Ethical approval: none sought. How discourse analysts report ethical issues around publicly available online data

Wyke Stommel[iD] and Lynn de Rijk

Radboud University, The Netherlands

Abstract

Although ethical guidelines for doing Internet research are available, most prominently those of the Association of Internet Researchers (www.aoir.org), ethical decision-making for research on publicly available, naturally-occurring data remains a major challenge. As researchers might also turn to others to inform their decisions, this article reviews recent research papers on publicly available, online data. Research involving forums such as Facebook pages, Twitter, YouTube, news comments, blogs, etc. is examined to see how authors *report* ethical considerations and how they *quote* these data. We included 132 articles published in discourse analysis-oriented journals between January 2017 and February 2020. Roughly one third of the articles (85 out of 132) did not discuss ethical issues, mostly claiming the data were publicly available. Quotations nevertheless tended to be anonymized, although retrievability of posts was generally not taken into account. In those articles in which ethical concerns were reported, related decisions appeared to vary substantially. In most cases it was argued that informed consent was not required. Similarly, approval from research ethics committees was mostly regarded unnecessary. Other ethical issues like consideration of users' expectations and intentions, freedom of choice, possible harm, sensitive topics, and vulnerable groups were rarely discussed in the articles. We argue for increased attention to ethical issues and legal aspects in discourse analytic articles involving online data beyond mentioning general concerns. Instead, we argue for more involvement of users/participants in ethical decision-making, for consideration of retrievability of posts and for a role for journal editors.

Corresponding author:
Wyke Stommel, Faculty of Arts, Radboud University, PO Box 9103, 6500HD Nijmegen, The Netherlands.
Email: w.stommel@let.ru.nl

Keywords

Ethics, Internet research, online data, methodology, anonymization, public data

Introduction

The ethics of Internet research have been a hot topic over the past decade (e.g. Fossheim and Ingierd, 2016; Heider and Massanari, 2012; Linabary and Corple, 2019; Nissenbaum, 2010; Roberts, 2015; Williams et al., 2017; Willis, 2019; Woodfield, 2018; Zimmer, 2010; Zimmer and Kinder-Kurlanda, 2017), also in the area of applied linguistics and discourse analysis (e.g. Page et al., 2014; Paulus and Wise, 2019; Spilioti and Tagg, 2017). In this article we focus on this area of research. Here, ethical challenges include the use of verbatim quotations which are inevitable in discourse analytic publications. This focus allows us to engage with a particular research community (our own) which may serve as an example for other fields.

The discussion of ethics for discourse analysis encompasses many different dimensions and concerns, including the public/private distinction, anonymity, informed consent, searchability/retrievability, sensitivity of data, and vulnerable groups. Generally, it is argued that ethics should be viewed as a contextualized process of decision-making at all critical junctures of a research project (Spilioti and Tagg, 2017). This implies that ethics should be of central concern to researchers who study social media, Internet discussions or other types of publicly available online data.

There have been attempts at creating ethical guidelines for Internet researchers and relevant research ethics committees, the most well-known are those developed by the Association of Internet Research (AoIR, www.aoir.org/ethics). These guidelines are based on general research ethics regulations which were put in place after World War II (Page et al., 2014). However, due to the diversity of the online landscape and Internet research approaches, many such guidelines leave room for interpretation. It is also not uncommon for research ethics committees to have difficulty making decisions about Internet research (e.g. Stevens et al., 2015: 17), since the Internet is so immense, diverse and ever-changing. Moreover, ethics and legal regulations vary across academic disciplines, countries and regions, and regulations and laws are adapted or replaced over time. This means that no guideline can cover Internet research generally, which leaves a lot of difficult decisions up to the researcher, especially when research ethics committees decide that protocols do not apply to the research. For example, weighing potential harm to participants is an ethical question but it may be difficult to decide what constitutes harm in case of research on online data. Does the risk of harming participants depend on the topic of online discussion, or on the assumed identity of the participants? Does harm relate to the research as such or mainly to quotations of the online messages?

Page et al. (2014) extensively discuss a number of such ethical dilemmas, raising awareness of sensitivities and concerns relevant to research involving social media. One such concern is the distinction between public and private data and whether participants may be assumed to anticipate that their messages may be used for research without their consent. Awareness of posting things publicly is not the same as consenting with research. The comparison of an online platform with conversations in a public square can also play a part in the consideration of the public-ness of the platform (Willis, 2019). Furthermore, the ethical dimension of accessing or collecting data includes considering the privacy of an online interaction and reflecting on the appropriateness of a situation (Nissenbaum, 2010). This means that in order to weigh up whether privacy will be infringed, it should be considered what is disclosed to whom in the particular environment. For example, what somebody discloses on a depression support forum, might not even be known to friends and family, implying this is rather private irrespective of whether the forum is publicly available. Furthermore, participants' expectations with regard to privacy may vary largely even within one single online community. Overall, the distinction between private and public data is not clear-cut despite public availability. Reflecting on an issue like privacy requires researchers' self-reflexivity rather than a statement like "the data were taken from a public site." Self-reflexivity has also been proposed as one of the priorities for those who study online data (Spilioti and Tagg, 2017). Scholars should:

> not only declare that ethical considerations were taken into account but . . . also explain how they were applied throughout the research process. For example, all scholars, whether they systematically undergo an ethics board screening or not, should share their application of ethical principles to their research to encourage ethics-related conversation (Locher and Bolander, 2019: 88).

With growing ethical awareness, comes a call for a more reflective, ethics-related conversation.

Furthermore, Page et al. (2014) state that "both informed consent and anonymization should be used to protect the privacy rights of participants" (Page et al., 2014: 75). However, this may be different when users of online platforms are regarded as authors (Bassett and O'Riordan, 2002). Given that quoting is standard practice in discourse analytic research, the question is which advice is followed and whether and how this is accounted for.

So, ethical dilemmas and decisions are inherent in research involving naturally-occurring, publicly available online data (posts, interactions, hashtags, clips, GIFs, emoji, etc.) that have not been generated for research purposes (Lester et al., 2017). What we lack is insight into whether and how the "conversation" about these issues is taking place in research articles. This is important to know because ethical guidelines tend to be rather general, so researchers might turn to published

journal articles to inform their decision making. We assume that how researchers report ethical considerations, including justifications and reflections on their decision-making, has implications for how ethical research conduct is understood by the research community (cf. Paulus et al., 2017). Hence, in this article we examine:

(1) How researchers reported ethical issues, how often and which ethical aspects?
(2) How researchers protected the identity of users behind the data, if they deemed this necessary (e.g. real names or pseudonyms, hyperlinking)?

Thus, our study is descriptive in focus, aiming to gain insight into which dimensions of ethics are most prevalent in discourse analytic articles. In the discussion of our findings, we suggest some "good practices" for reporting ethical issues for research on this type of online data.

We want to emphasize that it is by no means our intention to criticize fellow researchers for the ethical choices they made. We hope this article serves as input for the ongoing discussion of ethical conduct in research on online data (cf. Zimmer, 2010), which we consider crucial for understanding contemporary communication and social life.

Method

Our sample consisted of articles using publicly available online data of various sorts (see Table 1) published in journals with a focus on discourse (see Table 2 for the selected journals). We restricted our review to the field of discourse analysis, that is, empirical studies examining naturally occurring language-in-use, for three reasons. First, we are interested in quotations from online data, because they pose substantial ethical challenges, and quotations are standard practice in discourse analytic articles. Second, a focus on one field can link *reporting* practices to ethical *reasoning* in this particular field. Third, both authors work in the area of discourse analysis. Hence, this study engages with their own research community, while it ideally also sets an example for other fields.

We selected the analytical articles involving online data from all journal issues published between January 2017 and February 2020. Inclusion of articles in our corpus was based on the following criteria:

(1) the data were naturally-occurring,
(2) the data were reasonably accessible online by anyone (i.e. excluding data like e-mails or instant messaging, but including forums and social media for which an account is needed), and
(3) the article included quotations, whether or not verbatim.

Inclusion was not based on whether the authors presented their research as discourse analysis. The search resulted in a corpus of 132 articles, of which the majority employed qualitative methods (*n*=87), while others used mixed (*n*=31) or quantitative methods (*n*=14). Of these 132 articles, the methodological section was read in its entirety and the whole article was searched for a number of keywords: "ethic*", "public," "pseudo*," "anony*," "permission," "consent," "board," "IRB," "committee," and "agree*." We chose these keywords, as they were regularly deployed in sections of ethical considerations. When a keyword was identified in an article, the context for each keyword was checked for relevancy.

Author 2 and a research assistant coded the articles for codes that we extracted from discussions on the ethics of Internet Research (see Table 3), using Atlas.ti. The codes were extracted from Spilioti and Tagg (2017), the AoIR guidelines and Page et al. (2014) (see Table 3). The method of analysis was a rudimentary content analysis (Neuendorf, 2005). We added the code '*Referencing other articles/guidelines*' after having familiarized ourselves with the data. A second round of analysis was conducted per code, where sub-codes were added to do justice to variations within the coded phenomenon (e.g. within "informed consent" some articles claimed consent was unnecessary while others did seek consent). Codes were attached to relevant fragments using the QDAS Atlas.ti (version 8.4.5) in its simplest form, for quick overview of all relevant citations per code. Apart from explicit discussions of ethical issues, we also coded for statements that user consent or administrator permission was sought and about how the data were anonymized. When none of the dimensions were reported on, the article was coded as "Ethics unaddressed." Merely factual descriptions of anonymization procedures were not counted as ethical deliberations, as reasons for anonymization were not discussed (e.g. "(. . .) all complaints were anonymized in order to avoid identification as much as possible" (DeCock and Depraetere, 2018). Lastly, we analyzed how data were quoted, namely (1) if data were (non)-anonymized and (2) if data were quoted verbatim or modified in some way (we regarded translations as modifications, regardless of the reason for translation).

Next, we grouped some codes to form categories. For example, "asking for consent" (when users were approached) and "permission given" (when group administrators were approached) were aggregated for a more comprehensive discussion of involving others.

Findings

In approximately one third of our data set (47 out of 132), issues of research ethics were discussed, either minimally in one or two sentences, or more elaborately up to a paragraph. Discussion of ethical issues did not seem to depend on journal

Table 1. Data types in discourse studies of online data

Data type	Number of articles ($n = 132$)
Blogs	10
Facebook	22
Forum(s)	17
Instagram	7
Mumsnet	3
News websites	17
Reddit	4
Reviews (e.g. Airbnb, TripAdvisor, OpenRice)[1]	6
Tumblr	4
Twitter	28
Weibo	3
YouTube	6
Other platforms[2]	21

Note that some articles include multiple data types.

Table 2. Number of articles in which ethical issues were addressed/unaddressed per journal

Journal	Ethics addressed	Ethics unaddressed	Total
Applied Linguistics	0	1	1
Critical Discourse Studies	0	3	3
Discourse & Communication	7	8	15
Discourse & Society	1	6	7
Discourse, Context & Media	23	25	48
Discourse Studies	2	4	6
Journal of Computer-Mediated Communication	1	1	2
Journal of Language and Politics	0	9	9
Journal of Pragmatics	5	19	24
Language@Internet	4	8	12
Qualitative Health Research	4	0	4
Visual Communication	0	1	1
Total	47 (36%)	85 (64%)	132

policy, as most journals published articles with and without ethical discussion. See Table 2 for an overview per journal of how many articles did and did not address ethical issues.

None of the studies employing only quantitative methods ($n = 14$) discussed ethical issues. In the following, a quantitative study is indicated with an asterisk (#) to enhance transparency. We first focus on articles with no discussion of ethical

issues and then discuss the subset of our corpus in which ethical issues were reflected upon. This enables us to provide further details on which ethical dimensions were addressed and what these ethical accounts encompassed. We present some codes combined in one paragraph (e.g. users' expectations and users' intentions).

Ethics unaddressed

From the total set of articles, approximately two thirds (85 out of 132) did not report on ethical considerations when using online data. This means that ethical issues were not explicitly discussed nor was the obtaining of consent or permission mentioned.[3] The data in these articles were taken from platforms such as Facebook, Twitter, YouTube, and blogging sites like Tumblr. This suggests that it is common practice to use publicly available, online data for research purposes without making ethical issues explicit. Independent of ethical discussions, most articles anonymized quotes from the online data. The use of pre-collected data seemed to suggest to authors that considering ethical issues is not necessary. We discuss these two issues here.

Quotations and anonymization

Quotations of online data were typed text, sometimes including emoticons/emojis and rarely images. Overwhelmingly, quotations of the data were anonymized, even in the 85 articles in which ethical issues were not discussed. However, in many articles, it was not made explicit how the anonymity of the users was protected (68 out of 132). Although anonymity was frequently regarded as built-into the platforms with users being registered under self-invented nicknames, these nicknames were usually replaced by pseudonyms or identifiers in the articles. However, 36 articles did not (consistently) anonymize usernames, or offline identities. Also, usernames may be replaced in some quotations, but not in others. Adding hyperlinks to quoted posts was not uncommon in articles that did not anonymize, which implies that users/authors and their comments are retrievable.

As search engines can in some cases link quotations to authors/users, some researchers took measures beyond anonymization to prevent identification of users, for example by translating the cited data (Merrill and Åkerlund, 2018) or by paraphrasing rather than quoting (e.g. #Bright, 2018; Karlsen and Scott, 2019). We discuss two salient anonymization practices more extensively and one case in which citations were intentionally not anonymized.

In some articles, data were taken from online sites that were argued to be public in a more obvious way than sites of non-public persona. These included Twitter pages of politicians or a Facebook page of an organization. Nevertheless, ethical

decisions also vary in this subset of studies. In some cases anonymization was regarded as important, for instance in a quantitative study into citizens' interactions with politicians on Twitter (#Bright, 2018). Here, not only the usernames of all participants including the political parties and politicians were altered, but quotes were also slightly modified to "preserve the privacy of participants" (Bright, 2018: 21). In contrast, in a study of interactions on the Facebook pages of four Nigerian broadcasting news outlets with messages from citizens, literal, and non-anonymized quotes were provided for the following reason:

> To maintain spontaneity and originality, there was no indication to the community members that their comments were being captured for research purposes. The samples are presented as undoctored as decency and ethical constraints would allow (. . .). (Oyadiji, 2020: 183)

Leaving Facebook data non-anonymized, as Oyadiji (2020) did, is rare, probably because Facebook is often directly tied to users' offline identities. Thus, even in studies of online data with an arguably high public character ethical decisions appear to vary largely.

Even in articles in which anonymization is referred to, actual quoting practices may seem at odds with the policy given. We have identified a study on a highly sensitive topic and provided details to the editors of this journal. The study used pseudonyms for quotations, but still provided a footnote with the hyperlink to the original discussion, making the quotes identifiable. This shows that an anonymization account may not always be consistent with quoting practices.

In summary, anonymization was often not taken into account to such an extent that users were definitely irretrievable and the issue of searchability was rarely discussed.

Pre-collected data

Out of the 132 articles, 8 relied on a pre-collected dataset. Only one of these articles addressed ethical concerns (Jucker et al., 2018). This suggests that authors assumed that for datasets compiled prior to the research project an ethical discussion in the article is unnecessary. Many articles emphasized the public nature of the platform the data originated from or the public availability of the corpus they used.

However, even pre-collected datasets may involve apparently sensitive data. An online corpus of chat logs (details were made available to the editors) contained online grooming attempts by users who at the time believed they were communicating with an underage person. These interactions were part of decoy operations with the goal of prosecuting online groomers. The ethical side of using these data was not discussed in the article. One could therefore hypothesize that publicly available data or corpora seem to be treated as a justification for not discussing potentially relevant ethical concerns related to the data.

Table 3. Frequency of considering the ethical dimension in the articles

Ethical consideration	Number of articles (*n* = 47)
Ethics committee	10
Freedom of choice	1
Informed consent	22
Legal concerns	1
Permission given by administrator or moderator of platform	7
Possible harm (versus importance of study)	7
Public/private	33
Referencing other articles/guidelines	29
Sensitive topics	7
Terms and conditions	5
Users' expectations	6
Users' intentions	5
Vulnerable groups	2

Ethical issues addressed

In 47 out of the 132 articles, ethical implications of studying naturally-occurring, publicly available online data were discussed, although the comprehensiveness of these considerations varied substantially, with some using a few sentences and others devoting multiple paragraphs to research ethics questions. Ethical considerations included reference to other articles or research ethics committees and issues like informed consent, possible harm, user intentions and expectations, and sensitivity of the data. Similar to the articles that did not address ethical issues, the public aspect of data types was recurrently mentioned in articles that did discuss potential ethical concerns (33 out of 47), although this was not reported as justifying the use of the data without further steps or considerations. It was quite common (29 out of 47) to refer to how others dealt with or discussed ethical issues. Furthermore, informed consent was frequently (22 out of 47) deemed worthy of discussion, but rarely sought. Table 3 provides an overview of the frequency of considering the ethical dimensions in our data set.

What follows is a discussion of the different ethical considerations and actions taken as reported in the articles.

Public/private

The public availability of online data was the central argument used in justifying the inclusion of data in research. Of the 47 articles in which ethical issues were discussed, it was brought up 33 times, including in studies of news comments, Instagram posts, forums, reviews, Twitter, blogs, Facebook, etc. There does not

seem to be a consensus on whether certain data types are evidently public or not. Health-related discussions, for instance, were sometimes argued to be sensitive and therefore requiring ethical measures (Pounds et al., 2018), while others argued that they were public and thus freely usable (Sahota and Sankar, 2020).

Research ethic committees

In 10 out of 47 articles, references were made to a research ethics committee or Institutional Review Board. In six cases, approval or guidance from these committees was sought and obtained, while in the other four cases, no approval procedure was initiated. Researchers justified the absence of ethics approval by stating that:

- their data were considered "exempt from review" (Gordon and İkizoğlu, 2017: 268),
- their study was "a textual analysis and not human subjects research" (Sahota and Sankar, 2020: 296),
- "all data were gathered from publicly accessible venues" (Salzmann-Erikson and Hiçdurmaz, 2017: 287), or
- that their analysis was "a secondary review of pre-existing data" (Karlsen and Scott, 2019: 3).

At the same time, two of these statements were immediately followed by a list of precautions that were taken to protect the privacy of users. Thus, an exemption from ethics review was not treated as alleviating responsibilities regarding the use of the data for research altogether. But it was in some cases treated as the reason why it was understood that seeking users' consent or a more in-depth discussion of ethical issues was unnecessary. Overall, judging from the content of the articles analyzed, the discussion of ethical issues in relation to using online data seems to have taken place largely independent of research ethics committees. Researchers were much more likely to refer to other articles or guidelines than to their universities' research ethics committees when making ethical decisions (see paragraph 4.4).

Informed consent or permission by administrators

In 29 out of 47 articles in which ethical issues were discussed, the issue of permission to use data or informed consent was mentioned. Two aspects seemed to feed this ethical measure: whether the study was framed as human participant research and the extent to which the platform was seen as public. Sometimes, informed consent from participants was replaced by permission from platform administrators, such as forum administrators or Facebook group hosts. In 17 out of the 29 articles, it was argued that informed consent or permission to use the data was not required from an ethical perspective. Generally, asking for consent was argued to

be irrelevant due to the public availability of the data. A further justification focused on the participants and that it may be difficult to reach them or that it may impact on their forum behaviors.

In 8 out of the 29 articles, users were informed about the research and asked for consent. It was sometimes argued that a distinction could be made between users that did need to be asked for consent and others that did not. For instance, bloggers were asked for consent, but not the users that commented on these blogs. Or only the members of a *closed* Facebook group or *personal* Facebook profiles were asked for consent, not the participants of the other data sources that were used in these studies. Thus, where informed consent was sought, the necessity of this undertaking was often based on the public/private dichotomy, where public data was argued not to necessitate asking for consent. However, seeking informed consent *was* generally regarded as inevitable for analysis of images, particularly selfies (e.g. Koteyko and Atanasova, 2018; Matley, 2018; Veum and Undrum, 2018).

Another option was asking permission from forum or group administrators, which was reported in 3 out of 29 articles (Jaworska, 2018; Magaña and Matlock, 2018; Pounds et al., 2018) or from the company whose tweets were analyzed (Decock and Depraetere, 2018). This may entail a multifaceted process of ethical decision-making, including informing participants and creating an opportunity to "opt out" (Pounds et al., 2018).

Referencing other articles/guidelines

Many articles (29 out of 47) referred to other sources like the AoIR in explaining their ethical considerations and decisions. For instance, criteria postulated by other researchers were used to determine if consent from users had to be sought or whether additional precautions to protect users' identity had to be taken. Reference to others may be specific for distinct ethically motivated measures, such as the following:

> With respect to ethical concerns regarding data taken from online sources (BAAL, 2016; Page et al., 2014: 58–79), the posts were anonymized, and traceable unique hashtags were removed from the data. Equally, images and user names are given in this article only with explicit permission from the holders of the Instagram accounts. (Matley, 2018: 32)

In other cases, referencing others was used implicitly as justification for not taking certain measures:

> Consistent with recommended ethical practice for Internet research (Ess and Association of Internet Researchers, 2002; Wilkinson and Thelwall, 2011) and with prior studies relying on publicly available mental-health-related content from online message boards (Gajaria et al., 2011), this study is a textual analysis and does not constitute human subjects research. (Sahota and Sankar, 2020: 296)

Lastly, reference to other articles can also be a way to acknowledge differences of opinion regarding ethical issues. For instance: "Posts published on Facebook Pages are considered public information and can be accessed by anyone (but see Zimmer, 2010 for a discussion of the ethics of researching materials sourced from Facebook)" (Lawless et al., 2018: 46).

Researchers in our data set also referenced more generally, acknowledging that ethical issues for analysis of online data are not fully resolved (yet) (e.g. Gibson and Roca-Cuberes, 2019; Housley et al., 2017).

In summary, reference to others' ethical considerations in Internet Research was used to justify own conduct, in various degrees of extensiveness and more or less explicit in applicability to the study at hand, sometimes leaving it up to the reader to infer how the reference and the study are connected.

Users' expectations, intentions, and freedom of choice

In 10 out of 47 articles, the possible viewpoint and actions of the users were discussed as part of the ethical decision-making process, but in ways that did not align with each other. Three of these articles acknowledged that users might not expect their data to be used for research, despite the public nature of the platform, while three others argued that users were aware of their data being publicly available and could therefore be used for research without consent.

Reporting that data are publicly available seems to legitimize use of the data for many authors of articles in our data set on the following assumptions: "(. . .) users are aware of the potentially broad dissemination of their tweets." (Vladimirou and House, 2018: 153) or "(. . .) online observation should only take place when and where users "reasonably expect to be observed by strangers" [in reference to certain guidelines, but implying that their research falls within these boundaries]" (Kalim and Janjua, 2019: 75).

The articles that discussed users' expectations or intentions in direct relation to their study, assumed that users would most likely not expect this. These researchers either asked for consent (Atanasova, 2018) or decided to anonymize their data more extensively on the basis of these considerations (e.g. Zappavigna and Zhao, 2017).

One article referred to users' own freedom of choice in posting on a certain platform. The freedom users have to place a review without creating a profile, thus staying completely anonymous, is mentioned as further evidence that a certain platform could not be seen as a private group (Ren, 2018: 8) and could therefore be used for research purposes.

Hence, when user expectations and responsibilities were considered in the articles, these were often inferred from (other) posts on the platform the data were taken from.

Terms and conditions and legal concerns

Legally, terms and conditions stand for a contract between platform and user. When users are active on a platform, it is implied that they agree to the terms and conditions of that platform. Terms and conditions were occasionally (5 out of 47) referred to as relevant to potential ethical concerns. Mostly (four times), this was done to emphasize the public nature or availability of the data. For example, Jaworska (2018) states that "[t]he terms and conditions of Mumsnet stipulate that *Talk* is a public space and users are made aware that anyone can view their posts" (Jaworska, 2018: 27). Thus, the terms and conditions emphasize the public availability of posts. It is suggested by some authors in our data set that when terms and conditions include such a claim as the one given above, that this reduces ethical concerns.

Terms and conditions were also referred to as a justification for not modifying quotes. Giaxoglou (2018) explains the way tweets are presented in the article makes references to terms and conditions, referring to "Twitter's general principles for offline display/fit to print that posit the requirement to unmodified text" (Giaxoglou, 2018: 15). In contrast, data from an anti-immigration Facebook group were considered potentially illegal because they included hate speech (Merrill and Åkerlund, 2018). Therefore, the authors anonymized the data to protect the Facebook group users. None of the articles discussed the legal status of terms and conditions, nor of legal privacy issues related to the use of data for research and the quotation of posts.

Sensitive topics and vulnerability of participants

The sensitivity of topics or vulnerability of participants was mentioned as an ethical concern in 8 out of 47 articles. These articles either focused on a (mental) health topic, politics or on potentially sensitive topics. Sensitivity or vulnerability thus seemed related to participants revealing personal or private information about themselves, including illness or a political/ideological stance. However, our data set included many more articles in which data arguably included personal or private information, such as a Facebook discussion on dementia, forum discussions on weight loss and discussions on a cancer support forum. In these, no reference was made to sensitive topics and the vulnerability of participants. Hence, sensitivity of topic and vulnerability of participants are often *not* part of ethical accounts in articles.

Recognition of a topic as sensitive led to diverse ethical decisions. For example, one paper acknowledged the sensitivity of ideologically tainted comments on a news site, but it was not decided to act upon this in terms of anonymization or otherwise. In the health domain, personal obesity blog posts were recognized as sensitive (Atanasova, 2018), which lead to the decision to inform the users about

the research and ask personal bloggers, for consent. Some articles took even further measures. Pounds et al. (2018) who examined a Facebook support group on diabetes sought consent via the group moderators and tried to ensure that most participants of the group would know about the study and had the chance to object to the study or withdraw from the group during the observation period. It was reported that no one objected within the given time frame (which was decided upon in consultation with the group moderators) and that some participants responded positively to the notification of the study. Lundström (2018) acted upon the sensitivity of the topic in his research on suicide support on general forums by translating the data and removing all identifying information from the posts ("all references to specific discussion topics, user names, personal names, public institutions and geographical places," Lundström, 2018: 99) and by focusing only on support posts, not on posts about suicidal thoughts.

In summary, sensitivity was in some cases regarded a reason to involve site moderators and, as far as possible, users themselves.

Importance of study and/or possible harm

In 3 out of 47 articles indirectly referred to the importance of the research as a justification for ethical decision-making. Two articles argued in favor of the research because of importance in spite of ethical concerns. Housley et al. (2017) claimed that "the study of high-profile public social media accounts remains of analytic interest" (Housley et al., 2017: 569), while Deschrijver (2018) did not defend the research in general, but the importance of evaluation by fellow researchers to explain why they reproduced posts and usernames verbatim(Deschrijver, 2018: 15). Once, the analytical value of user names as important identity constructs was acknowledged, but this acknowledgement did not outweigh the ethical concern of "the risk of [users'] views being exposed outside of the particular context that they produced them in" (Gibson and Roca-Cuberes, 2019: 4).

In 4 out of 47 articles that addressed ethical issues, possible harm was mentioned, but generally not assumed to be directedly related to the study. For instance, a potential risk of harm was discussed as inherent to posting publicly online, and thus not specifically related to the research. Generally, the issue of possible harm (vs importance study) was rarely discussed.

Discussion and conclusion

This review illuminates that reporting on and acting upon ethical concerns and implications of discourse analytic research on naturally-occurring, publicly available online data is extremely varied. Generally, ethical issues seem under-discussed in research articles. When ethical issues are discussed, the comprehensiveness

of the discussion varies largely. The ethics reporting and quoting practices seemed loosely related to the type of data. Analyses of Facebook groups and forums for instance, tended to catalyze more ethical reflection than studies of Twitter and news comments.

Overall, it is remarkable that roughly two thirds of the articles in our sample did not discuss ethical issues at all, which implies that repeated calls for ethical reflection and conversation (AoIR; Page et al., 2014; Spilioti and Tagg, 2017) were not addressed. However, even when ethical issues were not topicalized in an article, the data were frequently anonymously quoted, although steps to make the authors of the posts irretrievable by search engines were rare. Use of a pre-collected data set appeared to make ethical accounting redundant.

When ethical concerns were reported in articles, the public availability of discussions on sites such as Twitter, Facebook, forums, and YouTube was overwhelmingly treated as *the* reason posts can be used for research, despite the fact that the dichotomy of private and public data has been recurrently claimed inappropriate to capture interactions on online platforms (Giaxoglou, 2017; Nissenbaum, 2010; Spilioti and Tagg, 2017). Reference to other articles or guidelines served to account for own practices, although often not related to specific decisions. Informed consent was regularly mentioned, but in most cases argued to be not required. The norm to seek informed consent when quoting posts from platforms (Page et al., 2014), did not seem to be adopted by the research community. Also, research ethics committees were mentioned, but mostly regarded unnecessary for the research at hand or ruled as such by the approached committee themselves. When a research ethics committee approved of a project, the committees' considerations were not made explicit in the research articles. Other ethical issues like consideration of users' expectations and intentions, freedom of choice, possible harm, sensitive topics, and vulnerable groups were hardly ever referred to in the articles.

In the following, we discuss some of our findings more extensively to highlight good practice and suggest future directions for ethical reporting and quoting in research articles involving online data.

First, ethical discussions that focus specifically on a concrete research project and the source of data, seem much more useful than general claims about what is ethical. For instance, arguing why a particular forum is thought to be public allows others to reflect on and respond to this argumentation, while a simple claim like "the data are publicly available" does not do justice to the multifaceted nature of online data. Researchers need to collectively make an effort to clearly report ethical considerations paving the way for those who follow. The fact that many articles refer to, and thus seem to rely on what others did/how others reasoned, underscores the value of making ethical deliberations more comprehensive and transparent (cf. Page et al., 2014; Spilioti and Tagg, 2017).

Second, in our corpus, the users or administrators of a site or platform were rarely involved in the research process. An analysis of users' and researchers' attitudes toward research of social media data (Golder et al., 2017) revealed that users' agreement with the terms and conditions of platforms was thought insufficient as a replacement of informed consent. The Golder et al. study also showed that although some users felt that a study's importance could trump individual privacy concerns, others argued that users' privacy interests should always take precedence over the researchers' goals. Additionally, some users in this study thought that researchers needed to gain permission from administrators *as well as* individual users, while others specifically opposed the possibility that list owners give permission on behalf of the users without user involvement in the process. Given that these issues were also recurrently reported in articles in our sample, it seems worthwhile to consider involving actual users in ethical decision-making (cf. Spilioti and Tagg, 2017).

Third, when a study benefits the greater good, users tend to think that research was more acceptable than if the benefit of the study is unclear (Golder et al., 2017). However, research articles only rarely or minimally explained the importance or benefits of their research in this context. Hence, making the assumed benefits of a study clearer could is recommendable in articles on online data.

Fourth, anonymization practices for quoting online data seem incomplete when traceability of posts is not considered or even intentionally prevented. The recommendation to seek informed consent when quoting (Page et al., 2014) is rarely followed, which makes anonymization even more important. It may be argued that informed consent is not required because the object of study is discourse rather than the human participant who produced the discourse (Bassett and O'Riordan, 2002). However, this is clearly too simplistic (cf. Stommel and Jol, 2016; Willis, 2019). As long as the topic of informed consent for the use of online data is unclear, it therefore seems advisable to consider retrievability of quotes (tweets, comments, etc.) and where data can be retrieved, to build in modifications where possible. This is particularly the case when the topic is sensitive (e.g. Lundström, 2018).

Fifth, judicial concerns related to ethics and/or copyright were rarely discussed in the articles with the exception of referring to terms and conditions of sites (cf. Pihlaja, 2017). This is surprising given the increasing attention for privacy online. Focusing on the European context, the General Data Protection Regulations (GDPR) require a legitimate ground for the processing of personal data. This can be consent, that should be given freely, informed, specific, and unambiguous (GDPR, 2016, art 6(1) sub a). Another ground, which can be invoked when sensitive data (such as health data, or data concerning sexual habits and attitudes) are in play, is that information is or was "manifestly made public" (GDPR, 2016, art 9(2) sub e), which could be the justification for some online platforms. Research may

rely on various exceptions when needed, to secure that the results of the research can be achieved (GDPR, 2016, art 89). Nevertheless, rights of subjects need to be safeguarded in a number of distinct ways, including that they should be notified about the research taking place and that they can exert specific rights. These judicial issues call for discussion among researchers and between legal experts and researchers, since their relevance to research on publicly available online data is obvious.

Sixth, our sample included one case in which it was decided that the use of data, namely usernames as a means of online identity construction, was unethical (Gibson and Roca-Cuberes, 2019) and one case in which it was argued that researchers should consider ethical issues before they choose which site to use (in this case for an analysis of online suicide discussions) (Lundström, 2018). Such considerations are valuable input to the ongoing conversation on ethics of research involving online data, as they point to the option of deciding against certain research questions, data types/online communities, etc. and thus to potential limits in research. We recommend that authors make explicit what research was *not* done, and/or *not* approved by ethics committees. This will help others to learn from preceding research projects.

Seventh, our findings show that individual journals do not, or not consistently require a discussion of ethical issues in the articles they publish. Journal editors could enhance self-reflexivity on ethical issues by explicitly requiring an account of ethical decision-making.

Finally, we want to note that our approach of analyzing the data through coding on various dimensions had a disadvantage. While the separate codes enabled us to systematize the analysis of the ethical discussions, it did not fully grasp that some discussions of ethical issues were much more comprehensive and/or convincing than others. The frequencies of mention of ethical issues are probably less telling than the close examination of ethical accounts in individual articles. We have tried to compensate somewhat for this shortcoming in the description of our findings per code.

Another limitation of our review is that our corpus did not include articles on contentious online phenomena like trolling and hate speech (cf. Rüdiger and Dayter, 2017). Also, it included very few cases of the use of images and only one study of the dark web. These specific data types presumably raise specific ethical dilemmas, such as additional risks for the researcher when asking consent for analysis of hate speech posts. This also holds for an allegedly truly anonymous platform like 4chan (Ludemann, 2018), whose user identities are unrelatable to any identity, online nor offline. Apparently, these phenomena are currently hardly studied by discourse analysts. Future research should therefore consider ethical issues for such data types and online phenomena.

We conclude that there is a discrepancy between research ethics in "theory" and "practice," more specifically, between ethical ideals and ethical justification/ discussion in journal articles. To resolve this discrepancy we propose more overt attention to ethical issues in discourse analytic publications, so that specific ethical challenges that are inherent to the multifaceted landscape of Internet research are more open for discussion between researchers. A useful future exercise would be to integrate our findings and the ensuing recommendations with existing ethical guidelines and frameworks (e.g. Page et al., 2014; Woodfield, 2018; Zimmer and Kinder-Kurlanda, 2017).

Acknowledgements

Funding

ORCID iD

Wyke Stommel https://orcid.org/0000-0003-2345-1691

Notes

1. Since reviews appear on a multitude of platforms, they were grouped by datatype. Reviews in this category appeared on the following platforms: Airbnb, Amazon (China), TripAdvisor, OpenRice, Xhiaozhu (Chinese Airbnb), and Yelp.
2. This category includes the following platforms: 4chan, Beyond Blue, CosmoVotes, the comment section of IS propaganda videos, Jodel, LinkedIn, chatrooms of Spanish sports newspaper Marca, online/social media marketing campaigns, Nairaland, Naijapals, Nigeria Village Square, OkCupid, Open Science notebooks, Periscope, Perverted-Justice.com, webpages associated with popular radio satire, Second Life, Sound Cloud, Taiwain e-Hospital (an official site run by Taiwan's Ministry of Health and Welfare), Twitch, Weibo, and Wikipedia talk pages.
3. When ethical were not discussed, this does not mean that the author(s) did not deliberate on ethical issues. It only means that their deliberations were not reported in that particular publication. For example, ethical issues were not mentioned in Georgakopoulou and Giaxoglou (2018), but extensively discussed elsewhere (Georgakopoulou, 2017), showing that the authors were aware of the ethical dilemmas in the field.

References

Note: references preceded by an asterisk were part of our dataset.

*Allington D (2018) 'Hitler had a valid argument against some Jews': Repertoires for the denial of antisemitism in Facebook discussion of a survey of attitudes to Jews and Israel. *Discourse, Context & Media* 24: 129–136. DOI: 10.1016/j.dcm.2018.03.004.

*Andersen EM and Rathje M (2019) Age and stage of life categorizations used to moralize in online social conflict. *Discourse, Context & Media* 28: 19–26. DOI: 10.1016/j.dcm.2019.02.001.

*Atanasova D (2018) "Keep moving forward. LEFT RIGHT LEFT": A critical metaphor analysis and addressivity analysis of personal and professional obesity blogs. *Discourse, Context & Media* 25: 5–12. DOI: 10.1016/j.dcm.2017.09.012.

Bassett EH and O'Riordan K (2002) Ethics of Internet research: Contesting the human subjects research model. *Ethics and Information Technology* 4(3): 233–247.

*Bondi M (2018) Try to prove me wrong: Dialogicity and audience involvement in economics blogs. *Discourse, Context & Media* 24: 33–42. DOI: 10.1016/j.dcm.2018.04.011.

*Brennan J (2017) Cruising for cash: Prostitution on Grindr. *Discourse, Context & Media* 17: 1–8. DOI: 10.1016/j.dcm.2017.02.004.

*Bright J (2018) Explaining the emergence of political fragmentation on social media: The role of ideology and extremism. *Journal of Computer-Mediated Communication* 23: 17–33. DOI: 10.1093/jcmc/zmx002.

*Carter-Thomas S and Rowley-Jolivet E (2017) Open science notebooks: New insights, new affordances. *Journal of Pragmatics* 116: 64–76. DOI: 10.1016/j.pragma.2016.12.003.

*Catalano T and Fielder GE (2018) European spaces and the Roma: Denaturalizing the naturalized in online reader comments. *Discourse & Communication* 12(3): 240–257.

*Coesemans R and De Cock B (2017) Self-reference by politicians on Twitter: Strategies to adapt to 140 characters. *Journal of Pragmatics* 116: 37–50. DOI: 10.1016/j.pragma.2016.12.005.

*Das R (2018) Populist discourse on a British social media patient-support community: The case of the Charlie Gard support campaign on Facebook. *Discourse, Context & Media* 24: 76–84. DOI: 10.1016/j.dcm.2017.11.005.

*Decock S and Depraetere I (2018) (In)directness and complaints: A reassessment. *Journal of Pragmatics* 132: 33–46. DOI: 10.1016/j.pragma.2018.04.010.

*Deschrijver C (2018) Defaulting to metalanguage: Financial concept negotiations in user comments as a case of mediatization. *Language@Internet* 16: article 3.

*Dominguez-Whitehead Y, Whitehead KA and Bowman B (2017) Confessing sex in online student communities. *Discourse, Context & Media* 20: 20–32. DOI: 10.1016/j.dcm.2017.05.002.

Fossheim H and Ingierd H (eds) (2016) *Internet Research Ethics*. Oslo: Cappelen Damm Akademisk.

*Gawne L and McCulloch G (2019) Emoji as digital gestures. *Language@Internet* 17: article 2.

Georgakopoulou A (2017) 'Whose context collapse?' Ethical clashes in the study of language and social media in context. *Applied Linguistics* 8(2–3): 169–189.

*Georgakopoulou A and Giaxoglou K (2018) Emplotment in the social mediatization of the economy: The poly-storying of economist Yanis Varoufakis. *Language@Internet* 16: article 6.

Giaxoglou K (2017) Reflections on Internet research ethics from language-focused research on web-based mourning: Revisiting the private/public distinction as a language ideology of differentiation. *Applied Linguistics* 8(2–3): 229–250.

*Giaxoglou K (2018) #JeSuisCharlie? Hashtags as narrative resources in contexts of ecstatic sharing. *Discourse, Context & Media* 22: 13–20. DOI: 10.1016/j.dcm.2017.07.006.

*Gibson W and Roca-Cuberes C (2019) Constructing blame for school exclusion in an online comments forum: Membership categorisation analysis and endogenous category work. *Discourse, Context & Media* 32: 100331. DOI: 10.1016/j.dcm.2019.100331.

Golder S, Ahmed S, Norman G, et al. (2017) Attitudes toward the ethics of research using social media: A systematic review. *Journal of Medical Internet Research* 19(6): e195.

*Gordon C and İkizoğlu D (2017) 'Asking for another' online: Membership categorization and identity construction on a food and nutrition discussion board. *Discourse Studies* 19(3): 253–271. DOI: 10.1177/1461445617701810.

*Gredel E (2017) Digital discourse analysis and Wikipedia: Bridging the gap between Foucauldian discourse analysis and digital conversation analysis. *Journal of Pragmatics* 115: 99–114. DOI: 10.1016/j.pragma.2017.02.010.

*Gruber H (2017) Quoting and retweeting as communicative practices in computer mediated discourse. *Discourse, Context & Media* 20: 1–9. DOI: 10.1016/j.dcm.2017.06.005.

Heider D and Massanari AL (2012) *Digital Ethics: Research & Practice*. New York: Peter Lang.

*Housley W, Webb H, Edwards A, et al. (2017) Membership categorisation and antagonistic Twitter formulations. *Discourse & Communication* 11(6): 567–590.

*Jaworska S (2018) 'Bad' mums tell the 'untellable': Narrative practices and agency in online stories about postnatal depression on Mumsnet. *Discourse, Context & Media* 25: 25–33. DOI: 10.1016/j.dcm.2017.11.002.

*Jucker AH, Hausendorf H, Dürscheid C, et al. (2018) Doing space in face-to-face interaction and on interactive multimodal platforms. *Journal of Pragmatics* 134: 85–101. DOI: 10.1016/j.pragma.2018.07.001.

*Kalim S and Janjua F (2019) #WeareUnited, cyber-nationalism during times of a national crisis: The case of a terrorist attack on a school in Pakistan. *Discourse & Communication* 13(1): 68–94.

*Karlsen AS and Scott KD (2019) Making sense of Starbucks' anti-bias training and the arrests of two African American men: A thematic analysis of Whites' Facebook and Twitter comments. *Discourse, Context & Media* 32: 100332. DOI: 10.1016/j.dcm.2019.100332.

*Koteyko N and Atanasova D (2018) Mental health advocacy on Twitter: Positioning in depression awareness week tweets. *Discourse, Context & Media* 25: 52–59. DOI: 10.1016/j.dcm.2018.04.007.

*Lawless M, Augoustinos M and LeCouteur A (2018) Dementia on Facebook: Requesting information and advice about dementia risk-prevention on social media. *Discourse, Context & Media* 25: 44–51. DOI: 10.1016/j.dcm.2018.01.011.

Lester JN, Muskett T and O'Reilly M (2017) Naturally-occurring data versus researcher generated data. In: O'Reilly M, Lester JN and Muskett T (eds) *A Practical Guide to Doing Social Interaction Research in Autism Spectrum Disorders (ASD): Communication, Discourse and Conversation Analysis*. London: Palgrave MacMillan, 87–116.

*Limatius H (2019) ''I'm a fat bird and I just don't care'': A corpus-based analysis of body descriptors in plus-size fashion blogs. *Discourse, Context & Media* 31: 100316. DOI: 10.1016/j.dcm.2019.100316.

Linabary JR and Corple DJ (2019) Privacy for whom? A feminist intervention in online research practice. *Information, Communication & Society* 22(10): 1447–1463.

Locher M and Bolander B (2019) Ethics in pragmatics. *Journal of Pragmatics* 145: 83–90.

*Lorenzo-Dus N and Di Cristofaro M (2018) 'I know this whole market is based on the trust you put in me and I don't take that lightly': Trust, community and discourse in crypto-drug markets. *Discourse & Communication* 12(6): 608–626.

*Lorenzo-Dus N and Izura C (2017) '"Cause ur special"': Understanding trust and complimenting behaviour in online grooming discourse. *Journal of Pragmatics* 112: 68–82. DOI: 10.1016/j.pragma.2017.01.004.

*Lorenzo-Dus N, Kinzel A and Di Cristofaro M (2020) The communicative modus operandi of online child sexual groomers: Recurring patterns in their language use. *Journal of Pragmatics* 155: 15–27. DOI: 10.1016/j.pragma.2019.09.010.

*Ludemann D (2018) /pol/emics: Ambiguity, scales, and digital discourse on 4chan. *Discourse, Context & Media* 24: 92–98. DOI: 10.1016/j.dcm.2018.01.010.

*Lundström R (2018) Spaces for support: Discursive negotiations of supporter positions in online forum discussions about suicide. *Discourse, Context & Media* 25: 98–105. DOI: 10.1016/j.dcm.2018.04.002.

*Lutzky U and Gee M (2018) 'I just found your blog'. The pragmatics of initiating comments on blog posts. *Journal of Pragmatics* 129: 173–184. DOI: 10.1016/j.pragma.2018.03.016.

*Lutzky U and Kehoe A (2017) '"Oops, I didn't mean to be so flippant"'. A corpus pragmatic analysis of apologies in blog data. *Journal of Pragmatics* 116: 27–36. DOI: 10.1016/j.pragma.2016.12.007.

*Magaña D and Matlock T (2018) How Spanish speakers use metaphor to describe their experiences with cancer. *Discourse & Communication* 12(6): 627–644.

*Matley D (2018) ''This is NOT a #humblebrag, this is just a #brag": The pragmatics of self-praise, hashtags and politeness in Instagram posts. *Discourse, Context & Media* 22: 30–38. DOI: 10.1016/j.dcm.2017.07.007.

*Merrill S and Åkerlund M (2018) Standing up for Sweden? The racist discourses, architectures and affordances of an anti-immigration Facebook group. *Journal of Computer-Mediated Communication* 23: 332–353. DOI: 10.1093/jcmc/zmy018.

*Musi E (2018) How did you change my view? A corpus-based study of concessions' argumentative role. *Discourse Studies* 20(2): 270–288. DOI: 10.1177/1461445617734955.

Nissenbaum H (2010) *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford: Stanford University Press.

*Oyadiji O (2020) Politeness in the interactions of selected Nigerian news- based virtual communities. *Discourse & Communication* 14(2): 175–201.

Page R, Barton D, Unger J W, et al. (2014) *Researching Language and Social Media; a Student Guide*. New York: Routledge.

Paulus TM and Wise AF (2019) *Looking for Insight, Transformation and Learning in Online Talk*. New York: Routledge.

Paulus T, Woods M, Atkins D, et al. (2017) The discourse of qualitative data analysis software (QDAS): Reporting practices of ATLAS.ti and NVivo users with implications for best practices. *International Journal of Social Research Methodology* 20(1): 35–47.

*Perrino S (2018) Narrative aftershocks: Digital retellings of an earthquake in Emilia-Romagna, Northern Italy. *Discourse, Context & Media* 25: 88–97. DOI: 10.1016/j.dcm.2018.04.001.

Pihlaja S (2017) More than fifty shades of grey: Copyright on social network sites. *Applied Linguistics* 8(2–3): 213–228.

*Pounds G, Hunt D and Koteyko N (2018) Expression of empathy in a Facebook-based diabetes support group. *Discourse, Context & Media* 25: 34–43. DOI: 10.1016/j.dcm.2018.01.008.

*Procházka O (2018) A chronotopic approach to identity performance in a Facebook meme page. *Discourse, Context & Media* 25: 78–87. DOI: 10.1016/j.dcm.2018.03.010.

*Recktenwald D (2017) Toward a transcription and analysis of live streaming on Twitch. *Journal of Pragmatics* 115: 68–81. DOI: 10.1016/j.pragma.2017.01.013.

*Reed DJ (2017) Performance and interaction on Soundcloud: Social remix and the fundamental techniques of conversation. *Journal of Pragmatics* 115: 82–98. DOI: 10.1016/j.pragma.2017.01.012.

*Ren W (2018) Mitigation in Chinese online consumer reviews. *Discourse, Context & Media* 26: 5–12. DOI: 10.1016/j.dcm.2018.01.001.

*Rizwan S (2019) National identity premises in Pakistani social media debate over patriotism. *Journal of Language and Politics* 18(2): 291–311. DOI: 10.1075/jlp.17020.riz.

Roberts LD (2015) Ethical issues in conducting qualitative research in online communities. *Qualitative Research in Psychology* 12(3): 314–325.

Rüdiger S and Dayter D (2017) The ethics of researching unlikeable subjects. *Applied Linguistics* 8(2–3): 251–269.

*Russo A (2019) Audience address on dating profiles: My desired-other and second person reference. *Language@Internet* 17(1).

*Ruzza C and Pejovic M (2019) Populism at work: The language of the Brexiteers and the European Union. *Critical Discourse Studies* 16(4): 432–448. DOI: 10.1080/17405904.2019.1605300.

*Sahota PKC and Sankar PL (2020) Bipolar disorder, genetic risk, and reproductive decision-making: A qualitative study of social media discussion boards. *Qualitative Health Research* 30(2): 293–302. DOI: 10.1177/1049732319867.

*Salzmann-Erikson M and Hiçdurmaz D (2017) Use of social media among individuals who suffer from post-traumatic stress: A qualitative analysis of narratives. *Qualitative Health Research* 27(2): 285–294. DOI: 10.1177/1049732315627364.

*Semino E, Demjén Z and Demmen J (2018) An integrated approach to metaphor and framing in cognition, discourse, and practice, with an application to metaphors for cancer. *Applied Linguistics* 39(5): 625–645. DOI: 10.1093/applin/amw028.

*Spilioti T (2019) From transliteration to trans-scripting: Creativity and multilingual writing on the internet. *Discourse, Context & Media* 29: 100294. DOI: 10.1016/j.dcm.2019.03.001.

Spilioti T and Tagg C (2017) The ethics of online research methods in applied linguistics: Challenges, opportunities, and directions in ethical decision-making. *Applied Linguistics* 8(2–3): 163–167.

Stevens G, O'Donnell V L and Williams L (2015) Public domain or private data? Developing an ethical approach to social media research in an inter-disciplinary project. *Educational Research and Evaluation* 21(2): 154–167.

Stommel W and Jol G (2016) Ethical considerations of secondary data use: What about informed consent? *Dutch Journal of Applied Linguistics* 5(2): 180–195. DOI: 10.1075/dujal.5.2.06jol.

*Van der Bom I, Paterson LL, Peplow D, et al. (2018) 'It's not the fact they claim benefits but their useless, lazy, drug taking lifestyles we despise': Analysing audience responses to Benefits Street using live tweets. *Discourse, Context & Media* 21: 36–45. DOI: 10.1016/j.dcm.2017.11.003.

*Veum A and Undrum LVM (2018) The selfie as a global discourse. *Discourse & Society* 29(1): 86–103.

*Vladimirou D and House J (2018) Ludic impoliteness and globalisation on Twitter: 'I speak England very best' #agglika_Tsipra, #Tsipras #Clinton. *Journal of Pragmatics* 134: 149–162. DOI: 10.1016/j.pragma.2018.05.005.

Williams M, Burnap P, Sloan L, et al. (2017) *The Ethics of Online Research: Advances in Research Ethics and Integrity* (ed K Woodfield). Bingley: Emerald Publishing Limited.

Willis R (2019) Observations online: Finding the ethical boundaries of Facebook research. *Research Ethics* 15(1): 1–17. DOI: 10.1177/1747016117740176.

Woodfield K (ed.) (2018) *The Ethics of Online Research*. Bingley, UK: Emerald Publishing Limited.

*Zappavigna M and Zhao S (2017) Selfies in 'mommyblogging': An emerging visual genre. *Discourse, Context & Media* 20: 239–247. DOI: 10.1016/j.dcm.2017.05.005.

Zimmer M (2010) "But the data is already public": On the ethics of research in Facebook. *Ethics and Information Technology* 12(4): 313–325.

Zimmer M and Kinder-Kurlanda K (2017) *Internet Research Ethics for the Social Age—New Challenges, Cases and Contexts*. New York: Peter Lang.