

“I will say the picture of the background is not related to the words”: using corpus linguistics and focus groups to reveal how speakers of English as an additional language perceive the effectiveness of the phraseology and imagery in UK public health tweets during COVID-19

Christian Jones*, David Oakey, Kay L. O'Halloran

University of Liverpool

ARTICLE INFO

Keywords:
 COVID-19
 phraseology
 multimodal
 EAL
 public health discourse
 tweets

ABSTRACT

This paper reports on an application of a multimodal corpus-based study into the effectiveness of public health information about COVID-19 for speakers of English as an additional language (EAL) in the UK. A corpus of information tweets from 13 UK public health agencies totalling 560,000 words, with concomitant images and videos, was collected between March 2020 and February 2021. The most frequent n-grams occurring across all 13 public health agencies, and sample images occurring alongside these, were identified. In this study, we examine how images and videos combine with the phraseology to shape these COVID-19 public health information messages. Following this, six illustrative tweets were used as prompts for three focus groups of EAL participants based in the UK representing a range of first languages and occupations. Data from the focus groups was analysed in order to identify how common public health phraseology and images were received, understood and responded to by participants and how they felt they could be amended to increase their effectiveness for EAL speakers. We conclude with suggestions for making the language of public health messages simpler and more direct, aligning images more clearly with the language used and removing linguistic ambiguity. These recommendations for how such messaging could be improved in future public health campaigns could ensure a more effective and inclusive public health response.

1. Introduction

This paper reports on application of a multimodal corpus-based study into the effectiveness of public health information about COVID-19 for speakers of English as an additional language (EAL) in the UK. Many studies into the language of COVID-19 already exist, such as early reports by the Oxford English Dictionary (OED, 2020) and Hunston (2020) and the special issue of IJCL (Mahlberg & Brookes 2021). Similarly the effectiveness of public health information has been widely studied in areas like Health Communication (Park et al, 2016) and Psychology, often in terms of message organisation and ‘framing’ (Greszczuk, 2020; Oliver, 2020; Banerjee & Meena, 2021), and the subject is commonly discussed in the mainstream media.

However, much of the health information produced to inform the public about the outbreak of COVID-19 was multimodal rather than exclusively text-based. Textual elements were accompanied by images and videos of medical staff and other frontline workers, patients, med-

ical facilities, politicians, and members of the public and public spaces (Oakey et al., 2022). While multimodality is widely researched in most areas of language studies, including education, digital communication, social interaction and other contexts involving language use (e.g. Bateman, Wildfeuer & Hiippala 2017; Jewitt, Bezemer & O'Halloran; Tan, O'Halloran & Wignell, 2020), the multimodal aspect of public health messaging is less often studied from the point of view of corpus linguistics (Baldry, Bianchi & Loiacono 2019). On the contrary, the text and images in public health messages are typically analysed qualitatively using a small data set (e.g. Aragbuwa & Adejumo, 2021; see also Tan & E, 2022). In addition, few studies elicit responses to findings from corpus data from language user groups such as second language learners.

In the case of EAL speakers in the UK, their understanding of public health messages during the pandemic was identified as an important area of research inquiry by the UK Parliamentary Office of Science and Technology (2020). The suggestion made was that there was a clear need to investigate the effects of public health messaging in relation to

* Corresponding author

E-mail address: Christian.jones2@liverpool.ac.uk (C. Jones).

the pandemic and particularly to understand how this was understood by underserved groups such as those who speak English as an additional language. Since EAL speakers comprise an estimated 8.9 per cent of the population in England and Wales (Office for National Statistics, 2021), we felt that this was an important focus of investigation. For this reason, we undertook a multimodal corpus study which examined the way textual and visual information representations combine in these public health messages and then applied the results of this study to see how messages are received, understood and responded to by EAL speakers. Such work is critical if we are to improve the effectiveness of current and future public health messages. The corpus-based section of the study (reported more fully in Oakey, Jones & O'Halloran, 2022) aimed to first understand the common forms and functions of language and images. These findings were then used as prompts for focus groups of EAL speakers to elicit and report their responses to these messages. We believe this approach represents an innovative application of corpus linguistics to a real-world problem.

Specifically, we sought to answer the following research questions:

- 1(a). What are the most frequent four-word n-grams in a corpus of tweets from UK public health agencies in the first 11 months of the COVID-19 pandemic?
- 1(b). What are the characteristics of images accompanying these frequent n-grams?
- 1(c). How do common n-grams and images work together to form public health messages?
- 2 (a). How do EAL speakers interpret the intended messages?
- 2 (b). What suggestions do they have regarding how the language or images could be made clearer for EAL speakers?

2. Literature review

2.1. Corpora and COVID-19

Corpus linguists were quick to quantify the impact on the English language of the SARS-CoV-2 virus, named "COVID-19" by the World Health Organization early in 2020 (WHO, 2020). The words "coronavirus" and "COVID-19" obviously became very frequent through their rapidly expanding use in news and social media reports in the first three months of the pandemic, as were their frequent collocations such as "outbreak", "infection", "spread", and "fear" (OED, 2020). New frequent collocates of other words were also identified through corpus linguistic work, illustrating almost in real time the dynamics of language change. Schuessler (2020), for example, reported that before the pandemic the most common collocates of "remote" in the OED were nouns like "control", "island" and "village", but by the end of 2020 the most frequent collocates of "remote" had suddenly become "working", "learning", and "work force", reflecting the very different working practices that evolved during the outbreak.

As described in the introduction to this special issue, a special edition of the *International Journal of Corpus Linguistics* (Mahlberg & Brookes, 2021) contained studies describing people's experiences during the pandemic, revealing how their "attitudes and beliefs are constructed and mediated through language use" (ibid, p. 441). The special issue also contained a description of a corpus specially collected to identify the effects of the pandemic on language use, the Coronavirus Corpus (Davies, 2021). This is comprised of COVID-related news articles which are included if they contain at least two occurrences of the words "coronavirus", "COVID", or "COVID-19", or other related words in their titles, such as, "contagious", "outbreak", "pandemic" and so on.

Despite this valuable corpus linguistic work into how news organisations, governments, and corporations reported and communicated the pandemic, there has been less work on how messages were received by their intended recipients and how well they were understood. The focus group data presented in this study, in the form of EAL speakers' responses to text and images in UK public health agency tweets in our

corpus, represent a more applied corpus linguistic approach, where not only the outgoing messages are described but also the responses to them. As we will discuss, we hope that the type of analysis we have undertaken could a) be developed into similar research relate to future public health campaigns and b) be applied to the formulation of future health messages by government agencies, in order to ensure they are as effective as possible in reaching EAL speakers.

2.2. N-grams

This study focuses on frequent n-grams, i.e. fixed strings of *n* words which occur repeatedly in a corpus of texts. These strings have been widely studied since the advent of corpus linguistics, and they have contributed to our understanding of phraseology, the study of word combinations and formulaic language. Phraseology before corpus linguistics largely focused on lexical and syntactic relations such as collocation and colligation (Carter, 1998; Firth, 1957), the semantic features of idioms and proverbs (Moon, 1998), and communicative acts such as greetings, prayers, and rituals (Glaser, 1998). Once the extent of the prevalence of n-grams was revealed through the introduction of corpus methods, explanations from language cognition and processing were sought relating n-grams to psycholinguistic choices by language users (Ellis, 2008; Schmitt, Grandage, & Adolphs, 2004; Sinclair, 1991). Particular n-grams such as the "lexical bundle", an n-gram occurring in at least 10 different texts at least 40 times per million words (Biber et al., 1999), have been widely analysed in both spoken and written corpora in many different registers of language use (Oakey, 2020).

2.3. Multimodal analysis

This corpus study is unusual in that the analysis of the corpus of tweets is not exclusively text-based. As mentioned above, tweets are comprised of text, images, and/or videos, and a multimodal aspect to the study is therefore required, something also noted as important by Bednarek and Caple (2014) in relation to news discourse analysis and Caple (2018) in relation to Instagram posts. Multimodal analysis considers how language, visual images and other resources (e.g. sound, music, and gesture) combine to construct meaning (e.g. Bateman, 2014; Kress & van Leeuwen, 2020; O'Halloran et al. 2016, Tan, O'Halloran, & Wignell, 2020; van Leeuwen, 2005). In this study we are interested in how the text and images (and in one case video) in the tweets work together to construct the messages and how EAL speakers perceive and understand them. We examine how the participants see the text and images working together, either reinforcing and expanding the meaning of the tweet or whether there is any semiotic clash (i.e. with diverging rather than converging meanings) between the two. As well as understanding responses to the textual n-grams, therefore, it is important to identify the influence of the image on understanding of the text and the influence of the text on understanding of the image (Oakey et al., 2022).

3. Methodology

This section details the decisions we took related to corpus design and linguistic analysis and the focus group participants, design and analysis.

3.1. Corpus design and analysis

Using our own Multimodal Analysis Platform (MAP), a cloud-based application for collecting, indexing, storing and analysing online media texts (O'Halloran, Pal, & Jin, 2021), we created a corpus of tweets sent from thirteen official UK public health accounts (e.g. @ p_h_s_official). These thirteen accounts were chosen as, at the time of the study, Public Health England (PHE) was divided into 13 regions and each region had a twitter account. The accounts and tweets therefore represent all the areas of the UK covered by PHE public health tweets at the time

of investigation. In this study, MAP was used to collect the Twitter data and identify the most common n-grams and their accompanying images. Tweets were collected between the period of 1 March 2020 to 17 February 2021. This period represented the first major phase of the pandemic in the UK and as such was a crucial time in public health messaging in this context. The image, text and URL for each tweet can be seen in Appendix 1. The final corpus contained 564,040 words and associated images and videos, which were used to create a list of the most frequent four-word n-grams contained within these tweets. In order to create such a list, tokens that were not syntactically part of the running text of the tweets, such as hashtags and the @ symbol, were removed from the word count while still being preserved in the corpus. In order to ensure that the full discourse community was represented, only n-grams with a range of 13 (i.e. they occurred in all sampled public health accounts) were included.

As we have made clear in our research questions, we are looking for n-grams in “a corpus of tweets from UK public health agencies in the first 11 months of the COVID-19 pandemic” rather than “a corpus of tweets from UK public health agencies about COVID-19.” This is because it is hard to draw the line between messages exclusively “about” the coronavirus and those about continuing UK public health concerns such as cancer, stroke, cardiovascular disease, mental health, diabetes and so on (Public Health England, 2014). We felt that if we used pre-selected linguistic search terms to include or exclude tweets from the corpus, we risked biasing the data in favour of our pre-conceived notion of which words are “about” COVID-19. And determining whether a message is or is not “about” COVID-19, in order to collect a corpus of tweets about COVID-19, is problematic. An example is shown by the tweet from our corpus in corpus extract 1.

3.1.1. Corpus extract 1

Finding cancer early makes it more treatable, so if you have been suffering from tummy troubles for three weeks or more please make sure you speak to your GP. The NHS and NHSNEY have put measures in place for you to access help safely. (phe_northeast, 25 November 2020)

While this tweet appears to be principally “about” early detection of cancer, rather than COVID-19, it reflects the pandemic conditions then prevailing in the community; it would not have been necessary to add the assurances of additional safe access measures if there had been no pandemic. Temporal references like “right now,” “at the moment” and “the current situation” similarly allude to the prevailing pandemic without mentioning it by name, as in Corpus extracts 2, 3 and 4, meaning that the tweets lack a COVID-19 related lexical referent that would show up in a keyword search:

3.1.2. Corpus extract 2

We all want to protect ourselves and our families right now, but that is especially important if you smoke. Smoking causes lung damage, weakened immune systems, cancer, heart attack, diabetes and stroke. (phe_northeast 22 October 2020)

3.1.3. Corpus extract 3

This #MentalHealthAwarenessWeek it's important to remember that feeling worried and anxious is perfectly normal, particularly while our lives are so different at the moment. Visit J1a0Wuube9 to find out about small steps you can take to help yourself (p_h_s_official, 18th May 2020)

3.1.4. Corpus extract 4

Are you struggling with existing mental ill-health? The current situation may be making it harder to cope. If you're in need of support, help is just a phone call away. Please reach out today if you need someone to talk to. (publichealthw, 1st January 2021)

These tweets, therefore, are still “about” COVID-19, in that they reflect public health issues under the prevailing COVID-19 conditions, and should not therefore be excluded from the corpus simply because they do not contain the search terms used by Davies (2021). A corpus collected

without recourse to external linguistic criteria such as search terms can bring to light language uses that the designer may not have predicted. This is why Sinclair stipulated the important corpus design requirement that “the contents of a corpus should be selected without regard for the language they contain, but according to their communicative function in the community in which they arise” (Sinclair, 2005, p. 5). So while we could possibly have followed the methodology used by Davies (2021) to collect the Coronavirus Corpus mentioned above, and select only tweets that were “about” COVID-19, or which contained particular words related to the pandemic, we included all UK public health tweets in our corpus.

For several reasons, only four-word n-grams were chosen. First, three-word n-grams are often embedded in four-word n-grams (e.g. *the spread of/the spread of covid*). Second, four-word n-grams are more likely to be syntactically whole when compared to nominal or clausal fragments. Lastly, much research in this area shows that three and four-word n-grams are often significantly higher in frequency than longer n-grams of, for example, five words (e.g. Jones, Byrne, & Halenko, 2018; O'Keeffe, McCarthy, & Carter, 2007).

In choosing the tweets used for the prompts for the EAL focus groups, we decided to move from n-gram to image by using MAP to help us search for the images which occurred alongside the most frequent n-grams. The reason for this choice was that it was not possible to base analysis on the frequent images because MAP finds images signalled either by appearing near a particular hashtag or tweet. Common images could therefore be associated with several n-grams, some of which may be very infrequent. For this reason, we used MAP to help us locate some images which occurred with common n-grams. We then searched for images through the Twitter site “advanced search” tool (Twitter.com). This ensured we were able to find and use an image with a frequent n-gram. An example of an image occurring next to two frequent n-grams *it is important that* and *the spread of* (highlighted in the tweet text) is shown in Fig. 1. In this case, the Tweet in in Fig. 1 contains a photograph of a woman working on a laptop at home. The tweet text and the photograph have been annotated using facilities in Multimodal Analysis Image software¹ (O'Halloran, Tan, & E, 2017) to highlight key features of the text (in this case, the n-gram) and the photograph (see section 4.2 and details in Oakey et al., 2022). For example, the woman is the focus of attention in the photograph, with a frontal perspective and a level camera angle with medium distance and her gaze is internal to the image. These selections function to position the viewer as an observer of this scene of ‘working from home’ This particular tweet was used as the first prompt in the focus group since the textual element of the message contains two very frequent n grams in the corpus i.e. *the spread of covid* and *it is important that* (see Table 4 for a complete list).

Following the above corpus-based multimodal method, we chose six tweets representing frequent n-grams and images used with them to use as prompts for the EAL speaker focus groups, the image, text and URLs for which can be seen in Appendix 1.

In order to analyse the images in the tweets, we examined the semi-otic characteristics of the images according to Halliday's (2009) meta-functional categories of ideational (experiential and logical), interpersonal and textual meaning. We followed O'Toole's (2011) framework where the metafunctional systems are organized at the rank of figure, episode and work (see Table 2). There were different types of images in the dataset, including photographs, infographics, and posters. For this reason, we annotated the elements which were specific for each image genre (for example, infographic headlines, questions, commands and icons) using Multimodal Analysis Image software (O'Halloran, Tan, & E, 2017). This permitted us to identify where the linguistic and visual elements of the text and image converged and diverged in terms of the meanings which were made. Table 2 summarises this.

¹ <http://multimodal-analysis.com/products/multimodal-analysis-image/software/index.html>

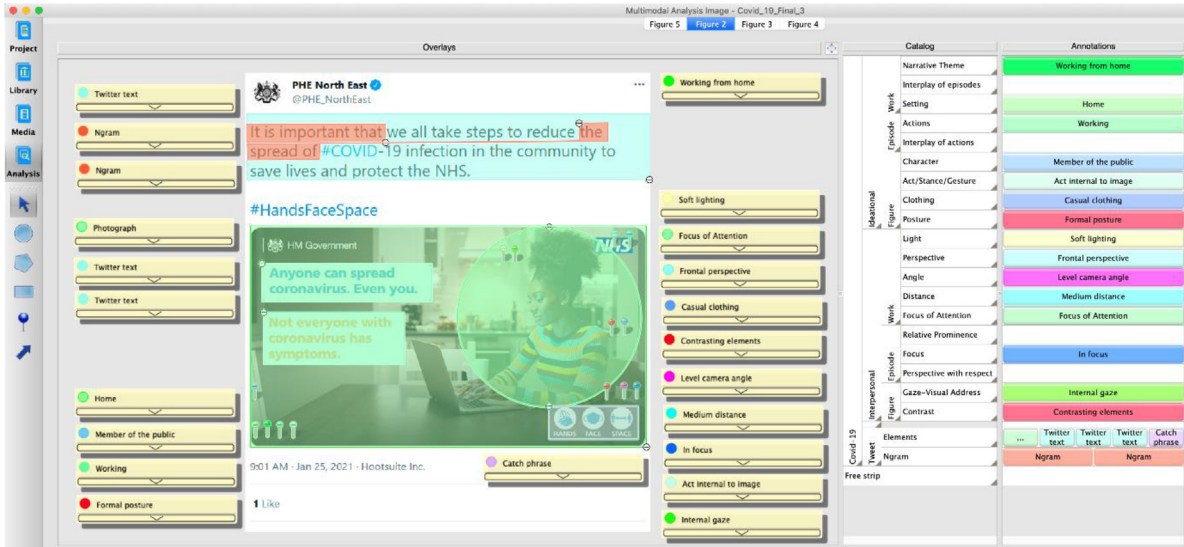


Fig. 1. Screenshot of sample focus group tweet (Oakey et al., 2022)

Table 1
Corpus of UK Public Health Agency tweets 1 March 2020 and 17 February 2021.

Public Health Agency	Twitter handle	Tweets	Tokens
Public Health England	PHE_uk	1,440	47,384
North of England: North East	PHE_NorthEast	1,521	59,102
North of England: North West	PHE_NorthWest	1,021	35,756
North of England: Yorkshire and the Humber	PHE_YorksHumber	726	35,080
Midlands and East of England: East Midlands	PHE_EastMids	1,201	43,872
Midlands and East of England: East of England	PHE_EoEngland	1,435	52,782
Midlands and East of England: West Midlands	PHE_WestMids	1,393	52,538
South of England: South East	PHE_SouthEast	452	15,176
South of England: South West	PHE_SouthWest	2,144	79,703
London	PHE_London	527	19,109
Public Health Scotland	P_H_S_Official	405	12,882
Public Health Wales	PublicHealthW	1,788	61,189
Public Health Agency	Publichealthni	1,346	49,467
Totals		15,399	564,040

Table 2
Image systems (adapted from Kress & van Leeuwen, 2020; and O'Toole, 2011; see Oakey et al. 2022).

Metafunction	Rank	System	Description
Ideational (experiential and logical)	Work	Narrative Themes; Interplay of Episodes; Setting	What is portrayed
	Episode	Actions; Interplay of Actions	Visual happenings, actions and relations (processes, participants and circumstances)
	Figure	Character; Acts/Stance/Gesture; Clothing; Posture	Participant characteristics
Interpersonal	Work	Light, Perspective, Angle; Distance	Visual effects and orientation
	Episode	Relative Prominence; Focus; Perspective	Happenings, actions and relations with respect to the whole image
	Figure	Gaze-Visual Address; Contrast	Direction of participant's gaze as internal to image or external to viewer; Contrasting elements
Textual	Work	Compositional Vectors; Proportion; Framing	The organisation of the parts as a whole, with visual markings of certain parts
	Episode	Relative Placement of Episode; Framing	Relative position of the happenings, actions and relations in relation to the whole image, with visual marking of certain aspects
	Figure	Relative Placement in the Episode; Arrangement; Framing	Relative position of figures and the visual marking of certain aspects of those figures

3.2. Participants

Following ethical approval for the project, 17 EAL speakers were recruited as participants to three focus groups. First languages represented were Chinese (11), Syrian (1), South Korean (1), German (1), Filipino (2), and French (1). There were six male and 11 female participants with a range of ages: 18-30 (11), 31-45 (4), 46+ (2). Participants had lived in the UK from six months to over a year at the time of the study and were here for the purposes of study, work and/or family reasons. They were recruited in response to an email advert asking for participants, which also explained the broad purpose of the study and its general aims. The advert was sent to contacts who had connections to the local EAL community in the north west of England, who then passed it on for us. This included, for example, the university Confucius Institute and colleagues working in health-related fields. Each participant self-reported that they had at least a CEFR B1 level of English, which we took to be reliable as this was needed for visa purposes in the UK. A B1 level English speaker is broadly said to have the following ability:

Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans (Council of Europe, 2022).

Once recruited, each participant received a ten pound shopping voucher for their participation.

Based on suggestions from researchers who have made extensive use of focus groups (e.g. Morgan, 1997; Barbour & Morgan, 2017) we decided on a focus group size of close to six members. Participants were therefore assigned to one of the three groups, meaning group one contained five participants, while groups two or three contained six participants. Our choice of three groups was partly dictated by convenience, based on EAL speakers living, working or studying in the UK at the time who responded to advertisements and gave their informed consent to participate. As a result, there is a larger proportion of Chinese L1 speakers when compared with other nationalities and this reflects the UK university setting. However, as will be seen below, the results do also show that a broad consensus was reached about the prompts discussed.

3.3. Focus group design

Discussions were based around six tweets containing the most frequent phrases and associated images (see Appendix 1 for the image, text and URL for each tweet), chosen based on the process explained in section 3.1, that is, we identified common four-word n-grams and then an image commonly used with one or more of these n-grams in a tweet. Based on suggestions from Morgan (1997), five basic prompts were used to facilitate the discussion. Firstly, participants were asked what public health messages they recalled about COVID-19, in the form of social media posts, posters etc. The intention of this prompt was simply to warm up participants and orient their thoughts towards the discussion topic. Next, each sample tweet was shown and participants were asked if they understood it, how it made them feel, how they felt it could be improved for other EAL speakers, and finally which tweet they found to be most and least effective. Follow up questions were used, depending on the responses given (see Appendix 2 for the full set of prompts). The aim of the focus groups was, as Morgan (1997, p.2) suggests, to enable 'group interaction to produce data and insights which would be less accessible without the interaction found in a group', something more likely to emerge when exploring topics which are familiar to participants (Macnaghten, 2017), as was the case here. As the essence of this form of data collection is based on the interaction between participants, it was vital that the interlocutor (one of the authors) ensured that all participants contributed to the discussion by, for example, nominating

some participants for their reactions to what other speakers said and to check for agreement and disagreement. We acknowledge that participants could have suggested different changes with a different set of prompts but the focus of this research was their response to how the language and images were used. There is no consensus on how many focus groups should be held for any research project but the aim is to reach saturation – whereby additional groups are essentially not providing any new answers and showing there is a broad consensus established.

As discussed, the original tweets can be found in Appendix 1 but for ease of reference, the six tweets are summarised in Table 3.

3.4. Focus group analysis

Due to social-distancing restrictions in place at the time, each group was held online using Zoom software. Each group discussion was recorded and transcribed, and NVivo software used to analyse the data. Computer Assisted Qualitative Data Analysis (CAQDAS) software such as NVivo is now commonly used with qualitative data of this nature. It allows researchers to interrogate the data in a much more systematic manner than we could if coding for themes manually (Kelle, 2002). For instance, if we make a code based on a named category, it is simple and quick to search the entire data set to find all mentions of this word. Kelle (2002, p.486) summarises these benefits as follows:

CAQDAS also helps with the systematic use of the complete evidence available in the data much better than any mechanical system of data organisation. If the data are methodically coded with the help of software, researchers will find evidence and counter-evidence more easily. This clearly reduces the temptation to build far-reaching theoretical assumptions on some quickly and arbitrarily collected quotations from the material.

Each transcription was coded in NVivo and three main themes emerged and were agreed by the researchers: overall understanding and impressions of tweets, the most and least favored tweets and suggestions for improving the language or images for EAL speakers. Examples were chosen based initially on frequent words used by participants within each code by employing the word frequency function in NVivo. For example, 'think' was the third most frequent word in the 'overall understanding and impressions' theme and searching for examples of this word enabled us to quickly find typical instances where participants expressed a viewpoint about each of the tweets and identify common, overlapping views. The examples chosen represent prototypical comments made by a number of participants.

These results will be discussed in relation to the second research question. However, first we will outline the results of research questions 1(a) to (c).

4. Results and discussion

In this section, we present the results as they relate to each research question and then discuss them briefly. Following this, we provide a summative discussion of all the results.

4.1. Most frequent 4-grams

RQ1(a). *What are the most frequent four word phrases in a corpus of tweets from UK public health agencies in the first 11 months of the COVID-19 pandemic?*

Table 4 lists the four-word n-grams occurring in tweets from all UK public health agencies in the corpus. Table 5 lists the items grouped together according to the most common overlapping patterns

What is noticeable here is firstly that in terms of form, many of the most frequent four-word phrases are taken from larger 'frames' and are frequent because they overlap with others and are used across many of these messages e.g. *the spread of covid* in *stop the spread of covid/help stop*

Table 3
A summary of the six tweets discussed in each focus group.

Tweet number and source	Text	Image	Text on image
1 Public Health England (PHE) North East 25/01/21	<u>It is important that we all take steps to reduce the spread of #Covid-19 infection in the community to save lives and protect the NHS.</u>	A photo of a woman working at home on a laptop.	Text on image: Anyone can spread the virus. Even you. Not everyone with coronavirus has symptoms.
2 Public Health Agency 14/11/20	Taking three simple steps can save lives. 1.Wash your hands regularly. 2.Keep social distancing (2 metres where possible). 3. Wear a face covering <u>Help stop the spread of COVID-19.</u> For more information visit: URL	Graphics showing hands being washed, two people with an arrow between them, someone wearing a mask. All pictures are within a red frame.	Wash your hands regularly, keep your distance, wear a face covering where needed.
3 Public Health England (PHE) 11/01/21	If you have symptoms or have tested positive, you must self-isolate for ten days. Stay home and protect lives. For more information on symptoms, testing and self-isolation visit: URL	Graphics showing a person sneezing, a test tube and swab and a house with a person in it and lock on it. The graphic is in three parts, the first yellow, the second orange and the third red.	Got symptoms? Tested positive? You must self-isolate for ten days.
4 Public Health England (PHE) North East 04/01/21	Vaccination is the most important thing you can do to protect the entire North East community against illness. When enough people get vaccinated, it's harder for a disease to spread to those who can't have vaccines. Read more: URL	Graphics showing four clipboards in bright purple, pink, orange and green colors.	Key facts on vaccines with a different fact on each clipboard.
5 Public Health England (PHE) 27/10/20	With Halloween coming up <u>it's important to stay safe while having fun:</u> Follow rules in your local area Wash your hands regularly and stay 2m apart Wear a face covering where appropriate Self-isolate and book a test if you experience any symptoms of COVID-19	Graphics showing four orange pumpkins as bullet points. A larger graphic of an orange pumpkin wearing a white face mask on a purple and green background	Cover your face this Halloween.
6 Public Health England (PHE) North East 02/02/21	It's important to follow the rules even after you've had the COVID-19 vaccination. Always remember to: Wash your hands Cover your face Keep your distance More on what to do after the vaccination: URL	An animated gif showing a vaccine dose bottle with graphics within the bottle such as of washing hands. The graphic is in navy blue on a grey background.	Wash your hands Cover your face Make space

the spread of covid/it's important that in it's important that you get tested. These frames form an important part of the messages contained in the corpus as the sample below shows, and confirm the formulaic nature of much language, as mentioned above (e.g. Sinclair, 1991; Biber et al., 1999). The following example is taken from the first tweet used in our focus groups (see Appendix 1) and the n-grams have been underlined here:

4.1.1. Corpus Extract 5

It is important that we all take steps to reduce the spread of #COVID-19 in the community to save lives and protect the NHS.

Secondly, phrases with *if* are very prevalent, showing that complex sentences (with main and subordinate clauses) are frequent. One example, from a tweet used as a prompt in our focus groups is *If you have symptoms or have tested positive, you must stay home and self-isolate*, but it was surprising just how many of the tweets contained these frequent n-grams within complex sentences.

Taken together, the most common n-grams form part of messages which seem to function to either give instructions or present information as a means of giving advice. An example of the first is *if you have symptoms, you must self-isolate* and an example of the second is *It is important that we all take steps to protect the community*. However, when we examine some of these n-grams in context, alongside the images used with them, the functions are not always so clear. We will discuss these issues in relation to research question 1(c).

4.2. Images accompanying these n-grams

RQ1(b). *What are the characteristics of images accompanying these frequent n-grams?*

As explained previously we identified the most frequent n-grams and examined the images accompanying those n-grams. Therefore, we have not identified the most frequent images, or those most frequently liked or retweeted. Nonetheless, several patterns emerge in relation to the images found with these common n-grams. In the first case, there are three predominant types of images: (1) photographs of members of the public and/or authority figures such as representatives of Public Health England; (2) procedural infographics with a series of questions and commands together with icons and drawings; and (3) posters with cartoon-like drawings.

Secondly, the images often contained linguistic elements. These textual elements are often parts of the Twitter text or they partially paraphrase the Twitter message. For example, procedural infographics are complex multimodal representations, with questions and commands which are accompanied by icons and/or drawings to reinforce the meaning, as displayed in Fig. 2. In many cases, the tweet text and the infographics are very similar, often repeating the same message. This raises the question of possible redundancy and information fatigue in the Tweets. However, it is possible that the infographics were also used

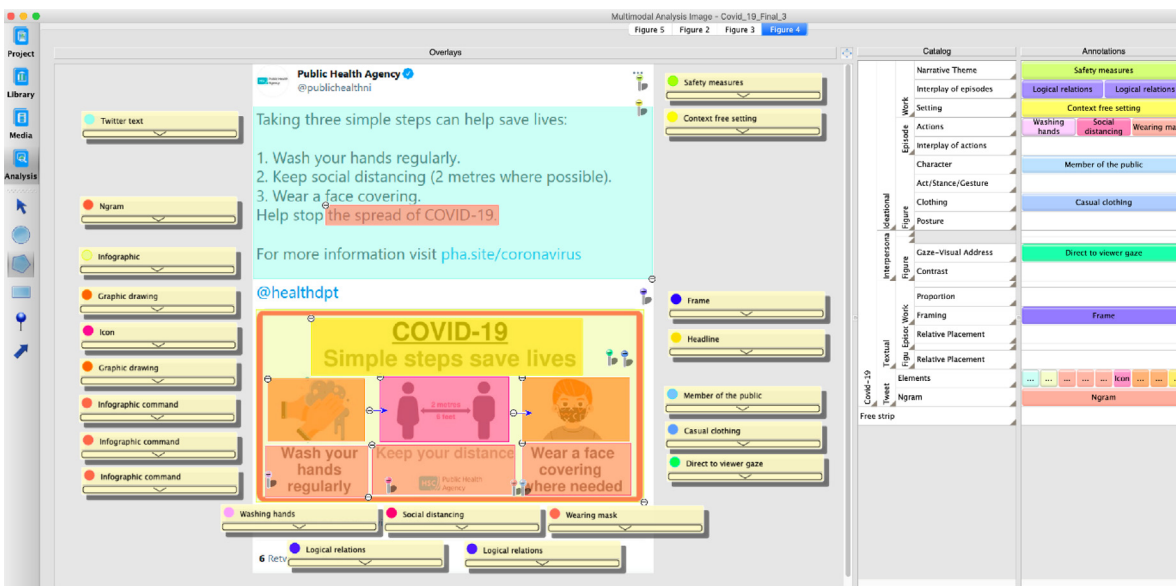


Fig. 2. Focus Group Prompt 3: Screenshot of image and text in MAP containing the common n-gram *stop the spread of* (Oakey et al., 2022)

Table 4

The twenty most frequent four-word n-grams in the COVID-19 corpus.

n-gram	Rank	Freq
<i>the spread of covid</i>	1	370
<i>it's important to</i>	2	279
<i>stop the spread of</i>	3	269
<i>to find out more</i>	4	215
<i>find out more about</i>	5	199
<i>if you have symptoms</i>	6	172
<i>help stop the spread</i>	7	150
<i>it's important that</i>	8	140
<i>if you have any</i>	9	122
<i>find out how to</i>	10	113
<i>prevent the spread of</i>	11	107
<i>if you have a</i>	12	97
<i>children and young people</i>	13	92
<i>you can do to</i>	14	91
<i>if you have been</i>	15	66
<i>more important than ever</i>	16	55
<i>there are lots of</i>	17	47
<i>you can take to</i>	17	47
<i>steps you can take</i>	19	37
<i>for the first time</i>	20	28

Table 5

Groups of overlapping four-word n-grams in the COVID-19 corpus.

	n-gram	Rank	Freq
Spread	<i>the spread of covid</i>	1	370
	<i>stop the spread of</i>	3	269
	<i>help stop the spread</i>	7	150
Important	<i>prevent the spread of</i>	11	107
	<i>it's important to</i>	2	279
	<i>it's important that</i>	8	140
If	<i>more important than ever</i>	16	55
	<i>if you have symptoms</i>	6	172
	<i>if you have any</i>	9	122
Find	<i>if you have a</i>	12	97
	<i>if you have been</i>	15	66
	<i>find out how to</i>	10	113
Can	<i>find out more about</i>	5	199
	<i>to find out more</i>	4	215
	<i>you can do to</i>	14	91
	<i>you can take to</i>	17	47
	<i>you can take to</i>	17	47
	<i>steps you can take</i>	19	37

in other contexts as standalone messages (for example, as printed posters for buildings etc).

4.3. Function of images and n-grams

RQ1(c) How do common n-grams and images work together to form public health messages?

As mentioned previously, when we examine the common n-grams, they seem to be part of larger frames which functioned to either give advice or instructions. These same functions are also what the text and images together often attempted to convey. At times, this seemed to be clear and text and image served to reinforce one another and construct convergent meanings. Fig. 2 shows an example where this seems to happen. As in Fig. 1, the text and infographic in Fig. 2 has been annotated using facilities in Multimodal Analysis Image software (O'Halloran, Tan, & E, 2017) to highlight key features. In this case, the linguistic commands in the tweet text and the procedural graphic are reinforced visually through icons and drawings which illustrate the main concepts: i.e.

washing hands, keeping social distance and wearing a face mask. The intersemiotic relations between the text and image are strongly convergent through the repetition of the linguistic elements (in this case, three commands) in both the text tweet and the infographic, which also contains iconic representations of the three requested actions in the same order.

Specifically in this example, the embedded textual elements in the infographic make small changes to the Twitter text: *Keep social distancing* becomes *keep your distance*, *wear a face covering* becomes *wear a face covering when needed*, and *taking three simple steps can help save lives* becomes *simple steps save lives*. In this way, the text and images are closely aligned to reinforce the central function saving lives through three main instructions. The use of the imperative form throughout reinforces this, keeping ambiguity to a minimum. In other examples, however, there appeared to be more evidence of ambiguity. This was particularly the case when advice giving and instruction functions were used together in the same tweet, for example as displayed in Fig. 3. In this case, the intersemiotic relations partially converged, given that only one lexical item (“#COVID19 vaccination”) and one command (i.e. “wash your hands”)



It's important to follow the rules even after you've had the #COVID19 vaccination.

Always remember to:

- ✓ Wash your hands
- ✓ Cover your face
- ✓ Keep your distance.

More on what to do after the vaccination:

[gov.uk/government/pub...](https://www.gov.uk/government/publications)



4:00 PM · Feb 2, 2021 · Twitter Web App

are visualised iconically as a vaccination bottle and the act of handwashing respectively.

In this example, the text begins by use of an indirect speech act (Searle, 1969) stating a fact as advice. *It's important to follow the rules...* strongly implies the rules should be followed. There is also, of course, an assumption that the reader knows what the rules are, in this case presumably the three commands with green ticks which follow: *wash your hands*, *cover your face*, and *keep your distance*. The imperative forms make it clear that these are instructions and not advice. Following this, there is an iconic image of washing hands with the embedded linguistic text *Wash your hands*. We could perhaps read into this that we should follow the three rules (even though *Keep your distance* is in itself unclear in terms of how much distance and from whom) but that we must wash our hands as a priority; this is in itself somewhat of a mixed message. However, the image is a video (GIF) which, when played, repeats the message from the tweet text but alters *keep your distance* to *make space*. While this section of the tweet seems fairly clear, reinforcing the instruc-

... **Fig. 3.** Focus Group Prompt 6 - screenshot of image and text in Tweet six containing *it's important to*

tions given, the use of a vaccine dose bottle, with images depicting text such as *wash your hands* inside it, is potentially also ambiguous. The link between the vial of vaccine and these instructions is assumed and implied (for example, *even if you have had the #COVID19 vaccination* which appears in the tweet text), rather than made clear. Altering the directive from *keep your distance* to *make space* could also lead to potential confusion. If I am directed to *make space*, for example, how much space is needed and to whom is it provided? The implicature (Grice, 1969) could clearly be interpreted in more than one way here.

These examples demonstrate the challenge of constructing public health advice via a medium such as Twitter. Space is limited and the time spent reading a tweet has been estimated, on average, to be less than five seconds (Counts and Fisher, 2011). Therefore, the language used needs to contain the minimum of ambiguity and text and images need to co-contextualise each other to reinforce a single message. In some of the examples we looked at it, this seems to happen. In others, there is semiotic dissonance and a divergence in meaning (as seen in

Fig. 3), with various implicatures possible when we combine the language and images (see Oakey et al, 2022). This means that the message may be interpreted in different ways by members of the public. This was, of course, an assumption we made based on our analysis. The next research question involved testing these ideas by eliciting responses from three groups of EAL speakers, who represent one section of the public.

RQ2 (a). *How do EAL speakers interpret the intended messages?*

RQ2 (b). *What suggestions do they have regarding how the language or images could be made clearer for EAL speakers?*

As noted in the methodology section, the focus group data is organized thematically, rather than looking at each tweet in turn. Data was coded into three main themes: overall understanding and impressions of tweets, the most and least favoured tweets, and suggestions for improving the language or images for EAL speakers. Each theme is shown in turn and the tweet referenced is given in brackets following the example. A summary of each tweet used with the frequent n-gram underlined is given in the methodology section and the image, text and URL for each tweet can be found in Appendix 1.

In general, participants understood the basic messages within the tweets but reactions to them varied in some cases. There was no precise pattern to this but it was possible for some participants to dislike an image (e.g. the bottle used to represent a vaccine vial in tweet six), while others were more ambivalent about it. Several tweets did, however, elicit a generally positive response, while several elicited a generally negative response. There was also broad agreement on the most and least favoured tweets and on the ways in which they could be improved for EAL speakers.

4.4. Overall receipt of the message - understanding and impressions

The results from focus groups participants were consistent with the results of our multimodal analysis. For example, where our analysis suggested that the message was more effective when images and text were aligned, focus group participants agreed:

This picture makes me feel like this situation is more serious do I have to pay attention to the content like 3 steps right in this picture and I can find the title covid-19. (Tweet two).

I think I think is is clear obviously and it's it's very eye catching because is is red and it looks like something very dangerous is going to happen so yeah yeah that's basically how I feel about it (Tweet two).

I feel it's quite cute the pumpkin looks really you know happy and the for me is just a warm hearted reminder that's while I'm enjoying Halloween don't forget to wear face mask (Tweet five).

Where our analysis suggested the message was less likely to be effective due to a misalignment of the images and text, focus group participants also agreed:

I think the messages above the image is very clear but as for the four colourful images I think it kind of it kind of stopped it kind of stopped me from reading (Tweet four).

When I first see this picture I think it's not very clear it's just like a advertisement selling the laptop (Tweet one).

I just totally ignore that woman and see the text and don't haven't connect between them that's the same with other people I can't find a connection between them maybe I need to think for a while and find a connection but at the if only give me one second then I can't find the connection (Tweet one).

I don't know what's the meaning of the bottle yeah (Tweet six).

I think it's a little bit confusing I mean I think I understand it because where I work but I think the picture of the face wouldn't relate with symptoms and test positive I think if I wasn't a medical professional wouldn't have idea what that means (Tweet three).

Because it's like is that the the the the symptom may not be seen but at first I was thinking girl is sick searching the information online [obscure] right hand on the the the hands and face mask (Tweet one).

4.5. The most and least favoured tweets

Overall, the majority of participants agreed that Tweet two was clearest and Tweet four the least clear and most in need of change. Typical comments on Tweet two are as follows:

The sentences are very simple and the the images matched the message of the sentences very well and it uses a red frame to highlight the importance of these messages (Tweet two).

These steps are simple yes we everyone can do it maybe simple steps protect ourselves or simple steps for keep ourselves healthy things like that (Tweet two).

I think is is clear obviously and it's it's very eye-catching because is is red and it looks like something very dangerous is going to happen so yeah yeah that's basically how I feel about it ok (Tweet two).

Typical comments on Tweet four were as follows:

The colour is little bit weird like the the green one sorry the green one is too bright I can't see the words on it (Tweet four).

And also repeating the words is too wordy (Tweet four).

I would just scroll up I wouldn't read the entire text I wouldn't mind reading the images it's very confusing (Tweet four).

Well, I think there are too many words and the words are too tiny (Tweet four).

4.6. Suggestions for improvements

Participants gave a number of suggestions around improving the messages for other EAL speakers. These were centred on matching the text and images, simplifying the language used and breaking the instructions into steps by the use of bullets or lists. Typical comments were as follows:

Maybe it's just me but the second sentence on the image starting with the word 'not' and it's a bit a complex sentence structure so maybe those who don't really have high level of proficiency might not understand what this sentence exactly means. (Tweet one).

I prefer the last one (Tweet three) it listed the key factors like the first one is symptoms the second one is tested positive like one point second point is blah blah blah this is a whole sentence so it's quite a little bit long for me (Tweet four).

I think the text is quite long although it's easy to understand for a foreign learner but I think like we can make it shorter like using just a slogan not blah blah is important because blah blah blah I think we can simply use the slogan at like like vaccine can stop (Tweet four).

I think the first like 3-4 seconds for the for the video I think it should be a lot slower because I think it comes too maybe too quick for some some second language speaker (Tweet six).

That's a little bit too much for me to be honest yeah 'cause i i was just like taking quick look so two sentences like that would be too much for me yeah (Tweet four).

Just use simple sentence like word take the three simple steps like the sentence in the second image like word (Tweet three).

Overall, these participants' responses show that the messages of the tweets were in some cases easily understood and participants felt that the messages were clear and that the text and images reinforced one another. One example where this convergence was particularly the case was in tweet two, where it was noted that the language was simple and clear, broken into numbered points and linked closely to the image used. In other cases, participants felt that the language and the images were not aligned and this was potentially confusing due to the divergent meanings which were made (tweets one and six for example), that the sentence structures were too long, repetitious or complex (tweet four, for example) or were not broken down into numbered lists as in tweet two. These comments provide useful pointers towards improving such public health messages for EAL audiences and these are discussed in brief in the conclusion.

5. Conclusion and Recommendations

5.1. Summary of findings

The research reported in this article has shown the responses by EAL speakers to common phrases and imagery in UK public health agency tweets during the pandemic. Frequent four-word n-grams, and their associated images, found in our corpus of tweets across all agencies became the basis for a series of focus group prompts, thus applying the results of corpus research to a real world context.

The focus group data we obtained in response to these prompts provides a number of ways in which we believe the language and imagery of future public health messaging could be improved for EAL audiences in the UK. These improvements are based on the following recommendations.

Firstly, it is important to ensure that texts and images align clearly so that the language used is supported by the images chosen, reinforcing a message which is likely to be read very quickly. Related to this, colours which match the importance of the message and can be easily read need to be chosen. Tweet four, for example, was criticised because the colours used made it almost unreadable. In addition, the bright colour contrasts function to simultaneously attract attention, making it difficult to realise a coherent message.

It is also important to recognise that participants from different cultural backgrounds may understand colours differently. For example in tweet four, the colour red may not always be associated with danger and therefore it is important that focus groups represent as broad a range of first language speakers as possible.

Secondly the phraseological patterns which form part of larger frame could employ more imperative forms such as *help stop the spread of covid*, *wash your hands* or *get tested*, making the function one of simple, unambiguous instruction, more often. Some of the more ambiguous tweets may have been designed in this way in order to try and 'nudge' readers rather than directly tell them what to do. Nudging is a term from the field of psychology and was first coined by [Thaler and Sunstein \(2009\)](#). [Mulderigg \(2018\)](#) suggests that nudging is an attempt to change people's behaviour without them realising that this has happened. In a discussion of this phenomenon in relation to UK public health policy on obesity, she suggests that a key element of nudging lies in its attempt to feel non-coercive and lead us to make better choices of our own volition rather than feeling we are being forced to do something. Given that the UK government at the time of these tweets was known to employ a nudge unit ([Mulderigg, 2018](#)), this may explain why some of the tweets have semiotic dissonance, with different implicatures possible and do not always directly tell readers what they need to do. Based on the evidence of these focus groups, we would suggest that this is not a strategy that works well for EAL speakers. The nuance required in crafting messages whose illocutionary force is that of a nudge rather than a stronger directive may be too subtle for EAL speakers.

Thirdly, texts containing the frequent n-grams could be amended to ensure as many sentences as possible are simplified and use one clause at a time. For example, tweet two contains the n-gram *you can do to*, in the text *Vaccination is the most important thing you can do to protect...* Such a construction could easily be simplified to: *Vaccination is important. It protects us all.* Complex sentences with several clauses such as the following from tweet four made the messages too long and harder to understand for these participants: *When enough people get vaccinated it's harder for the disease to spread to those who can't get vaccinated.* This finding reinforces a point made by [Houts, Doak, Doak and Lozaldo \(2006\)](#), in a wide ranging review of the evidence for using images to support healthcare communication. Finding a positive effect for the use of pictures in areas such as patient recall of the message, they suggest that simple language should be used and that it is closely linked to the images chosen.

Lastly, bullet points and numbers can be used to break messages into clear, simple steps which can be understood at a glance. Tweet two was

the best understood and liked by all three focus groups precisely because it does this.

These suggestions may seem like common sense but, as mentioned above, the texts and images likely to produce effective messages for EAL speakers may not be common sense to politicians or communications units intent on using 'nudge' tactics. Such messaging may result in language and imagery which, in its attempt to appear non-coercive, reads as ambiguous or unclear to an EAL speaker. Tweet one, for example, may seek to show the reader an image of someone working from home while simply stating the risks from COVID-19 (*Anyone can spread the virus. Even you*) in an attempt to nudge the reader into changing their behaviour, perhaps by working from home where possible. However, as evidenced in these focus groups, EAL speakers were unclear about what the image was trying to convey and how it linked to the text. In short, they were confused about what they were supposed to do and so would be unlikely to act. [Mulderigg \(2018\)](#) discusses these issues in relation to public health campaigns on obesity aimed at children and families, and shows that language in such campaigns can be similarly poorly chosen. She also makes the point that the intention behind such communication (and thus the use of ambiguity) may also be to place the emphasis on changes to individual behaviour and thus blame can be apportioned to individuals (not governments) if this does not change. This could be another explanation for the use of language and images in some of these tweets.

5.2. Future directions

One obvious limitation of this study is that our sample was based on a limited sample of L1 speakers. That is, there were a larger proportion of Chinese L1 speakers when compared with other nationalities, as noted in the methodology section. Future studies could attempt to address this by trying to employ a sample which better represents the most common first languages spoken in the UK. According to the [UK office for national statistics \(2021\)](#) this would currently include. Polish, Romanian, Punjabi and Urdu speakers.

A similar study design to the one used here could also be employed to understand how other public health campaign messages are received, understood and responded to by EAL speakers. This might include, for example, campaigns aiming to increase take up of influenza vaccinations or anti-smoking campaigns. A multimodal corpus of tweets can be created and analysed and then focus groups conducted, recorded and analysed. The findings could then be used to inform the formulation of future messages in a bid to make them more effective for EAL speakers.

Finally, [UK Parliamentary Office of Science and Technology \(2020\)](#) also identified the need to know how other marginalised groups (such as people with learning disabilities) receive, understand and respond to public health messages. While their focus was on COVID-19, future work in this area could also address other types of public health messages with such groups.

5.3. Final thought

In this article we have shown how multimodal corpus analysis can be applied to the real world problem of constructing effective public health messages for EAL speakers via the use of focus groups. Such application shows that this kind of methodology can produce clear insights which could be used to improve the messaging of future public health campaigns for EAL speakers and other members of the public. In addition, educational programmes aimed at developing multimodal literacy skills (e.g. [Danielsson & Selander, 2016](#); [Holsanova, 2020](#); [Jewitt & Kress, 2003](#); [Lim, 2018](#); [O'Halloran, 2023](#); [Unsworth & Macken-Horarik, 2015](#); [Walsh, 2010](#)) would enhance EAL speakers' ability to read and understand text and images and how they combine to create meaning in the digital world of today.

There is clearly no such thing as a perfect public health message. However, this study shows that a combination of multimodal corpus

analysis and focus group data can be used to effectively analyse how such messages are received, understood and responded to by groups such as EAL speakers. We hope that in future such analysis can then be used to help formulate more effective public health messages with a strong convergence of meaning which can be readily understood by a variety of speakers from different communities.

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.

Acknowledgements

Support for this project was provided from the Research Development Initiative fund in the School of the Arts at the University of Liverpool.

We acknowledge the contribution of our collaborators on this project:

Dr Nathan Bray, School of Health Sciences, Bangor University

Dr Kate Fleming, Institute of Population Health Sciences, University of Liverpool

Mr Gautam Pal, Department of Computer Science, University of Liverpool

Dr John Tulloch, Institute of Infection and Global Health, University of Liverpool

Mr Minhao Zhang, Department of Computer Science, University of Liverpool

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.acorp.2023.100053](https://doi.org/10.1016/j.acorp.2023.100053).

References

- Aragbuwa, A., Adejumo, V., 2021. Covid-19 #Takeresponsibility: a multimodal discourse analysis of select NCDC's online public health advertising campaign. *Journal of Language and Literature* 21 (2), 267–280. doi:10.24071/joll.v21i2.3044.
- Baldry, A., Bianchi, F., Loiacono, A., 2019. *Linguaggi specialistici e Traduzione Tecnica* (LiSpeTT – Specialised Languages and Translation). Representing and redefining specialised knowledge: Medical discourse. Università del Salento, Italy.
- Bateman, J., Wildfeuer, J., Hiippala, T., 2017. Multimodality: Foundations, research and analysis – A problem oriented introduction. De Gruyter Mouton, Berlin, Boston doi:10.1515/9783110479898.
- Banerjee, D., Meena, K.S., 2021. COVID-19 as an “Infodemic” in public health: Critical role of the social media. *Frontiers in Public Health* 9 (231). doi:10.3389/fpubh.2021.610623.
- Barbour, R.S., Morgan, D.L. (Eds.), 2017. *A new era in focus group research: Challenges, innovation and practice*. Palgrave Macmillan.
- Bateman, J., 2014. *Text and image: A critical introduction to the visual/verbal divide*. Routledge.
- Bednarek, M., Caple, H., 2014. Why do news values matter? Towards a new methodological framework for analysing news discourse in critical discourse analysis and beyond. *Discourse & Society* 25 (2), 135–158. doi:10.1177/0957926513516041.
- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., 1999. *Longman grammar of spoken and written English*. Longman.
- Caple, H., 2018. Lucy says today she is a “Labordoodle”: How the dogs-of-Instagram reveal voter preferences. In: *Social Semiotics*, pp. 1–21. doi:10.1080/10350330.2018.144358.
- Carter, R., 1998. *Vocabulary: Applied linguistic perspectives*, 2nd ed. Routledge.
- Council of Europe, 2022. *CEFR Global scale: common reference levels*.
- Counts, S., Fisher, M., 2011. *Taking It All In? Visual Attention in Microblog Consumption*. Association for the Advancement of Artificial Intelligence.
- Danielsson, K., Selander, S., 2016. Reading multimodal texts for learning – A model for cultivating multimodal literacy. *Designs for Learning* 8 (1), 25–36. doi:10.16993/dfl.72.
- Davies, M., 2021. The coronavirus corpus: Design, construction, and use. *International Journal of Corpus Linguistics* 26 (4), 583–598. doi:10.1075/ijcl.21044.dav.
- Ellis, N.C., 2008. *Phraseology: The periphery and the heart of language*. In: Meunier, F., Granger, S. (Eds.), *Phraseology in foreign language learning and teaching*. John Benjamins, pp. 1–13.
- Firth, J.R., 1957. *A synopsis of linguistic theory 1930-1955*. In: Firth, J.R., Firth, J.R. (Eds.), *Studies in linguistic analysis*. Basil Blackwell, pp. 1–32.
- Glaser, R., 1998. The stylistic potential of phraseological units in the light of genre analysis. In: Cowie, A.P. (Ed.), *Phraseology: Theory, analysis, and application*. Clarendon Press, pp. 125–143.
- Greszczuk, C., 2020. Making messages work. *Health Foundation*, 21st April 2020. <https://www.health.org.uk/what-we-do/a-healthier-uk-population/thinking-differently-about-health/making-messages-work>.
- Grice, H.P., 1969. Utterer's meaning and intention. *The Philosophical Review* 78 (2), 147–177. doi:10.2307/2184179.
- Halliday, M.A.K., 2009. *Collected works of M.A.K. Halliday (10 Volumes)*. Continuum.
- Holsanova, J., 2020. Uncovering scientific and multimodal literacy through audio description. *Journal of Visual Literacy* 39 (3-4), 132–148. doi:10.1080/1051144X.2020.1826219.
- Houts, C.C., Doak, L.G., Doak, M.J., Loscalzo, M.J., 2006. The role of pictures in improving health communication: A review of research on attention, comprehension, recall, and adherence. *Patient Education and Counseling* 61 (2), 173–190. doi:10.1016/j.pec.2005.05.004.
- Hunston, S.E., 2020. Changing language in unprecedented times. *University of Birmingham, Department of English News*, 8th April 2020. <https://www.birmingham.ac.uk/schools/edacs/departments/englishlanguage/news/2020/changing-language.aspx>.
- Jewitt, C., Kress, G., 2003. *Multimodal literacy*. Peter Lang.
- Jones, C., Byrne, S., Halenko, N., 2018. *Successful spoken English: Findings from learner corpora*. Routledge.
- Kelle, U., 2002. *Computer-Assisted Qualitative Data Analysis*. In: Seale, C., Gobo, G., Gubrium, J.F., Silverman, D. (Eds.), *Qualitative Research Practice*. Sage Publications, p. 473–489.
- Kress, G., van Leeuwen, T., 2020. *Reading images: The grammar of visual design*, 3rd ed. Routledge.
- Lim, F.V., 2018. Developing a systemic functional approach to teach multimodal literacy. *Functional Linguist* (13) 5. doi:10.1186/s40554-018-0066-8.
- Macnaghten, P., 2017. Focus groups as anticipatory methodology: A contribution from science and technology studies towards socially resilient governance. In: Barbour, R.S., Morgan, D.L. (Eds.), *A new era in focus group research: Challenges, innovation and practice*. Palgrave Macmillan, pp. 343–363.
- Mahlberg, M., Brookes, G., 2021. Language and Covid-19: Corpus linguistics and the social reality of the pandemic. *International Journal of Corpus Linguistics* 26 (4), 441–443. doi:10.1075/ijcl.00043.mah.
- Moon, R., 1998. *Fixed expressions and idioms in English: A corpus based approach*. Clarendon Press.
- Morgan, D.L., 1997. *Focus Groups as Qualitative Research*, Second Edition Sage Publications.
- Mulderrig, J., 2018. Multimodal strategies of emotional governance: a critical analysis of ‘nudge’ tactics in health policy. *Critical Discourse Studies* 15 (1), 39–67. doi:10.1080/17405904.2017.1382381.
- O'Halloran, K.L., 2023. Developing competence in multimodal discourse. In: Tomlinson, B. (Ed.), *Developing Materials for Language Teaching*, 3rd edition. Bloomsbury.
- O'Halloran, K.L., Pal, G., Jin, M.H., 2021. Multimodal approach to analysing big social and news media data. *Discourse, Context and Media* 40. doi:10.1016/j.dcm.2021.100467.
- O'Halloran, K., Tan, S., Smith, B., 2016. Multimodal approaches to English for academic purposes. In: Hyland, K., Shaw, P. (Eds.), *The Routledge handbook of English for academic purposes*. Routledge, pp. 256–269.
- Oakey, D.J., 2020. Phrases in EAP academic writing pedagogy: Illuminating Halliday's influence on research and practice. *Journal of English for Academic Purposes* 43, 1–15. doi:10.1016/j.jeap.2019.100829.
- Oakey, D.J., 2022. What can a corpus tell us about lexis? In: O'Keefe, A., McCarthy, M. (Eds.) *The Routledge handbook of corpus linguistics*. Routledge, pp. 185–203.
- Oakey, D.J., Jones, C., O'Halloran, K.L., 2022. Phraseology and imagery in COVID-19 public health messages. In: Tan, S., E, M.K.L. (Eds.), *Discourses, modes, media and meaning in an era of pandemic: A multimodal discourse analysis approach*. Routledge, pp. 89–114.
- O'Halloran, K. L., Tan, S., & E, M. K. L. (2017). Multimodal analysis for critical thinking. *Learning, Media and Technology*, 42(2), 147–170. doi:10.1080/17439884.2016.1101003.
- O'Keefe, A., McCarthy, M., Carter, R., 2007. *From corpus to classroom - Language use and language teaching*. Cambridge University Press.
- Oliver, D., 2020. Covid-19 highlights the need for effective government communications. *British Medical Journal* 369, m1863. doi:10.1136/bmj.m1863.
- O'Toole, M., 2011. *The language of displayed art*, 2nd ed. Routledge.
- Oxford English Dictionary, 2020. *Corpus analysis of the language of Covid-19*. OED Blog, 14 April 2020. <https://public.oed.com/blog/corpus-analysis-of-the-language-of-covid-19/>.
- Park, H., Reber, B.H., Chon, M.G., 2016. Tweeting as health communication: Health organizations' use of twitter for health promotion and public engagement. *Journal of Health Communication* 21 (2), 188–198. doi:10.1080/10810730.2015.1058435.
- Public Health England (2014). *Health matters: public health issues: Facts, resources and information on major public health issues for public health professionals, local authorities and CCG commissioners*. <https://www.gov.uk/government/collections/health-matters-public-health-issues>.
- Schmitt, N., Grandage, S., Adolphs, S., 2004. Are corpus-derived recurrent clusters psycholinguistically valid? In: Schmitt, N. (Ed.) *Formulaic sequences*. John Benjamins, pp. 127–151.
- Schuessler, J., 2020. Oxford's 2020 Word of the Year? It's too Hard to Isolate. *New York Times*, 22nd November 2020. <https://www.nytimes.com/2020/11/22/arts/oxford-word-of-the-year-coronavirus.html>.
- Searle, J., 1969. *Speech acts*. Cambridge University Press.
- Sinclair, J.M., 1991. *Corpus, concordance, collocation*. Oxford University Press.

- Sinclair, J.M., 2005. Corpus and text - Basic principles. In: Wynne, M. (Ed.), *Developing linguistic corpora: A guide to good practice*. Oxbow Books, pp. 1–16.
- Tan, S., E, M.K.L. (Eds.), 2022. *Discourses, modes, media and meaning in an era of pandemic: a multimodal discourse analysis approach*. Routledge.
- Tan, S., O'Halloran, K.L., Wignell, P., 2020. Multimodality. In: De Fina, A., Georgakopoulou, A. (Eds.), *Handbook of Discourse Studies*. Cambridge University Press, pp. 263–281.
- Thaler, R.H., Sunstein, C., 2009. *NUDGE: Improving decisions about health, wealth, and happiness*. Penguin.
- Twitter Advanced Search, nd. <https://twitter.com/search-advanced?lang=en-GB>.
- UK Office for National Statistics, 2021. *Language, England and Wales: census 2021. Main language, English language proficiency and household language in England and Wales, Census 2021 data. The proportion of adults using English as their main language and a breakdown of other languages used in the UK*.
- UK Parliamentary Office of Science and Technology (2020) Area of Research Inquiry 10.3. <https://post.parliament.uk/covid-19-areas-of-research-interest/>.
- Unsworth, L., Macken-Horarik, M., 2015. Interpretive responses to images in picture books by primary and secondary school students: Exploring curriculum expectations of a 'visual grammatics'. *English in Education* 49 (1), 56–79. doi:10.1111/eie.12047.
- van Leeuwen, T., 2005. *Introducing social semiotics*. Routledge.
- Walsh, M., 2010. Multimodal literacy: What does it mean for classroom practice? *Australian Journal of Language and Literacy* 33 (3), 211–239.
- World Health Organisation (WHO) (2020) Naming the Coronavirus Disease (COVID-19) and the Virus that Causes it., [https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-\(covid-2019\)-and-the-virus-that-causes-it](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it).