algorithms of Oppression*. New York University Press.
- Olier, J. S., and Spadavecchia, C. (2022). Stereotypes, disproportions, and power asymmetries in the visual portrayal of migrants in ten countries: an interdisciplinary AI-based approach. *Human. Soc. Sci. Commun.* 9:410.
- Poole, D. (2005). An excess of description: Ethnography, race, and visual technologies. *Annual Review of Anthropology* 34: 159-179.
- Pourcel, S. (2010). Motion: A conceptual typology. In V. Evans and P. Chilton (eds.), *Language, Cognition and Space: The State of the Art and New Directions*. Equinox. pp.419-450.
- Powell, T.E., Boomgaarden, H.G. De Swert, K. and de Vreese, C.H. (2015). A clearer picture: The contribution of visuals and text to framing effects. *Journal of Communication* 65 (6): 997-1017.
- Putland, E., Chikodzore-Paterson, C. and Brookes, G. (2023). Artificial intelligence and visual discourse: A multimodal critical discourse analysis of AI-generated images of “dementia”. *Social Semiotics* 1-26.
- Quinn S., Stark P. and Edmonds R. (2007). *Eyetracking the News: A Study of Print and Online Reading*. The Poynter Institute.
- Richardson, J.E. 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.
- Romano, M. and Dolores Porto, M. (2021). Framing CONFLICT in the Syrian refugee crisis: Multimodal representations in the Spanish and British press. In L. Filardo- Llamas, E. Morales- López and A. Floyd (eds.) *Discursive Approaches to Sociopolitical Polarization and Conflict*. Routledge. pp.153-173.

- Santa Ana, O. (1999). 'Like an animal I was treated': Anti-immigrant metaphor in US public discourse. *Discourse & Society* 10 (2): 191-224.
- Santa Ana, O. (2002). *Brown Tide Rising: Metaphors of Latinos in Contemporary American Public Discourse*. University of Texas Press.
- Schemer, C. (2012). The influence of news media on stereotypic attitudes toward immigrants in a political campaign. *Journal of Communication* 62: 739–757.
- Schemer, C. and Meltzer, C.E. (2019). The impact of negative parasocial and vicarious contact with refugees in the media on attitudes toward refugees. *Mass Communication and society* 23 (2): 230-248.
- Scherman, A., Etchegaray, N., Pavez, I. and Grassau, D. (2022). The influence of media coverage on the negative perception of migrants in Chile. *International Journal Environmental Research and Public Health* 19 (13): 8219.
- Schoonjans, S. (2018). *Modalpartikeln als multimodale Konstruktionen: Eine korpusbasierte Kookkurrenzanalyse von Modalpartikeln und Gestik im Deutschen*. Berlin: De Gruyter.
- Slobin, D.I. (2004). The many ways to search for a frog: Linguistic typology and the expression of motion events. In S. Strömquist and L. Verhoeven (eds.), *Relating Events in Narrative, Vol. 2. Typological and Contextual Perspectives*. Lawrence Erlbaum Associates Publishers. pp.219-257.
- Steen, F. and Turner, M. (2013). Multimodal construction grammar. In M. Borkent, B. Dancygier and J. Hinnell (eds.), *Language and the Creative Mind*. CSLI Publications. pp. 255-274.
- Stocchetti, M. (2011). Images: Who gets what, when and how? In M. Stocchetti and K. Kukkonen (eds.), *Images in Use: Towards the Critical Analysis of Visual Communication*. John Benjamins. pp.11-38.
- Stöckl, H. (2020). Linguistic multimodality – multimodal linguistics: A state-of-the-art sketch. In J. Wildfeuer, J. Pflaeging, J. Bateman, O. Seizov, and C. I. Tseng (eds.), *Multimodality: Disciplinary Thoughts and the Challenge of Diversity*. DeGruyter. pp. 41–68.
- Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language* 2 (1): 23-55.
- Stubbs, M. (1996). *Text and Corpus Analysis*. Blackwell.
- Stubbs, M. (2002). *Words and Phrases: Corpus Studies of Lexical Semantics*. Blackwell.
- Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (ed.), *Language Typology and Syntactic Description*. Cambridge University Press. pp. 57–149.
- Talmy, L. (2000). *Toward a Cognitive Semantics*. MIT Press.
- Uhrig, P. (2022). Hand gestures with verbs of throwing: collostructions, style and metaphor. *Yearbook of the German Cognitive Linguistics Association* 10: 99–120.
- Utych, S.M. (2018). How dehumanization influences attitudes toward immigrants. *Political Research Quarterly* 7 (2): 440-452.
- Van Leeuwen, T. (2005). *Introducing Social Semiotics*. Routledge.

- Van Leeuwen, T. (2008). *Discourse and Practice: New Tools for Critical Discourse Analysis*. Oxford University Press.
- Van Leeuwen, T. (2014). Critical discourse analysis and multimodality. In C. Hart and P. Cap (eds.), *Contemporary Critical Discourse Studies*. Bloomsbury. pp.281-296.
- Vezonik, A. (2018). Securitising migration in Slovenia: A discourse analysis of the Slovenian refugee situation. *Journal of Immigrant & Refugee Studies* 16 (1): 39-56.
- Westberg, G. and Kvåle, G. (2024). The generic uniqueness of AI imagery: A critical approach to Dall-E as semiotic technology. *Discourse & Society*.
- Wilmott, A.C. (2017). The politics of photography: Visual depictions of Syrian refugees in U.K. online media. *Visual Communication Quarterly* 24 (2): 67-82.
- Zhang, L. and Rui, Y. (2013) Image search – from thousands to billions in 20 years. *ACM Trans. Multimedia Comput. Commun. Appl.* 9: 1–20.
- Ziem, A. (2017). Do we really need a multimodal construction grammar? *Linguistics Vanguard* 3: 20160095.
- Zima, E. (2017). On the multimodality of [all the way from X PREP Y]. *Linguistics Vanguard* 3: 20160055.
- Zima, E. and Bergs, A. (2017). Towards a multimodal construction grammar. *Linguistics Vanguard* 3 :20161006.
- Zlatev, J., Blomberg, J. and David, C. (2010). Translocation, language and the categorization of experience. In V. Evans and P. Chilton (eds.), *Language, Cognition and Space: The State of the Art and New Directions*. Equinox. pp.389-418.