Content Analyses of User Comments in Journalism: A Systematic Literature Review Spanning Communication Studies and Computer Science

Julius Reimer, Marlo Häring, Wiebke Loosen, Walid Maalej & Lisa Merten

To cite this article: Julius Reimer, Marlo Häring, Wiebke Loosen, Walid Maalej & Lisa Merten (2021): Content Analyses of User Comments in Journalism: A Systematic Literature Review Spanning Communication Studies and Computer Science, Digital Journalism, DOI: 10.1080/21670811.2021.1882868

To link to this article: https://doi.org/10.1080/21670811.2021.1882868

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 31 Mar 2021.

View supplementary material
Submit your article to this journal
View related articles
View Crossmark data
Content Analyses of User Comments in Journalism: A Systematic Literature Review Spanning Communication Studies and Computer Science

Julius Reimera, Marlo Haringb, Wiebke Loosena, Walid Maaleb and Lisa Merten a

aLeibniz Institute for Media Research | Hans-Bredow-Institut (HBI), Hamburg, Germany; bDepartment of Informatics, University of Hamburg, Hamburg, Germany

ABSTRACT
Different disciplines have studied the content of online user comments in various contexts, using manual qualitative/quantitative or (semi-)automated approaches. The broad spectrum and disciplinary divides make it difficult to grasp an overview of those aspects which have already been examined, e.g. to identify findings related to one’s own research, recommendable methodological approaches, and under-researched topics. We introduce a systematic literature review concerning content analyses of user comments in a journalistic context. Our review covers 192 papers identified through a systematic search focussing on communication studies and computer science. We find that research predominantly concentrates on the comment sections of Anglo-American newspaper brands and on aspects like hate speech, general incivility, or users’ opinions on specific issues, while disregarding media from other parts of the world, comments in social media, propaganda, and constructive comments. From our results we derive a research agenda that addresses research gaps and also highlights potentials for automating analyses as well as for cooperation across disciplines.

KEYWORDS
User comments; journalism; qualitative content analysis; quantitative content analysis; automated content analysis; systematic literature review; computer science; interdisciplinary research

Introduction
In recent years commenting on journalistic stories has become one of the most prevalent participation options that news media offer their users on their websites and social media channels (Stroud, Scacco, and Curry 2016b). Not only are comment spaces increasingly widespread, so is the audience’s use of them, making the online discussion of news a popular way for people to engage with socio-political issues (Heise et al. 2014a, 2014b; Loosen et al. 2013a, 2013b; Reich 2011; Reimer et al. 2015; [154: Ruiz et al. 2011]; numbers in brackets indicate the studies analyzed, which are listed...
The sheer volume of comments flooding into a newsroom every day can be overwhelming (Braun and Gillespie 2011), especially when a large share is uncivil or “toxic” in tone and content (Chen and Pain 2017; Loke 2012). Unsurprisingly, a growing number of media organizations decided to shut down comment sections on their own websites and instead refer their audience to their accounts on social media such as Facebook or Twitter, where commenting is still possible (cf. the list compiled by Cherubini 2016). In general, the experiences of incivility, hate speech, and propaganda in comments appear to have disillusioned many journalists to the extent that they no longer think of comments as furthering public debate. On the contrary, journalists tend to regard them as a threat to deliberation ([95: Ksiazek, Peer, and Zivic 2015]), or at best, “a necessary evil […] to attract audiences” (Reich 2011, 103).

The relevance of user comments for journalism and public debate has invariably led to a wave of related research. Scholars have, for instance, investigated how journalists perceive comments and commentators (e.g. Heise et al. 2014a; Loke 2012), how media organizations (should) moderate discussions on their websites (e.g. Ziegele and Jost 2020), how comments influence users’ assessments of a news item or a media brand (e.g. Houston, Hansen, and Nisbett 2011; Kümpel and Springer 2016), or what motivates audience members to voice their opinion and engage in online news discussions (e.g. Heise et al. 2014a; Springer, Engelmann, and Pfaffinger 2015).

Researchers have also looked at the comments themselves: one strand of literature leverages user comments as a proxy for measuring public opinion on various issues (Naab and Sehl 2017) from climate change ([90: Koteyko, Jaspal, and Nerlich 2013]) and the financial crisis ([10: Baden and Springer 2014]) to breastfeeding ([72: Grant 2016]). Another strand of studies look at user comments for their own sake, for instance, by analyzing their inherent architecture and deliberative potential ([55: Freelon 2015]; [154: Ruiz et al. 2011]) or by focussing on (in-)civility in user comments ([27: Coe, Kenski, and Rains 2014]). Moreover, comment analyses can differ in terms of the methods applied: comments can be analyzed quantitatively or qualitatively in a manual process (Krippendorff 2013; Kuckartz 2014) as well as quantitatively in a (semi-)automated way (Boumans and Trilling 2016; Lewis, Zamith, and Hermida 2013).

Additionally, user comments are not only an object of investigation in the social sciences, but also in computer science and computer linguistics where they are used as a data source for gathering information on product requirements, innovative new features or issues (Maalej et al. 2016) and as empirical data for the development of new analytical methods (e.g. [39: de-la-Peña-Sordo et al. 2015]; [53: Foster et al. 2011]).

Due to this broad thematic, methodological, and disciplinary spectrum of user comment analyses with reference to journalism, it is difficult to obtain a comprehensive overview of which specific aspects have been analyzed and to identify under-researched topics or the limitations present in current work. This however appears desirable, if not necessary, in order to facilitate the advancement of knowledge in this area. This in turn is important as in the long run it can help journalists manage the flood of comments, identify, make use of, and respond to valuable and insightful statements among them, as well as generally drive back incivility, hate speech, and propaganda, and further constructive public debate.
Aim and Structure of the Study

Against this background, this study aims to build a bridge between the disciplines of communication studies and computer science and their respective stocks of content analyses of user comments referring to journalism. We focus on these two disciplines not only because they are the authors’ fields of expertise and are both particularly interested in user comments. More importantly, these two disciplines can be fruitfully brought together in this field of research: in communication studies, in which content analysis has been fundamentally developed as an empirical method, comments are often examined to understand their very content. Computer science, on the other hand, focuses more on the development of automated approaches and tools for analyzing large volumes of comments. This is particularly relevant given the above-mentioned challenges which the sheer volume of comments as well as the incivility, hate speech, and propaganda found in them pose to the journalistic profession and public discourse in general. We focus on user comments that occur in a journalistic context because it is a domain of particular societal relevance and central to public debates. It is also a traditionally prominent subject in communication research and—with the emergence of online media, social networks, and computational journalism—is becoming increasingly relevant for computer scientists as well (e.g. [67: Gonçalves et al. 2016]; [86: Kabadjov et al. 2015]; [166: Sood and Churchill 2010]; [189: Zhou, Zhong, and Li 2014]). Additionally, journalism’s topical breadth makes user comments in this domain particularly interesting as well as challenging when it comes to the development of automated approaches, as opposed to more homogeneous comments, for example, those found in app stores or on travel websites. Finally, we had to limit our selection to studies using the same empirical method. First, including studies with different methodological approaches would have increased the number of studies to analyze beyond a manageable amount. Second, it would have made coding too complex since different methods are characterized by different parameters. Their samples, for example, consist of different kinds of cases: comments, users, journalists, newsrooms. Not only would we have ended up with a lot of incomparable data, but also would a codebook that accounts for all this have been even more extensive than the one we used (see Appendix 3).

We compare the two disciplines in regard to their thematic and methodological foci to reveal gaps in current research and to establish the grounds for mutual inspiration and interdisciplinary cooperation that can advance the field of study as a whole. In particular, we aim to:

- provide an overview of the two disciplines’ bodies of research with respect to the different features scrutinized by user comment analyses and provide scholars with a well-informed starting point for their own studies;
- highlight qualitative work that can be used for developing quantitative, possibly automated, methods to investigate particular aspects of user comments;
- identify the potential for and limitations of automation of comment analyses, especially in communication studies, by highlighting particularly promising approaches that make use of machine learning;
- identify gaps in current research and under-investigated features of comments; and
- integrate these insights in an agenda for joint future research between communication scholars and computer science researchers.
To this end, we conducted a systematic literature review of content analyses investigating user comments in online journalism and identified what variables or constructs (e.g. opinions/sentiment, incivility, information that comments add to the original article) have, so far, been analyzed in what way (qualitatively/quantitatively; manually/ (semi-)automatically) and from which disciplinary perspective.

In the following section we first elaborate on the context of our research, describing the development, use, and management of user comments in journalism, placing a particular focus on findings from survey-, interview-, and observation-based studies, which we will use in the conclusion to complement and contextualize the insights gained from our systematic review of content analyses. After that, we explain the methodology of our review, including our codebook, the systematic search for studies to be analyzed, and their coding. We continue with the presentation of our results, placing a special focus on the comparison between the two disciplines. Finally, we conclude the paper with a discussion of the identified research gaps and the potentials for collaboration on user comments in order to develop the aspired agenda for joint research on user comments by computer science researchers and communication scholars.

User Comments in Journalism

In recent years, user comments in journalism have developed from being a mere sub-aspect of participatory journalism into a distinct field of research of its own, reflecting the “tension between professional control and open participation” (Lewis 2012) which journalism still struggles with: the early years of online journalism saw a growing number of news websites providing features for users to publicly comment on news stories, e.g. to not fall behind competitors who already offer comment sections, to strengthen users’ loyalty, or to obtain information about them and their preferences (Heise et al. 2014a, 2014b; Karlsson et al. 2015; Loosen et al. 2013a, 2013b; Reimer et al. 2015; Singer et al. 2011). But for a few years now, an increasing number of media outlets have shut down their comment sections, referring users who want to comment to their social media accounts, often because they fear the questionable tone of comments—the incivility or toxicity—may damage their brand and/or they lack the resources necessary to monitor and moderate the sometimes overwhelming amount of comments flooding into newsrooms every day (Bergström and Wadbring 2015; Cherubini 2016; Karlsson et al. 2015; Loke 2012). Other news organizations only delete hate speech, rude language, spam, propaganda, and other incidents of “dark participation” (Frischlich, Boberg, and Quandt 2019) as well as off-topic statements and do not pay too much attention to the rest of the comments that may be insightful ([167: Sood, Churchill, and Antin 2012]). Correspondingly, journalists’ attitudes towards comments appear to have changed: while at first, they were viewed as “a space for a new ‘deliberative democratic potential’” ([31: Collins and Nerlich 2015]), now they are often considered as a rather shallow and aggressive form of audience participation (Bergström and Wadbring 2015; Loke 2012), a threat to deliberation ([95: Ksiazek, Peer, and Zivic 2015]), or at best, “a necessary evil […] to attract audiences” (Reich 2011, 103).
Studies suggest that between a quarter and half of users have commented on a news story at least once, but that only a small share of them does so regularly or frequently (Heise et al. 2014b; Karlsson et al. 2015; Loosen et al. 2013a, 2013b; Purcell et al. 2010; Reimer et al. 2015; [154: Ruiz et al. 2011]; Springer, Engelmann, and Pfaffinger 2015; Ziegele et al. 2013). However, comments tend to be widely appreciated as the percentage of “lurkers”, i.e. users who don’t comment themselves, but read others’ comments, is usually much larger (Heise et al. 2014b; Karlsson et al. 2015; Loosen et al. 2013a, 2013b; Reimer et al. 2015). Lurkers engage with comments to gather more information on a particular story—e.g. to sense the climate of opinion—and for entertainment (Springer, Engelmann, and Pfaffinger 2015; [45: Diakopoulos and Naaman 2011]). The motives for active commenting, on the other hand, are more manifold: studies found them to relate to public debate and information (express an opinion, bring topics deemed important onto the agenda, correct errors and perceived misrepresentations in reporting, learn from dialogue with others), to identity and emotion management (present oneself, vent anger, spend time) as well as to a feeling of belonging through interaction with other users or journalists, whose participation in comment sections was also found to be a motivating factor (Aschwanden 2016; [45: Diakopoulos and Naaman 2011], Heise et al. 2014a, 2014b; Loosen et al. 2013a, 2013b; Meyer and Carey 2014; Reimer et al. 2015; Springer, Engelmann, and Pfaffinger 2015). By contrast, incivility and a lack of deliberative quality have been found to keep users from participating in comment sections (Engelke 2019; Weber 2014).

Possibly related, studies have found active users and their comments not to be representative of the general population and public opinion (Friemel and Dötsch 2015; Hölig 2018): commenting users, for example, tend to be overwhelmingly male as well as rather extroverted and narcissistic.

The amount of comments a news story receives seems to be independent of general audience interest, but depend on its topic and related news factors as most-clicked stories are not necessarily those with the most comments and commenters were found to prefer political content (Tenenboim and Cohen 2015; Ziegele et al. 2013) as well as stories characterized by proximity, impact, and frequency (Weber 2014).

While surveys suggest that discussions are an important objective for commenters (Heise et al. 2014a, 2014b; Loosen et al. 2013a, 2013b; Reimer et al. 2015; Springer, Engelmann, and Pfaffinger 2015; Ziegele et al. 2013), content analyses show that comments often lack the necessary interaction between users ([101: Len-Ríos, Bhandari, and Medvedeva 2014]; [153: Rowe 2015]; Taddicken and Bund 2010). Instead, Ruiz et al. ([154: 2011, 8]) describe them as a “dialogue of the deaf”, finding that comments mostly contain users’ personal opinions in the form of a reaction to the article and not so much to other comments. This is problematic because, as Weber (2014, 942) states, “the potentials for quality discourse emerge only […] when there is a certain degree of interactivity among the users’ comments”. However, some studies found that, at least in moderated comment sections, substantial, albeit smaller, portions of comments refer to other users’ statements, provide alternative perspectives or information on the commented story’s topic, give arguments for stated opinions, contain humour or point towards an error or misrepresentation in the story (Aschwanden 2016; Heise et al. 2014b; Loosen et al. 2013a, 2013b; Reimer et al. 2015; Witschge 2011).
One aspect related to the content of comments that is researched particularly often is their (in-)civility. Coe, Kenski, and Rains ([27: 2014]), for instance, found that the share of uncivil comments does not necessarily increase when the amount of incoming comments does. Instead, it increases when “weightier” topics are discussed. Correspondingly, studies found that the topics more strongly related to uncivil discourse include politics, society, crime and justice, disasters and accidents, the environment, and feminism (Diakopoulos and Naaman 2011; [58: Gardiner et al. 2016]). Moreover, Coe, Kenski, and Rains ([27: 2014]) show that article-inherent aspects, such as the sources cited, influence the level of civility in comments. Various studies also suggest that anonymity has a negative influence on civility and other indicators for comment quality ([54: Fredheim, Moore, and Naughton 2015]; [158: Santana 2014]).

Additionally, the article author’s gender is of importance as female journalists received more comments that had to be blocked than their male colleagues ([58: Gardiner et al. 2016]). In this regard, a whole strand of research shows how much journalists—especially those of female or non-binary gender, of colour, and/or with an immigrant background—suffer from hateful comments addressed to them. This is particularly important because it has been shown that the anticipation of hateful audience feedback can impair journalists’ work, and open hostility towards them can negatively affect the general public’s sentiment towards the profession as a whole (Obermaier, Hofbauer, and Reinemann 2018; Post and Kepplinger 2019).

The quality and civility of comments appear to differ depending on the platform they are posted on: in a case study of the Washington Post, Rowe ([152: 2015]) found that comments on the newspaper’s website show more deliberative quality than those on the outlet’s Facebook page. This is consistent with the findings from our own previous studies in which we found that comments from three German media’s Facebook pages contain less interaction with other comments, arguments to back stated opinions, or alternative perspectives on the discussed topic, but more personal attacks than comments to the same articles on the media’s websites (Heise et al. 2014b; Loosen et al. 2013b; Reimer et al. 2015). Furthermore, in surveys the outlets’ users found the comments on the media’s Facebook pages were not as good a supplement to the commented article, less informative and, generally, of lower quality than those on their websites (Heise et al. 2014b; Loosen et al. 2013b; Reimer et al. 2015).

Because of their overwhelming volume and their often-toxic tone, monitoring and moderating comments requires considerable resources (Diakopoulos and Naaman 2011; [71: Graham and Wright 2015]; Heise et al. 2014a, 2014b; Loosen et al. 2013a, 2013b; Reimer et al. 2015). Yet, a survey by Stroud, Alizor, and Lang (2016a) showed that most comment sections are moderated, and nearly all media organizations respond in one way or another to their audience via the comment section or on social media. While these tasks are often fulfilled by specific newsroom personnel such as social media editors, studies found that a considerable share of “ordinary” journalists also respond to commenters at least occasionally or even engage in discussions in comment sections (Chen and Pain 2017; Heise et al. 2014b; Loosen et al. 2013a, 2013b; Reimer et al. 2015; Stroud, Alizor, and Lang 2016a). In any case, the high workload and difficulty in identifying “good” or insightful comments within the vast amounts of user statements keep journalists from using the information in comments
for their work and from actually engaging with their audience (Braun and Gillespie 2011; [71: Graham and Wright 2015]; Heise et al. 2014a, 2014b; Loosen et al. 2013a, 2013b; Reich 2011; Reimer et al. 2015). The situation is further complicated by the multiplication of media channels—such as websites, social media, blogs—which journalists are active on and receive feedback through (Heise et al. 2014a; Kramp and Loosen 2018; Loosen et al. 2017; Neuberger, Nuernbergk, and Langenohl 2019). This is due to the different subgroups of the audience and “commenting cultures” represented on these platforms.

Against this background of challenges that comments pose to journalism and public discourse in general, a considerable share of research has focussed on automating analyses to help with the moderation of uncivil statements especially and, less frequently, to identify comments that are constructive or contain information potentially valuable to journalists or other users ([44: Diakopoulos 2015]; Loosen et al. 2017; Løvlie 2018).

In relation to uncivil comments, different moderation strategies have been found to have different effects: Ziegele and Jost (2020) show that factual responses raise other users’ willingness to participate in the discussion while sarcastic responses decrease the perceived credibility of the commented story and the news medium publishing it, but increase discussions’ entertainment value. Additionally, Stroud et al.’s ([171: 2015]) study suggests that if, instead of an unidentified staff member, a recognizable reporter responds to comments, the deliberative quality of comments increases. Ruiz et al. ([154: 2011, 482]) ascertain that a strict moderation style leads to a lower amount of comments, but also conclude that overall “the different solutions adopted (pre-/post-moderation, in-house/outsourced) do not seem to direct the quality of the debate in a clear direction”.

**Method**

To achieve the study goals, we conducted a systematic (or structured) literature review (Greenhalgh et al. 2004). Essentially, this is “a form of content analysis whereby the unit of analysis is the article” (Massaro, Dumay, and Garlatti 2015, n.p.). Such an approach appears particularly appropriate when there is a wide range of research on a subject (Petticrew and Roberts 2006). Here, systematic literature reviews give “an overview of the scope of existing research, the prevalence of the procedures used, and the identified problems” (Naab and Sehl 2017, 1257) as well as identify research gaps (Petticrew and Roberts 2006, 2). As such, systematic reviews are useful sources to consult “before embarking on any new piece of primary research” (Petticrew and Roberts 2006, 21; authors’ emphasis). The systematic procedure also reduces the risk “that seminal articles may be missed” (Massaro, Dumay, and Garlatti 2015, n.p.) and “minimize[s] bias through exhaustive literature searches” (Tranfield, Denyer, and Smart 2003, 209). This is of particular importance in areas such as user comment analyses where it is difficult to obtain a comprehensive overview of the large body of research in any other way. Consequently, the method stands and falls with how one searches for and selects the relevant literature and what aspects one analyzes. We will explain both in the following.
Literature Selection and Sample

Based on our research aims, we defined three criteria for studies to be included in our sample: they have to be 1) based on a manual quantitative or qualitative, or a (semi-)automated analysis of the content of 2) user comments that 3) refer to journalistic stories.

We then created a search string that combined synonyms for each of the above-mentioned inclusion criteria (see Appendix 1 for the development and resulting search string). In December 2016, we used this search string to search the titles, keywords, and abstracts of all entries in arguably the most comprehensive literature database in communication science (EBSCO Communication and Mass Media Complete) as well as in its counterparts in computer science (ACM Digital Library and IEEE Xplore). In addition, we consulted the reading lists of the Coral Project (2016; Diakopoulos 2016) and of our own research project ‘Systematic Content Analysis of User Comments for Journalists (SCAN4J)’. Due to the advanced internationalization of both disciplines, we can assume that researchers of all nationalities (also) publish in English-language publications. To broaden our view and complement the sample with comment analyses performed in other disciplines, we also searched four extensive multidisciplinary repositories (Springer Link, ScienceDirect, Web of Science, Google Scholar; see Appendix 1 for the list of repositories and exact search procedures).

The search resulted in a list of 2,219 potentially relevant studies. Based on title, abstract and keywords, in cases of doubt on the full text, we identified the 192 studies which meet all of the three above-mentioned inclusion criteria (see Appendix 2 for more details on the inspection of the initial search results and the final list of studies included).

Codebook and Coding Procedure

To review these studies, we developed a detailed codebook (see Table 1 for an overview and Appendix 3 for the complete codebook). At its core are categories to capture what comment-related properties the studies are concerned with. To develop these, we initially collected comment-related aspects based on the knowledge we had gained in previous research projects on audience participation in journalism and on

Table 1. Codebook overview.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Example (sub-)categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bibliographical information</td>
<td>Authors, publication year, discipline, etc.</td>
</tr>
<tr>
<td>Methodology</td>
<td>Comment analysis method applied, additional methods applied, features/algorithms in automated approaches, reliability/evaluation scores, etc.</td>
</tr>
<tr>
<td>Sampling</td>
<td>Media brands &amp; news stories comments refer to, number of analyzed comments, etc.</td>
</tr>
<tr>
<td>Construct categories</td>
<td></td>
</tr>
<tr>
<td>Quantitative aspects</td>
<td>Length of comments, number of comments per story, etc.</td>
</tr>
<tr>
<td>Kinds of content</td>
<td>Personal opinion/attitude, argument for opinion, media criticism, propaganda, etc.</td>
</tr>
<tr>
<td>Incivility</td>
<td>Offensive language, personal insults, racism, sexism, etc.</td>
</tr>
<tr>
<td>Addressees of comments</td>
<td>Other users, journalist, forum moderator, etc.</td>
</tr>
<tr>
<td>Emotionality</td>
<td>Anger, hatred, fear, surprise, humour, etc.</td>
</tr>
<tr>
<td>Readability</td>
<td>Sentence length, technical/foreign terms, etc.</td>
</tr>
<tr>
<td>Facticity</td>
<td>Correctness of facts stated in comments</td>
</tr>
<tr>
<td>Other variable/construct</td>
<td>(Open category)</td>
</tr>
</tbody>
</table>
comments mining in software engineering. Then we complemented this list based on a close examination of studies that we already knew met our three inclusion criteria and represent a variety of different perspectives on comment analyses: they refer to news media from a variety of countries (e.g. Bulgaria, Finland, Germany, Romania, the U.S., UK), focus on diverse news topics (e.g. politics, a plane crash, celebrities, data visualisations) as well as research interests (e.g. incivility, opinions, deliberation, media criticism, emotions, neo-Nazi propaganda) and come from different years (2008–2016) as well as disciplines (communication studies, computer science, linguistics, and political science). This was important to account for the diversity of perspectives on user comments as well as for potential shifts in interests over time. We stopped as soon as the three last studies examined did not investigate any new comment features, which was the case after having analyzed 15 studies in total (i.e. [35: Craft, Vos, and Wolfgang 2016]; [79: Hille and Bakker 2014]; [83: Hullman et al. 2015]; [95: Ksiazek, Peer, and Zivic 2015]; [102: Loke 2012]; [107: Macovei 2013]; [129: Neagu 2015]; [142: Pond 2016]; [144: Quinn and Powers 2016]; [152: Rowe 2015]; [164: Slavtcheva-Petkova 2016]; [169: Strandberg 2008]; [171: Stroud et al. 2015]; Taddicken and Bund 20102; [182: Van den Bulck and Claessens 2014]). Finally, we added potentially relevant aspects from survey and interview studies on comments in journalism (e.g. Heise et al. 2014a, 2014b; Loosen et al. 2013a, 2013b, 2017; Reimer et al. 2015; Springer, Engelmann, and Pfaffinger 2015). This way we are able to include variables that had not yet been the focus of content analyses, but whose investigation might be of particular interest to researchers, journalists, users, or protagonists of news stories. Examples of such variables include who is addressed in a comment (“addressees of comments”) or whether it contains propaganda (“kind of content: propaganda”). Through thematic clustering, we collectively and consensually grouped the identified objects of user comment analyses into seven “construct categories” with similar thematic focus and added an open residual category to document any “unforeseen” aspects analyzed in the sample studies.

The coding was carried out by five coders. As suggested by Lombard, Snyder-Duch, and Bracken (2002), eighteen random papers (approx. 9%) were coded by all coders. On this basis, we calculated Holstí’s coefficient for intercoder agreement ($r_H$), as proposed, for example, by Früh (2007). Of the 127 variables in our codebook 117 reached an acceptable to very good reliability score (58 variables with $r_H \geq 9$; 42 with $0.9 > r_H \geq 0.8$; 42; 17 with $0.8 > r_H \geq 0.7$). Of the 10 variables with reliability values of less than 0.7, we still included four due to their high level of relevance to our research questions. These are: “other methods applied in study”, “number of comments analyzed”, “kind of content: personal opinion/attitude”, “kind of content: other”. In these cases, we report the Holstí-score in a note.

**Results**

We first describe our sample in terms of the originating discipline of the studies, their authors, year of publication, and venue. Then we report which media, countries, languages, and platforms have been at the focus of comment analyses. Following this, we look at the methodology of the studies, including what aspects of comments have
been analyzed automatically using machine learning techniques. Finally, we turn towards the seven construct categories, i.e. the examined quantitative aspects and kinds of comment content as well as the incivility, addressees of comments, emotionality, readability, and facticity of postings. In accordance with our research aims, we place special emphasis on comparing the two fields of communication studies and computer science throughout this section.

It is worth noting that the coding and counting were done on the level of papers and not on that of research projects. Consequently, we also use the word “study” as a synonym of “paper”, and not of “research project”. The differentiation on paper level was the only one feasible because authors rarely state exactly what research project(s) the presented data originate(s) from. It is possible, therefore, that one paper reports on the results from more than one project as well as that the findings of one project are reported on in more than one paper.

**Publication Disciplines, Authors, Years, and Venues**

Nearly half of the studies in our sample can be attributed to communication science (92 studies = 47.9%). Papers from computer science and from other disciplines make up approximately a quarter each (52 studies = 27.1% and 48 studies = 25.0%, respectively).

The relatively small amount of studies from computer science is, in part, due to our inclusion criterion that a paper needed to present at least some results exclusively related to comments on journalistic content (see above). The most prominent other disciplines, as inferred from the publication venues in our sample, are health sciences (12 studies), language studies/linguistics (6), behavioural science/psychology (5) as well as political science, sociology, or ecology-related science (four each).

In sum, 139 different publication venues are represented in our sample, the vast majority of them (81.3%) with one paper only. This reflects the research fragmentation resulting from the thematic breadth of user comments that makes them interesting to a variety of disciplines. In view of the significantly larger number of communication studies in the sample, it is not surprising that the “top three” publication venues are: *Journalism* (eight studies), *Journalism Studies* (5), and *New Media & Society* (5). The top computer science venues are: the *ACM Conference on Computer Supported Cooperative Work and Social Computing* (CSCW) (four studies), the *International Conference on Social Informatics* (SocInfo) (4), and the *Annual ACM Web Science Conference* (WebSci) (2). The sample also comprises some journals one would not necessarily associate with analyses of user comments in online journalism, e.g. *Clinical Obesity*, the *Journal of Human Lactation*, *Crisis – The Journal of Crisis Intervention and Suicide Prevention*, *Stem Cell Reviews and Reports*, and the *Urban Water Journal*. The studies published in these and similar venues mostly investigated comments on a specific news topic that is of relevance to the respective field of research in order to ascertain users’ opinions on the topic.

The disciplinary divide becomes apparent in the publishing activities of the 433 authors in our sample: although their insights are of potential interest to more than one discipline, only eight of them (1.8%) have written for venues representing
different disciplines, and in only three such instances (0.7%), these were communication and computer science publications.

Across disciplines, comment analyses have gained popularity from 2003 to 2015 as Figure 1 shows. The rather low numbers for 2016, especially for computer science, may result from proceedings for conferences in that year, which had not yet been published.

Media Brands, Countries, Languages, and Platforms Covered

The 192 studies in our sample analyze comments referring to 295 different media brands. On average, each study looked at about four brands ($M = 3.9$, $n = 165$), the maximum being 35. However, more than half of the 165 studies that mention how many media they included analyzed comments from only one brand (51.5%).

The most prevalent brands in the sample are the New York Times and the Guardian (26 studies each), supposedly due to their international reputation as well as the convenient APIs of their websites. Other media organizations of particular prominence are the BBC (17 studies), the Washington Post (15), the Daily Mail (10), the British Daily Telegraph (9), and the Wall Street Journal (8). Especially the “top group” of the 21 brands that were included in studies four or more times shows a strong tendency towards broadsheet newspapers, whereas broadcasters (TV: 4, radio: 1), digital natives (3), and tabloids (2) are rarely the focus of research.

The media brands come from 43 different countries and all continents except Antarctica. However, comment analyses strongly focus on Anglo-American countries. Nearly half the studies are concerned with US or UK media (30.1%, 18.6%, $n = 183$). The next large category contains studies that look at media from more than one country (12.6%), including another five papers combining UK and US media only. Another
recurring country combination is Qatar and Saudi Arabia/United Arab Emirates for comparing comments related to Al Jazeera and Al Arabiya (three studies). The least attention has been devoted to African media, with only five studies.

The focus on Anglo-American media is reflected in the languages of the comments analyzed. In total, the studies examine comments in twenty-six different languages. Nearly two thirds of papers, however, are concerned with English comments (62.4%, \( n = 186 \)). A distant second is the group of studies that investigate comments in more than one language (5.9%), with only one of these eleven works not also including English comments. Comments in other languages than English are examined in less than five percent of the cases each.

A striking difference between the disciplines is that more widely spoken languages are analyzed in an automated manner more often (Arabic: four of six studies involve automated analyses; Spanish: 4 of 9; Chinese: 3 of 5; English: 33 of 116). This suggests that analysis software and lexica for less common languages might still be rare or immature.

In ninety percent of cases, the comments analyzed are gathered from the comment sections of the respective media’s websites, while comments from Facebook, Twitter and other social platforms play a much smaller role despite the attention they receive in the public discourse about online discussions (see Table 2). Comparing disciplines, we find that Facebook comments are examined even less in computer science than in communication studies, while for tweets it is the other way around. Similarly, only 6.6% of studies look at comments from more than one platform and only two studies include three platforms.

**Methodological Aspects**

Looking at the analysis methods applied in the studies, we realized an expected methodological divide between communication science and other disciplines, which predominantly use manual methods, and computer science in which automated analyses are by far the most prevalent (see Table 3). However, computational methods appear to be slowly becoming more prominent among communication scholars with the

| Table 2. Comment analyses by commenting platforms (multiple coding possible). |
| :-------------: | :-------------: | :-------------: | :-------------: |
| % &nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&n

| Table 3. Comment analysis methods applied (multiple coding possible). |
| :-------------: | :-------------: | :-------------: | :-------------: |
| % &nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&n


share of communications studies including automated approaches rising from ten percent in 2013 \((n=10)\) to 22.2\% in 2016 \((n=18)\).

Studies in which two or even all three different content analysis methods were applied, for instance when a qualitative pre-study is conducted to prepare a manual or automated quantitative analysis, make up less than a fifth of our sample (17.2 and 1.6\%, respectively). Multi-method designs are prevalent, though, since nearly half of the studies (47.4\%), in addition to analyzing comments, also investigate the corresponding journalistic stories or include interviews or surveys with users, journalists or comment moderators.\(^4\) Such approaches are significantly less common in computer science than in communication studies or other disciplines (26.9\%, \(n=52\) vs. 52.7\%, \(n=91\), \(\chi^2=12.52, df = 2, p<.01\)).

On the other hand, procedures and programmes for automated analyses appear to be created and tested almost exclusively in computer science. In that discipline, nearly all studies involving automated methods also elaborate on the development, implementation, and/or evaluation of software programmed especially for the analysis (90.9\%, \(n=44\)). In communication studies, this kind of work is virtually non-existent (14.3\%, \(n=7\)).

In our corpus, two thirds of automated analyses employ machine learning approaches (supervised: 46.3\%, unsupervised: 14.8\%, both: 5.6\%, \(n=54\)). Because one of our research objectives is to identify potentials for automating comment analyses, we further inspected these studies qualitatively. First, we grouped the different aspects of comments that have been examined in a (semi-)automated manner into six larger categories. (Note that the categories often overlap because aspects can be analyzed on their own as well as as one part of another category. Categorizing comments as on- or off-topic, for instance, can also be one step towards identifying troll comments.) We then identified particularly promising (semi-)automated approaches within each of the six categories that we briefly present now:

- **Sentiment:** Fifteen studies apply different procedures to determine the sentiment users express in their comment, i.e. their positive, neutral, or negative attitude towards a topic, usually that of the news story. Sentiment analysis is often used as one means among others (e.g., named entity recognition, part of speech tagging, vector space models) in a larger framework designed to achieve a higher goal. The respective sentiment analysis performance is not assessed individually. The approaches are also developed for different languages. For instance, Tumitan and Becker ([181: 2014]) test different procedures to predict Brazilian election results based on sentiments expressed in comments (in Brazilian Portuguese) before the polls. Kabadjov et al. ([86: 2015]) describe different approaches to summarizing forum discussions (in English and Italian) including argumentation mining and sentiment analysis.

- **Trolling, hate speech, and spam:** Five papers are concerned with identifying destructive user contributions including hate speech, spam, or troll comments. Supervised machine learning approaches reach high accuracy scores, but they also depend on pre-labelled training data. Contrarily, de-la-Peña-Sordo et al. ([39: 2015]; [40: 2014a]; [41: 2014 b]) employ semi-supervised machine learning and compression models (for comments in Spanish). Their procedures perform nearly as well as supervised approaches, while depending on less labelled data and being incrementally updateable.
On/off-topic: In seven studies, machine learning is used to detect whether or not a comment is related to the original article or fits into the discussion context. This is important for tasks such as discourse and argumentation analysis, troll detection, or forum moderation. Here, de-la-Peña-Sordo et al.’s ([40: 2014a]; [41: 2014b]) unsupervised approach of comparing the comments’ vector space model with that of the news story’s lead seems promising. This approach, developed for the Spanish language, was only part of a larger framework and its performance was not tested individually.

Discussion structure: Three studies try to determine the structure of the discussion between users in a forum thread. For instance, Schuth, Marx, and de Rijke ([160: 2007]) propose a method for the precise detection of (Dutch-language) comments referring to other comments. De-la-Peña-Sordo et al. ([39: 2015]) show that a Random Forest algorithm performs best in identifying whether a (Spanish-language) comment refers to the news story or to another comment.

Topics: Three studies seek to cluster topics discussed in comments or identify “hot topics”, i.e. themes that spark considerable discussion. Both supervised ([189: Zhou, Zhong, and Li 2014]) and unsupervised methods ([3: Aker et al. 2016]; [115: Meguebli et al. 2014]) have been used for these tasks, with the former, more labour-intensive approaches performing (expectably and) considerably better.

Diversity and anomaly: Six studies deal with assessing the diversity of comments or with detecting unusual user contributions. For instance, Giannopoulos et al. ([63: 2012]) seek to identify within an English-language discussion thread a subset of comments that are most heterogeneous with regard to the aspects of the news article they refer to and the sentiments expressed. These diverse comments, they argue, could be highlighted for other users in order to counter the risk of filter bubbles and echo chambers.

In the case of manual quantitative content analyses, researchers usually calculate inter- or intracoder reliability scores to account for the robustness of their method (Lombard, Snyder-Duch, and Bracken 2002). However, although most studies in our sample were published in high-quality academic journals, only half of the papers including a manual quantitative analysis provide such metrics (52.6%, n = 78). With less than a fifth, the share is even smaller for studies employing a qualitative method (16.1%, n = 87), which might be due to the fact that reliability scores are not undisputed in qualitative research. Similarly, only about a fifth of studies with a manual quantitative approach and a quarter of those including qualitative comment analyses report to have exercised consensual or peer coding (22.2%, n = 81; 24.7%, n = 93) which is a common measure for strengthening validity and reliability (Kuckartz 2014; Kurtanović and Maalej 2017; Schmidt 2004).

Number of Stories and Comments

The average and maximum number of news stories a study is concerned with seem extremely high: M = 6,569.3; Max = 200,000. However, this is due to some extreme outliers, the median is “only” 50 stories. Comparing methods reveals that the high mean
value is connected with automated analyses that on average deal with significantly more news stories than studies only using other methods \( (M_{\text{automated}} = 26,338.6, \ n = 34 \) vs. \( M_{\text{other}} = 824.3, \ n = 117; \) Welch-test: \( t = 2.53, \ p < .05 \)). Unsurprisingly, the result is similar when comparing computer science and communication research \( (M_{\text{comp}} = 26,797.6, \ n = 32 \) vs. \( M_{\text{comm}} = 398.4, \ n = 77; \) Welch-test: \( t = 2.52, \ p < .05 \)). Entirely qualitative analyses on average examine the comments of a much smaller number of news stories than other manual analyses with a purely quantitative approach; however, the difference while large is not statistically significant \( (M_{\text{qual}} = 71.5, \ n = 60 \) vs. \( M_{\text{quant}} = 2,118.0, \ n = 41 \)).

Concerning the actual number of comments analyzed, the high mean value \( (M = 11,946,993.3, \ n = 165) \) is largely due to seven studies that analyze extremely large amounts of comments from approximately 2.5 million up to 1.8 billion individual examples.\(^5\) The median is “only” 1,795 comments, and the minimum number of comments analyzed is only 25.

**Construct Categories: Aspects of Comments Analyzed**

We organized the aspects of comments that studies may investigate in categories of similar constructs. The first category is *quantitative aspects* which were investigated in nearly half of the studies (95 studies = 49.5%). Here, the most prevalent aspect was the number of comments per individual news story, media brand, platform, etc. (see Table 4). Other aspects repeatedly examined were the number of individual commentators engaged, how many comments each of them posted, and how long these comments were. We found only minor differences between the disciplines in regard to the percentage of studies including quantitative aspects.

The category most frequently researched comprises *particular kinds of content* the occurrence and/or nature of which were examined in nearly all studies in our sample (170 studies = 88.5%). As expected, personal opinions, e.g. on the topic of the related news story, are the most frequently researched aspect of this category (see Table 5). This includes sentiment analyses that categorize comments into positive, neutral, and negative opinions. User comments often seem to be used as a proxy to determine “public opinion” although commenting users are not representative of the general population (Naab and Sehl 2017).

Other kinds of content that studies often investigated include instances where users provide an argument for the opinion they state, what frames or perspectives on a topic comments contain, and whether users react to each other’s comments or merely

<table>
<thead>
<tr>
<th>Table 4. Quantitative aspects researched in comment analyses.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Number of comments per individual news story, media brand, platform, etc.</td>
</tr>
<tr>
<td>Number of individual commentators</td>
</tr>
<tr>
<td>Length of comments</td>
</tr>
<tr>
<td>Number of comments per commentator</td>
</tr>
<tr>
<td>Development of amount of comments over time</td>
</tr>
</tbody>
</table>
post their own thoughts. The prevalence of these aspects might be due to their assumed relation to the (deliberative) quality of user discussions which is a frequently researched topic (e.g. [55: Freelon 2015]; [154: Ruiz et al. 2011]).

Interestingly, computer science studies look at “constructive” content that might also be of interest for other users or journalists considerably less frequently than communication research (or other disciplines for that matter). This applies to media criticism that could help improve reporting as well as aspects that could result or be included in future stories, such as personal experiences, additional pro/contra arguments, new information and leads, or untouched frames and perspectives (Loosen et al. 2017). Similarly, studies involving automated content analysis rarely focus on these kinds of content.

More than a third of papers in the sample examined the incivility of comments (68 studies = 35.4%). This may be because it is negatively related to the (deliberative) quality of user debates and a much-discussed topic among practitioners (Loosen et al. 2013a, 2013b, 2017; Reimer et al. 2015; Ziegele and Jost 2020). By far the most researched forms of incivility are general hostility and personal insults (57.4%) as well as profanity, such as the use of swear words (55.9%). Hostility and personal insults are examined more often when automated methods are involved (61.1% vs. 42.9%), while more specific forms of hate speech are rarely a topic of automated analyses: sexism, racism, religious or political intolerance were all investigated in only one such study which was conducted by an interdisciplinary team of journalism practitioners including a communication scholar, a data scientist, and graphic editors ([58: Gardiner et al. 2016]). The picture regarding hate speech is similar for manual quantitative analyses. By contrast, if a study involves a qualitative approach, it is more likely to also look at these forms of hate speech (sexism: 16.1 vs. 2.8%; racism: 32.3 vs. 13.9%; religious: 16.1 vs 2.8%; political: 12.9 vs. 8.3%; n = 31 vs n = 36).

In our sample, forty-five studies, or 23.4 percent, dealt with emotions expressed in comments or their overall emotionality. The most frequently studied forms of emotion are irony, sarcasm, and cynicism, which are central to similar extents in both

<p>| Table 5. Kinds of content researched in comment analyses. |
|-------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>%</th>
<th>Total (n=170; 88.5% of corpus)</th>
<th>Communication studies (n=82)</th>
<th>Computer science (n=41)</th>
<th>Other discipline (n=47)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal opinion, attitude, evaluation, judgement, verdict</td>
<td>70.6</td>
<td>68.3</td>
<td>68.3</td>
<td>76.6</td>
</tr>
<tr>
<td>Argument for opinion</td>
<td>34.1</td>
<td>37.8</td>
<td>4.9</td>
<td>53.2</td>
</tr>
<tr>
<td>Frame, perspective, etc.</td>
<td>34.1</td>
<td>37.8</td>
<td>7.3</td>
<td>51.1</td>
</tr>
<tr>
<td>Reaction to other comment</td>
<td>25.3</td>
<td>31.7</td>
<td>24.4</td>
<td>14.9</td>
</tr>
<tr>
<td>On-/off-topic</td>
<td>23.5</td>
<td>25.6</td>
<td>29.3</td>
<td>14.9</td>
</tr>
<tr>
<td>Personal experience</td>
<td>22.4</td>
<td>30.5</td>
<td>4.9</td>
<td>23.4</td>
</tr>
<tr>
<td>Additional information, leads, material, etc.</td>
<td>16.5</td>
<td>24.4</td>
<td>7.3</td>
<td>10.6</td>
</tr>
<tr>
<td>Reference or link to external source</td>
<td>14.1</td>
<td>19.5</td>
<td>9.8</td>
<td>8.5</td>
</tr>
<tr>
<td>Media criticism</td>
<td>12.4</td>
<td>15.9</td>
<td>2.4</td>
<td>14.9</td>
</tr>
<tr>
<td>Mentioning of specific persons</td>
<td>11.8</td>
<td>11.0</td>
<td>9.8</td>
<td>14.9</td>
</tr>
<tr>
<td>Additional frame, perspective, etc.</td>
<td>10.6</td>
<td>14.6</td>
<td>2.4</td>
<td>10.6</td>
</tr>
<tr>
<td>Additional argument</td>
<td>7.6</td>
<td>11.0</td>
<td>4.9</td>
<td>4.3</td>
</tr>
<tr>
<td>Propaganda</td>
<td>0.6</td>
<td>1.2</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Other kind of content, e.g. questions, personal information about user, style/structure (rhetorical, interactional, etc.), meta-discourse on comments</td>
<td>38.8</td>
<td>39.0</td>
<td>34.1</td>
<td>42.6</td>
</tr>
</tbody>
</table>
communication and computer science (36.8% vs. 33.3%). This is interesting because automated analyses are considered error-prone when recognizing these particular variants of human emotion and language (e.g. [121: Moreo et al. 2012]; [175: Thelwall, Buckley, and Paltoglou 2012]), which might explain computer scientists’ attention to this topic. However, the absolute numbers of studies concerned with them are low (7 vs. 3). Positive emotions (pity/sympathy, surprise, curiosity/interest, love, happiness/joy, enthusiasm, humour: 43.2%) are researched nearly as often as those feelings with a negative connotation (anger, hatred, contempt/disgust/nausea, fear, sadness, shame/guilt: 52.3%).

A quarter of studies (48 studies = 25.0%) look at variables indicating who is addressed in a comment (Häring, Loosen, and Maalej 2018). Research examining whether specific users are addressed occurs most often (see Table 6). Much less frequently, scholars were interested in finding out if protagonists of the story or other people affected—e.g. obese people in studies on emotions towards obesity or weight loss surgery—were addressed. This also applies to individual journalists, the newsroom in general, or the general public. Interestingly, no computer science study looked at whether newsrooms, individual journalists, or community managers were addressed.

In nearly one tenth of our sample (18 studies = 9.4%), scholars also determine aspects of readability or comprehensibility of comments. In these studies, researchers identify if comments contain technical, foreign, or other terms that users might not understand (38.9%), how complex and long the sentences in it are (38.9% and 33.3%, respectively), and if there are typos or other errors (22.2%).

Only twelve studies (6.3%) are concerned with checking the facticity of comments, i.e. they somehow check if comments contain factual statements or not. Interestingly, of these studies only one comes from computer science.

Finally, around eleven to thirteen percent of studies across disciplines also examined comments in relation to a variable we could not attribute to one of our construct categories, for instance the location attached to a post.

### Table 6. Addressees of comments investigated.

<table>
<thead>
<tr>
<th>%</th>
<th>Total (n = 48; 25.0% of corpus)</th>
<th>Communication studies (n = 33)</th>
<th>Computer science (n = 5)</th>
<th>Other discipline (n = 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific user(s)</td>
<td>70.8</td>
<td>66.7</td>
<td>80.0</td>
<td>70.0</td>
</tr>
<tr>
<td>Protagonist(s) of journalistic story or other external figure(s)</td>
<td>31.3</td>
<td>30.3</td>
<td>40.0</td>
<td>30.0</td>
</tr>
<tr>
<td>Individual journalist(s)</td>
<td>25.0</td>
<td>24.2</td>
<td>–</td>
<td>40.0</td>
</tr>
<tr>
<td>Newsroom</td>
<td>20.8</td>
<td>21.2</td>
<td>–</td>
<td>30.0</td>
</tr>
<tr>
<td>Audience as a whole/general public</td>
<td>18.8</td>
<td>15.2</td>
<td>40.0</td>
<td>20.0</td>
</tr>
<tr>
<td>Community manager(s)/forum moderator(s)</td>
<td>4.2</td>
<td>6.1</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Other addressee(s)</td>
<td>16.7</td>
<td>24.2</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

**Conclusion: An Agenda for Future (Interdisciplinary) Research**

This article presents a systematic literature review of 192 analyses of user comments with reference to journalism identified in a systematic database search. It shows that user comments referring to journalistic stories on various topics—from elections, to climate change, to breastfeeding—are analyzed in order to answer a wide range of
research questions—for example, to determine users’ opinions on specific topics or the deliberative quality of online discussions—through the employment of manual quantitative or qualitative as well as (semi-)automated methods. In some cases, comments are the empirical data required for developing and testing automated approaches.

What all these studies have in common is that ultimately, they aim to understand who communicates about what and how in user comments. In view of the inundation of user comments posted every day and their diverse nature, there is an evident urge to (partially) automate their analysis. This is true for both academia and the journalistic field, and also for all other domains in which users share comments, such as streaming services, shopping websites, or app stores (e.g. Maalej et al. 2016). Whether automated analyses are created to support comment moderation or for scientific research, we consider it useful that they are designed collaboratively by the fields of social science and computer science. For this reason, rather than just summarizing our results in this conclusion we draw on them to develop a joint agenda for future research that we believe should be addressed in close collaboration by communication studies and computer science.

Our findings show that interdisciplinary cooperation is still rare: for instance, only three of the 433 authors represented in the sample (0.7%) have published in both communication and computer science venues. However, our results show that computer scientists and communication scholars obviously share many research interests as nearly all aspects of user comments have been investigated separately in both disciplines.

This suggests considerable potential for cooperation in theoretical, methodological, and research-practical terms: even though the use of computational comment analysis is slowly gaining importance in communication science, the tools for these analyses are predominantly developed by the field of computer science. However, particularly precise procedures can probably be best produced if they are guided by a clearly contoured research question and a profound understanding of the associated phenomena. Here, communication science offers a rich stock of elaborate theories and empirical findings which can help identify and formulate pressing research questions as well as operationalize the relevant aspects methodologically: communication scholars can contribute their expert knowledge of journalism, public debate, opinion formation, and other communicative phenomena to precisely tailor methodological approaches to the complexity of the aspect under consideration as well as to appropriately interpret the sometimes unclear results of, for instance, topic models. Similarly, the many existing communications studies investigating certain comment aspects qualitatively can be consulted to assess if and how their analysis processes can be automated or, at least, how particular aspects of interest may be operationalized for computational approaches. The automated detection of hate speech, for instance, could be improved to differentiate between its different forms—racism, sexism, religious and political intolerance, etc.—like qualitative analyses do. Another research desideratum would be to develop software that can identify comments with “constructive” content that is likely to be of interest to other users or even of use to journalists: media criticism that could help improve reporting as well as aspects that could result or be included in future stories on the topic at hand, such as users’ personal experiences related to the topic, additional pro/contra arguments, new information and leads, or untouched frames and perspectives (Loosen et al. 2017).
Second, we found that it is significantly less common in computer science studies to combine the analysis of comments with other methods and data, such as surveys of users or journalists. In other words, computer scientists often look at comments in isolation. By contrast, as we show in the overview on existing research, communication scholars are often also interested in the relationships between comments and other factors, such as the article commented on or the attitudes and motivations of commenting users. This may represent an important research gap within computer science because including such possibly independent or moderating variables factors could prove effective in refining automated approaches for the analysis of comments. An example could be that computer science researchers have so far only looked at whether a comment addresses other users or protagonists of the news story, but have not taken into account that comments may also be intended for the journalist who authored the piece, the newsroom as a whole or the forum moderators, which is likely to make a difference in terms of the comment’s content and moderation.

Third, content analysis is considered one of the core methods for communication science and has been developed decisively in this field of study (Krippendorf 2013). The acquired methodological expertise could help generate better training data and truth sets for supervised machine learning, i.e. manually labelled data used to train an automated approach and to evaluate its performance. Computer science papers seldomly make transparent how the variables under consideration were theoretically derived and operationalized in order to create this training and benchmark data. Additionally, the task of coding is regularly assigned to untrained annotators of services like Amazon’s Mechanical Turk (e.g. [22: Cheng, Danescu-Niculescu-Mizil, and Leskovec 2015]; [61: Ghorbel 2012]; [167: Sood, Churchill, and Antin 2012]) with no reliability scores reported to determine coding accuracy (a notable exception being, for example, the study by Carvalho et al. [19: 2011]). More rigorous procedures could generate more valid and reliable training and benchmark data and help improve an automated approach.

Fourth, we found a slight difference in regard to the disciplines’ predominant epistemological interests, which also speaks for their complementarity. Researchers from the field of computer science often seek to detect if a comment contains a certain aspect: does it, for example, include racism, or criticism of the news story? Communication scholars on the other hand usually want to analyze how they do so: what kind of racist statements are made (and which potential others are not)? What exactly is being criticized: the selection of the topic (and disregard of others), the presentation, alleged bias, etc.? Consequently, in interdisciplinary projects, computer scientists may determine which comments to include in a sample in order to study a particular phenomenon while communication scholars provide insights into the very nature of that phenomenon.

In terms of more concrete research gaps that should be addressed in the future, we observe that research tends to concentrate on comments referring to Anglo-American newspaper brands while simultaneously disregarding others, especially African media. While this may, in part, be due to our own focus on English-language studies, there certainly is a need for further investigation, as commentary cultures, the diversity of opinions expressed in comments, and the (in-)civility of discourse, etc. are likely to vary in different countries with different political and media systems. In addition, many
automated methods have been developed for the English language so that there are still considerable gaps with respect to other languages that are less common.

Comment analyses are seldomly concerned with the positive or useful aspects of user contributions. This is especially true for studies in computer science and research involving automated analyses in particular. There is more work needed, therefore, on automatically identifying “constructive content”, e.g. to help journalists “mak[e] sense of user comments” (Loosen et al. 2017) and use them for newwork. The highlighted valuable comments could serve as positive examples for other users to follow and consequently improve the overall discussion. Moreover, as shown in the research overview, this would likely enhance users’ perception of and loyalty to the media brand. Additionally, more constructive comments could encourage journalists to join the debate, which, as other studies have found, can in turn motivate lurkers to participate actively as well.

Nevertheless, the strong focus on negative aspects—most notably on incivility and hate speech—is justified by their exceptional relevance for journalism as illustrated in the research overview. In this regard, automated analyses developed jointly by communication scholars and computer scientists have great potential: first, they can help overburdened moderators manage the inundation of comments, for example by identifying comments that likely have to be moderated or deleted. Existing research suggests that this can have a positive effect on the perception of the media brand and also raise the overall quality of the comment section. This in turn could grant moderators the time to also check comments on their medium’s profiles on Facebook and other social media and improve debate there. Second, automated approaches can help further test two aspects which, as shown in the research overview, have been found to affect the civility and deliberative quality of user comments: different moderation strategies and diverse designs of comment section interfaces on the journalists’ and moderators’ as well as on the users’ side (Loosen et al. 2017; Løvlie 2018; Schneider and Meter 2019). With the help of automated approaches, both moderation strategies and interface designs could be tested in field experiments in real comment sections with large amounts of comments and evaluations carried out in real-time. Computer science researchers often have the know-how when it comes to developing the interfaces to test. Communication scholars on the other hand cannot only contribute their domain expertise on how moderation strategies and interfaces that improve user discussions might look, but also point out the typical problems of automated procedures and examine the results with relation to these: their difficulties related to “accounting for context, subtlety, sarcasm, and subcultural meaning” as well as their tendency to “lay[] the burden of error on underserved, disenfranchised, and minority groups” (Gillespie 2020, 3).

All of the above has the potential to reduce incivility and strengthen the constructive voices in comment sections and, thereby, motivate more journalists and users to participate, which is crucial as “the potentials for quality discourse emerge only when a substantial amount of users participate in commenting on a news item” (Weber 2014, 942). This way, the deliberative potential theoretically associated with user discussions could be promoted in practice.

Despite the focus on incivility, some negative aspects of user comments apparently receive too little attention. There are, for instance, hardly any comment analyses focusing on identifying propaganda or determining if comments are based on facts. In view
of the current debates about these kinds of “dark participation” (Frischlich, Boberg, and Quandt 2019), these represent fundamental research desiderata.

Instead we see a pervasiveness of studies examining the occurrence or nature of personal attitudes or sentiments expressed by users. This shows that researchers—just like lurkers in comment sections—often use comments as a “proxy” to fathom “public opinion” on a particular issue. Based on our sample, this is the case especially in disciplines other than communication or computer science. This may, of course, be problematic due to the lack of representativeness of active users and their comments in relation to the general population (e.g. Friemel and Dötsch 2015; Hölig 2018)—a fact that communication scholars can point out when working across disciplinary boundaries.

A particular shortcoming of comment analyses, so far, is the near-complete focus on user posts in comment sections and the small number of studies investigating comments made on social media, not to mention comparative analyses of comments from different platforms and media brands which also attract different groups of users. This is unsatisfying because of the outstanding real-world relevance of these points: first, as outlined in the research overview, most media also maintain accounts in social networks, especially Facebook, or have even “outsourced” comments to them entirely. Second, there is much debate among practitioners about how much comments on platforms differ in tone and “quality” (Loosen et al. 2017; Reimer et al. 2015), something that is very much worth scrutinizing empirically; and third and most importantly, social networks have been discussed as a main distribution channel for disinformation, and it is highly relevant that we empirically determine what role comments play in framing and (re)interpreting “real and true” as well as “fake and false” news, for example when someone shares a journalistic post with their own introductory comment. Automated approaches that process far more content appear to have the potential to scale up comment analyses and, in so doing, also advance cross-platform comparisons.

Through our literature review we identified particularly promising approaches that can be adopted and adapted in the contexts discussed above in order to detect trolling or spam, “hot topics”, and exceptional statements as well as determine the sentiments, relatedness to the story’s topic, discussion structure, and diversity of comments.

Our study, of course, has some limitations. For instance, since we collected our data at the end of 2016, other relevant studies have been published; however, due to the considerable effort required, we were unable to research and code these as well. Furthermore, this review does not consider studies that also analyze comments to news stories (besides other content), but do not present the journalism-related results separately (see Appendix 2). Additionally, we have not included papers that solely report on the development and testing of a software tool for automated analyses. As a consequence, computer science papers and those studies that make use of automated methods may be generally underrepresented in our sample. As a result, some relevant tools and procedures may not be included in this review. Of equal importance, as this article is in itself a manual quantitative content analysis, it provides only little insight into what the studies in our sample actually discovered about the comment properties they focus on.

Despite these points, our review produced valuable insights. We identified underresearched aspects of user comments, hopefully point researchers towards studies from either discipline upon which they can build their own research, and we were
able to develop a research agenda of particularly pressing research gaps and how these could be addressed jointly by communication studies and computer science. We would like to see these efforts be considered as the foundations of a bridge that seeks to overcome the disciplinary divide in a bid to meaningfully and productively advance the analysis of user comments in journalism.

**Notes**

1. The literature database and further information can be found on the project website: https://scan.informatik.uni-hamburg.de
2. This study meets our inclusion criteria but was excluded from the sample because it was not found through our systematic research and is in German.
3. There are two ahead-of-print articles in our sample ([4: Al-Rawi 2016]; [93: Ksiazek 2016]), for which we coded the year of online publication as the year of publication.
4. In the reliability test, this variable only had a Holsti-score of \( r_H = 0.62 \).
5. In the reliability test, this variable only had a Holsti score of \( r_H = 0.56 \).
6. In the reliability test, this variable only had a Holsti-score of \( r_H = 0.53 \).
7. In the reliability test, this variable only had a Holsti-score of \( r_H = 0.69 \).

**Acknowledgements**

The authors would like to thank the peer reviewers for their valuable feedback on previous drafts of this article. Our special thanks go to Volodymyr Biryuk, Hannah Immler, Anne Schmitz, and Louise Sprengelmeyer for their great coding efforts.

**Disclosure Statement**

No potential conflict of interest was reported by the authors.

**Funding**

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: this work is partly funded by the Hamburg programme Ahoi Digital (project “Forum 4.0”) as well as the German Federal Ministry of Education and Research under grant 01UG1722AX.

**ORCID**

Julius Reimer [http://orcid.org/0000-0002-5961-909X](http://orcid.org/0000-0002-5961-909X)
Marlo Häring [http://orcid.org/0000-0002-2551-5065](http://orcid.org/0000-0002-2551-5065)
Wiebke Loosen [http://orcid.org/0000-0002-2211-2260](http://orcid.org/0000-0002-2211-2260)
Walid Maalej [http://orcid.org/0000-0002-6899-4393](http://orcid.org/0000-0002-6899-4393)
Lisa Merten [http://orcid.org/0000-0002-7624-2490](http://orcid.org/0000-0002-7624-2490)

**References**


Appendix 2

Inspection of Initial Search Results and Final List of Studies Analysed

1. Inspection of Initial Search Results and Identification of Actually Relevant Studies

The first two columns of Table A1 show the numbers of initial search results by the sources searched, totalling 2,219 potentially relevant studies (see Appendix A for details on the search procedure). Of these we first excluded:

- forty-nine publications from before 1999 as they are highly unlikely to be concerned with online comments;
- eight publications in languages other than English due to a lack of language proficiency;
- eleven publications that could not be found or obtained from databases or local libraries;
- eleven monographs and theses because these publication types are not provided systematically by all repositories and too voluminous to be analysed in full.

Each of the remaining studies was assigned to one of six researchers who, based on title, abstract and keywords, decided whether or not the study met the three inclusion criteria. In cases of doubt, we checked the full text and decided consensually.

The search string was designed to be as inclusive as possible so that we did not overlook any relevant papers. This, however, also means that it produced a large number of false-positive search results that we needed to identify and exclude to build the final sample: for instance, articles whose abstracts contain the three terms “journalism,” “content analysis,” and “comments” often do not analyse user comments but journalistic commentary.

We also excluded analyses of mixed material that mingle comments on journalistic stories with other web content if they do not report results for journalism-related comments separately. For example, Twitter studies that examine tweets with certain hashtags often include tweets related to journalistic stories, but not exclusively.

The third column of Table 1 shows the results of the inspection, totalling 192 studies actually relevant for the review. A list of the 192 studies can be found below.

Table A1. Overview of studies analysed by sources searched.

<table>
<thead>
<tr>
<th>Database, journal, reading list searched</th>
<th>Search results/potentially relevant studies (without duplicates)</th>
<th>Final corpus/relevant studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBSCO</td>
<td>1,065</td>
<td>66</td>
</tr>
<tr>
<td>ACM</td>
<td>695</td>
<td>18</td>
</tr>
<tr>
<td>IEEE</td>
<td>107</td>
<td>13</td>
</tr>
<tr>
<td>Springer Link</td>
<td>40</td>
<td>21</td>
</tr>
<tr>
<td>Web of Science</td>
<td>71</td>
<td>41</td>
</tr>
<tr>
<td>ScienceDirect</td>
<td>203</td>
<td>8</td>
</tr>
<tr>
<td>Google Scholar (first 200 search results)</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>The Coral Project reading list</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>(78 entries)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Literature database of research project</td>
<td>16</td>
<td>3</td>
</tr>
<tr>
<td>“SCAN4J” (280 entries)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2,219</td>
<td>192</td>
</tr>
</tbody>
</table>
List of Studies

Case ID


50. Fan, Wen, Shutao Sun, and Guohui Song. 2011. “Sentiment Classification for Chinese Netnews Comments Based on Multiple Classifiers Integration.” 2011 Fourth International Joint Conference on Computational Sciences and Optimization, 829–834. DOI:10.1109/CSO.2011.239


69. Graham, Todd. 2010. “The Use of Expressives in Online Political Talk: Impeding or Facilitating the Normative Goals of Deliberation?” International Conference on Electronic Participation, 26–41. DOI:10.1007/978-3-642-15158-3_3


German Federal President’s Affair (2011/2012).” The European Communication Research and Education Association’s (ECREA) 4th European Communication Conference (ECC), Istanbul, October 24–27.


181. Tumitan, Diego, and Karin Becker. 2014. “Sentiment-Based Features for Predicting Election Polls: A Case Study on the Brazilian Scenario.” 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 126–133. DOI:10.1109/WI-IAT.2014.89


