Digitally mediated communication

1. Introduction

This chapter will begin by considering what is unique about digitally mediated discourse, but will also outline some of the continuities with other forms of discourse. It will present the basic tenets of three main frameworks (which also overlap to a considerable extent): computer-mediated discourse analysis (based on the work of Herring, e.g. 2004), discourse-centred online ethnography (based on Androutsopoulos 2008) and critical approaches to digitally mediated-discourse (based on KhosraviNik & Unger 2016, also called social media critical discourse studies in KhosraviNik's more recent work, e.g. 2017). It will then go on to suggest how elements of each can be adopted to form a coherent methodology in investigations in digital media contexts. The ethical issues specific to digital media research will be discussed in some detail, with reference to examples of different approaches to e.g. anonymization, inclusion of usernames/identities, and inclusion of verbatim data. Towards the end of the chapter, a case study of the US politician Alexandra Ocasio-Cortez's tweets will illustrate how how to carry out an achievable investigation into digitally mediated discourse that nevertheless draws on the latest theoretical and methodological insights in more advanced scholarly work.

2. Outline of approach

At one level, there is nothing special about digitally mediated language. Huge sections of the internet are formal written documents such as government reports, newspaper articles, company mission statements. Other than being viewed on a screen rather than on paper, these can be said to be almost identical to their printed counterparts. A key notion here is *affordances*, defined by Gibson (1977) as the action possibilities of objects (or technologies). For example, a large computer screen allows viewers to see a whole page, or perhaps two side-by-side. This is similar to paper on a desk, which can be laid out side-by-side. Paper texts do not allow readers to click on hypertexts, or search for particular words by typing them into a search box. On the other hand, pieces of paper can be used to make paper aeroplanes. These two technologies (computer screens vs. paper) thus have different affordances.

On social media platforms, there is a lot of informal written language which has some similarities with older, pre-digital texts. For instance, Gillen (2013) argues that posts on Twitter have many similarities with Edwardian postcards. Increasingly social media posts incorporate images, audio and video elements, which again have some similarities to pre-digital texts. Some vlogs, for example, are not unlike home movies that people used to create using video cameras. Herring (2013) identifies three different kinds of genres (text types)/genre features in social media texts:
- Familiar: genres that do not differ substantially from their pre-digital counterparts, such as the government reports mentioned above. Familiar genre features include the use of emoticons or non-standard spelling in texts, which have existed in various forms for centuries.
- Reconfigured: genres that have been repurposed or reshaped in line with digital affordances, for instance status updates. These do exist in other forms (for instance,

the signs some university staff put on their doors saying whether they are available or in a meeting), but social media status updates have considerably extended and changed the function of such texts.

- Emergent: genres that are not found in pre-digital or pre-social media contexts, and include examples such as collaborative knowledge-sharing platforms like Wikipedia, which allow multiple users to write and edit text.

There are some fairly well-established features and affordances of digitally mediated communication that may differ somewhat from non-digital communication, and this should be borne in mind when undertaking a project (for further detail, see Page et al. 2014, or Herring's 2007 work on medium and situation factors). For instance, digitally mediated communication can:

- often be faster than written, but slower than spoken exchanges
- have multiple participants (e.g. in a Whatsapp group) and involve different relationships of addresser to addressee, for instance one-to-many (Twitter posts); many-to-one (online petition sites)
- be addressed to an unseen audience, but increasingly is personalised via algorithms, allowing social media platforms and advertisers to target people based on specific identity characteristics, habits and behaviours.
- be a "narrower" channel than face-to-face communication, which involves not only verbal but also visual, auditory and gestural information, but increasingly the use of video and other multimodal features mean digital communication is becoming more similar to face-to-face interactions

Still, there are more continuities and similarities between non-digital communication and digital communication than there are differences. With the above points in mind, which may lead to some adaptations, the other approaches outlined in this book can therefore be applied to digitally mediated communication.

There is no single correct or best way to research digitally mediated communication. Which specific frameworks and tools researchers use for a particular project depends very much on the context, the kind of data available, the particular questions being asked (see next section), and finally, but importantly, the specific interests and capabilities of the researcher. While digital communication today is ubiquitous, and it is scarcely possible to imagine that a person below retirement age does not use any kind of digital communication in their daily life, this was not always the case. A number of linguists, primarily in the US initially, started analysing how people use digital technology to communicate from the very early era of digital computing in the 1970s onwards. The focus was primarily on verbal text, which was typed into a device (usually a terminal and later PC) using a keyboard and then appeared on a screen (usually monochrome, in the early days) on the recipient's device. This incorporated research into so-called bulletin board systems (akin to today's online discussion forums and platforms like Reddit) and email lists (still in existence today). This early work used the label "computer-mediated communication" (CMC) to describe its main object of analysis. The particular focus was on how the medium (typing, electronic devices) affects the message. There was a belief that CMC was generally "between" speech and writing, with some spoken and some written features – although this belief does rest on

particular assumptions about writing being generally formal and planned, and speech being informal and spontaneous, whereas it is easy to find counter-examples. CMC research often was largely descriptive, giving accounts of linguistic behaviours, and focussed particularly on differences to other written (or spoken) genres, such as the use of emoticons (☺ 'smiley face') or abbreviations (*CU L8R* 'see you later') (see Barton and Lee 2013 for an overview of these).

More recently (from the 1990s onwards) researchers, in particular Herring, started to develop a framework for computer-mediated discourse analysis (CMDA), which focussed on language as an online, social behaviour. Later still (from the mid-2000s onwards), researchers started to take an interest in what was then an emerging group of digital platforms which can loosely be grouped under the label "social media". There are different understandings of what social media includes, ranging from only "classic" social networking sites such as Facebook or Twitter, perhaps with the addition of peer-to-peer instant messaging platforms, to a much broader understanding that would also incorporate, platforms such as online review and advice sites, newspaper discussion pages, Wikipedia, and even email (see Page et al. 2014: 5 for further discussion). During this period, Androutsopoulos (e.g. 2008) developed the discourse-centered online ethnography (DCOE) approach, in which he emphasised that observation of digitally mediated behaviours should be accompanied with direct contact and interviews with the people producing and consuming these texts. Towards the middle of the 2010s, a number of critical discourse researchers started to take a greater interest in digital and social media, and started to consider how the traditional theories and frameworks of critical discourse studies could be applied or reframed for digitally mediated contexts. While this research is still very much in development, and different labels are used for it, I will here refer to it as digitally mediated discourse studies (DMDS). The three more recent approaches mentioned above (CMDA, DCOE, DMDS) are discussed in more detail in the remainder of this section.

2.2 Computer-mediated discourse analysis (CMDA)

Herring (2001:612) describes CMDA as concerned with "language and language use in computer-networked environments". Importantly, like discourse studies, CMDA is an eclectic approach with no single unified methodology or theoretical assumptions. Instead, Herring (2004:342) argues that "CMDA as an approach to researching online behaviour provides a methodological toolkit and a set of theoretical lenses through which to make observations and interpret the results of empirical analysis." While CMDA emerged as a sub-field of CMC in the 1990s, it eventually developed its own identity, and a particular focus on the first four categories set out in Table X.1, below: *structure*; *meaning*; *interaction*; and *social behaviour*. The tables lists the main issues that researchers concerned themselves with (e.g. formality, what is communicated, coherence, community), the linguistic and discursive phenomena that relate to these issues (e.g. syntax, speech acts, turn-taking, conflict management), and the specific methods and linguistic frameworks that can be used

to study them (e.g. text analysis, pragmatics, conversation analysis, interactional sociolinguistics). The final two categories, *participation* and *multimodal communication*, were added in Herring's later work (2013), and also indicate a shift away from the understanding of computer-mediated discourse as primarily text-based, and "free from competing influences from other channels of communication and from physical context" (Herring 2001:612).

[Table X.1 around here]

It is beyond the scope of this chapter to explain every single issue, phenomenon and method in this table, but each of them is described by Herring and also in a large number of other textbooks, such as the accessible overviews in Culpeper et al. (2018), and some are mentioned in the examples below.

2.3 Discourse-centred online ethnography (DCOE)

While CMDA is a broad framework that allows for the analysis of different linguistic and discursive phenomena at different levels, DCOE is a particular methodological approach to research into digitally mediated texts and practices. Androutsopoulos (2008) draws on the tradition of ethnographic research, which involves researchers immersing themselves in a community or culture to attempt to gain a better understanding of the behaviours that they observe. Androutsopoulos (2008) proposes the following six guidelines for observation of digitally mediated practices and actors (people), reproduced here with some additional explanation and suggestions that adapt the framework for more contemporary digital media platforms in parentheses:

1) Examine relationships and processes rather than isolated artefacts (in other words, researchers should not just treat digitally mediated texts as static, but think about them in terms of part of a set of practices that change over time)
2) Move from the core to the periphery of a field (try to identify the most important sites, sections, pages, etc. of a digitally mediated context, and look at those first, then consider more marginal ones)
3) Repeat observation (observe the same people/groups on multiple occasions and different people/groups in comparable contexts)
4) Maintain openness (avoid approaching the data with a fixed mindset, and don't assume that what is observed the first time will always be true)
5) Use all available technology (consider how different people use a particular platform, with all its features and affordances, also how they might engage with digital media with different devices, operating systems, etc., and how actors involved might make use of different technologies as part of a set of practices, e.g. sharing images across multiple platforms)
6) Use observation insights as guidance for further sampling (let what is learnt about the sites/platforms/actors, such as what kinds of texts are written, or who the main actors are, inform further research, and be prepared to change plans if it seems appropriate as more information is uncovered)

These are followed by a further six guidelines for contact with the actors responsible for or involved in the texts and practices that have been observed:

1) Contacts should be limited, non-random, and include various participation formats (this means researchers should be focussed in their choice of contacts, based on observations of practices, and should have a clear goal in mind in their information-gathering. Participation formats here means the position/role of the participant, e.g. professional, amateur, moderator, frequent user, newbie, etc.)
2) Pay attention to the initial contact (first impressions count! It can be difficult for researchers to motivate participants outside their own friendship networks, so give yourself the best chance to engage participants in your research)
3) Formulate and customise interview guidelines (It is not necessary, or even desirable, to have a rigid script for interviews and ask everyone the same questions: interview questions should be based at least in part on the observed data and prior knowledge about the participant).
4) Confront participants with (their own) material (this allows for a greater understanding of texts and practices – by asking participants how and why they use digital media, researchers can gain valuable insights)
5) Seek repeated and prolonged contacts (for ethnographic research, long-term contact between the researcher and participants is extremely valuable and allows for exploration of whether and how practices change over time)
6) Make use of alternative techniques whenever possible (this could include questionnaires, surveys, or public chat sessions (where not ethically problematic) as well as individual interviews.

While CMDA on its own allows researchers to take an *etic* (from an outside observer's) perspective, DCOE allows the addition of an *emic* (from within the community/culture being studied) perspective, which can strengthen analyses and provide a rich source of insights and analytical detail. Androutsopoulos is of course not the only researcher to conduct digitally mediated ethnographies. For instance, Hine (e.g. 2015) provides in-depth accounts and guidelines on how to conduct digital ethnographies.

2.3 Digitally mediated discourse studies (DMDS)
As other chapters in this book show, discourse studies is a varied and eclectic field with no single methodological or theoretical framework. KhosraviNik & Unger (2016) argue that many of the same linguistic and discursive phenomena can be analysed in digital media as in more traditional contexts of language use. What is particularly needed to account for digitally mediated communication, however, is an understanding of media dynamics, and how these shape and are shaped by the way language is used in digital contexts. For the specific case of social media platforms, Figure X.1 shows the interactions between users, institutions (i.e. companies and public institutions), and contexts.

[Figure X.1 around here]

Other approaches to digital media often ignore the corporate institutional contexts in which digital media platforms are provided. When someone posts on Twitter, Instagram or Snapchat, for instance, platforms that are free to end users, this potentially generates data that linguists could, if appropriate ethical procedures are followed, analyse in various ways. However, the data that is generated is then provided, usually indirectly and via particular algorithms that factor in personal information such as age, sex, sexuality and more, to advertisers, who can then in turn target users with specific adverts (whether for products,

services or political parties), that are based on their identities and preferences. Critical analyses of social media (and other digital media) should take into account the role of these institutions, even if only by describing them as part of the context in which communication takes place. More nuanced analyses take account of the role of these institutions in promoting or suppressing particular instances or kinds of communication, such as for example activism, hate speech, or advertising. Fuchs (2014) provides an accessible introduction to the theoretical underpinnings of this idea.

3. Identifying research questions

The most important consideration when choosing research questions is that they should be appropriate for the context and data (or vice versa – appropriate data and contexts should be chosen for specific research questions). Research questions in studies of digitally mediated discourse should ideally incorporate all of the following elements:

1) A question word or phrase (e.g. Why, how, what, do/does, to what extent)
2) A linguistic or discursive phenomenon
3) An issue
4) A particular social or institutional context in which digitally mediated communication occurs, or a group of people within such a context
5) A question mark

Examples of 2) and 3) are given in Table X.1, above, though this should not be seen as a definitive list. Based on these guidelines, a typical question might therefore be:

> How[1] does face-work [2] contribute to community cohesion [3] in YouTube discussions around makeup vlogs [4]?[5]

Several things are worth noting about this question: First, it contains a presupposition. Saying "How does X contribute to Y" presupposes that X *does* contribute to Y. Another way of phrasing this question would therefore be "Does face-work contribute to community cohesion in YouTube discussions around makeup vlogs and if so, how?" It is generally a good idea to avoid presuppositions in research questions, but sometimes they are unavoidable, or not problematic; for instance, research that is specifically concerned with hate speech, with data that has been selected because it is hate speech, does not need to ask whether hate speech occurs in the given context. The second notable feature is that everything before "YouTube" could apply in numerous contexts, including those that do not directly involve digital communication. This is important because it reinforces what I have stated above: that many of the linguistic and discursive features, and by extension ways of analysing them, and the questions we might ask about them, are similar or the same in digital and non-digital contexts. Nevertheless, there will be interesting things to say about the digital context, and the way it interacts with the issues and phenomena under investigation. This can be left implicit, as in the question above, or could be turned into a separate question (perhaps for a different project!), such as:

> How do vloggers construct their identities in written YouTube discussions around their vlogs, and how does this differ from their identity construction in their videos?

This is potentially quite a broad question, which could incorporate a number of discursive phenomena, and may thus be a starting-point for a more detailed investigation of specific phenomena once these have been identified as salient in an initial investigation. There could also be an ethnographic element. Following the guidelines outlined above as part of the

DCOE approach, the observation of the vloggers' behaviour could be accompanied by interviews with vloggers in which they are asked about their identity construction practices). It is worth noting that research questions, while they are often outlined near the beginning of an academic text, can evolve and change in the course of a project. The examples so far suggest that research into digitally mediated discourse focusses on the "higher-level" issues and phenomena in the CMDA framework (see Table X.1), such as social behaviours and multimodality. This is not necessarily the case, however – any of the phenomena and issues could be incorporated into a discourse studies project, e.g. investigating the use of specific orthographic conventions as an in-group identity marker on a social networking site, and how newcomers who are not familiar with the conventions are socially sanctioned or excluded.

It is also possible to pose a different sort of research question, which is concerned more with evaluating a particular theory or methodology in a digital context, e.g.

> Do the typical rules of turn-taking found in spontaneous spoken conversation between friends also apply in instant messaging?

This would allow the researcher to evaluate the conversation analysis framework, which was developed for non-digital interactions (see e.g. ten Have 2007), in digital contexts. Again, this is a rather broad question, so it may be advisable to narrow it down by including additional issues or phenomena. For instance:

> Do the typical rules of turn-taking, *in particular the turn allocation component*, found in spontaneous spoken conversation between friends also apply in instant messaging?

Or

> Do typical *gendered turn-taking strategies* found in spontaneous spoken conversation between friends also apply in instant messaging?

It is of course important to remember that there is now a substantial body of research into digitally mediated discourse from many different perspectives and research traditions. As such, any study that includes a more methodological or theoretical question should include a thorough literature search for recent journal articles and other scholarly publications that use data from digitally mediated contexts. This also allows for an *abductive research process*, i.e. applying the questions and concerns posed in existing scholarly literature to new data and contexts, and then, in light of your study, developing new insights about and explanations of both the theories you are applying and the contexts you are investigating. This is in contrast to more *inductive* forms of research, which would involve drawing conclusions from specific hypotheses about the data that can be tested in experiments or through systematic observation (abduction is discussed further in KhosraviNik & Unger 2016).

4. Data collection methods, including ethics

4.1 Ethical considerations

The different elements of the three frameworks described in Section 2 above provide the beginnings of a methodological framework for digital communication. I have only briefly mentioned ethical considerations so far, which should however be at the forefront of researchers' minds when undertaking any project involving human participants or their data. And, due to the participatory nature of many digitally mediated texts, and in particular

social media texts, there is a huge amount of human-generated data that can be difficult to categorise as "public" or "private". Page et al. (2014: 58) summarise the overarching ethical principle as "avoid harm to yourself and others". While this is a good start, it can be difficult to work out what might cause actual or potential harm. While ideally all human-generated data would only be collected and analysed after informed consent has been granted by the participants, this would make it almost impossible to study most publicly available social media data, because researchers are unlikely to get a high number of responses when they contact the people whose data they want to use. There is a useful overview in Page et al. (2014: Chapter 4), and a set of overarching guidelines published by the Association of Internet Researchers (Markham & Buchanan, 2012). However, I will here outline several key concerns that researchers should consider before embarking on a digitally mediated discourse studies project:

- **Public vs. private:** The first judgement to make is whether participants had any expectation of privacy when they posted their data online. Clearly, a private Whatsapp group set up for friends to communicate with one another would be inappropriate for a researcher to use without getting consent from everyone involved. In fact, even asking for consent after being part of a context like this can create harm; participants may feel they were being surveilled without their knowledge by someone who was acting under false pretences. At the other extreme, a publicly available Tweet by a well-known public figure or company is unlikely to have been posted with any expectation of privacy.

- **Anonymity and verbatim data**: There is little point in anonymising participants when publicly available data is quoted verbatim. Searching the internet for the quoted text will often turn up the original source very quickly. But unlike researchers in other disciplines and fields, discourse analysts cannot necessarily just aggregate data or change words around to make data more difficult to find, because we need to evidence our analyses with empirical data. So researchers have to make a choice around whether the topics in the data, the contents of data, or the participants' identities are sensitive, which means that they should not be exposed. For example, when researching a potentially sensitive topic like eating disorders, a researcher may not choose to include individual disclosures and descriptions of people's experiences, even if they are found in a public online discussion forum, because this may expose them further and potentially removes their ability to delete data about themselves. Digital media articles by journalists may be considered less problematic, by contrast. Where there is a potential risk to the people involved in digitally mediated data, texts should not be included verbatim (and perhaps not used at all, if too sensitive). However, it is also possible that they can be used with little risk of harm to the participants that would supersede any possible risks from the original posting (in which case they can be used without anonymisation, if institutional ethics rules allow for this). This would apply, for instance, if someone's social media posts had been widely reported in the media. There are two additional ways of resolving or mitigating this problem: first, raw data can be withheld except from anyone involved in supervising or assessing work, or from other researchers, and can be made available only on request. Second, when working with translated data, it is often much harder for readers to trace the text back to an original source.

- **Informed consent:** Some university ethics committees are considerably behind the times; for instance, they may require official names and signatures on paper forms

for all participants, which may be completely impractical with digitally mediated research where participant and researcher are not physically co-present. Furthermore, by asking for official names, a layer of anonymity made possible by participants choosing their own screen names is removed. Fortunately, this practice seems to be changing, and ethics committees are becoming more flexible in this regard. If researchers do endeavour to seek informed consent (where appropriate), this should be done using the same means of communication as they are using to collect data, e.g. if recruiting participants on Facebook, via a direct Facebook message.

- **Access to data:** Although it is trivially easy to access huge volumes of data, this does not mean that it is all 'fair game', even if public. In particular, data that relates to sensitive or contentious subjects (e.g. political beliefs, eating disorders, criminal activity) should be treated with great care. This is rarely a problem with academic texts that are only read by markers or dissertation supervisors, but can become a problem if the work is then made available online, or is seen by peers, friends or family. The relative anonymity of digital identities means that researchers may not know if someone whose data they are collecting is vulnerable (e.g. a child).

- **Unintended effects:** Following on from the previous point, there may be numerous unintended consequences of collecting and using participant data, for instance data may still be available even if it has been deleted from the original social media platform; participants may gain more visibility than they would have had posting in some obscure forum known only to a small community of users.

As a rule of thumb, any data that can only be accessed by someone with a personal username and password (e.g. Facebook) or that is between individuals or within a small group (e.g. WhatsApp) should never be used without gaining the participants' informed consent. Any data that is publicly available should be assessed carefully in terms of how sensitive it is, and how much risk there is to participants (and the researcher themselves) before being used.

4.2 Data collection

Before deciding what data to collect, it is first necessary to consider how much data is required to answer a particular research question and to be able to make claims with an appropriate degree of certainty. For instance, collecting only one tweet from a political leader will allow very little to be said about anything other than that specific tweet, which could however still have highly interesting features or be particularly important. Collecting all that politician's tweets, on the other hand, allows stronger claims to be made, but could be difficult to analyse within the scope of a time-limited research project. Ideally, to gain an insight into digitally mediated communication practices, a variety of texts should be collected, from different actors, over a specific time period, and from different genres. In Unger et al. (2016), we suggested that after reviewing existing theoretical knowledge (and after completing any required ethics processes), researchers should systematically collect available data and find out more about the context. This will give an overview of the research context, as well as indicating which types of data are available and what kinds of

insights might be gained from them. This initial data collection could also include participant interviews (following the DCOE approach) alongside observations of participants' practices. Once this initial dataset has been collected, a subset can then be selected for specific analysis – for instance, choosing a smaller sample of social media texts that will help answer the research questions, based on transparent criteria. As an example, a study of racist ideologies on Wikipedia might start with a large number of contentious articles, and then gradually narrow this down to just one or a few for closer analysis. Measures used to downsize the sample could include text-external features (e.g. specific date ranges, presence on a list of widely-read articles) or text-internal features (e.g. presence of certain keywords, articles of a specific length, articles linked from a particular page, etc.).

Digitally mediated contexts lend themselves to the collection of large bodies of texts, though this can be difficult to reconcile with ethical requirements (see above). There are various sophisticated tools available (e.g. FireAnt for Twitter, WebCorp for general web corpora, NodeXL for various platforms). These involve some time investment to learn to use effectively. Page et al. (2014: Chapter 8) outline a number of techniques for the collection and selection of social media texts. However, simpler tools, such as using the advanced search functionality of search engines can also be a useful way of collecting a variety of data, providing certain precautions are taken to avoid the "filter bubble" effect caused by search engine algorithms tailoring content to specific users. Using a private browser window should ensure this effect is limited, and should also prevent researchers from accidentally collecting content that is only available to users who are logged in to a platform with a specific account. For critical discursive analyses of digitally mediated communication, often simple tools such as copying and pasting content from a platform into a word processor or spreadsheet, taking screenshots on mobile devices, or saving pages as PDFs can be sufficient. Where possible, data should be saved with its full set of multimodal features (including visuals, sound etc.), and then transcribed where necessary (e.g. written transcripts of videos) to be presented in an accessible format that conforms to ethical guidelines in academic texts.

4.3 Choosing data for a study of young successful female politicians
In most of the remainder of this chapter, I will illustrate some key points about analysis of digitally mediated communication using a specific example. In this case, I am interested in the political phenomenon of young, female politicians who have quickly risen to fame, and seem to have a strong traditional and social media presence. I would like to know more about how they achieve this success, how they discursively construct their and others' identities, and how they respond to media and social media attention. Although hate speech, misogyny and trolling directed against these politicians (or other female public figures) would be a worthwhile topic to study, this is something that is already the subject of considerable scholarly attention (e.g. Hardaker & McGlashan 2016), and I would prefer to focus more on their own strategies of positive self-representation, and how they position

themselves to interact with their social media followers. As suggested above, any digitally mediated communication project should begin with an overview of theoretical frameworks and concepts, followed by systematic collection of a range data and as much information about the topic and context as possible. In this case, we will assume that I have read up on US politics, progressive populism, gender and politics, gender and social media, and other relevant areas, and have tried to establish what the latest relevant theoretical frameworks relating to discourse in social media contexts are (e.g. Fuchs 2014; KhosraviNik & Unger 2016). A number of politicians immediately spring to mind; for instance, Mhairi Black, the Scottish National Party Member of the UK Parliament (MP) who, at age 20, became the youngest MP in almost two centuries, and Jacinda Ardern, the Prime Minister of New Zealand, who at age 37 is the youngest female leader of a government in the world. However, I settle on Alexandra Ocasio-Cortez (AOC), a Congresswoman for New York who was elected in the US midterm elections in 2018 after beating a long-term incumbent for the seat in the Democratic primaries, and then took office in early 2019. While each of these three politicians is interesting in her own right, and could be the subject of a comparative analysis, I choose just AOC because her success is recent, making her the most topical choice, and because she is a highly active (and interactive) social media user. Any claims I make are therefore limited just to AOC, and not to the other two politicians, which limits the claims I can make about the topic in general (young successful female politicians). My initial data search tells me there is a huge amount of media coverage of her successes, from the primary win onwards (e.g. Embury-Denis 2018), and also that she is an active social media user herself. I start looking through her social media profile and narrow my search down to Twitter, on the grounds that the data here are easily accessible (her tweets are public) and that the risk of my conducting research on her tweets is very low, because she has a much more visible public profile than any research I do is likely to achieve, and is a public figure. For the purposes of this analysis, I will focus on two tweets which illustrate many of the features I find in her other posts. However, I note that were I completing a full-scale project, I would analyse her data systematically and make more transparent choices. To do this, I could use corpus techniques (see Chapter X, in this volume) to identify particular aspects of content or form that make her tweets characteristic compared to other Twitter users, such as her use of particular words, emoticons or syntactic constructions, and then choose specific examples that are "typical" or "atypical" of her style. Or, even without using corpus tools, I could select a larger number of random tweets and manually note down their features, to get a sense of all the different ways in which AOC uses twitter, and then try to find tweets that illustrate these differences, to give a rich description of her social media practices. Alternatively, I could also focus on all the tweets within a specific date/time window, based on external events such as a policy announcement or an attack on her in the media. This could allow me to show how she uses her Twitter account in relation to specific political or media events. All of these approaches should be possible within the scope of a student project, bearing in mind the principle that the more data that is analysed, the less

detailed the analysis can be: this is why the process of data selection (i.e. down-sampling) following a transparent process is so important.

5. Analysing and Interpreting Data

Having collected the data, I am now ready to begin a preliminary analysis. Figures X.2 and X.3, below show the two tweets by AOC. They are both examples of quoted tweets – a feature introduced by Twitter in 2015, whereby an existing tweet is shown in an inset box with the user's comment/response above it. Both tweets share a number of features: the standard layout of Twitter includes the profile photo, name and @username, a number of buttons for replying, retweeting etc. AOC's profile picture is partially overlapped by a tick in a blue rosette, which indicates she is a verified user (a public figure). The photo looks professional, and she is smiling directly at the camera, which Kress and van Leeuwen (2006) would describe as placing a *demand* on the viewer to engage with the image. Her very short username (only three characters), which she acquired in late 2018, suggests that she is serious about her social media identity – such short usernames are not easy to come by, and it seems likely she or her campaign bought it for a not inconsiderable sum of money. Without inside knowledge of her campaign, there is no way of knowing to what extent she personally controls her social media account, though it is not uncommon for politicians to manage at least some of their social media accounts at present.

In the first tweet (Figure X.2), AOC replies to @derekwillis's comment about a political science conference. AOC here is doing some nuanced positioning work: on the one hand, her use of "I believe" could indicate uncertainty, and the use of quotation marks around "paradigm-shift" and indeed the hyphen in the term, might indicate she is distancing herself from the term as if not fully familiar with it. On the other hand, her use of a winking emoji suggest she is making a joke or being satirical. The term "paradigm-shift" was coined by the philosopher Thomas Kuhn (1962) and has been widely used to describe shifts of thinking in academic fields such as economics, including for the radical departure from orthodox economics of John Maynard Keynes, who AOC is presumably familiar with as a reported proponent of Modern Monetary Theory. In other words, AOC is poking fun at the political scientists by playfully appropriating "their" language, while also presenting herself as someone who wishes to bring about a marked change in US politics.

[Figure X.2 around here]

In the second tweet (Figure X.3), AOC quotes a tweet about a press release from a (supposedly non-partisan) think-tank. The press release is about her approach to campaign funding, which eschews large donations and instead relies on grassroots members. Rather than just retweet this arguably rather positive article about her, AOC first tells a "small story" (see Page 2013). Numbers, and particularly large amounts of money, seem to often be used in political communications as a way of being persuasive (compare, for instance,

Donald Trump's frequent tweets about how much money he has). Here, instead AOC is describing a very small amount ($1/month) because her aim appears to be to show that small donations can make a difference. She personalises the description of her campaign in the quoted tweet by describing a "dishwasher in New Jersey" – this is a well-established technique in political speeches (compare for instance US Presidential candidate's John McCain's invocation of "Joe the Plumber" in his 2008 campaign against Barack Obama). AOC's tweet uses empty lines to separate the story into three parts: the personal opening, then a more general philosophical point which follows from the dishwasher's selfless act, and finally a conclusion, followed by a link to her campaign donations page. Despite the white space, each subsequent part is linked to the opening through cohesive links: "That gesture…" and "Because of him", and finally the story is linked with the present "I can do more now" and a colon finally makes the connection with the donations page.

[Figure X.3 around here]

Overall, these two tweets appear to follow the pattern pioneered by Barack Obama and other Democratic US politicians: she engages with individual users, sends positive messages about herself and her campaign, uses a mix of light-hearted humour, appeals to pathos (emotion) and tries to differentiate herself from "politics as usual" in various ways. Although there is not enough space to describe them in full here, it is worth briefly considering the next steps in a study of this kind. For example, this could involve observing reactions (in the form of replies, retweets and media reactions) to AOC's tweets. Or an ethnographic element could be added, involving interviews with some of AOC's followers (supporters or detractors) along the lines suggested by Androutsopoulos (2008). Finally, although even looking at just two tweets can yield some interesting findings and interpretations, clearly a greater number of AOC's social media posts is needed, including examining her presence on different platforms, to present a fuller picture of her digitally mediated practices and their relation with the overall political context. This might also allow some reflection on the media dynamics at play here – popular and/or contentious politicians like AOC generate a huge amount of attention, which translates into revenue for private social media companies and their shareholders, not just political donations.

6. Issues and limitations

There are a number of limitations and issues in approaching digitally mediated communication as a researcher, some of which have already been outlined earlier in the chapter. First, there are considerable ethical barriers, which must be carefully negotiated to keep both participants and researchers safe. Second, there are some inherent features of digital media platforms that make researching them interesting, but sometimes challenging. One major challenge is the dynamic nature of many digital media texts. In many digital media contexts, including platforms like Facebook, Wikipedia, and online news sites, texts do not necessarily stay the same longer term or sometimes even for more than a few

minutes. Researchers thus need to decide whether they take a snapshot and analyse the text as "static at a particular time", or whether they attempt to capture this dynamism in their analysis for instance by documenting and discussing changes. However, not all platforms are as transparent as Wikipedia, which keeps all changes that have been made accessible to users, and it may be difficult for researchers to keep coming back to check for changes. A further issue, perhaps a limitation of existing research rather than a difficulty, is that it is very tempting for researchers to chase after new platforms, rather than focus on users' discursive practices within and across a range of platforms. Each new popular platform such Facebook, Twitter, Snapchat, has generated a raft of excited articles by academics documenting the "next big thing", while the actual practices of users may not actually change all that much over time (Herring 2004 warns of this danger). Finally, a problem for researchers is that they may wish to research a context that is relatively unfamiliar to them. This means not only that their lack of familiarity with conventions may be a barrier to gaining acceptance, but also that specific meanings may be obscure to them, due to the use of in-jokes, irony, particular registers and specialised lexis, or unfamiliar orthographies, to mention just a few problems. Nevertheless, through careful application of observations within a discursive framework, and the use of participant interviews where possible/desirable, these problems are surmountable.

Further Reading

Page et al. (2014) provide an accessible introduction to language and social media. With specialised chapters focussing on ethics, qualitative and quantitative approaches and ethnography, this is a good place to start for researchers new to researching in digitally mediated contexts.

Unger et al. (2016) is a chapter in a qualitative methods handbook which shows in more detail how the discourse-historical approach to critical discourse studies can be applied to social media data.

Fuchs (2014) is an excellent introduction to social media from a critical, sociological perspective. It is not a linguistics textbook, but rather examines some of the broader societal and philosophical issues related to social media.

Barton & Lee (2013) provides a highly accessible introduction to digitally mediated language, primarily from an ethnographic perspective.

References

Androutsopoulos, J. (2008). 'Potentials and Limitations of Discourse-Centred Online Ethnography'. *Language@Internet 5,* article 8.

Barton, D. and Lee, C. (2013) *Language Online: Investigating Digital Texts and Practices,* London: Routledge.

Culpeper, J., Kerswill, P., Wodak, R., McEnery, A., & Katamba, F. (Eds.). (2018). English language: Description, variation and context. Palgrave: London.

Embury-Dennis, T. (June 27, 2018). 'Everything you need to know about Alexandria Ocasio-Cortez, the 28-year-old socialist hailed as the "future of the Democratic Party"'. *The Independent*. London: Independent Print.

Fuchs, C. (2014). *Social media: a critical introduction*. London: Sage.

Gibson, J. (1977) 'The theory of affordances', in Shaw, R. and Bransford, J., eds., *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, Hillsdale, NJ: Lawrence Erlbaum, 67-82.

Gillen, J. (2013), Writing Edwardian postcards. *Journal of Sociolinguistics 17*: 488-521. doi:10.1111/josl.12045

Hardaker, C. & McGlashan, M. '''Real men don't hate women'': Twitter rape threats and group identity'. *Journal of Pragmatics 91*, 80-93.

Herring, S. C. (2001). Computer-mediated discourse. In D. Schiffrin, D. Tannen, & H. Hamilton (Eds.), *The Handbook of Discourse Analysis* (pp. 612-634). Oxford: Blackwell Publishers.

Herring, S. C. (2004). Computer-mediated discourse analysis: An approach to researching online behavior. In S. A. Barab, R. Kling, & J. H. Gray (Eds.), *Designing for Virtual Communities in the Service of Learning* (pp. 338-376). New York: Cambridge University Press

Herring, S. C. (2007) 'A faceted classification scheme for computer-mediated discourse', *Language@Internet 4,* article 1.

Herring, S. C. (2013) 'Discourse in web 2.0: Familiar, reconfigured, and emergent', in D. Tannen, D. and Trester, A. M., eds., *Discourse 2.0: Language and New Media,* Washington, DC: Georgetown University Press, 1-25.

Hine, C. (2015). *Ethnography for the Internet: Embedded, Embodied and Everyday.* London: Bloomsbury

KhosraviNik, M. (2017). Social media critical discourse studies (SM-CDS). *Handbook of critical discourse analysis*, 582-596.

KhosraviNik, M., & Unger, J. W. (2016). Critical discourse studies and social media: Power, resistance and critique in changing media ecologies. *Methods of critical discourse studies*, London: Sage, 205-233.

Kress, G. & van Leeuwen, T. (2006). *Reading Images: The Grammar of Visual Design.* London: Routledge.

Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Markham, A. and Buchanan, E. (2012) *Ethical Decision-making and Internet Research v 2*, available: http://aoir.org/reports/ethics2.pdf.

Page, R. (2013). *Stories and social media: Identities and interaction*. Routledge.

Page, R., Barton, D., Unger, J. W., & Zappavigna, M. (2014). Researching language in social media. In *Researching language and social media.* London: Routledge.

ten Have, P. (2007) *Doing Conversation Analysis* (2nd edn). London: Sage.

Unger, J., Wodak, R., & KhosraviNik, M. (2016). Critical discourse studies and social media data. *Qualitative research*, London: Sage, 277-293.