Applied Corpus Linguistics 3 (2023) 100043

Contents lists available at ScienceDirect

Applied Corpus Linguistics

journal homepage: www.elsevier.com/locate/acorp

Articles

Creating and analysing a multimodal corpus of news texts with Google Cloud Vision's automatic image tagger

Paul Baker*, Luke Collins

Department of Linguistics and English Language, Lancaster University, Lancaster, UK

ARTICLE INFO

Keywords:
News
Annotation
Visual
Image
Obesity
Discourse

ABSTRACT

This study describes the creation and analysis of a small multimodal corpus of British news articles about obesity, where tags were assigned to images in the articles using the automatic tagger Google Cloud Vision. In order to illustrate the potential for analysis of image tags, the corpus analysis tool WordSmith was used to identify differences between newspapers in the ways that obesity was framed. Three forms of analysis were carried out – the first simply compared keywords across the newspapers, the second examined key visual tags and their collocates associated with each newspaper, while the third incorporated a combined analysis of words and image tags. The three analyses produced complementary findings, indicating the value in using Google Cloud Vision in creating and analysing multimodal corpora. The paper ends by reflecting on the method undertaken, while considering how additional research could improve our understanding of image tagging.

1. Introduction

Representations of people with obesity in the press have been shown to stigmatise and contribute to shaming discourses while using a 'personal responsibility' frame and offering minimal coverage of body positive messaging (Brookes and Baker, 2021). Stigmatising representations of obesity have been identified in both the language content and the visual materials that are presented alongside the verbal commentary in news content (Heuer et al., 2011). Corpus linguistics has been applied in the identification and examination of language patterns that establish and reinforce dominant views about obesity (Bączkowska, 2020; Brookes and Baker, 2021), however large-scale analyses of media discourses using corpus approaches have typically investigated only the written text of news articles, as a result of the challenges of documenting image (and other visual) content in a machine-readable format. While some studies have conducted multimodal analyses of news texts, these have tended to analyse text and image independently, before considering the interrelations between different modes (e.g., Bednarek and Caple, 2017), or have oriented towards more detailed, manual qualitative analysis that is difficult to extrapolate to a representative sample (Heuer et al., 2011). In this study we present an approach to corpus creation that facilitates a multimodal, cross-publication investigation of representations of obesity in the UK press. We compiled a small corpus of news articles on the topic of obesity and demonstrate how automatic annotation of the images included in those articles can be incorporated into the text files in order to enable corpus analysis of text and image relations. We consider the different strategies of national UK newspapers and reflect on how our approach can be applied further as a form of corpus-assisted multimodal discourse analysis.

2. Literature review

Systematic studies of the coverage of obesity across a range of national news contexts have long-established that the news media focuses on individual behaviours rather than societal factors (such as limited access to healthy food options in low-income neighbourhoods), and on behaviour-change-oriented solutions (such as dieting and exercising) rather than system-level or policy changes (Heuer et al., 2011; Gollust et al., 2012; Nimegeer et al., 2019). Brookes and Baker (2021) remark that the pervasive 'personal responsibility' frame (which blames individuals for obesity) was particularly characteristic of politically right-leaning tabloids, and increasing over time while left-leaning broadsheets were more likely than other newspapers to attribute obesity to social causes like economic inequality. This indicates that discussions of obesity are situated within wider ideological discourses and that readers may receive different messages about obesity, depending on which publication they engage with.

Stigmatising representations of obesity are also found in visual depictions of people, objects and concepts. Heuer et al. (2011), for example, analysed 500 online news website photographs from five major news

* Corresponding author.
E-mail address: p.baker@lancaster.ac.uk (P. Baker).

https://doi.org/10.1016/j.acorp.2023.100043
Received 2 December 2022; Received in revised form 13 January 2023; Accepted 24 January 2023
Available online 26 January 2023

websites and found that 72% of images met at least one of the criteria they determined to be stigmatising. These criteria included emphasising the abdomen, images of people not wearing clothes or wearing ill-fitting clothes, not showing heads/faces, or showing people eating unhealthily and being sedentary. They found that such images are less likely to show people with obesity as wearing professional clothing, fully clothed or exercising compared to pictures of people who do not have obesity (Heuer et al., 2011). In a study of online news videos about people with obesity/overweight, Puhl et al. (2013, p.686) similarly found that 65% of adults and 77% of young people were portrayed in a negative, stigmatizing manner.

The communicative potential of images is well-acknowledged; as Rafiee et al. (2021, p.1) argue, '[n]ews images call attention and raise awareness with an immediacy that text cannot easily achieve' in that they can offer a clear representation of some aspect of the story that is immediately recognisable. Furthermore, images are effective in 'increasing [a] mass audience's emotional reaction toward social events' (Rafiee et al., 2021, p.1). McClure et al. (2011) have shown that the selection of an image to accompany a 'neutral' news article can affect audiences' wider perceptions of people with obesity. In their experimental study, they found that when participants viewed a photograph that negatively portrayed a person with obesity, they subsequently reported higher levels of (negative) weight bias and that even when presented with a more positive image, respondents reported levels of weight bias consistent with what has been reported elsewhere e.g. by Bacon et al. (2001), McClure et al. (2011). As with language, we can critically examine the choices that text creators make in offering particular images that create and promulgate certain kinds of social practice. In this study, we consider the range of representations offered by news publications in the UK, taking into account both the written and image content.

2.1. Image annotation

One of the challenges of annotating non-verbal elements like images, highlighted by Knight and Adolphs (2020), is that they are not ready-made 'textual units' in the way that we can conceptualise tokens, lemma, or n-grams. Indeed, Rashtchian et al. (2010) highlight issues with quality control in their crowdsourced attempts to conduct large-scale image annotation, which related not only to how the images were captioned but also with subsequent crowdsourced views on the quality of those captions. Others have also highlighted issues with consistency in manual coding, such as Gatt et al. (2018), who crowdsourced annotators to document features of images of faces but who varied in the extent to which they described physical, emotional and inferential attributes.

Manual approaches to image annotation informed by social semiotic theory (Halliday, 1978) tend to apply some form of thematic classification that reduces the images to a small set of distinguishable codes. Tan et al. (2020), for example, apply a superordinate categorisation that identifies 'protest', 'political/financial', 'poverty/austerity' and 'miscellaneous' images in a corpus of news texts concerned with austerity measures in the UK. They then determine subcategories according to the inclusion of various social actors, actions and portrayed scenes. Collins (2020) demonstrates a simple classification for images in Facebook posts, along with annotation of emoji, where the use of a single summative label facilitates keyword and collocation analysis according to standard corpus procedures. Such an approach is necessarily reductive of the various image elements and the means and findings of the study have little generalisability beyond the study data. These simplified categorisation schemes – along with crowdsourcing projects – go some way to minimising the labour-intensiveness of the annotation process. However, in addition to issues with homogeneity, they also tend to be bespoke taxonomies that would not be applicable to other datasets. Thus, there would be value in developing a coding scheme that finds a balance between offering sufficient detail to document meaningful elements of image data but that can be applied generally across studies. Furthermore, Christiansen et al. (2020) also advocate that an effective approach to quantitative image analysis must support contextual analysis, in affording a view of the interplay between text and image, as well as simplicity. So there should be minimal re-adjustment of existing (manual) practices in order to encourage widespread adoption (as people are unlikely to use schemes that are too labour-intensive). In pursuit of these aims, Christiansen et al. (2020, p.150) discuss their Visual Constituent Analysis (VCA), as a means of corpus construction that uses 'a machine learning approach to annotating images and constructing large-scale multimodal text-image corpora'. Our study builds on this work by utilising the same automated tool, Google Cloud's Vision (discussed further below), to our corpus of news articles about obesity, in order to test its applicability to corpus-assisted discourse studies (Partington et al., 2013) and demonstrate how the interplay between text and images can be analysed.

2.2. Text-image relations in corpus analysis

Having established that verbal and image content both have important contributions to make with respect to the representation of important health and social issues like obesity, we must consider how we can record and investigate the relations between them, within a given text. In corpus linguistics, we see the development of tools such as the UAM CorpusTool (O'Donnell, 2008), which have been designed to support the custom annotation of files during corpus construction (e.g. with metadata or text mark-up) and which have been used by, for example, Bednarek and Caple (2017) as part of a relational database which aligns two modes that supports corpus-assisted multimodal discourse analyses (CAMDA). Their (independent) coding of text and images according to news values enabled them to assess how the news values activated in one mode reinforced, complemented or clashed with those activated in another mode (Bednarek and Caple, 2017), further validating the proposition that different conclusions would be drawn if these elements were analysed separately. Bednarek and Caple's (2017) relational database does not include information about the position of each element in the original text. However, since their analysis focuses on the content that appears at the head of the news text – they discuss headlines, opening paragraphs and images that appear at the top of articles – they can assume close proximity.

McGlashan (2021, p.218) reviews the literature in which researchers have theorised the relationship between text and image, such as 'anchorage (where words make the contents of an image more specific), illustration (where images support or provide more details about cooccurring text), and relay (symmetrical relationships wherein words and images work together equally)'. Through his exploration of what he calls 'collustration' (a portmanteau of 'collocation' and 'illustration') in children's books featuring same-sex partnerships, McGlashan (2021, p.231) asserts that conventional forms of dress associated with (Western) weddings 'index [an] ideologically homonormative version of same-sex relationships, including an affirmation of hegemonic constructions of gender'. This concept of collustration emphasises the spatial relations that result from the juxtaposition of verbal content and image and this will inform how corpus files that include multimodal elements are prepared.

Christiansen et al.'s (2020) VCA approach to studying a large database of tweets involves concatenation of the outputs from the automated annotation tool Google Cloud Vision, alongside the typed content of the tweets in which they appeared, thus resembling concordance lines. Compared to our study, the issue of the spatial relationship is less consequential for them since they are looking at tweets and can include the full verbal content alongside the image tags. A tweet can reasonably be represented as: full text plus image, with no interruption to the verbal content. Our case is different. Furthermore, they explain that the positioning of these text elements in their concordance lines is not consequential for their analysis. Christiansen et al. (2020) do not analyse

Table 1
Articles about obesity.

Publication	Articles	Words	Image files
Mail	48	45,861	160
Express	25	13,972	65
Star	12	5563	49
Mirror	18	10,386	43
Independent	17	10,038	15
Guardian	8	6862	13
Times	11	5304	8
Telegraph	10	6132	5
	149	**104,118**	**358**

the Vision image tags, rather these allow them to create a downsampled corpus (on the basis of including specified elements, i.e., 'Donald' and/or 'Trump'), with corpus analysis of the verbal content. Our work extends beyond this by analysing the tags, in combination with the verbal content of the news articles. We also take into account their position within the article for collocation analysis. Our analysis of news articles thus operates on the presumption that the ordering and positioning of images is consequential for how the article is read and that readers are encouraged to perceive correspondences between the image and the typed text that immediately precedes and follows it i.e., the co-text. As such, we discuss the practical challenges of incorporating multiple tags for image content in its appropriate position in relation to verbal content (based on how it appeared in the original article). This is one aspect of our corpus construction that has implications for how we investigate text-image relations in the representation of obesity and part of our methodological approach, which we now turn to.

3. Methods

3.1. Corpus construction

Our aim was to construct a corpus that would enable us to test the applicability of an automated image annotation scheme to a multimodal investigation of obesity in news texts. We therefore set out to collect a modest corpus of articles about obesity that we could reasonably process and analyse in sufficient detail with respect to verbal content, image content and text-image relations. We searched the online news repository *LexisNexis* for articles containing the search terms 'obese' or 'obesity', restricting our query to UK National Newspapers (online daily and Sunday editions) and covering the month of September 2022. Brookes and Baker (2021) have shown that there is monthly variation in the UK news coverage of obesity, according to religious traditions and the weather, for example. September was shown to be a month less-conspicuously characterised by weather and calendar events – the only timely factors appearing to be the return to school and the annual conference for major political parties – and so we anticipated that our data would not be affected too significantly by seasonal trends.

Following manual deduplication and relevance-checking, we identified a corpus of 149 articles related to obesity, which were distributed according to publication as shown in Table 1. The *LexisNexis* archive does not include images and so we subsequently had to manually source the images that appeared in the articles from our corpus. We encountered some accessibility issues in this respect, in that some articles were behind a paywall (which was frequently the case for articles from the Times). However, in those instances, it was typically possible to collect the image that appeared at the head of the article (15 instances, 10.07%). For 121 articles (81.21%), we were able to determine and collect the full number of images that appeared, even if the number was zero. In twelve instances (8.05%), we were unable to retrieve the original article and therefore unable to determine how many images it included. Our resulting image corpus comprised 358 images, which we processed using the automatic image annotation tool, Google Cloud Vision (henceforth Vision), to generate the tags that would be added to the text files and included in our subsequent multimodal analysis.

3.2. Automatic image annotation

Vision constitutes one of a range of services that utilises Google's proprietary machine learning algorithms to, in this instance, provide automatic image annotation. A free demo allows images to be tagged, one at a time, although signing up for an account enables batches of images to be tagged (currently, the first 1000 images can be tagged for free, after that, it costs $1.50 to tag each subsequent batch of 1000 images).[1]

The tool supports a machine-learning service that enables users to upload files and annotations to 'train' the platform to apply the annotation to a wider dataset. However, it already has in place a range of annotation categories including facial recognition, object localisation and text detection. With a view to testing the applicability of a ready-made, automated system for the purposes of multiple research projects, we applied Vision's existing coding categories (for a demonstration of the creation and application of custom annotation, see Bisong, 2019).

The standard coding categories currently provided by Vision are as follows:

- Face detection: locating faces with bounding polygons (which identify the perimeter of the detected element according to the coordinates of the image) according to facial landmarks, such as eyes, ears, nose, mouth, etc. Also returns likelihood ratings for the emotion categories joy, sorrow, anger, surprise and for general image properties i.e., underexposed, blurred, headwear present.
- Image properties: returning the dominant colours of an image.
- Label detection: providing a generalized, textual description for the image as a whole and a confidence score for each element.
- Landmark detection: providing the name of a landmark, a confidence score and coordinates for the detected entity.
- Logo detection: providing a textual description, confidence score and bounding polygon for recognised logos in the image.
- Object localisation: providing a textual description, confidence score and bounding box annotation for multiple objects recognised in a single image.
- SafeSearch annotation: providing likelihood ratings for the explicit content categories adult, spoof, medical, violence and racy.
- Text detection: providing optical character recognition.

One of the challenges with working with proprietary tools is that it is difficult to get a sense of how these categorisations have been developed. For instance, in relation to Label detection, Google documentation tells us that the tool can identify 'general objects, locations, activities, animal species, products and more' (https://cloud.google.com/vision/docs/labels) but since these descriptors correspond with Google's Knowledge Graph (an algorithm that is continuously and automatically gathering information and collating factual information), the potential number of label options is indeterminate.

For illustration, the image shown below in our corpus is from a Daily Mail article (5 September 2022) which gives advice on 'bizarre' weight loss techniques, including sitting at the table to eat. Vision assigned 43 labels to this image with a confidence of 50% or above (the default cut-off point). From a manual analysis of the image, comparing its content to the tags assigned, they are reasonably accurate, apart from a small number like Desk and Restaurant.

[1] See https://cloud.google.com/vision/pricing/

LABEL_DETECTION
Watch: 98%
Food: 98%
Table: 97%
Tableware: 96%
Furniture: 93%
Sharing: 89%
Smile: 88%
Plate: 87%
Dishware: 87%
Cuisine: 86%
Chair: 85%
Dish: 84%
Food Craving: 84%
Event: 75%
Picture Frame: 73%
Fun: 73%
Bowl: 71%
Recreation: 70%
Conversation: 70%
Television: 70%
Eating: 69%
Cooking: 68%
Comfort Food: 65%
Room: 65%
Leisure: 63%
Meal: 61%
Vegetable: 60%
À La Carte Food: 59%
Party: 59%
Couch: 58%
Supper: 58%
Spoon: 57%
Desk: 57%
Dessert: 56%
Brunch: 56%
Breakfast: 56%
Lunch: 54%
Culinary Art: 54%
Side Dish: 53%
Dinnerware Set: 53%
Restaurant: 52%
Home Appliance: 51%
T-shirt: 50%

Richards and Tunçer (2018) used Vision to process a large number of photographs posted to the photo-sharing website Flickr of a specific cultural ecosystem in Singapore. They aimed to analyse the image tags to determine what was of most interest to visitors to the area. They restricted their integration of Vision Labels to five key words for each image. Nevertheless, this enabled them to generate clusters of topics that highlighted what was most likely to be photographed. Furthermore, Gonçalves and Paiva (2021) have shown that Vision's automatic processing of text, logos and landmarks were particularly effective in a prototype application designed to assist visually impaired people in navigating their environments.

Table 2
Tags assigned to the images in the corpus.

Category	Occurrences	Number of files	Range of tags
LABEL	10,250	358	4–50
OBJECT	1454	338	0–28
COLOUR	358	358	1
FACE	328	168	0–37
SAFE	130	130	1
LOGO	89	58	0–4
LANDMARK	4	4	0–1

For our research purposes, we are interested in how the features of Vision's automatic annotation support our critical investigation of representations of obesity and allows us to draw distinctions between sets of images provided by different news publications. As described below, what we found, however, was that certain coding categories were not particularly effective for our collection of images.

Table 2 demonstrates the frequency of tags in our data according to each coding category. For certain categories, there is a greater potential for multiple tags for one image file; for instance, a single image could generate as many as 50 Label tags (the fewest being four). Other categories, by design, were only recorded as one tag. With respect to COLOUR, we only documented the dominant colour and SafeSearch annotations were recorded as a combination of the pertinent elements i.e., Adult-Racy. With respect to the categories LOGO and LANDMARK, we found that not only were these relatively infrequent in the data but that they were also prone to errors. Our manual checking indicated that the category LOGO was only used correctly in 31.46% of cases, while LANDMARK (which identifies 'popular and human-made structures') was only accurate in two out of the four cases that it was used.

For this trial analysis we have only included the LABEL category as this provided the highest number of tags, as well as being the most wide-ranging category. Adhering to Christiansen et al.'s (2020) notion of 'simplicity' for effective annotation schemes, we retained only the textual description for each Label, we did not record confidence scores. Vision only returns descriptors with a 50+% confidence score. Nevertheless, as we have discussed above, there is no clear indication of the range and granularity of the Label schema**.** In our data for instance, there are 1425 different Label types appearing in at least one of the images with a confidence score of 50+%. Vision provides a JavaScript Object Notation (JSON) output[2] for all of its annotation which we converted in Notepad++ so that the Label tags appeared in the following format:

<LabelWatch>
<LabelFood>
<LabelTable>
<LabelTableware>

The tags were then (manually) inserted into the appropriate places in their corresponding newspaper article text files. The reason for adding Label at the start of each tag was to enable our corpus analysis tool, WordSmith to distinguish between cases of the word *watch* in the corpus and its label, as for some of the calculations described below, WordSmith will ignore the < and > symbols, effectively combining the frequency of <Watch> with the word *watch*. When we refer to the tags below, we will elide the Label part, to make the tags easier to read in this paper.

A full-stop was inserted at the end of each set of image tags in order to act as a 'sentence break'. This ensures that when we try to identify

[2] JSON is a standard text-based format for representing structured data based on JavaScript object syntax.

Table 3
Keywords across and within the corpus.

Texts	Positive Keywords	Negative keywords
Mail	loneliness, Mailonline, children, army, Americans, ADHD	highlight, cholesterol
Mirror	surgery, his, fight, surgeon, Turkey, ml, surgeons, VTE, after, rate	
Star	her, she, love, influencer, monks, Woody, Abby, followers, so, Sophie, lingerie, Instagram, wear, Pam, bunny, dog, curvy, bra, Sienna, Harvey, comments, model, trolls, positive, herself, me, haters, Katie, Ann, Elloise, she's	the, of
Express	cholesterol, visceral, protein, fat, reduce, COVID, percent, belly, gum, guar, levels, fibre, help, lipoprotein, chickpeas, semaglutide, oats, Mosley, whey, plant, decrease, LDL, soluble, almonds, placebo	cent, I
Guardian	Hart, inflation, strategy	you, fat
Independent	government, ban, steps, junk, restrictions, advertising, policies	I
Times	I, Ozempic, I've, Matt, Lucas	
Telegraph	juice, fruit, Telegraph, minutes, Redfern, medium	I

collocates of tags (in order to determine combinations of tags in the same image), we do not consider combinations of tags from two different images that occur next to one another in a single text.

Previous research on images (Heuer et al., 2011) and video content (Puhl et al., 2013) in news on obesity regularly features elements that contribute to stigmatising representations of people with obesity and so we are interested in determining if the automated system – via its Label tags – can highlight elements like this, that make a significant contribution to the characterisation of those depicted. Following Bednarek and Caple (2017), we are also interested in how representations in different modes reinforce, complement, or clash with one another, as part of our wider investigation into how UK national newspapers differ in the way that they report on obesity.

4. Analysis

4.1. Verbal content

The corpus analysis tool WordSmith 7 (Scott, 2016) was used to carry out the analysis.[3] This tool was chosen due to its high level of functionality involving tags. For example, in its settings WordSmith can be instructed to ignore or include certain tags in its analytical procedures, depending on whether a particular tag is present or absent in a file.

To identify differences between newspapers at the linguistic level, the keywords approach was taken, combined with concordance analyses, where the articles of each newspaper were compared against the other articles in the corpus. WordSmith enables keywords to be obtained, first by allowing users to create frequency lists (called WordLists) of a corpus, then comparing two WordLists together and running statistical tests (such as the log-likelihood test) against all the words in each list. The resulting keywords are shown in Table 3 (positive keywords are more likely to appear in a particular newspaper, negative keywords are less likely to appear).

Three publications – the Express, the Star and the Independent – were shown to have particularly distinctive keywords that related to their framing of obesity and so we will focus on those for our discussion of the verbal content of the news articles.

The Express tends to strongly favour stories which focus on weight reduction and how this can be achieved by changes to diet or use of supplements. These stories tend to combine a bio-medical frame by quoting from research, with a personal responsibility frame – which implies that readers might benefit by adopting diet changes based on the results of the research (keywords are shown in bold in the quoted examples).

> Adding more **chickpeas** to diet could 'drastically' lower **visceral fat**, research suggests.
>
> (Express, 29 September 2022)

> Apart from **fibre**, strawberries also **help** to support the production of two fat-burning hormones called leptin and adiponectin.
>
> (Express, 19 September 2022)

The Star features a different set of keywords which relate obesity to female experience (keywords include *her, she, herself* and various female names). These articles frequently refer to social media influencers who have obesity and advocate a body positive attitude towards weight. The articles often feature descriptions of the women posing in clothes which reveal their bodies, and they are appreciatively described as *curvy*.

> A plus size babe transformed into a "body **positive bunny**" for a **lingerie** photoshoot.
>
> (Star, 25 September 2022)

The women are also described as having received criticism, although these comments are evaluated as cruel and attributed to *haters* and *trolls*, indicating an authorial stance which is sympathetic towards people with obesity.

> While she is an inspiration to many of her **followers**, she receives cruel comments from **haters**.
>
> (Star, 15 September 2022)

Finally, the Independent tends to view obesity through a social responsibility frame, with keywords which refer to government policies aimed at reducing obesity. The following example is from a quote by the head of policy and public affairs at the World Cancer Research Fund International,

> However, it's vital that our **Government** presses ahead with important measures to help people access healthier diets and to **ban** TV advertising of **junk** food before 9pm.
>
> (Independent, 16 September 2022)

As with the bio-medical frame, articles in the Independent tend to quote academic research, although this tends to be based on studies which are critical of government policy.

> In January 2021, research by the University of Cambridge concluded that obesity **policies** in England across the last three decades had been inadequate.
>
> (Independent, 29 September 2022)

4.2. Image tags

The WordList function of WordSmith was used to identify the most frequent Label tags in the corpus. Again, using the keywords procedure, we identified sets of key tags associated with each newspaper (see Table 4), by comparing WordLists from each newspaper against the remainder of the corpus (e.g. to obtain key tags in the Mail, we used the texts from the other seven newspapers). Our observations about which image tags emerge as prominently associated with each newspaper are shown in this section.

Looking back to their corresponding images, the Mail's key image tags tended to be used in charts as part of infographics. <Rectangle> was used in bar charts, whereas <Slope> occurred in line charts. <Map> appeared in informational graphics, for example, one article contained 19 images of maps which showed prevalence of self-reported obesity across US adults by state. This unusually-high number of images inevitably affected the tags that were generated for the Mail sub-corpus, nevertheless we would be cautious about making claims that such figures are char-

[3] WordSmith 5, which has similar functionality to WordSmith 7, is currently free to download online.

Table 4
Key tags across newspapers.

Newspaper	Key tags
Express	<Meat> <Recipe> <ServeWare> <Produce> <Tableware> <Ingredient> <Nail> <StillPhotography> <Whiskey>
Guardian	<Shelf>
Independent	N/A
Mail	<Rectangle> <Slope> <Map>
Mirror	N/A
Star	<Undergarment> <Swimwear> <ArtModel> <Lingerie> <Brassiere> <Fur> <Trunk> <Thigh> <Abdomen> <Chest> <Bikini> <SwimsuitTop> <HumanLeg> <LingerieTop> <PhotoShoot> <Waist> <DogBreed> <Canidae> <CompanionDog> <Carnivore> <Dog> <SportingGroup> <Underpants> <FashionDesign> <Blackhair> <WorkingAnimal> <Snout> <SwimsuitBottom> <LongHair> <Shoulder> <Jewellery> <Brownhair> <Necklace> <Briefs> <FetishModel> <WorkingDog> <AncientDogBreeds> <Blond> <FashionModel> <Sitting> <Navel> <Smile> <TerrestrialAnimal> <Sportswear> <Wolf> <SledDog> <Canis> <Knee>
Telegraph	N/A
Times	N/A

acteristic of the Mail's representation more generally, since so many of these appeared in just one article.

The Express foregrounded a focus on food (consumption), and images were often tagged as <Meat>, <Recipe>, <ServeWare>, <Produce>, <Tableware> or <Ingredient>. Such images tended to be found in informational articles, for example, that describe which foods contain high amounts of cholesterol.

Some of the key tags for the Star related to animals, for example, <Dog>, <Carnivore>, <CompanionDog>, <DogBreed> and only occurred in a set of images from a single news story about a dog, which did not appear to have been picked up by other news venues. More indicative of images across a range of articles in the Star are the labels <ArtModel> and <FetishModel>, which reflect the regular use of images of people posing for photographs that they have then distributed through their social media profiles, for example. The 'Model' part of these labels suggests that these images are presented as exemplars, in that although there is objectification and fetishisation, the person – and the body – is presented in artistic or fashionable terms. Such images also tended to be tagged with words relating to underwear or body parts e.g. <Lingerie>, <Brassiere>, <Bikini>, <Thigh>, <Abdomen>, <Chest>.

As well as identifying key tags across the articles, it was possible to use WordSmith to identify co-occurrences of tags across the same image. This was done by first entering the 'Advanced Settings' and removing the code <*> from the text box called 'Mark-up to Ignore'. Then, a concordance search was carried out on a tag e.g. <HumanBody> and the Collocates tab was selected, to show frequent collocates. As some images contained more than 40 tags, we changed the Horizons (or span) to L25 and R25, to minimise the potential that collocating tags were missed. We also ensured that the 'stop at sentence break' option was activated, to ensure that tags from one image were not identified as collocating with other tags from a second image which appeared close by.

Table 5 shows the 10 most frequent tag collocates of <Abdomen>, <Thigh>, <Waist> and <HumanLeg> and <Chest> (these are among the 10 most frequently appearing tags in the corpus) across the whole corpus. The table shows that all of these tags frequently collocate with one another, ergo they are associated with the same images.

The Guardian had one key image tag: <Shelf>. Images with this tag tended to include pictures of supermarket aisles and appeared to be linked to articles where the role of the food industry or social factors were viewed as central to obesity. For example, articles with images tagged as <Shelf> contained headlines like 'Truss plan to axe sugar tax runs into legal and parliamentary hitches' and 'Liz Truss could scrap anti-obesity strategy in drive to cut red tape'. The latter article describes how Truss might 'lift the ban on sugary products being displayed at checkouts as well as "buy one get one free" multi-buy deals in shops.' Similar to the Independent then, the Guardian uses the social responsibility frame, a finding which was revealed through analysis of a key Label tag.

Several newspapers did not have any key Label tags, which, to an extent, is likely to be due to the small numbers of articles collected, along with difficulties in obtaining images from some newspapers. Although the Mirror did not contain any Label tags which were key, a closer look at the kinds of images in this newspaper revealed a high proportion of celebrity images which were used to focus on individual stories in relation to weight loss/gain. Similarly, news stories which related obesity to political news were often accompanied by images of the Prime Minister. The recognisability of high-status individuals is not built into the automated annotation we have used, however. This could be as a limitation of the approach, though the Vision documentation does refer to beta testing a Celebrity recognition tool and has previously had a Web entities function that linked people to web searches.

4.3. Text-image relations

We again used the keywords procedure to identify the relationship between words and images in the corpus. So, for example, taking the most frequent label tag <Event>, we could enquire whether news articles containing images which had this label were more likely to feature certain words, compared to articles which did not contain <Event> images. Working with a single corpus, WordSmith can be instructed to only consider certain files when it creates WordLists, through altering the settings under Tags & Markup which are found in the Advanced Settings. Opening a window called 'Only if Containing', WordSmith can be set to only consider files if they contain a particular tag – in this case, the tag <Event> (note, this is case-sensitive). This allowed us to build a WordList which only included texts containing <Event> images. We then went back to the Advanced Settings and built a second Wordlist, this time instructing the tool to only consider texts which did *not* contain the <Event> tag. The two Wordlists were compared together using the KeyWords facility in WordSmith. This process was repeated for the ten most frequent tags in the corpus, resulting in ten sets of positive and negative keywords.

A second way of identifying the relationship between image and text was to consider which words tend to occur near (or collocate with) a particular type of image. In the Concordance window, we searched for the tag <Event> and then used the Collocates tab to identify a list of the most frequent words which co-occurred with the <Event> tag. As described in the previous section, we also obtained the tag collocates of the tags, removing <*> from the box marked 'Mark-up to ignore' and setting the horizons (span) at L25 and R25.

Table 6 shows the positive and negative keywords associated with the 10 most frequent image label tags in the corpus, as well as showing the 5 most frequent lexical (i.e. nouns, verbs and adjectives) and image tag collocates.

In the sample analysis below, we look in more detail at a selection of tags: <Event>, <Font>, <Abdomen> and <Gesture>. These were chosen as each one revealed something different and interesting in terms of the analysis.

4.3.1. The <Event> tag

<Event> was the most frequent Label, occurring in 143 images across 78 files. It was most common in images in the Mail and Express but is proportionally over-represented in the Mirror and the Times. <Event> images tend to show people posing, often in smart clothes (hence the collocation with the <FormalWear> and <FashionDesign> tags) or performing (for example, singing on stage). Such images tend to show well-groomed people in formally-posed photographs and so offers an alternative – potentially more 'agreeable' or aspirational – representation in contrast with depictions of people with obesity in ill-fitting

Table 5
Most frequent collocates of the top body part tags.

<Abdomen>	<Thigh>	<Waist>	<HumanLeg>	<Chest>
<Waist>	<HumanLeg>	<Abdomen>>	<Thigh>	<Abdomen>
<Chest>	<Abdomen>	<Thigh>	<Abdomen>	<Trunk>
<Thigh>	<Waist>	<HumanLeg>	<Waist>	<Undergarment>
<Trunk>	<Undergarment>	<Trunk>	<Trunk>	<Waist>
<HumanLeg>	<Trunk>	<Pattern>	<Undergarment>	<Thigh>
<Undergarment>	<Chest>	<Chest>	<Chest>	<HumanLeg>
<Pattern>	<Knee>	<Undergarment>	<Knee>	<Elbow>
<Knee>	<Pattern>	<Sleeve>	<Pattern>	<Neck>
<Elbow>	<Swimwear>	<FashionAccessory>	<Elbow>	<Swimwear>
<Neck>	<Elbow>	<Knee>	<Sitting>	<Shoulder>

Table 6
Keywords and collocates of the most frequently occurring Label tags.

	Label	Frequency	Positive Keywords	Negative Keywords	Top 5 lexical Collocates	Top 5 Tag Collocates
1	<Event>	143	I, me, my	UK, sugar, protein	weight, according, body, research, said	<Fun> <Gesture> <FormalWear> <Font> <ElectricBlue>
2	<Font>	121	COVID, drug, percent	me, my, I	face, said, study, research, obesity	<Rectangle> <Brand> <Event> <ElectricBlue> <Pattern>
3	<Pattern>	107		cholesterol	said, people, department, Florida, study	<Top> <Waist> <Abdomen> <HumanLeg> <Thigh>
4	<ElectricBlue>	101	drug		said, diet, women, people, results	<Pattern> <Magenta> <Sleeve> <Event> <Font>
5	<Gesture>	101	visceral, belly	strategy, products, review, food	said, health, results, women, weight	<Event> <FashionAccessory> <Sleeve> <Abdomen> <Chest>
6	<Abdomen>	90	body, love, followers	food	researchers, people, influencer, time, body	<Waist> <Chest> <Thigh> <HumanLeg> <Undergarment>
7	<Thigh>	88	cancer	food	influencer, time, body, said, Jess	<HumanLeg> <Abdomen> <Waist> <Undergarment> <Trunk>
8	<Waist>	87	waist, her, BMI, height, love, belly	drinks, food	body, said, people, time, waist	<Abdomen> <Thigh> <HumanLeg> <Trunk> <Pattern>
9	<HumanLeg>	85	belly	food	obesity, time, influencer, body, said	<Thigh> <Abdomen> <Waist> <Trunk> <Undergarment>
10	<Chest>	84	she, her, body, followers, love, instagram	obesity, food	said, time, body, diet, researchers	<Chest> <Trunk> <Undergarment> <Waist> <Thigh>

Table 7
<Event> occurring with first person pronouns.

Text	(Manual[4]) description of image containing <Event> tag
A holiday was **my** worst nightmare	Professionally staged image of a middle-aged woman, smiling, wearing make-up and a black and white patterned top.
By the time **my** 50th birthday came round my weight had hit a staggering 21 stone	A young man in glasses and a suit, posing in a group photograph outside a building.
He said: "**I** have Asperger's Syndrome and had some other health issues in the past but **I** grew up very obese. High school was a tough time for **me** so **I** always turned to binge eating.	
Brendan Fraser: role taught **me** severely obese people are 'incredibly strong'	The film star Brendan Fraser, wearing a smart jacket and shirt, photographed against a blue background with writing on it.
Speaking in an interview with The Sun, Robbie said: "**I** have lost weight but it is a constant fight. Inside **me** there's a giant person.	The pop star Robbie Williams, holding a microphone and appearing to be singing, against an abstract backdrop.

Table 8
<Event> occurring with *research*.

Text	Manual description of Image containing <Event> tag
The **research**, compiled by the University of South Florida, says statins should not be given to healthy people even if they have high levels of both good and bad cholesterol.	Two smartly-dressed young women, both eating.
And in the **research**, 44 per cent of people agreed PACE would make it easier to avoid scoffing snacks, with Professor Daley adding: "PACE food labelling may reduce the number of calories selected from menus and decrease the number of calories consumed by the public.	Infographic showing the calories, fat, sugars and salt content of a pizza.
The **research** team is concerned by these findings, as they indicate that Twitch viewers are more likely to purchase the foods advertised to them. Many of these foods are highly unhealthy too.	A young man wearing a black t-shirt with the logo G-Fuel, in an urban location, holding a food carton while sticking out his tongue and making a gesture with one hand.
Separate **research** by the COVID Symptoms Study app has found that having multiple early symptoms at the onset of infection may increase the risk of the condition 3.5 times.	An infographic showing symptoms for COVID.
The **research** showed that eating dark chocolate and cocoa alone didn't appear to have a major effect on heart health.	An infographic showing a range of types of food that have high cholesterol.

clothes observed by Heuer et al. (2011) and Puhl et al. (2013). The accompanying text has a high proportion of first-person pronouns (*I, me, my*). In articles containing <Event>, *my* collocates most often with the nouns *life* (33 times), *weight* (16) and *body* (16), indicating that these articles tend to be focussed upon narratives about people's weight. *Me* tends to collocate with phrases noting difficulty or stress e.g. *hard for me, bothers me, makes me so sad*. Concordance analyses of such keywords occuring near an image tagged with <Event> (examples are shown in Table 7) indicates that these pronouns usually occur within narratives about people's struggles with weight, as the following examples indicate.

These articles are humanising, usually placing their protagonists in a positive light, presenting a sympathetic representation of people with obesity. Losing weight is seen as difficult although tends to be accepted as a worthwhile endeavour. Collectively, the articles could be said to access a 'brave struggle' discourse around obesity (we have not seen this discourse explicity related to obesity in other literature, although Adelman (2018) provides a criticism of it, in a different context).

However, images containing <Event> also tend to appear in contexts that contain the word *research*. Arguably, this could be due to Google Cloud Vision mistagging some images. For example, three of the images in the Table 8 are of infographics which provide information about food. Because these images show different foods, it is likely that Google Cloud Vision has incorrectly associated such cases with events where buffets are sometimes available.

4.3.2. The <Font> tag

This was the second most common tag occurring 121 times across 63 articles. Proportionally, it is most likely to be found in the Guardian. It tended to occur in articles containing the words *COVID*, *drug* and *percent* and was less likely to appear in articles with first person pronouns like *I, my* and *me*, indicating that <Font> images did not typically appear with stories involving first person perspectives.

Images tagged as <Font> tend to appear in informational articles, as part of infographics. Other uses of <Font> in images include numbers on a tape measure or wording on a box of medication (see Table 9). Thus, images with the <Font> tag are likely to appear in articles which use the bio-medical frame of obesity, focussing on scientific research as a way of combatting obesity.

4.3.3. The <Gesture> tag

Images tagged <Gesture> occurred in 101 images spread across 70 articles. Proportionally, such images are most likely to be found in the Times or the Mirror. These images tend also to be tagged with body part tags like <Neck>, <Waist>, <Finger> and <Thigh> as well as the clothing tag <Sleeve>. Such images also tend to occur in texts that also mention the words *visceral* and/or *belly* (see Table 10). These are references to body fat, and the gesture is often a hand squeezing or measuring belly fat, as some of the following examples indicate (it is negligible that these images actually do contain gestures, although they do at least involve human hands). Such articles tend to be informational and are linked to weight loss advice, as opposed to human-interest stories.

4.3.4. The <Abdomen> tag

This image tag occurs 90 times across 48 articles. Proportionally, it is most likely to be found in the Mirror. It tends to appear in images that are also tagged with other body parts like <Trunk>, <Chest> and <HumanLeg> as well as <UnderGarment>. <Abdomen> images occur in articles containing the words *body, love* and *followers* (see Table 11). The keyword *love* appears in clusters which relate to loving oneself or one's body like *love themselves* (2), *love the skin they're in* (2), *love ourselves* (1), *love my body* (3), *love their bodies* (1) and *love herself* (2) – these comprise 38% of uses of *love* in these texts. Similarly, the keyword *body* occurs in clusters which suggest a positive attitude: *body positive* (8), *body positivity* (7) *love my body* (3). However, *body* also occurs in clusters which reference the bio-medical frame: *body mass index* (13), *body fat* (10), *of their body weight* (4), *when the body* (5). Finally, *followers* tends to refer to social media followers of the protagonists of articles about body positivity e.g. *followers on Instagram* (3), *her followers* (7).

Many of the <Abdomen>-tagged images show people with obesity. These articles tend to appear in body positivity news stories, sometimes featuring people described as online influencers, although can also sometimes appear in information-based articles which include definitions of obesity through measures like waist size.

[4] Due to the difficulty in obtaining permission to reproduce images from newspapers, we have instead provided a description of the image ourselves. This is not based on Google Cloud Visions' output, but a human interpretation.

Table 9
<Font> occurring with related keywords.

Text	Image containing <Font> tag
People who were overweight or obese had 20 percent and 36 percent greater odds of long **Covid**, respectively.	Image of a person measuring their waist with a tape measure.
Officers across the country are failing to meet recruitment numbers as fewer youth are qualified to serve, which some blame on the **COVID**-19 pandemic.	US army recruiting poster.
The Food and **Drug** Administration (FDA) will now consult on them, before coming up with proposals.	Chart showing trends in obesity among adults and young people in the US.
An insulin regulator, the **drug** is injected once a week and can reduce a patient's appetite, triggering weight loss.	A box of the drug Ozempic, which appears to be on a doctor's desk.
The changing risk scores, the researchers noted, mirrored weight loss, which was 17 **percent** on average for those subjects given semaglutide and 3 **percent** for those in the placebo group.	An infographic relating to Type 2 Diabetes.
Cases of ADHD in American children increased 30 **percent** from 2011 to 2018, though the reason why has not been pinpointed.	A bar chart showing risk of developing ADHD by weight of the mother during pregnancy.

Table 10
<Gesture> occurring with related keywords.

Text	Image containing <Gesture> tag
"**Visceral** fat was surrounding my liver and causing scarring which was putting me at risk of getting diabetes, hepatitis, cirrhosis and other things.	A woman kissing a man's cheek. Both are well-dressed.
But there's specific foods that might be able to improve the restriction of **belly** fat.	A hand squeezing body fat, with an inset image of a jar of gherkins.
In other words, you need food first thing to power you through the morning, and you don't want to be filling your **belly** at night when your body is preparing for sleep.	A pair of running shoes.[5]
When it comes to busting **visceral** fat, the benefits of fibre are well-established.	A close up of someone squeezing the fat around their stomach, with various images of food super-imposed.
'WHR, however, better reflects levels of abdominal fat, including **visceral** fat, which wraps around the organs deep inside the body.	A medical practioner measuring the waist of a person with obesity.

Table 11
<Abdomen> occurring with related keywords.

Text	Image containing <Abdomen> tag
'Thanks Jess. I struggle with this every day. We need to learn to **love** ourselves more. Your posts inspire me,' one woman wrote.	"Before" and "after" pictures of a woman who has lost weight.
Much of his fanbase is understood to come from his TikTok account which now has more than 140,000 **followers** and 2 million likes, whilst his Instagram has 52,000 followers.	A smiling man with obesity in water, holding a beach ball.
So I hit 50k **followers** yesterday! It is mind-blowing to me that anybody wants to hear about my life or my fashion choices.	An woman with obesity, wearing a low-cut dress and sun-glasses, smiling and posing.
A body positivity influencer is on a journey to unlearn the hatred she has towards her **body** - despite trolls calling her obese.	A young woman wearing a low-cut top.
The waist-to-hip ratio is a measure indicating healthy levels of **body** fat.	A medical practioner measuring the waist of aperson with obesity.
Despite her new found **love** for herself, the influencer has received some cruel comments online -particularly from men.	Two pictures of a young woman wearing a bikini, showing how she looks with and without an Instagram filter.

Articles which contain <Abdomen> images tend not to feature the word *food*, indicating that the articles tend to focus less on the relationship between diet and obesity but more on the bodies of individuals and how they feel about them. These kinds of articles tend to represent a body positivity frame and unlike analyses of earlier news articles (e.g. Brookes and Baker, 2021), they are more associated with a narrative which shows people learning to be happy with their weight, and thus providing them with social acceptance, as opposed to sympathetic weight-loss narratives.

[5] This was a case where the tagger was utterly wrong – there isn't even a human being present. We have included it to remind readers that the online tagger is not perfect.

5. Conclusion

The inclusion of image tags in the analysis of articles about obesity provided an additional dimension to the analysis, indicating how particular frames or discourses around obesity were reinforced or subtly altered through the combination of written text and image, and how newspapers did this differently. We note, for example, that the more sympathetic articles about obesity, which focussed on their narratives where people describe struggling to lose weight, tended to feature pictures of such people dressed in smart or formal clothing (hence the <Event> tag), whereas the body positivity articles which focussed on loving one's body, tended to feature pictures of women in clothing that revealed their bodies and are perhaps more provocatively posed. An analysis which only focused on the written text might therefore miss

some of the ways that messages about obesity are embedded in the news articles.

When used with a small corpus of news articles about obesity, Vision's Label coding performed well, assigning tags which were largely accurate and were able to provide useful analytical insights. The tagging system was by no means perfect, although this is also the case for part-of-speech taggers. Additionally, Vision did not include tags like <Celebrity> which would have been useful. Considering the small size of our corpus, we could have added these tags ourselves, or removed inaccurate ones by hand. It is also possible to 'train' Vision to include additional tags, by presenting it with pre-coded images, so more work could be done to ascertain the value of such training.

Christiansen et al. (2020, p.151) emphasise that automatic image annotation will 'never be as in-depth or as critical as its qualitative counterpart' and certainly, the annotation generated by Vision operates at a level of generality. But it is in that regard that we see potential for a wider application across different data types, since it is not determined by the specifics of a particular research study. Furthermore, the tool was not developed in relation to an archive of obesity-related images, so its contribution to our analysis here in allowing us to identify meaningful differences between subsets of images is indicative of its potential use in other, topic-focused contexts. A potential limitation to Vision's wider adoption is that it is not open source and users remain at the mercy of its developers in terms of what features remain available and how it is developed. Granularity is also an issue – our analysis indicated that fine-grained distinctions were made for body parts but this was not the case for other aspects of images. As different elements are updated and removed, there will be difficulties in establishing the replicability of the methods outlined here. In terms of corpus creation, there is also an amount of initial manual work in terms of sourcing articles and images and inserting image tags into the right places in files. We were able to automate some of these tasks, although this is an area where more work would need to be carried out.

As it is, we have demonstrated how researchers can use the current functionality to conduct a multimodal analysis and there is certainly scope to make adjustments to these procedures, towards different research objectives. We included all of the tags assigned to each image, based on the default confidence score cut-off of 50%. Future work could involve experimenting with higher cut-offs to determine the extent to which this would improve tagging accuracy without removing accurate tags. Further analysis could also be carried out in terms of exploiting some of the other tag categories, like FACE, COLOUR and SAFE which Vision assigned to images. For example, FACE assigns tags based upon emotional expressions, whereas SAFE identifies the likelihood that an image contains explicit content. These kinds of tags would be useful for examining the extent to which certain types of articles about obesity contain images of happy or sad people, or sexualised images. However, we believe that we have successfully demonstrated the application of an online image tagger to aid corpus analysis, and we note its exciting potential to make new discoveries about the interplay between text and image.

Funding

This work was undertaken within the ESRC Centre for Corpus Approaches to Social Science (CASS) at Lancaster University (Grant reference ES/R008906/1).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Adelman, D., 2018. Disability is an Art": Disability as Discourse/Counter-discourse Master's Thesis. University of South Carolina.

Bacon, J.G., Scheltema, K.E., Robinson, B.E., 2001. Fat phobia scale revisited: the short form. Int. J. Obes. Relat. Metab. Disord.: J. Int. Assoc. Study Obes. 25, 252–257. doi:10.1038/sj.ijo.0801537.

Bączkowska, A., 2020. Framing the conceptualization of obesity in online Chinese and British quality newspapers: a corpus-assisted study. In: Lewandowska-Tomaszczyk, B. (Ed.), Cultural Conceptualizations in Language and Communication. Springer, pp. 129–154. doi:10.1007/978-3-030-42734-4_8.

Bednarek, M., Caple, H., 2017. The Discourse of News Values. Oxford University Press.

Bisong, E., 2019. Building Machine Learning and Deep Learning Models on Google Cloud Platform. Apress doi:10.1007/978-1-4842-4470-8_42.

Brookes, G., Baker, P., 2021. Obesity in the News: Language and Representation in the Press. Cambridge University Press.

Christiansen, A., Dance, W., Wild, A., 2020. Constructing corpora from images and text: an introduction to visual constituent analysis. In: Rüdiger, S., Dayter, D. (Eds.), Corpus Approaches to Social Media. John Benjamins, pp. 149–174. doi:10.1075/scl.98.07chr.

Collins, L.C., 2020. Working with images and emoji in the Dukki Facebook Corpus. In: Rüdiger, S., Dayter, D. (Eds.), Corpus Approaches to Social Media. John Benjamins, pp. 175–196. doi:10.1075/scl.98.08col.

Gatt, A., Tanti, M., Muscat, A., Paggio, P., Farrugia, R.A., Borg, C., Camilleri, K.P., Rosner, M., van der Plas, L., 2018. Face2Text: collecting an annotated image description corpus for the generation of rich face descriptions. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). European Language Resources Association (ELRA), pp. 3323–3328.

Gollust, S.E., Eboh, I., Barry, C.L., 2012. Picturing obesity: analyzing the social epidemiology of obesity conveyed through US news media images. Soc. Sci. Med. 74 (10), 1544–1551. doi:10.1016/j.socscimed.2012.01.021.

Gonçalves, J., Paiva, S., 2021. Inclusive mobility solution for visually impaired people using Google Cloud Vision. In: 2021 IEEE International Smart Cities Conference (ISC2), pp. 1–7. doi:10.1109/ISC253183.2021.9562892.

Halliday, M.A.K., 1978. Language as Social Semiotic: the Social Interpretation of Language and Meaning. Edward Arnold.

Heuer, C.A., McClure, K.J., Puhl, R.M., 2011. Obesity stigma in online news: a visual content analysis. J. Health Commun. 16 (9), 976–987. doi:10.1080/10810730.2011.561915.

Knight, D., Adolphs, S., 2020. Multimodal corpora. In: Paquot, M., Th, S. (Eds.), A Practical Handbook of Corpus Linguistics. Springer International Publishing, pp. 351–369. doi:10.1007/978-3-030-46216-1_16.

McClure, K.J., Puhl, R.M., Heuer, C.A., 2011. Obesity in the news: do photographic images of obese persons influence antifat attitudes? J. Health Commun. 16 (4), 359–371. doi:10.1080/10810730.2010.535108.

McGlashan, M., 2021. Linguistic and visual trends in the representation of two-mum and two-dad couples in children's picturebooks. In: Moya-Guijarro, A.J., Ventola, E. (Eds.), A Multimodal Approach to Challenging Gender Stereotypes in Children's Picture Books. Routledge, pp. 213–238.

Nimegeer, A., Patterson, C., Hilton, S., 2019. Media framing of childhood obesity: a content analysis of UK newspapers from 1996 to 2014. BMJ Open 9, e025646. doi:10.1136/bmjopen-2018-025646.

O'Donnell, M., 2008. Demonstration of the UAM CorpusTool for text and image annotation. In: Proceedings of the ACL-08: HLT Demo Session (Companion Volume). Association for Computational Linguistics, pp. 13–16.

Partington, A., Duguid, A., Taylor, C., 2013. Patterns and Meanings in Discourse: Theory and Practice in Corpus-Assisted Discourse Studies (CADS). John Benjamins.

Puhl, R.M., Peterson, J.L., DePierre, J.A., Luedicke, J., 2013. Headless, hungry, and unhealthy: a video content analysis of obese persons portrayed in online news. J. Health Commun. 18 (6), 686–702. doi:10.1080/10810730.2012.743631.

Rafiee, A., Spooren, W., Sanders, J., 2021. Framing similar issues differently: a cross-cultural discourse analysis of news images. Social Semiotics doi:10.1080/10350330.2021.1900719, Online first.

Rashtchian, C., Young, P., Hodosh, M., Hockenmaier, J., 2010. Collecting image annotations using Amazon's Mechanical Turk. In: Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk. Association for Computational Linguistics, pp. 139–147.

Richards, D.R, Tunçer, B., 2018. Using image recognition to automate assessment of cultural ecosystem services from social media photographs. Ecosystem Services 31 (C), 318–325. doi:10.1016/j.ecoser.2017.09.004.

Scott, M., 2016. *WordSmith Tools* Version 7. Lexical Analysis Software.

Tan, S., O'Halloran, K., Wignell, P., Lobinger, K., 2020. Images of austerity in the British press and in online media. In: Griebel, T., Evert, S., Heinrich, P. (Eds.), Multimodal Approaches to Media Discourses: Reconstructing the Age of Austerity in the United Kingdom. Routledge, pp. 134–162.