Counter narratives against antisemitic hate speech. A comparative study of French and British online comments sections

Laura Ascone, Karolina Placzynta (Centre for Research on Antisemitism (ZfA)) & Chloé Vincent (ZfA, Universiteit Gent)

The number of laws recently enacted across Europe are evidence of the governments' willingness to both limit and counter hate speech. On a micro-level, online comments sections show that some users try to counter hate speech as well, namely by taking part in a discussion and/or reporting hate content. This paper explores the way web users counter antisemitic hate speech in online comments sections triggered by ten domestic and international discourse events, as reported by British and French media. A total of 48,251 web comments were collected and analysed adopting a mixed methods approach. The quantitative analysis sheds light on both the proportion and evolution of antisemitic and counter speech comments in the different sub-corpora. A deeper qualitative analysis allowed us to investigate the link between the stereotypes expressed in the antisemitic comments and those responded to in counter speech comments. We also discuss possible reasons for the overarching patterns revealed by the data, namely the differences between types of discourse events and the language communities.

Suggested citation:

Keywords:

counter speech, antisemitism, mainstream media, comments sections, mixed-method approach

„ON AVERAGE, THE UK DATA CONTAINS MORE COUNTER SPEECH THAN THE FRENCH; HOWEVER, THE SHARE OF COUNTER SPEECH IN THE UK DOES NOT INCREASE WITH ANTISEMITIC COMMENTS AS MUCH AS IN FRANCE."

Laura Ascone, Karolina Placzynta & Chloé Vincent

Introduction

The sheer number of laws recently enacted across Europe (among others, the NetzDG in Germany or the Loi Avia in France) are evidence of the willingness to both limit and counter hate speech. On a micro-level, online comments sections show that some users try to counter hate speech as well reporting hate content and/or taking part in a

> **The sheer number of laws recently enacted across Europe are evidence of the willingness to both limit and counter hate speech.**

discussion. This paper explores the latter, concentrating on antisemitic hate speech. Counter speech, or a counter narrative, is understood here as a discourse countering, in an explicitly antagonistic way (Mouffe 2010), what has been stated elsewhere, e.g. in other comments within the online comment thread under analysis or reported in a media article on which web users are commenting. Here, we considered a counter narrative to be only the comments explicitly countering antisemitic discourse: more precisely, antisemitic stereotypes and tropes.

To evaluate these patterns, the analysis was conducted on comments taken from the websites and social media platforms of major media outlets in the United Kingdom and France. The corpus was built in the context of the transnational and interdisciplinary research project "Decoding Antisemitism: An AI-driven Study on Hate Speech and Imagery Online", whose goal is to investigate the content, structure and frequency of antisemitism in online spaces (Becker et al. 2021).

After presenting the corpus and the methodology, attention will be paid to the proportions of both antisemitic and counter speech comments in the two language communities under investigation, and to the link between the discourse event and the amount of counter speech. This quantitative investigation will be followed by a qualitative analysis of the connections between the antisemitic and counter speech comments. Furthermore, the impact of the discourse event on the way web users counter antisemitic comments was also investigated.

Research background and design

Contemporary antisemitism in online spaces

The advent of interactive web, and of social media in particular, has shifted the communication dynamics on the internet. As web users can often enjoy relative anonymity, the views they articulate may be more extreme than those expressed offline (Monnier and Seoane 2019). Additionally, communication often happens within filter bubbles made up of like-minded people, creating the echo chamber effect, confirming web users' cognitive biases, and making extreme views seem closer to

the norm (Cinelli et al. 2021, Strippel et al. 2023). This can be further accelerated if a controversial statement comes from a well-known figure, for example from the field of politics or entertainment, especially if they have considerable influence and following online. As a result, boundaries of what is acceptable to say in public are pushed ever further. This phenomenon does not exclude hate ideologies, and in the case of antisemitism it seems to be on the rise, especially in its implicit forms (Schwarz-Friesel and Reinharz 2017).

In our analysis, we have often seen that a spike in the number of antisemitic comments is typically triggered by the news coverage of global or local events in mainstream news. The resulting debate in online comments, though moderated, inevitably crosses the line of legitimate critique and spills into the territory of antisemitic ideation, and can target Jewish or Israeli people, Israel itself, as well as non-Jewish actors – including other web users.

> **In our analysis, we have often seen that a spike in the number of antisemitic comments is typically triggered by the news coverage of global or local events in mainstream news.**

Countering hate speech online

There is rich literature on discourses openly, or even antagonistically countering hate speech from another person or group (Mouffe 2010) and attempting to prevent or mitigate its potential harm (Cepollaro et al. 2022). Counter speech often takes the form of an activity planned and ran by organisations dedicated to fighting hate, but it can also be spontaneously actualised by individuals. Such citizen-generated counter speech is certainly visible in comments sections of the mainstream media and social media platforms. Online comments countering antisemitism vary from straightforward disagreements, to more elaborate structures that show both knowledge of classic or modern antisemitic concepts and a range of argumentative strategies. Nevertheless, measuring their efficacy and uncovering the trends that may be governing counter narratives remains difficult without extensive studies and large datasets (Garland et al. 2022).

Methodological framework

The findings presented here are based on the analyses conducted in the course of the project *Decoding Antisemitism: An AI-driven Study on Hate Speech and Imagery Online*, which aims to build a comprehensive picture of contemporary antisemitic discourse through the examination of the leading media in the UK, France and Germany. In the qualitative analysis stage, the data is examined using content analysis software, which allows for clear, consistent annotation of the downloaded comment threads, and a quick collation of annotation results. The comments are annotated according to a classification system developed by the project team – a list of over a hundred conceptual

and linguistic categories, based on the IHRA definition of antisemitism[1] and a pragmalingustic approach (Mayring 2015), which takes into consideration the impact of the context and co-text on the utterance (cf. Chapelan et al.).

The quantitative analysis is based on the annotation produced in the qualitative analysis stage. In order to evaluate the amount of counter speech in the online discussions, we consider the proportion of comments labelled as counter speech for each comment thread. As mentioned above, we considered a counter narrative to be only the comments explicitly countering antisemitic discourse, more specifically antisemitic stereotypes and tropes. We use the programming language R (package lme4, Banta et al. 2010) to build linear regression models with the proportion of counter speech comments as the dependant variable.

> **As mentioned above, we considered a counter narrative to be only the comments explicitly countering antisemitic discourse, more specifically antisemitic stereotypes and tropes.**

While both stages were rigorously carried out by the research team, there is, of course, scope for error. Potential limitations will have resulted from the different readings of a single comment; we tried to mitigate this with frequent discussions about the classification system and regular validation. Furthermore, if a comment seemed to have more than one interpretation, we erred on the side of caution and chose the non-antisemitic reading, which may have resulted in false negatives. We did not investigate the identity of web users, instead focusing on the possible impact of the comment itself, and on the overall picture of antisemitic discourse it contributed to.

Collecting the data and building the corpus

The data was collected from news outlets positioned within the political mainstream of the UK and France, using their websites and official media accounts as the source of comment threads, collected with a custom-designed data crawling tool. Although the content had already been moderated by either human or automatic moderators, it still contained a significant number of antisemitic comments, often expressed implicitly.

The web comments were posted in response to ten discourse events, both domestic and international, covered by British and French media in 2021 and 2022. For the UK we analysed reactions to media reports of the claims made by Professor David Miller, who alleged that students from the University of Bristol's Jewish Society were "political pawns by a violent, racist foreign regime engaged in ethnic

---

[1] https://www.holocaustremembrance.com/resources/working-definitions-charters/working-definition-antisemitism.

cleansing" (Liphschiz, 2021); we also examined the case of Irish novelist Sally Rooney, who refused to grant an Israeli publishing company translation rights to her best-selling novel as part of a cultural boycott of Israel. The third discourse event dealt with the US company Ben & Jerry's decision to stop selling their products in Israeli settlements. The three French discourse events we investigated were the reactions to the ban of comedian Dieudonné M'bala M'bala's and political essayist Alain

> **In total, the threads comprised 48,251 web comments, out of which 6,797 were considered antisemitic (either explicitly or contextually), while 4,100 contained counter speech.**

Soral's Facebook and YouTube accounts, to the emergence of antisemitic slogans in demonstrations against the Covid-19 'health pass', as well as to the alleged use of Pegasus spyware, developed by the Israeli company NSO Group, to collect intelligence on French politicians. Finally, the four international events examined in both French and British setting dealt with the Covid-19 vaccination campaign in Israel, an escalation phase of the Arab-Israeli conflict, terrorist attacks in Israel, and the Russian invasion of Ukraine. Overall, 310 threads were annotated for both countries. For the quantitative analysis stage, we eliminated the shortest and longest 5% of these threads, which resulted in a dataset of 278 threads containing from 71 to 264 comments (with the mean of 147). In total, the threads comprised 48,251 web comments, out of which 6,797 were considered antisemitic (either explicitly or contextually), while 4,100 contained counter speech.

Quantitative analysis: the big picture

In the process of annotation, we gained some understanding of how the volume of counter speech evolves depending on the number of antisemitic comments, and since the data is very diverse – covering two language communities and ten discourse events – quantitative analysis was a helpful approach to building a more comprehensive picture of the patterns within the corpus. A linear regression model was built, using the R package lme4 (Banta et al. 2010), with the proportion of counter speech comments in a thread as the dependent variable and the proportion of antisemitic comments, the language community and the discourse events as the independent variables. The p-value of the one-way ANOVA of the models are used to evaluate the significance of the relation under scrutiny.

As expected, the proportion of counter speech in a thread was revealed to be strongly correlated with the proportion of antisemitic comments (p-value = 2.1e-7); however, the amount of antisemitism is generally higher than the amount of counter speech: the slope of the linear model is 0.29, meaning that the proportion of antisemitism is around 3.4 times the proportion of counter speech, as shown below in Figure 1.

It seems that when the proportion of antisemitism in a thread is below 10%, the level of counter speech increases at a rate similar to antisemitic speech. However, when the latter reaches the level of 10% or higher, the proportion of counter speech does not increase in parallel with the proportion of antisemitism anymore; instead, the gap between the two grows. A further study could look at a more complex regression model.
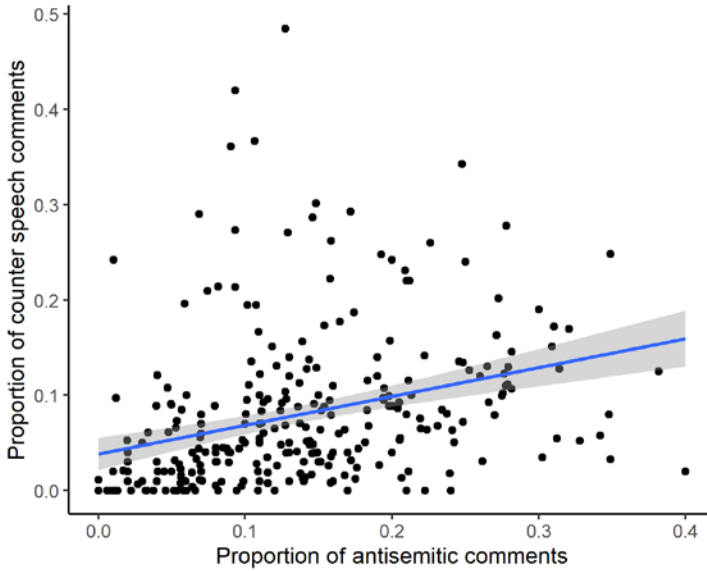

Figure 1: Proportion of counter speech comments in a thread as a function of the proportion of antisemitic comments

Moreover, we found that the proportion of counter speech and the language community from which the data was sourced are related in a significant way (p-value=0.0009). Given that the level of antisemitism reported in the various discourse events studied differs from one country to another, we evaluated the model fitting the proportion of counter speech in relation to either the UK or France, and the proportion of antisemitic comments within the threads. As shown in Figure 2, we found significantly more counter speech in the data from the UK comments sections than those from France. In addition, the interaction between the proportion of antisemitic comments and the country is also significant (see Table 1): when the proportion of antisemitism increases, the proportion of counter speech increases in both the UK and France. However, this relation is significantly larger in French data.

Predictors	Df	F-value	P-value
Speech community	1	12.5	0.00047
Proportion of antisemitic comments	1	24.5	1.3e-6
Interaction term	1	10.3	0.0015

Table 1: Results of one-way ANOVA of the linear model for counter speech proportion in Facebook threads
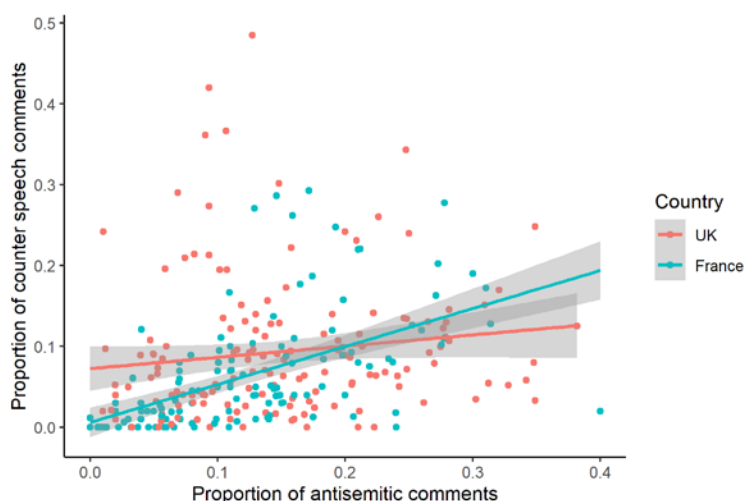
A

Figure 2: Proportion of counter speech comments as a function of antisemitic comments and country

Finally, we wanted to investigate whether specific discourse events are meaningful in determining the proportion of counter speech online. The Analysis of Variance (ANOVA) of the model, that is comparison of variances across the means of different discourse events, shows that the level of counter speech differs from one event to another (p-value = 1.4e-6). Three events stand out, as shown in Figure 3: the cases of Dieudonné in France, and Sally Rooney and David Miller in the UK – all three with a public figure at the centre. We also noted that the cases of Miller and Rooney are the only events for which we annotated more comments as counter speech than as antisemitic.
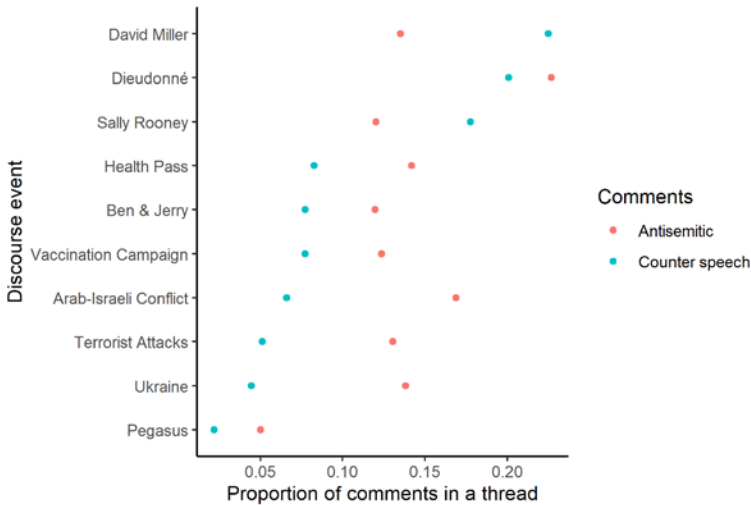
Figure 3: Mean proportion of antisemitic and counter speech comments for each discourse event

Qualitative analysis: a closer look[2]

While the quantitative analysis allowed us to determine both the proportion and evolution of antisemitic and counter speech comments throughout the different sub-corpora, a qualitative analysis was necessary to understand how they relate to each other conceptually. Our analysis revealed that most counter speech comments challenged the stereotype expressed in the respective antisemitic comment, as shown in a small sample of selected examples. Examples below are only a small representation of the large and complex counter discourse present in online comments sections. The selection aims to reflect some of the common concepts being countered, and the main argumentative and rhetorical strategies used to delegitimise antisemitic statements.

A: „So indicating a list of actors in the health crisis whose actions or positions are disapproved is punishable because some of them (7/12 I think) are Jewish? FN style ladies aren't my cup of tea, but isn't there a problem here?"

B: „Please show us the role of Soros and Rotschild in the health crisis. To see" (TWITT_20210810)

---

2   Please note that this chapter contains quotes from antisemitic hate speech. Even though the selected quotes do not contain offensive language, the logic and content presented here can be harmful to the directly affected individuals and communities, particularly in today's context.

In the first comment, User A questions the accusation of antisemitism levelled at Cassandre Fristot for naming supposed actors in the Covid-19 crisis on a placard she carried during street protests. Moreover, by referencing the Front National party the user suggests this DENIAL OF ANTISEMITISM[3] is not due to their political views. By asking to explain the alleged role of George Soros and the Rothschild family, both widely identified as Jewish, in the crisis, User B indirectly debunks User A's argument according to which Fristot listed these names unrelated to their Jewish background.

A: „Named terrorists by whom? Israel took over the land by force of arms."

B: Proscribed terrorists by most Western countries ... And they have not started one war in the last 73 years…So the moral of the story is don't declare war if you don't want to lose land…[…] And you have the cheek to accuse Israel of "taking over the land by force of arms [...]". (FB-TELEG_20211104)

Similarly, in a comment from the Rooney corpus, User A accuses Israel of being an EVIL state and expanding through violence. These stereotypes are then countered by User B, who transposes the responsibility to Palestine, argues that land was taken not by force but as a result of the conflict, and employs logic ("don't declare war if you don't want to lose land") to give weight to their statement and discredit Palestine.

However, some comments did not counter the actual stereotype evoked in the antisemitic comments:e:

A: „The problem in your story is that the aggressor and the occupying power is Israel"

B: „The problem in your story is that Israel is a sovereign and legitimate nation, over the whole Jerusalem and, in the long run, from the see to the Jordan." (FB-MONDE_20210512)

Here, in reaction to coverage of the 2021 escalation phase of the Arab-Israeli conflict, User A describes Israel as the aggressor and occupier, an EVIL entity acting against the Palestinian population. Instead of responding to this antisemitic stereotype, User B counters the DENIAL OF ISRAEL'S RIGHT TO EXIST by describing it as "a sovereign and legitimate nation". While these two tropes are distinct, B may have understood "occupying power" as a more indirect suggestion that Israel is illegally occupying territories outside its borders. To turn the argument around, B opens with the same expression ("The problem in your story is that…"). The analysis of this and other examples has shown that such

---

[3] Since stereotypes are phenomena that exist on the conceptual, i.e. mental, level and can be reproduced using language, stereotypes are given in small caps in accordance with the conventions of cognitive linguistics.

comments may prompt their addressees to react and even to reaffirm their antisemitic position. In other terms, though online counter speech is needed, its spontaneous forms may paradoxically fuel the emergence of further antisemitic comments. Additionally, some comments counter stereotypes and valorise Israel by diminishing another out-group, as if their goal was to find another scapegoat.

Closing comments

Unsurprisingly, our analysis of online comments reacting to selected media reports confirmed that the comments countering antisemitic discourse tend to be more numerous when antisemitism is more visible, suggesting that web users are aware of what constitutes antisemitic content – even when it is implicit. The question remains whether counter speech rises in reaction to antisemitism, whether it fuels it, or whether the two are not, in fact, directly

> **The quantitative examination of the data revealed there is proportionally more counter speech in the events involving a controversial personality accused of antisemitism.**

correlated, but instead dependent on the type of a media report they react to. As noted earlier, the quantitative examination of the data revealed there is proportionally more counter speech in the events involving a controversial personality accused of antisemitism, namely Dieudonné in France, and Sally Rooney or David Miller in the UK.

Perhaps personalisation of antisemitic ideas in those figures makes them easier to oppose. Another possible reason could be the figures' celebrity – unlike Fristot, Dieudonné, Rooney and Miller had already been known for their work in their respective professional fields, with a large following as well as a solid opponent base. It may also result from the manner in which news is reported: media coverage centred around one person is more likely to already include reports of antisemitism accusations made against them. Thus, online comments responding to such coverage are more likely to contain counter speech in reaction to the news article, rather than just to antisemitic comments within the thread. It is also possible that a well-known figure attracts more attention from web users, prompting a more dynamic debate in the comments sections.

Moreover, the profile of counter speech seems to vary from one country to the other: on average, the UK data contains more counter speech than the French; however, the share of counter speech in the UK does not increase with antisemitic comments as much as in France, suggesting that web users in the two countries use different strategies when producing antisemitism and counter speech. As the annotation process continues and the dataset grows, more quantitative and qualitative analyses can provide insights on these phenomena. Additionally, findings on the efficiency of counter

speech and the presence of other forms of hate in the comments countering antisemitism open the way to more complex questions: is spontaneous counter narrative efficient enough? Are the positive effects of counter speech jeopardised by the negative ones? Can counter speech be considered as such if it also conveys hate speech? All these questions require further studies, which would help counter hate ideologies in a more efficient way, and provide practical solutions to web users, media and social media platforms, or even policy makers.

> **There ist a need for further studies, which would help counter hate ideologies in a more efficient way, and provide practical solutions to web users, media and social media platforms, or even policy makers.**

Laura Ascone, Ph.D., focuses her research on computer-mediated communication, the expression of emotions, and hate speech. She defended her doctoral title in Linguistics at the Université Paris-Seine. Her dissertation on „The Radicalization through the Expression of Emotions on the Internet" dealt with the rhetorical strategies used in both jihadist propaganda and institutional counter-speech. Currently, she is a postdoctoral researcher at the Center for Research on Antisemitism (ZfA) at the Technical University of Berlin, involved in the international project „Decoding Antisemitism: An AI-driven Study on Hate Speech and Imagery Online."

Karolina Placzynta is a linguist and political scientist with an interest in pragmatics, sociolinguistics, and critical discourse analysis. Her research focuses on the normalization of prejudice, the mainstreaming and marginalization of discourses in the media, and the intersections of discriminatory discourses. Before joining the „Decoding Antisemitism" project at the Center for Research on Antisemitism (ZfA) at the Technical University of Berlin, she analyzed patterns in the discursive representations of immigration in the British press.

Chloé Vincent ist spezialisiert auf Soziolinguistik, wobei sie sowohl quantitative als auch qualitative Methoden anwendet. Sie interessiert sich für Diskriminierung und wie diese in und durch Sprache aufrechterhalten wird, insbesondere für verdeckte Manifestationen von Diskriminierung in der Sprache. In ihrer Magisterarbeit untersuchte sie die Einstellung französischer Muttersprachler zu regionalen französischen Akzenten. Sie hat außerdem einen Master of Engineering von der Grenoble INP Graduate School of Engineering.

References

Banta, Joshua A./Stevens, Martin H./Pigliucci, Massimo (2010). A comprehensive test of the 'limiting resources' framework applied to plant tolerance to apical meristem damage. Oikos 119(2), 359–369.

Becker, Matthias J., Ascone, Laura, Bolton, Matthew, Chapelan, Alexis, Krasni, Jan, Placzynta, Karolina, Scheiber, Marcus, Troschke, Hagen, Vincent, Chloé (2021). Decoding Antisemitism: An AI-driven Study on Hate Speech and Imagery Online. Discourse Report 2. Berlin: Technische Universität Berlin. Centre for Research on Antisemitism.

Cepollaro, Bianca/Lepoutre, Maxime/Simpson, Robert Mark (2022). Counterspeech. Philosophy Compass 18(1). https://doi.org/10.1111/phc3.12890.

Chapelan, Alexis/Ascone, Laura/Becker, Matthias J./Bolton, Matthew/Haupeltshofer, Pia/Krasni, Jan/Krugel, Alexa/Mihaljević, Helena/Placzynta, Karolina/Pustet, Milena/Scheiber, Marcus/Steffen, Elisabeth/Troschke, Hagen/Tschiskale, Victor/Vincent, Chloé (2023). Decoding Antisemitism: An AI-driven Study on Hate Speech and Imagery Online. Discourse Report 5. Berlin: Technische Universität Berlin. Centre for Research on Antisemitism. Online verfügbar unter https://decoding-antisemitism.eu/publications/fifth-discourse-report/ (abgerufen am 05.12.2023).

Cinelli, Matteo, De Francisci Morales, Gianmarco, Galeazzi, Alessandro, Quattrociocchi, Walter, Starnini, Michele (2021). The echo chamber effect on social media. Proceedings of the National Academy of Sciences 118(9).

Garland, Joshua, Ghazi-Zahedi, Keyan, Young, Jean-Gabriel, Hébert-Dufresne, Laurent, Galesic, Mirta (2022). Impact and dynamics of hate and counter speech online. EPJ Data Sci. 11(3). https://doi.org/10.1140/epjds/s13688-021-00314-6.

Liphschiz, Cnaan (2021). Nearly 200 scholars back UK lecturer who called Jewish students Israel 'pawns'. The Times of Israel, 27.02.2021. Online verfügbar unter https://www.timesofisrael.com/nearly-200-scholars-back-uk-lecturer-who-called-jewish-students-israel-pawns/ (abgerufen am 26.07.2021).

Mayring, Philipp (2015). Qualitative Inhaltsanalyse. Grundlagen und Techniken [Qualitative content analysis. Basics and techniques]. 12th ed. Weinheim/Basel, Beltz.

Monnier, Angeliki/Seoane, Annabelle (2019). Discours de haine sur internet. Publictionnaire. Dictionnaire encyclopédique et critique des publics. Online verfügbar unter https://publictionnaire.huma-num.fr/notice/discours-de-haine-sur-linternet/ (abgerufen am 05.12.2023).

Mouffe, Chantal (2010). Politique et agonisme. Rue Descartes 67(1), 18–24.

Rosenfeld, Alvin H. (Hg.) (2015). Deciphering the New Antisemitism. Indiana, Indiana University Press.

Schwarz-Friesel, Monika/Reinharz, Jehuda (2017). Inside the Antisemitic Mind: The Language of Jew-Hatred in Contemporary Germany. Waltham, Brandeis University Press.

Strippel, Christian/Paasch-Colberg, Sünje/Emmer, Martin/Trebbe, Joachim (eds.) (2023). Challenges and perspectives of hate speech analysis: An interdisciplinary anthology. Berlin: Digital Communication Research.

„OUR ANALYSIS OF ONLINE COMMENTS REACTING TO SELECTED MEDIA REPORTS CONFIRMED THAT THE COMMENTS COUNTERING ANTISEMITIC DISCOURSE TEND TO BE MORE NUMEROUS WHEN ANTISEMITISM IS MORE VISIBLE, SUGGESTING THAT WEB USERS ARE AWARE OF WHAT CONSTITUTES ANTISEMITIC CONTENT — EVEN WHEN IT IS IMPLICIT."

Laura Ascone, Karolina Placzynta & Chloé Vincent