5 Phraseology and imagery in UK public health agency COVID-19 tweets

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

Introduction

A large amount of public health information was produced to inform the public about the outbreak of COVID-19. Textual elements were accompanied by the visual: Images and videos of medical staff and other frontline workers, patients, medical facilities, politicians, and members of the public and public spaces. Many studies exist on the language of COVID-19 (e.g. Hunston, 2020; Oxford English Dictionary (OED), 2020) and the effectiveness of public health information in terms of message organisation and "framing" (Banerjee & Meena, 2021; Greszczuk, 2020; Oliver, 2020). What has been less studied so far has been the way textual and visual information representations combine in various media formats to impact on how the messages are received, understood and acted upon. In this regard, it is critical to first understand the common forms and functions of language and images as a contribution to understanding the effectiveness of current and future public health messages. However, we are not aware of multimodal studies that have explored in detail the language and imagery used in public health messages related to the COVID-19 pandemic.

The first stage of this project therefore, as reported in this chapter, uses corpus linguistics and social semiotic analysis to examine the phraseology and imagery of public health information to understand how linguistic and visual representations combine to shape COVID-19 public health information messages. Discovering and disseminating this new knowledge requires multidisciplinarity and multimodality. In this case, we draw on our expertise in the field of applied linguistics and multimodal social semiotics (e.g. Halliday, 1978; O'Halloran, Tan, & Wignell, 2019; van Leeuwen, 2005). In linguistics, "phraseology" seeks to describe how words tend to form clusters that appear to be the result of a single choice (e.g. *I don't know*). Such patterns carry with them particular meanings and thus form the narrative of particular forms of information. At the same time, multimodal social semiotic analysis identifies the functions of major messaging components: For example, happenings constructed

DOI: 10.4324/9781003168195-8

through processes, participants and circumstances; the logical relations between these happenings; and the stance towards them. We combine these approaches and use corpus linguistics and multimodal social semiotic analysis to examine the phraseology and imagery of public health information to understand how linguistic and visual representations combine to shape COVID-19 public health information messages.

General research aims/questions

As already mentioned, this research project looks at how people understand the language and images in public health information about COVID-19. Many researchers in health communication, psychology and the media have already studied how well public health messages are received by the public, but we do not know enough about how the language and images used in this public health information work together to convey certain meanings. Crucially, we do not yet understand how frequent phrases combine with visual representations to shape the overall meaning of public health messages. Such understanding is extremely important both now and for the improvement of future COVID-19 public health messages. The broad aims of this chapter are therefore as follows:

1. To identify the most common phrases used by UK public health agencies to provide public health information about COVID-19 on Twitter
2. To identify the common images used with the aforementioned common phrases
3. To understand how both phrases and images function in this public health discourse

As mentioned, this study is part of a broader project, which has a main aim of applying this analysis to focus groups made up of speakers of English as an additional language (EAL) in the UK. The aim here is to investigate how well these common phrases and images are understood by this group and how they could be improved for future public health information campaigns. This chapter focuses on the first part of this study; namely, the identification of the most common phrases and images used by UK public health agency information about COVID-19 on Twitter.

Literature review

This section reviews previous work in the areas drawn upon for this study: It briefly describes work on language, corpora and COVID-19, then discusses corpus design issues, phraseology and multimodal social semiotic analysis.

Language, corpora and COVID-19

The disease caused by SARS-CoV-2 virus was given the name "COVID-19" by the World Health Organization early in 2020 (WHO, 2020). This name soon appeared as a lexical item in written and spoken news and social media texts. As these texts became archived in online corpora, studies by corpus linguists soon revealed the rapid expansion in use of the words "coronavirus" and "COVID-19" between December 2019 and March 2020, and the associated rise in frequent collocations of these terms: *Outbreak, infection, spread* and *fear* (OED, 2020). Frequent collocations of other words also changed rapidly. Prior to December 2019, for example, the *Oxford English Dictionary* (OED) corpus showed that the most common collocates of *remote* were nouns such as *control*, *island* and *village*, but following the rise in the numbers of people working from home by the end of 2020, the most frequent collocates of *remote* had suddenly become *working, learning* and *work force* (Schuessler, 2020, cited in Oakey, 2022). Such analysis is revealing about recent language changes and serves to illustrate the dynamic nature of phraseology. A corpus such as the British National Corpus compiled in the early 1990s (BNC World, 2001), for example, has no examples of *remote learning*.

Beyond immediate studies of changing language use, corpus linguists then produced studies of how language use reflected people's experiences during the COVID-19 outbreak, and how their "attitudes and beliefs are constructed and mediated through language use" (Mahlberg & Brookes, 2021, p. 441). Some researchers have collected corpora specifically to identify the effects of the pandemic on language use. For example, the Coronavirus Corpus (Davies, 2021) contains COVID-related news articles that are included if they contain at least two occurrences of the words *coronavirus*, *COVID* or *COVID-19*, or have one of several words in their titles, such as *at-risk*, *cases*, *confirmed*, *contagious*, *hydroxychloroquine*, *outbreak*, *pandemic*, *stay-at-home* and so on. Such corpora are of course useful and they enable us to better understand how news discourse about this pandemic operates linguistically speaking, and as Davies (2021, p. 587) points out, "by creating a stand-alone corpus dealing just with COVID-19, the corpus will be much smaller, and searches will be faster".

However, to some degree, a specialised corpus "about" COVID stands in opposition to Sinclair's (2005, p. 5) dictum 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". Collecting a corpus on linguistic grounds by using search words fixes the epistemological horizon the researchers already have in mind before they interrogate the corpus, and thus pre-imposes "aboutness" on the data. Sinclair (2005), by contrast, observed that the whole point of a corpus collected without prior reference to linguistic criteria was that it could reveal unpredicted usages that researchers could

not or would not have thought of themselves beforehand in the form of search words. In collecting the data for the corpus in this study, therefore, we have sought to avoid using pre-imposed linguistic criteria (see "Methodology" section for more details). Moreover, it is difficult to establish precisely whether something is or is not "about" COVID-19. The tweet in corpus extract 1, for example, is about cervical screening rather than COVID-19 per se, and yet it relates to conditions imposed by the pandemic:

> Corpus extract 1.
>
> Cervical screening is still happening and is considered essential travel, so please don't ignore your next invite.
>
> (P_H_S_official, 22 January 2021)

While the context around other n-grams in our data sometimes contained words related to cancer, flu or mental health, these words were in tweets sent only because of circumstances arising from the pandemic. So, while these tweets, therefore, may not have contained the specific lexical items used by Davies, they can still be said to be "about" COVID-19.

Phraseology and n-grams

The principal textual unit of analysis used in this chapter is the *n-gram*, a fixed combination of *n* words that occurs repeatedly in a corpus. The study of n-grams belongs to the area of phraseology, which is concerned with how words combine and make up users' repertoires of formulaic language. Since the only practicable way to identify and count n-grams is to use a computer corpus, the description and study of these items has been a relatively recent addition to the field since the application of computers to the study of language. The traditional focus of phraseological study was on lexical and syntactic relations between words, as in the case of collocation and colligation (Carter, 1998; Firth, 1957), the somewhat quirky semantic behaviour of language features such as idioms and proverbs (Moon, 1998), and socially conventionalised acts such as greetings, prayers and ritual formulae (Glaser, 1998). Phraseologists were interested in how the interplay between spoken and written formulaic utterances leads to words tending to cluster in frequently occurring patterns that appear to be the result of a single psycholinguistic choice by the speaker or writer, and which are formed appropriately to meet the expectations of particular discourse communities (Howarth, 1998; Pawley & Syder, 1983; Wood, 2002; Wray, 2002).

The introduction of computer corpora in phraseology revealed how common n-grams were in language production and this ubiquity led to insights into language cognition and processing, suggesting that n-grams were part of a single psycholinguistic choice by the user (Ellis, 2008;

Schmitt et al., 2004; Sinclair, 1991). This observation led to two decades during which the n-gram and, a closely related form, the lexical bundle (an n-gram occurring in at least ten different texts at least 40 times per million words) (Biber et al., 1999) were widely analysed in both spoken and written corpora (Oakey, 2020). N-grams first were used to illustrate differences in form between registers of English. A key observation, for example, was that in academic prose lexical bundles (Biber et al., 1999, p. 1012) comprised mostly noun phrase elements such as *the end of the*, while lexical bundles in conversation usually contained fragments of clauses or verb phrases, e.g. *I don't know what* (Biber et al., 1999, p. 1002).

In addition to their structural forms, n-grams were then analysed in terms of their discourse function (Biber, 2006; Biber et al., 2004; Cortes, 2004; Hyland, 2008) aligning with Halliday's (Halliday & Matthiessen, 2014) metafunctional categories of the ideational, interpersonal and textual (Oakey, 2020). In Cortes's taxonomy (Cortes, 2004, p. 41) referential or ideational n-grams such as *at the beginning of* and *the end of the* function as time, place or text markers. Interpersonal stance n-grams such as *I don't know why* and *are more likely to* express attitudes that frame some other proposition, while text organizers carry the message in relation to the surrounding discourse, such as marking contrast, e.g. *on the other hand*, causation, e.g. *as a result of*, or emphasis, e.g. *it is important to*.

Both these formal and functional aspects of n-grams are used in the research in the current chapter to reveal more about the textual side of frequent messages in UK public health agency tweets. That said, tweets are not conveyed just using words, but by the juxtaposition of pictures, text and video. For this reason, this study also investigates "how language, images and other resources work together to create meaning" (O'Halloran et al., 2016, p. 256) to identify the influence of the image on understanding of the text and the influence of the text on understanding of the image.

Multimodality

As long recognised in multimodal studies (e.g. Bateman, 2014; Kress & van Leeuwen, 2020; Tan et al., 2020; van Leeuwen, 2005), meaning is the result of the integration of different semiotic resources; in this case the text and images (and videos), which work together to construct the UK Public Health agency messages. In what follows, we examine how the text and images work together to co-contextualise or re-contextualise the meanings made by each semiotic resource. That is, do the text and images function together to reinforce and expand the meaning of the public health messages, or is there some form of semiotic dissonance between the two? The text and images in the Public Health England (PHE) tweets about COVID-19 were investigated using the Multimodal

Analysis Platform (MAP),[1] a cloud-based application for collecting, indexing, storing and analysing online media texts (O'Halloran et al., 2021). In this study, MAP was used to collect the Twitter data and identify the most common n-grams and the accompanying images.

Specific research questions

Bringing together the separate strands of work on the areas on phraseology, corpus linguistics and social semiotics, this study aims to answer the following questions in relation to the UK public health agency tweets:

RQ1 What are the most frequent n-grams in a corpus of tweets from UK public health agencies in the first 11 months of the COVID-19 pandemic?
RQ2 Which images are associated with these n-grams?
RQ3 How do these frequent n-grams and images typically function in this data?

Methodology

Data collection

Public health organisations have long used Twitter for public health messages (Park et al., 2016). For this study a corpus of all tweets sent by UK public health agencies during the first 11 months of the pandemic (i.e. 1 March 2020 to 17 February 2021) was collected without prior reference to linguistic content, as mentioned in the literature review. The reason for this was to try and avoid pre-judging the results based on a definition of what words are and are not related to COVID-19. Since we are looking at the functional role of n-grams, as well as their formal structures, it is important to collect all texts from all Twitter accounts so that we can examine them, following Sinclair (2005, p. 5), "according to their communicative function in the community in which they arise".

For this purpose, data were collected, indexed and stored automatically by the web crawling application in our Multimodal Analysis Platform (MAP). MAP contains dashboards for visualising the results (O'Halloran et al., 2021). Figure 5.1 shows an example of a MAP dashboard output for the n-grams in this study.

MAP was used to collect all tweets sent from all 13 UK public health agency Twitter accounts. These included Public Health England (reorganised as the UK Public Health Security Agency (UKHSA) in 2021) and other regions around England, as well as Public Health Scotland, Public Health Wales and the Public Health Agency in Northern Ireland (see Table 5.1). The data collection included the textual and visual components of the messages in written and image formats, and comprised approximately 15,400 tweets and 564,000 tokens, as shown in Table 5.1.

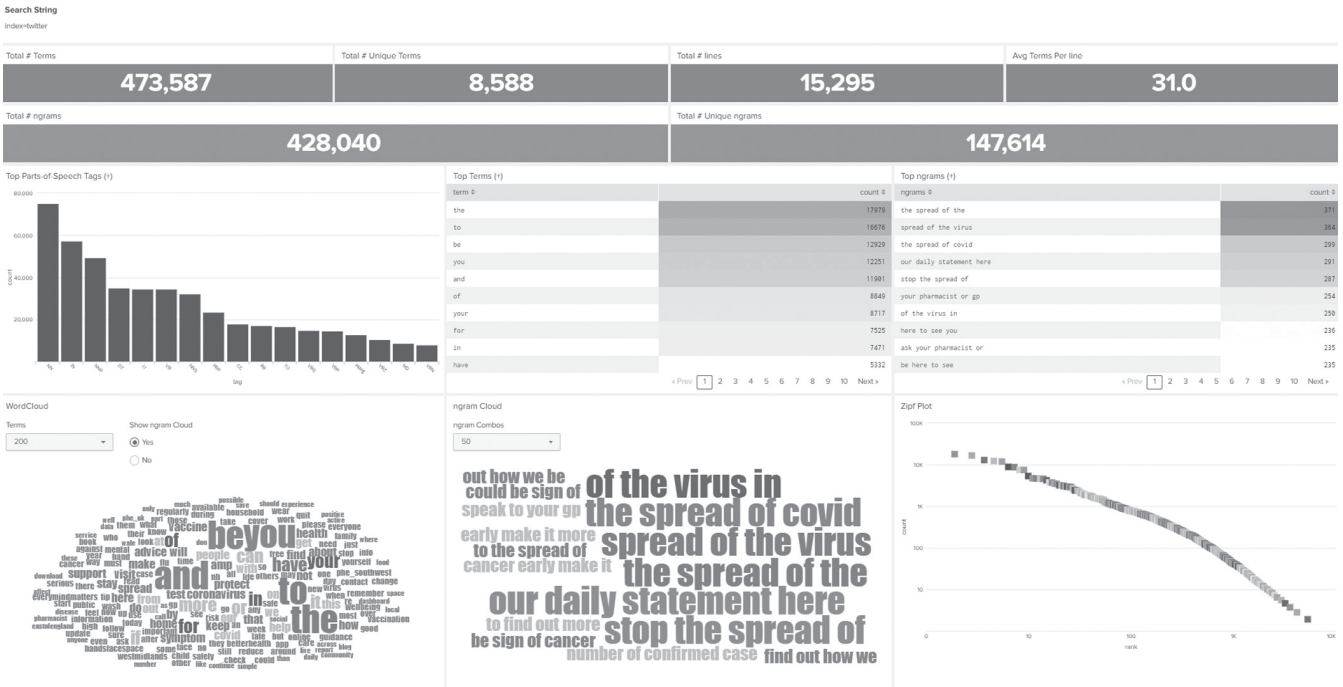

Figure 5.1 MAP dashboard output.

Table 5.1 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**

Language analysis

For linguistic analysis, we focused on the n-gram, a phraseological feature consisting of frequent recurring fixed strings of words (see "Phraseology and n-grams" in the literature review for a fuller discussion). While the MAP platform was initially used to identify the n-grams and surrounding linguistic context, MAP indexes n-grams so that all forms of a lemma are counted as one occurrence of the base form, for example *it was important* and *it is important* are both counted as occurrences of *it be important*. Such lemmatisation affects the position of this n-gram in the frequency list, and so Antconc 4.0.5 (Anthony, 2022) was therefore used to reidentify the n-grams and surrounding linguistic context so that all separate instances of n-grams containing distinct forms of each lemma were counted.

In this study, we chose to focus on only four-word n-grams for several reasons. Firstly, three-word n-grams tend to be embedded in four-word n-grams (e.g. *the spread of/the spread of COVID*) and so will already be included in the analysis. Second, four-word n-grams are at least more likely to be syntactically whole and thus more meaningful than nominal or clausal fragments often found in three-word n-grams. Lastly, in most corpora, three- and four-word n-grams tend to be the significantly higher in frequency than longer n-grams (e.g. Jones et al., 2017; O'Keeffe et al., 2007). To ensure that the full discourse community was represented, only n-grams with a range of 13, namely they occurred in all

thirteen public health accounts, were included. HTML was deleted and the initial hashtags for individual tweets were listed in a separate column and not included in counting the number of tokens. Other hashtags were retained in the data if they played a syntactic role in a clause element, for example "it's important to remember that #COVID19 is still with us". These quantitative data were used to answer the first research question concerning the most frequent n-grams in the corpus.

There were two possible directions for matching the images and the frequent n-grams; either to move from n-gram to image, or image to n-gram. We chose to move from n-gram to image by using MAP to help us search for the images that occurred alongside the frequent n-grams. The rationale for this choice was that it was not really possible to find the frequent images because MAP finds images signalled either by appearing near a particular hashtag or tweet and images are also represented by file names. The same image could therefore have several file names and be used with a number of different n-grams. For this reason, we used MAP to give us an indication of some images that occurred with common n-grams. We then searched for images through the Twitter site "advanced search" tool (Twitter.com, n.d.). This ensured the juxtaposition of 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) found using the Twitter site "advanced search" tool is shown in Figure 5.2. In this case, the image in Figure 5.2 is a photograph that has been annotated using facilities in Multimodal Analysis Image software[2] (O'Halloran et al., 2017) according to the framework in Table 5.2 (see "Image analysis" and "Results and Discussion" sections below).

These data were used to answer research question two concerning the images that were found associated with the most frequent n-grams.

Image analysis

To understand the text and image relations, we undertook a social semiotic analysis of the images that accompanied the most frequent n-grams using the framework for image systems displayed in Table 5.2. Halliday's (2009) metafunctional categories of ideational (experiential and logical), interpersonal and textual meaning are realized through systems organized at the rank of work, episode and figure in the images, following O'Toole (2011). The images in the UK Public Health Twitter dataset consist of various genres, including photographs, infographics and posters. Therefore, the elements specific to each genre (for example, infographic headlines, questions, commands and icons) were also included in the analysis. Multimodal Analysis Image software (O'Halloran et al., 2017), which has facilities for entering systems and annotating the images according to the generic elements and the framework in Table 5.2, was used to conduct the analysis (see Figures 5.2 to 5.5).

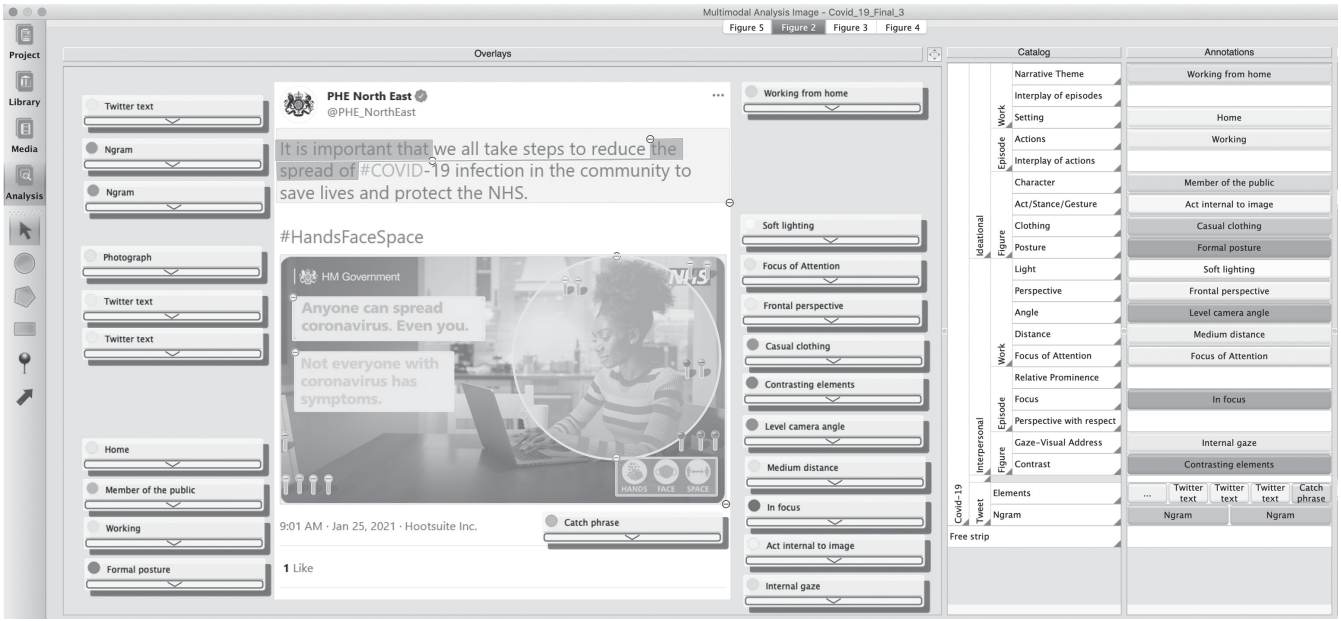

Figure 5.2 A sample tweet with an n-gram and accompanying image.

Table 5.2 Image systems (adapted from Kress & van Leeuwen, 2020; and O'Toole, 2011)

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

Language and image analysis

Finally, to answer the third research question concerning the text and image relations, we explored the n-grams and images in context via concordance lines, to understand the relationship between form and function. Such analysis is necessarily qualitative as it involves subjective statements about the likely intended illocutionary force (Searle, 1969) and implicature (Grice, 1975) on behalf of the message writer. This mixture of quantitative and qualitative data analysis is one widely employed in corpus analysis (e.g. Biber, 2006; Hyland, 2008; Jones & Waller, 2015;

Figure 5.3 Procedural infographic used with the n-gram *if you have symptoms*.

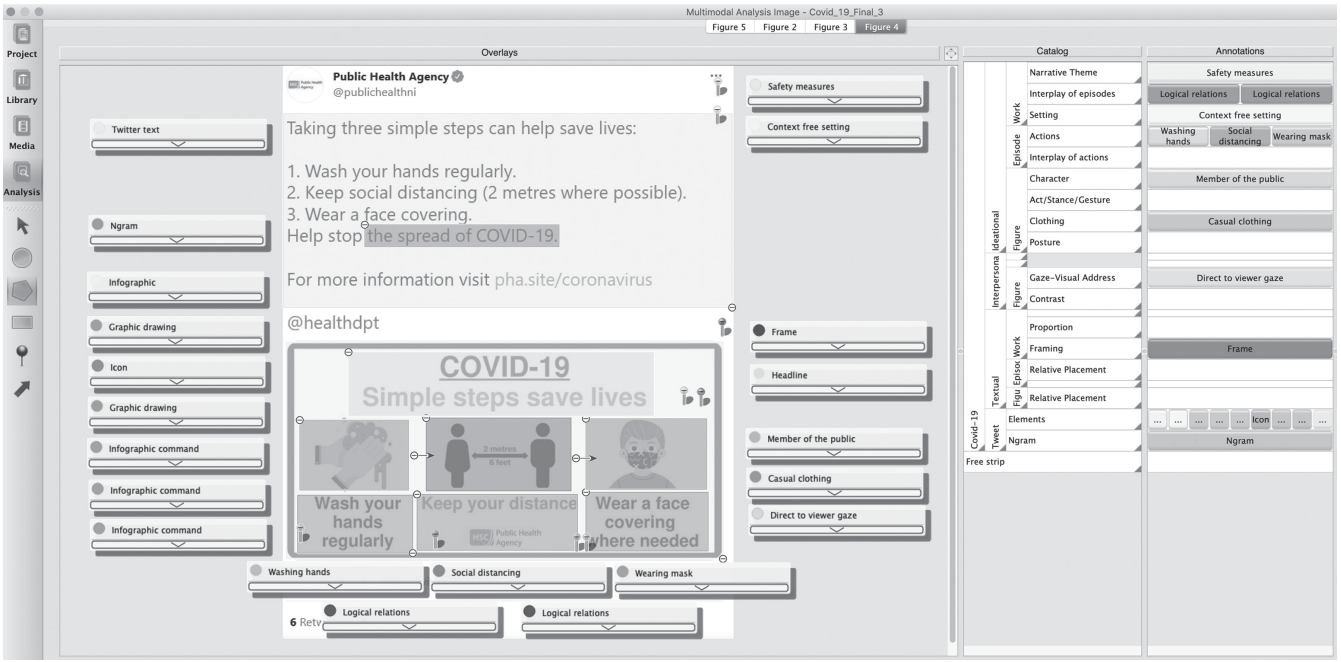

Figure 5.4 Procedural infographic used with the n-gram *the spread of Covid*.

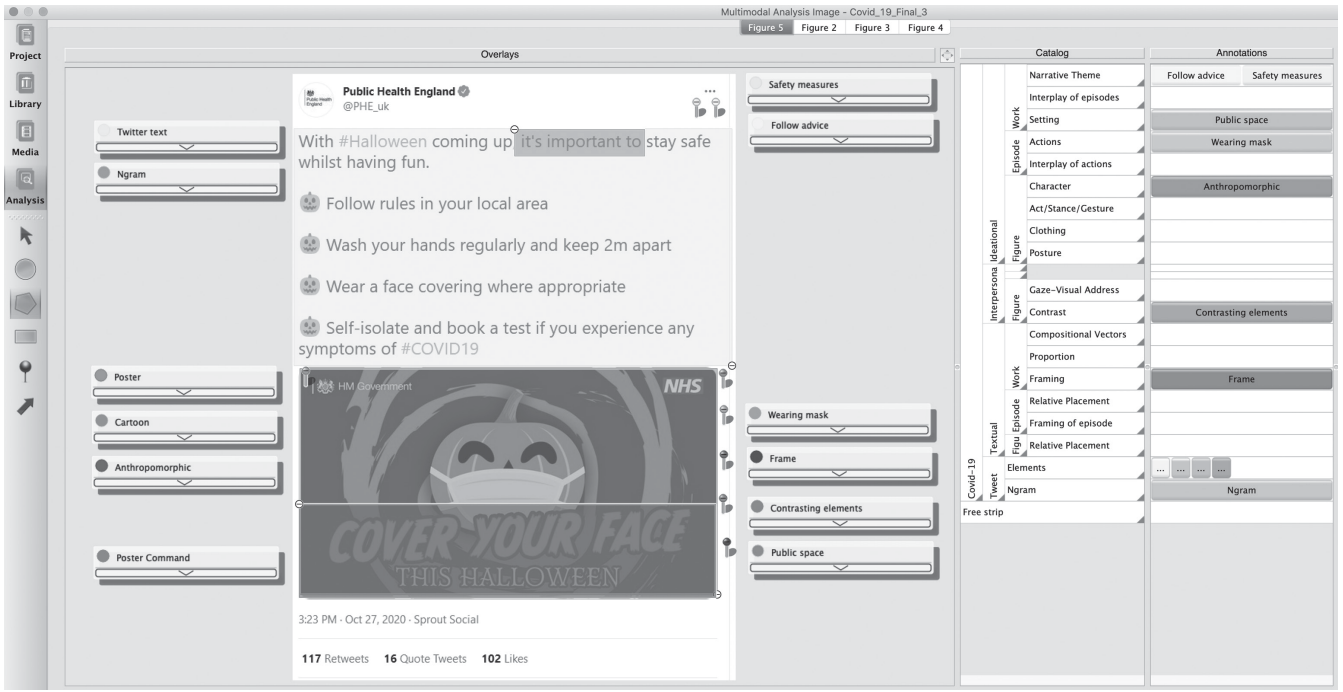

Figure 5.5 Poster used with the n-gram *It is important to*.

Oakey, 2020). In this case, the approach includes multimodal social semiotic analysis of the images, which themselves consist of embedded text and images.

Results and discussion

RQ1 What are the most frequent n-grams in a corpus of tweets from UK public health agencies in the first 11 months of the COVID-19 pandemic?

A total of 20 four-word n-grams occurred in all 13 UK public health agency Twitter accounts and are shown in Table 5.3.

It is clear from Table 5.3 that the forms of several of these n-grams overlap with each other, and so they and their co-text can be considered together rather than described separately. For example, three of these contain *the spread of* (ranked number 1, 3 and 11) and three contain *important* (ranked number 2, 8 and 16). This overlapping is one we would expect and has been observed in analyses of n-grams in other corpora. O'Keeffe, McCarthy and Carter (2007), for example, showed that common three-, four-, five- and six-word n-grams in a spoken corpus often contained the same core element such as *the end of*, *at the end of*, *at the end of the* and *at the end of the day*. In our data, the same phenomenon occurs with examples such as *the spread of*. Such common items

Table 5.3 Four-word n-grams occurring in tweets from all 13 Public Health Agencies 1 March 2020 to 17 February 2021

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

typically formed part of larger frames, for example *help stop the spread of COVID* was a common example of this. Other verbs and noun phrases such as *reduce*, *prevent* and *the virus* also occurred very frequently with *the spread of*. The concordance lines in corpus extract 2 show some examples of this.

Corpus extract 2. *The spread of* in UK public health agency tweets

1. With the government calling on people to stay at home to help stop **the spread of** coronavirus, we're asking people who drink alcohol to think about their drinking.
2. Wear a face covering where social distancing is difficult to maintain. Help stop **the spread of** COVID-19.
3. If you're out shopping, remember to stay 2m apart. Do your bit to stop **the spread of** the virus across the North East.
4. If you're out shopping, remember to stay 2m apart. Do your bit to stop **the spread of** the virus across the North West.
5. If you're out shopping, remember to stay 2m apart. Do your bit to stop **the spread of** the virus across Yorkshire and the Humber.
6. We must each take personal responsibility to limit **the spread of** the virus and protect our loved ones, particularly if they are vulnerable.
7. Our actions over the coming weeks will be crucial in containing **the spread of** the virus and ensuring we do not overwhelm our NHS in January.
8. Every person that is following the stay at home guidance is playing their part in slowing **the spread of** the virus and saving lives, so thank you.
9. Find out how we are responding to **the spread of** the virus in our daily statement here:
10. • Wash your hands. • Cover your face. • Make space. These are the three most effective ways we can control **the spread of** the virus.

While the frequent occurrence of n-grams involving the noun *spread* reveals the ideational aspect of these messages, it is also notable in this data that the conjunction *if* occurs often (ranked 6, 9, 12 and 15). This is somewhat surprising as we expected the nature of public health tweets to preclude the use of *if* in a subordinate clause as part of a complex sentence structure. That is, while in a general corpus of tweets Boot et al. (2019) found that subordinating conjunctions were quite common, our expectation was that n-grams in public health agency tweets would form part of simple sentences, allowing for a short, directly stated message. Nevertheless, there were a number of examples of patterns containing subordinate clauses. The most common three noun phrase collocates placed directly after the n-gram *if you have a* in Table 5.3 were *long term health condition*, *high temperature* and *new continuous cough*.

The *if you have a* n-gram containing an *if* clause fragment has been identified in other studies of n-grams in different registers, for example Biber et al. (2004, p. 381). It is classed as a discourse organiser, which reflects the relationship between prior and coming discourse, and introduces a topic or focus. In our data this n-gram was used to make advice or an instruction conditional on these other health conditions being present. Normally, the main clauses accompanying such examples contained a modal form, expressing possibility, or some kind of obligation to act. Examples of this can be seen in corpus extract 3 with the n-gram in bold and the modal underlined.

Corpus extract 3. *if you have a* in UK public health agency tweets

1. **If you have a** high temperature, a new continuous cough or loss of taste or smell, you <u>need</u> to get a coronavirus test.
2. **If you have a** new continuous cough OR a high temperature (37.8 degrees or higher), you <u>should</u> stay at home for at least 7 days from when your symptoms started.
3. **If you have a** new, continuous cough and/or a high temperature, you <u>should</u>: • stay at home • not leave your house for 7 days from when your symptoms started
4. **If you have a** persistent cough, high temperature or loss of taste or smell, you <u>may</u> have coronavirus. Stay home. Book a test. Get a test. Stop the spread.
5. You <u>can</u> get an employment and support allowance **if you have a** disability that affects how much you can work. If you're ill or self-isolating find out if you can apply.

Another example of modality in the data was n-grams containing the modal verb *can*. These were also frequent and occurred three times (ranked 14, 17 and 19). These n-grams generally comprise parts of clauses post modifying a noun such as *thing* or *steps* as in the examples in corpus extract 4.

Corpus extract 4. *you can do* in UK public health agency tweets

1. The latest analysis from PHE shows that more people from the #WestMidlands in their 20s are testing positive for #COVID19 than any other age group in the city. Make sure you're up to date on what **you can do** to help keep your friends and family safe
2. #SelfIsolation and #SocialDistancing can be hard. But there are simple things **you can do** to look after yourself.
3. Vaccination is the most important thing **you can do** to protect the entire #EastofEngland community against serious illnesses. When enough people get vaccinated, it's harder for a disease to spread to those who can't have vaccines.
4. There are lots of things **you can do** to take care of your family's mental health during the #coronavirus (#COVID19) outbreak.

5. COVID-19 hasn't gone away. There are things **you can do** when you're out and about to #KeepWalesSafe

The final noticeably frequent n-grams in the tweets are those containing the adjective *important* (ranked 2, 8 and 16). The n-gram *it's important to* has also been observed to be frequent in studies of other registers and classified as expressing a writer's "impersonal attitudinal modality stance" (Biber et al., 2004, p. 385) that "frames" some other proposition (Biber et al., 2004, p. 384). Examples can be seen in corpus extract 5.

Corpus extract 5. *it's important to* in UK public health agency tweets

1. Around 1 in 3 people have #COVID19 with no symptoms, so **it's important to** always remember #HandsFaceSpace.
2. Cornwall will move up to tier 2 from midnight on 26 December. **It's important to** be clear on what you can and cannot do.
3. We've produced a blog on why **it's important to** continue to follow #SocialDistancing rules during this holy month of #Ramadan to protect against #COVID19
4. #COVID19 can live indoors for more than 24 hours. That's why **it's important to** follow the latest guidelines and prevent spreading infection. Remember hands, face, space and ventilate.
5. **It's important to** get some exercise each day. Go out in the fresh air close to where you live in the #EastOfEngland & practice #SocialDistancing to protect against #COVID19.
6. #Coronavirus lingers in enclosed spaces so **it's important to** keep your home ventilated. Letting fresh air indoors can reduce the risk of infection by over 70%.
7. **It's important to** maintain social distancing even when you're exercising. Exercise alone or with one person from another household.
8. **It's important to** only share information from trusted sources like the NHS, public health agencies, government, universities or your local council. If in doubt, fact check the claims before you share them.
9. **It's important to** remember that #COVID19 is still with us and we need to do all we can to protect our families and friends.
10. #COVID19 can be spread through the air by droplets that are exhaled from the nose and mouth of an infected person. **It's important to** ventilate indoor spaces if someone in your home has the virus as this can help prevent transmission to others.

The functions of these n-grams will be further discussed in the answer to research question 3.

RQ2 Which images are associated with these n-grams?

As noted, when looking for images, we had to decide either to be language driven or image driven when looking for images in the public health tweets. We decided to be led by the most frequent n-grams and look at the images accompanying those n-grams. This means that some images used may be less frequent than if we had led by image frequency, perhaps by the number of likes or retweets. With this in mind, there are several observations we can make about the images we found with these common n-grams. Firstly, there were three predominant types of images. The first were photographs, as displayed in Figure 5.2 where an image of member of the public accompanies the n-gram *it is important that*. The featured participants in the photographs were either members of the public (as seen in Figure 5.2) or an authority figure such as somebody from Public Health England. The second type were procedural infographics, which contained a series of questions and commands together with icons and drawings, often with the use of bright and bold colours. Figures 5.3 and 5.4 show typical examples of the procedural infographics in the dataset. Lastly, posters were also featured in the public health messages, with cartoon like drawings, as displayed in Figure 5.5.

The images often contained embedded text elements, as seen in Figures 5.2 to 5.5. This often either added more to the Twitter text or partially paraphrased it, making small additions to the initial message. Figure 5.4 is a typical example of this, in an image accompanying *the spread of Covid*. Here we can see some small changes made from the initial text in the image. *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*. Procedural infographics are themselves complex forms of visual representation, consisting of questions and commands that form relational identifying relations with their respective icons and/or drawings to reinforce the meaning. For example, the rhetorical questions *Got symptoms? OR Tested positive?* in Figure 5.3 are juxtaposed with icons that visualise this state, namely, someone coughing and a COVID-19 test kit. These two questions, which are logically connected to each other, are followed by the command = *You must self-isolate for 10 days*. Therefore, the text/icon implication sequence, organised from the top to the bottom reading path, forms a procedural text with a command if certain conditions are met. Similarly, there are icons attached to the various commands in the infographics in Figure 5.4, which realise a series of commands that are (implicitly) logically connected by "and": *Wash your hands regularly*, *Keep your distance* and *Wear a face covering where needed*. In this case, the infographic contains a headline *COVID-19, Simple steps to save lives* and the reading path is left to right. In both cases, the tweet text and the infographics co-contextualise and reinforce each other, with the possible risk of repetition.

A number of questions emerge about the choice of images in the tweets, some of which will be discussed in relation to research question three. Amongst these are why infographics were chosen so often. Perhaps the intention was to convey the message as simply as possible, in a way that it was hoped most people would understand given that the rhetorical questions and commands are accompanied by simple icons and drawings. The use of bright colours was presumably also chosen to engage people with the messages and photographs used to either create empathy (for example, in the case of Figure 5.2, which depicts a member of the public working at home) or ensure readers would listen to an authoritative voice. This raises the question as to why so much text in the tweet was chosen to accompany each image and why this text often made small changes to what was depicted visually in the infographics, which themselves consist of linguistic and visual elements. This also raises the question of redundancy where the tweet text repeats the information in the infographic, leading to possible information overload and fatigue.

RQ3 How do these frequent n-grams and images typically function in this data?

When exploring the n-grams and images together, there seem to be two common functions: text and images serving as indirect speech acts and tweets using the same strategy (stating factual information) but accompanied by more direct speech acts, in the form of instructions. An example of the first – text and images as indirect speech acts (Searle, 1969) – we might see when a statement such as *It's cold in here* serves as a request for somebody to close the window. In these tweets a common means of achieving this was to advise by simply stating information, in what Halliday (1994, pp. 363–367) calls a "mood metaphor" (for example, the speech function of command is realised by a declarative rather than an imperative). The image and text shown in Figure 5.2 is a good example of this. Here the n-gram is used in a text that appears to simply state factual information, *It is important that we all take steps to reduce the spread of Covid-19,* with an objective stance in terms of modality of the statement (*it is important*). The image is then accompanied by more factual information in the form of several declarative statements: *Anyone can spread coronavirus. Even you* and *Not everyone with coronavirus has symptoms.* The image of somebody representing a member of the public appears to be working on a laptop computer at home, given that this person is casually dressed in a softly lit room, which looks like a kitchen. The person is looking at the computer, so the viewer is positioned as an observer to the scene, which depicts the narrative theme of "working from home". In this case, the lighting functions to highlight the worker in the foreground and the neat and well-equipped kitchen in the background. The worker is prominent, given the medium camera distance, her position in

the image, and her brightly coloured clothing. The photograph depicts an ideal home scene, with orderly and quiet working conditions.

However, the illocutionary force (Searle, 1969) of the written message in this case is clearly not simply to state facts. Instead, it is to advise the reader to perhaps stay at home, if possible, get tested, and at least wash their hands, wear a face mask and maintain social distancing. This observation about the function of the n-gram is at odds with Biber's assigned function of *it is important* as expressing a writer's "impersonal attitudinal modality stance" (Biber et al., 2004, p. 385). It seems from our analysis of the n-gram, when juxtaposed with an image in a tweet, that the impersonal nature of the n-gram is only a surface feature and that its actual function is to give advice. However, there also appears to be a mismatch between the text, image and illocutionary force here. In short, although we can suggest the function here is "advising", it is not really clear what the Twitter text and the embedded text in the photograph are actually advising the reader to do. The image is of someone working at home, but does this then mean we can take steps to reduce the spread of COVID-19 by staying at home? What then, are we to make of the text *Anyone can spread coronavirus. Even you* and *Not everyone with coronavirus has symptoms* in the same space as the image? Here, the implied advice is perhaps that we should get tested regularly to avoid spreading the virus. This example illustrates the difficulty of constructing public health advice in this way – the implicature can clearly be interpreted in several ways and, we would suggest, is likely to be given different interpretations by different readers. Therefore, the text and images do not co-contextualise each other to reinforce a single message. Rather, there is semiotic dissonance, with the result that the message may be interpreted in different ways by members of the public. Moreover, the image is an idealised version of what it means to work at home; clearly not everyone enjoys an orderly home environment as depicted in the photograph.

The second common function was for tweets to use the same strategy (stating factual information) but accompanied by more direct speech acts, in the form of instructions, as seen in Figures 5.3 and 5.4. Figure 5.5 is another example of this form of tweet text, where explicit instructions are provided before Halloween: *Follow the rules in your local area*; *Wash your hands regularly and keep 2m apart*; *Wear a face covering where appropriate*; *and self isolate and book a test if you experience any symptoms of #COVID-19*. The accompanying image is a poster with the command to cover faces during Halloween, as displayed in Figure 5.5.

In this example, at first glance, the illocutionary force of the tweet seems clearer. We are told what to do and the imperative form makes it clear that these are instructions and not suggestions. However, despite this, there is still some ambiguity here. In the text, the instruction *wear a face covering* is accompanied by *where appropriate*. The implicature here is that the reader is presumed to know when and indeed where this might be. A similar issue is with *follow rules in your local area*. The implicature

is that the reader knows what these rules are and how they may differ from other rules in place. Finally, the poster seems to contradict the instruction in the text. Here we are simply told to cover our face through the command *COVER YOUR FACE THIS HALLOWEEN*, which is displayed underneath a cartoon drawing of an anthropomorphised pumpkin with a face mask. The implicature is different to that in the text, where the use of *where appropriate* suggests we do not always need to cover our face and that we are aware of instances when it is not important. In this case, the Twitter text and poster function to co-contextualise each other, but the circumstantial conditions for the wearing of face masks remain somewhat ambiguous.

Conclusion

The research reported in this chapter has revealed that in UK public health agency tweets, linguistic and visual elements are, predictably, interconnected through the common themes of preventing the spread of COVID-19 through safety measures (i.e. working from home, getting tested, keeping socially distanced, and washing hands). However, we have observed that the tweet texts and embedded linguistic text are not always perfectly aligned with the images, especially in cases where information is provided in the form of factual statements that are only vaguely related to the image. In this case, the most frequent n-grams form core aspects of the messages, but these often function indirectly, and their illocutionary force can be interpreted in various ways, unlike the procedure graphics where the conditions and logical implications are clearly articulated. However, procedure infographics are complex multimodal texts and need to be directly linked to the accompanying tweet text, to be more effective. Additionally, there is a clear risk of redundancy and information overload if the tweet text merely repeats what is depicted graphically, and a risk of misinterpretation if the text is repeated with minor changes within the graphics. While public health tweets of this nature need to communicate their message quickly and with minimal ambiguity, overall, the examples identified in this study suggest that the manner in which the most common n-grams and accompanying images are used reveals that there is a great deal of ambiguity present.

This stage of the project has described the language and images used in public health messaging and the nature of the text/image relations. As seen in this study, the text and images can function to co-contextualise each other, in this case using the tweet text and images such as photographs, procedural infographics, and posters. Key to such an approach is to analyse the message through a multimodal lens, analysing common forms and functions of language and how these interact with images to carry ideational, interpersonal and textual messages. Although our focus has been limited and reasonably narrow here, the language/image analysis methodology developed in this project can be applied to all sources of

information about COVID-19 (mainstream news, social media, government reports and so forth). Insights into how phraseology and imagery in public health information are understood and acted upon will lead to increased effectiveness of the linguistic and visual choices made in ensuing messages in future pandemics.

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

Notes

1 http://35.246.69.64:8000/en-GB/account/login?return_to=%2Fen-GB%2Fapp%2FMultimodal%2Fwelcome

2 http://multimodal-analysis.com/products/multimodal-analysis-image/software/index.html

References

Anthony, L. (2022). AntConc (Version 4.0.5) [Computer Software]. Tokyo, Japan: Waseda University. Available from www.laurenceanthony.net/software

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). https://doi.org/10.3389/fpubh.2021.610623

Bateman, J. (2014). *Text and image: A critical introduction to the visual/verbal divide*. Routledge.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. John Benjamins.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371–405.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.

Boot, A. B., Tjong Kim Sang, E., Dijkstra, K., & Zwaan, R. A. (2019). How character limit affects language usage in tweets. *Palgrave Communications*, *5*(1), 76. https://doi.org/10.1057/s41599-019-0280-3

British National Corpus, version 2 (BNC World). 2001. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. www.natcorp.ox.ac.uk/

Carter, R. (1998). *Vocabulary: Applied linguistic perspectives* (2nd ed.). Routledge.

Cortes, V. (2004). Lexical bundles in published and student writing in history and biology. *English for Specific Purposes*, *23*(4), 397–423.

Davies, M. (2021). The coronavirus corpus: Design, construction, and use. *International Journal of Corpus Linguistics*, *26*(4), 583–598. https://doi.org/10.1075/ijcl.21044.dav

Ellis, N. C. (2008). Phraseology: The periphery and the heart of language. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 1–13). John Benjamins.

Firth, J. R. (1957). A synopsis of linguistic theory 1930–1955. In J. R. Firth (Ed.), *Studies in linguistic analysis* (pp. 1–32). Basil Blackwell.

Glaser, R. (1998). The stylistic potential of phraseological units in the light of genre analysis. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and application*. (pp. 125–143). Clarendon Press.

Greszczuk, C. (2020) Making messages work. Health Foundation, 21st April 2020. Accessed on 1 May 2020 from https://www.health.org.uk/what-we-do/a-healthier-uk-population/thinking-differently-about-health/making-messages-work

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics* (pp. 41–58). Academic Press.

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. Edward Arnold.

Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). Edward Arnold.

Halliday, M. A. K. (2009). *Collected works of M.A.K. Halliday* (10 Volumes). Continuum.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar* (4th ed., revised by C. M. I. M. Matthiessen ed.). Routledge.

Howarth, P. A. (1998). Phraseology and second language proficiency. *Applied Linguistics*, *19*(1), 24–44.

Hunston, S. E. (2020). Changing language in unprecedented times. *University of Birmingham, Department of English News*, 8th April 2020. Accessed on 23rd April 2020 from https://www.birmingham.ac.uk/schools/edacs/departments/englishlanguage/news/2020/changing-language.aspx

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for specific purposes*, *27*, 4–21.

Jones, C., Byrne, S., & Halenko, N. (2017). *Successful spoken English: Findings from learner corpora*. Routledge.

Jones, C., & Waller, D. (2015). *Corpus linguistics for grammar: A guide for research*. Routledge.

Kress, G., & van Leeuwen, T. (2020). *Reading images: The grammar of visual design* (3rd ed.). Routledge.

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. https://doi.org/10.1075/ijcl.00043.mah

Moon, R. (1998). *Fixed expressions and idioms in English: A corpus based approach*. Clarendon Press.

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.

Oakey, D. J. (2022). What can a corpus tell us about lexis? In A. O'Keefe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (2nd ed.,) (pp. 185–203). Routledge.

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*. https://doi.org/10.1016/j.dcm.2021.100467

O'Halloran, K. L., Tan, S., & E, M. K. L. (2017). Multimodal analysis for critical thinking. *Learning, Media and Technology*, *42*(2), 147–170. https://doi.org/:10.1080/17439884.2016.1101003

O'Halloran, K. L., Tan, S., & Smith, B. (2016). Multimodal approaches to English for academic purposes. In K. Hyland & P. Shaw (Eds.), *The Routledge handbook of English for academic purposes* (pp. 256–269). London: Routledge.

O'Halloran, K. L., Tan, S., & Wignell, P. (2019). SFL and multimodal discourse analysis. In G. Thompson, W. L. Bowcher, L. Fontaine, & D. Schönthal (Eds.), *The Cambridge handbook of systemic functional linguistics* (pp. 433–461). Cambridge University Press.

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. https://doi.org/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. Accessed on 23rd April 2020 from 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. https://doi.org/10.1080/10810730.2015.1058435

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191–226). Longman.

Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? In N. Schmitt (Ed.), *Formulaic sequences* (pp. 127–151). John Benjamins.

Schuessler, J. (2020) 'Oxford's 2020 Word of the Year? It's too Hard to Isolate', New York Times, 22nd November 2020. Accessed on 17th March 2022 from 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 M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice* (pp. 1–16). Oxbow Books.

Tan, S., O'Halloran, K. L., & Wignell, P. (2020). Multimodality. In A. D. Fina & A. Georgakopoulou (Eds.), *Handbook of discourse studies* (pp. 263–281). Cambridge University Press.

Twitter.com. (n.d.). *Twitter Advanced Search*. https://twitter.com/search-advanced?lang=en-GB

van Leeuwen, T. (2005). *Introducing social semiotics*. Routledge.

Wood, D. (2002). Formulaic language in acquisition and production: Implications for teaching. *TESL Canada Journal*, *20*(1), 1–15.

World Health Organisation (WHO) (2020). 'Naming the Coronavirus Disease (COVID-19) and the Virus that Causes it'. Accessed on 17th March 2022 from https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.