# Humorous self-censorship strategies on YouTube: Semiotic structure and social-semiotic functions

punctum.gr

BY: **Alexander Brock** and **Merle Willenberg**

ABSTRACT

This article looks at the semiotics of humorous self-censorship. To this end, selected examples from a corpus of YouTube commentary videos and their respective comment sections are presented and discussed. On the one hand, the analysis focuses on structural aspects of humorous self-censorship signs from various modes (written, spoken, images, emojis, etc.) in their multimodal interplay. On the other hand, we analyze socio-semiotic aspects of our examples: their anchoring in specific speech communities, marked by background knowledge and shared communicative practices. The analysis shows that (1) structural manipulations of the spelling and/or phonetic shape of lexical items and of images etc. serve to secure both the understanding of the censored item as well as plausible deniability, while generating potentially humorous incongruities, (2) various positions in the participation process are exploited to participate in this process, including trigger warnings and other metacommunicative actions and levels, (3) humorous self-censorship serves a number of social-semiotic functions, such as the negotiation of group norms of sayability, the expression of group solidarity, and – importantly – entertainment.

# 1. Introduction

Attempts at regulating online communication, especially under conditions of anonymity, have led to several measures, such as the formulation of community guidelines for social network platforms, and various forms of punishment, including expulsion from platforms. A point especially interesting for linguists, however, is that regulation of digitally mediated communication (henceforth DMC) has not only led to censorship by the respective platform but also to self-censorship by users trying to avoid sanctions (Calhoun and Fawcett 2023; Marwick and boyd 2011; Willenberg 2023; 2024a; 2024b, among others). For our purposes, censorship is understood broadly as any form of (formal) suppression of language expression in awareness of existing or presumed communicative norms (cf. Allan and Burridge 2006:27). We differentiate between censorship as an external suppression of language expression, often by an authoritative or institutional body (like platform providers), and *self*-censorship, which describes the suppression or modification of language expression by its producers.

Our empirical observations have shown that self-censorship is sometimes performed in humorous ways, which is evident in the semiotic forms of censoring and in the recipients' reactions in the comment sections. Moreover, both the forms and the content of humorous self-censorship seem to react to the complex, multimodal and multi-level environments within which it takes place.

These observations lead to a number of research questions, which this article addresses:

- How does humorous self-censorship work in a multimodal, multi-level environment within a specific social context?
- How does humorous self-censorship reflect and shape this context?
- How do humorous forms of self-censorship on a micro-level relate to function in a social-semiotic framework and to language ideologies and norms of sayability on online platforms?

In connection with the observations mentioned initially, the video-sharing platform YouTube offers a unique research location for studying the intersection of self-/ censorship and humor signs, mainly in how multimodality is used as a social-semiotic resource in the pursuit of humor. This is the case not only due to its significance in the contemporary media landscape, but primarily due to its complex multimodal communication form (Androutsopoulos and Tereick 2015; Schmidt and Marx 2019), as described below (Section 2.1).

Regarding the specific genre under consideration here – commentary videos – this intersection is particularly evident, as such videos comment primarily on other platform creators or their content and usually revolve around potentially controversial topics that may go against YouTube's community guidelines. Debates instigated in the videos may be continued in the respective video's comment section. This way, a complex, multi-level communicative environment is created, extending across different videos (original and commentary videos) and comment sections. This environment will here be referred to as the *commentary sphere*, which subsumes these different text types and participants involved in their production and reception (Willenberg 2024b:89). A crucial point of interest lies in the potential uptake of self-/censorship techniques and humorous forms between videos and comments, as this would reveal shared norms and semiotic practices among the participants.

In this article, we analyze various occurrences of humorous self-censorship on YouTube commentary channels, examining their micro-semiotic forms and social-semiotic functions. We aim to represent not only the wide variety of techniques but also the multi-functionality of humor in self-censorship.

## 2. Social semiotics and humor theory

### 2.1. Micro- to macro-semiotics on YouTube – The censorship sign

Both self-/censorship and humor are relevant phenomena on different levels of meaning-making, ranging from the micro-semiotic level of the individual sign to the macro-semiotic level of social-semiotic processes. This section concentrates on censorship on various levels, while Section 2.2 introduces the humorous aspect.

The micro-level analysis consists of basic semiotic resources available for meaning-making in a given sign system, or communicative mode (e.g., Bateman 2016; Jewitt 2015; Stöckl 2019). For the communicative mode of writing, this would include the written mode in general but also its individual resources such as typography, orthography, and lexicogrammar, among others (Stöckl 2019:49). Their combined potential for meaning-making can be described as the mode's modal affordance (Jewitt 2015:72).

Among other things, semiotic resources can be used to create censorship signs. Taking the potentially offensive sign of the word *fuck* as an example, we can consider its censored counterpart *f\*ck*. Here, the modal affordances of writing are used to obscure the original sign through typographic substitution (Willenberg 2024b:56-57): the grapheme <u> has been replaced by the typographic symbol of the asterisk <\*>. This creates a new, ambiguous sign – a *censorship sign* – whose main characteristic is plausible deniability concerning the original sign body. *Plausible deniability* refers to the fact that while recipients may recover the original potentially offensive signifier, the sign-makers have created the possibility to deny using such a signifier.

The use of different semiotic resources in the creation of censorship signs can also achieve various degrees of concealment (Willenberg 2024b:188-193): staying with the example of *fuck*, substituting more graphemic material with alternative symbols, as in *f\*\*\**, leads to a more comprehensive concealment of the original sign body and thus to a higher degree of plausible deniability.

In addition to performing the censorship, the creation and presence of a censorship sign highlight the very presence of censorship in a communicative setting. Figure 1 provides an overview of the concept.
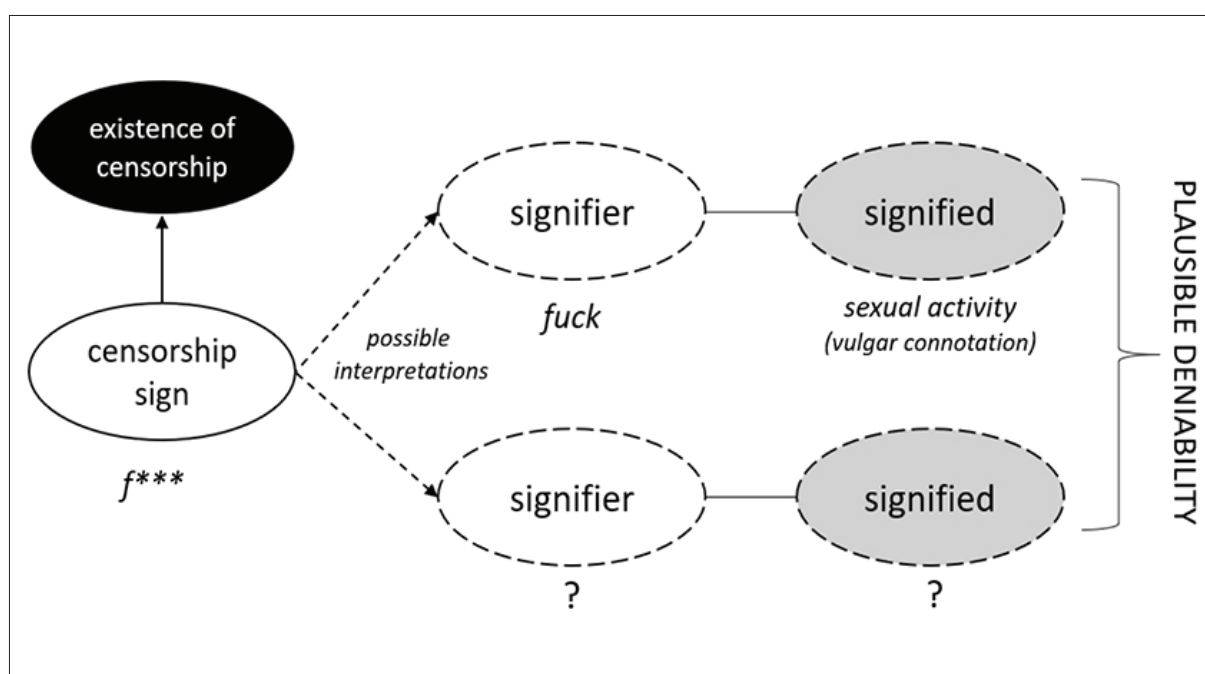


**Figure 1.** Graphic - Censorship Sign

The use of censorship signs may fulfil a variety of social functions, including politeness, (group) identity construction (including maintenance of anonymity) and more on the macro- or social-semiotic level of analysis. Meaning is always made on this macro level in relation to communicative constellations and prevailing norms. On social network platforms, these norms may be established either by codified rules (in the form of community guidelines), enforced by various sanctions that their violation may bring about (e.g., demonetization of content), or in the form of emergent community norms. In this normative environment self-censorship may become necessary and humor thrives.

The communicative constellations which constitute the communication form (Brock and Schildhauer 2017) of YouTube commentary videos are distributed across different communicative levels (Brock 2016; Chovanec 2022; Dynel 2014; Schmidt and Marx 2019). Using this conceptual ordering allows us to describe the distinct semiotic potentials of each communicative level (CL), with different affordances available between the video, the comments and the "hosting space" (Androutsopoulos and Tereick 2015:356). It also allows us to consider the different participation frameworks (Goffman 1981) present on the CLs and how they may affect humor or censorship processes. We follow Schmidt and Marx's (2019) conceptualization of four distinct CLs for YouTube, which are briefly summed up in Table 1:

**Table 1.** Communicative levels on YouTube, based on Schmidt and Marx (2019), see also Willenberg (2024a)

| CL 1 | video interaction | • basic participation framework, including speaker, ratified and non-ratified participants (Goffman 1981, Levinson 1988), as seen in the videos<br>• podium / platform formats (Goffman 1981:138-140) |
|---|---|---|
| CL 2 | sender-recipient interaction | • video production / author<br>• video distribution / (collective) sender (Dynel 2014:43) / releaser<br>• recipients / (mass) audience (Goffman 1981:137-138) |
| CL 3 | comments | • alternation of production and recipient roles |
| CL 4 | website-user interaction | • platform (YouTube) as communicator<br>• user as producer or recipient = *produsers*[1] |

---

[1] Due to the potentially shifting participation roles, participants on YouTube will here be generally referred to as *produsers* (Bruns 2008). The term *YouTuber* is reserved for content creators who upload videos to the platform.

## 2.2. **Humor theories and functions of humor**

### 2.2.1 *Humor theories*

According to the incongruity theory of humor (Dynel 2009:41-68; Ritchie 2004:46-58, and many others), humor is generated when an ill-fitting, incongruous element is introduced into a horizon of expectation. This is the most widely accepted theory of humor, and elements of it can be found in other humor theories, including Bergson's (2013[1900]; see below). A much-discussed question remains whether an incongruity must be resolved to generate humor (e.g., Martínez 2023; Suls 1983). As demonstrated in the literature (Brock 2004:213-259), incongruities can be constructed in any dimension of the communicative process, from individual sounds and words, via sentence constructions and text patterns, to realms of social and group norms. Ambiguity (Chamizo-Domínguez 2018:80) – which was shown to be a central aspect of the censorship sign (see Section 2.1) – may also be used to generate humor, for instance, when alternative readings of the same surface form are triggered by the punchline of a joke (Suls 1972:82).

For this investigation, the central questions are: a) What constitutes a horizon of expectation for individual recipients in a DMC context, and b) when is a piece of communication regarded as ill-fitting and incongruous? These two aspects are closely connected. While it falls into social semiotics to reconstruct general aspects of a communicative context, it is impossible for researchers – as for introspective recipients – to reconstruct exactly what a complete state of expectation is at any given time. Deviations from expectations, however, are readily noticed, allowing us to reconstruct at least one particular aspect of the previous horizon of expectation – the one that fails to be fulfilled by the incongruous element. This makes an analysis of incongruities a methodologically plausible enterprise, supported in the case of comment sections in DMC by possible reactions and metacommunication by the recipients. In the context of censorship in DMC, there are numerous possibilities for incongruities to emerge. Among others, censorship and its degree may be marked as incongruous, and the censorship sign can be incongruous in itself.

Bergson (2013[1900]:37) proposed a humor theory in which he saw the mechanical behavior in humans – "something mechanical encrusted upon the living"– as the primary source of humor. In his theory, human behavior is funny inasmuch as it resembles a blind mechanism. It should be noted that mechanical behavior can be considered a special case of incongruous behavior. In the context of DMC and self-censorship in the *commentary sphere*, censorship can be executed mechanically and profusely, even affecting elements of communication which are not censorable, such as YouTubers' names (see Section 4.3 below).

2.2.2 *Functions of humor*

The literature on functions of humor is extensive (Fry 1977; Hay 2000; Husband 1977; Long and Graesser 1988; Marone 2015; Martineau 1972; Zijderveld 1976 and many others), and some of its functions are more situation-specific than others. Functions mentioned in the literature are the following:

- rebellion against the powers that be and liberation (Zijderveld 1976:97, 173, 200-201),
- bolstering of morale and hope (Martineau 1972:104),
- 'safety valve' (ibid.) and catharsis (Husband 1977:267), where humor is used by people to vent their frustration, for instance, under repressive regimes,
- adaptation to circumstances (Martineau 1972:106, 111; Zijderveld 1976:74),
- conveying social norms (Long and Graesser 1988:54; Martineau 1972:108),
- undermining social norms (Zijderveld 1976:178),
- communication management (Hay 2000:725; Long and Graesser 1988:55; Marone 2015; Martineau 1972:103, 106; Zijderveld 1976:181-185),
- expression of group solidarity (Hay 2000:716-717; Martineau 1972:108, 116-119; Zijderveld 1976:186),
- power and control (Hay 2000:716-717; Holmes 2000:165; Long and Graesser 1988:53-54; Martineau 1972:106-107),
- self-disclosure and social probing (Martineau 1972:113; Long and Graesser 1988:53; Hay 2000:716),
- (positive) face work (Haugh 2011; Holmes 2000:167; Long and Graesser 1988:54-55)
- the *aesthetic* function (Martin 1987; Nunez-Ramos and Lorenzo 1997; Tsur 1994), which arises out of the play with semiotic elements.

While any combination of these functions may be relevant on the social-semiotic level of the analysis, some of them may be particularly important in the context of self-censorship in the social media. Among them might be the expression of group solidarity, for instance, if a number of *produsers* of a commentary channel applies the same humorous censorship techniques. Conveying social norms might also play a role in using humorous censorship signs. Here, "the speaker focuses attention on norms that society is unclear about, and points out the direction of his or her opinion" (Long and Graesser 1988:54). This function draws its significance from the fact that the *produsers* united on a YouTube channel commonly do not know each other offline and the norms conveyed online are the only ones available to regulate their social contact (Tsakona 2017:183). Finally, the self-disclosure and social probing function may also play a central role in the communicative environment of self-censorship. With each post, there is a risk that the values expressed in them do not meet with the *commentary sphere*'s approval.

Through humorous statements or language forms, participants can test the community's reaction to specific views, language forms, etc. In the case of a lack of approval, they can withdraw the form in question and mark it as non-serious and something they did not really mean. This way, a participant can cautiously convey their opinions and probe the community for their stance on things without taking full responsibility for what was communicated.

## 2.3. The humorous censorship sign

In Section 2.1, it was pointed out that the censorship sign typically works on two functional levels simultaneously: it performs the censorship and thereby creates plausible deniability, and it signals the very fact of censorship at the same time. The humorous censorship sign adds another functional level to these by not only performing and indicating censorship, but by doing so in a humorous way. This is shown in Figure 2:
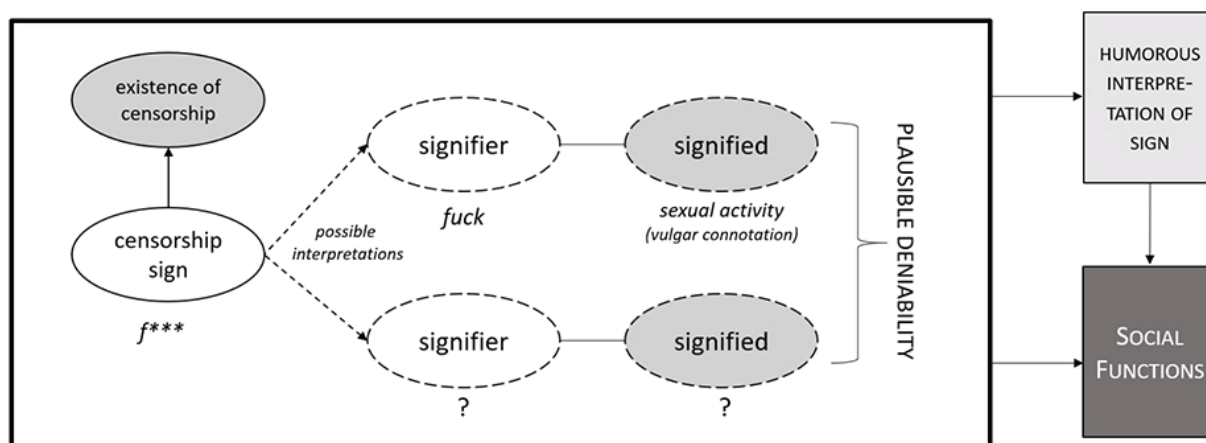


**Figure 2.** Graphic - Humorous Censorship Sign

Consequently, there are three levels of description when it comes to humorous censorship signs:

1. *Performing censorship*: The micro-semiotic processes that create plausible deniability concerning the original, potentially offensive sign body.
2. *Indicating censorship*: A censorship sign points to the institutional censorship regime, or in other words, a need for self-censorship.
3. *Introducing humor with its potential functions*: A censorship sign's form and/or presence may be perceived as humorous.

The following section introduces both corpus and methods used for the present investigation.

## 3. Corpus and methods

The corpus for this investigation consists of 42 YouTube commentary videos (1,244 minutes in total) and their corresponding comments (169,526) collected for a project on censorship and self-censorship (Willenberg 2024b). This material was qualitatively coded for occurrences of censorship in the context of that project. These coded instances were then investigated for cases of humor.

Occurrences of humor and their potential functions were identified in three steps:

1. The authors subjectively identified places of potentially humorous incongruities.
2. These potential instances of humor were then confirmed by identifying humor responses in the videos' comment sections. Possible indicators of humorous reception are emojis and digital laughter (e.g., *hehe*, *haha*, abbreviations like *lmao – laughing my ass off*) or metacommunicative passages.
3. Functions of these humorous segments were identified via recipients' reactions and the communicative environment. The primary source of insight here was metacommunicative comments, but other indicators were also considered. For instance, repetition of the same censorship technique is qualified as a possible sign of solidarity, norm-building / acceptance, and positive facework. An example is the uptake of the censorship sign *eating d*, which will be discussed in detail below (see Section 4.2).

Methodologically, it must be pointed out that we regard the functions of humor mentioned above only as *potential* functions carried mainly by the third semiotic layer of the humorous censorship sign (Section 2.3). We consider any function fulfilled only if explicit evidence of it can be found in the *commentary sphere*, for instance in the form of metacommunicative comments by recipients or in the videos themselves.

In the interest of completeness, we include some examples in our discussion which are clearly incongruous but did not get an explicit humor reaction in the comments. An example is the replacement of the word *fuck* by a duck emoji (see Section 4.1 below). This censorship technique works on the rhyme between the two words. At the same time, replacing a taboo lexeme with a harmless image is incongruous in the contrast between content and micro-semiotic form.

Overall, the selected examples are meant to illustrate how humorous self-censorship is performed on various levels, using several semiotic resources for both humor and self-censorship.

# **4.** Examples of humorous self-censorship on YouTube commentary channels

The analysis of the material shows that censorship signs used on YouTube commentary channels often fulfil a humorous function in addition to their censorship function (see Section 2.3). The following examples from the corpus illustrate this.

## 4.1. **Humorous censorship signs in the written mode**

Several words in our corpus are (self-)censored by replacing a part of their graphemic form with a different letter, resulting in seemingly innocent words being used in place of dispreferred items (see Willenberg 2024b:199-204). One instance of this is the micro-semiotic censorship sign of duck as a substitute for the taboo word *fuck* (Example 1), occasionally in the form of an emoji (Example 2).

Example 1: Comment on "I watched the haunting of Shane Dawson so you didn't have to…" [2]

> Shane can't redeem himself. He's gross. Shane and Trisha have one thing in common, a complete lack of giving a duck 😉 about how ppl feel about them.

Example 2: Comment on "These family vloggers just made the worst video on YouTube"

> Short answer? No.
> Long answer? No, 🦆 you. […

As pointed out above (Section 3), the innocent form of a duck emoji to stand for the taboo word fuck constitutes an incongruity which undermines social norms of sayability and – if taken up by other *produsers* – creates group solidarity. In addition, the playful manipulation of the signifier introduces an aesthetic dimension into the communication. In Example 1, the author marks their non-serious dimension with an emoji.

Other rhyming forms used for censorship and humor are so-called shm-reduplications (Nevins and Vaux 2003).

---

[2] All comments are presented in their original form, including possible typos or grammatical errors.

Example 3: Comment on "These family vloggers just made the worst video on You-Tube"

> Does it rhyme with smish smortion?

Here, the poor quality of the phonetic replacement of *abortion* adds another layer of incongruity to the censorship: an ill-conceived and therefore incongruous realization of a shm-reduplication. At the same time, an element of Bergson's (2013 [1900]) mechanical can be traced here, in that the form in Example 3 constitutes a link in the endless chain of mechanical shm-reduplications, be they well-formed or deficient.

Example 4: Comment on "Family Vlog Channels Are Sad"

> @commenter Charlie drip is hit but this is not cornhub

The replacement of *Pornhub* by *cornhub* potentially fulfils all the functions identi-fied for Examples 1 and 2. In addition, the form presupposes an in-group of *produsers* familiar with the pornography website Pornhub. As well as group solidarity of people in the know, this usage conveys a social norm for this particular group by suggesting this to be a possible form to refer to the website.

Example 5: Comment on "Nikocado Avocado is Eating Himself to the Grave"

> Heart Attack Grill is filled to the absolute brim with sin/gluttony & I honestly hope that it is burned to the ground 1 of these days as an almost symbolic gesture of it going up in he 🏒 🏒 fire & thereby cleansing the ungodly grip it has on the fools of society.

Here, the aesthetic play with two innocent emojis in the form of hockey sticks standing for similar-looking graphemes may be central beyond the fact that the manip-ulation renders the word hellfire unrecognizable to the YouTube content moderation algorithm. Further, an incongruity is introduced on the micro-semiotic level, as emojis are typically used to substitute entire lexemes rather than individual graphemes (see Example 2 above).

Example 6: Comment Thread on "Shane's Comeback"

| Comment | Shane dropping the oops "I almost kermited herbaside" for sympathy is such a manipulative behavior it brings back memories of every shitty guy I dated | commenter1 |
|---|---|---|
| Reply 1 | 💀💀 especially when he basically almost herbacide baited James Charles | commenter2 |
| Reply 2 | i'm surprised no one pointed out he said that jc was too egotistical to commit herbacide when tati pointed out that this might send james over the edge. i definitely would have commented that lol<br><br>edit: changed "sudoku (i wrote the real word) to herbacide | commenter3 |
| Reply 3 | took me a while to understand that quote 😂 | commenter4 |
| Reply 5 | What is herbacide | commenter5 |

Example 6 shows that censorship signs cannot be taken for granted, even in a tight-knit *commentary sphere*. While it is evident that at least some *produsers* consider the form *kermited herbaside* funny,[3] at the same time, a new group norm of referring to suicide is being discussed, with candidates such as *kermit/commit* and *herbaside/herbacide* as potential candidates. The playfulness introduced via the humorous censorship sign in the top comment leads to further language play, as evident in the edited section in Reply 2, where *sudoku* is introduced as another possible substitute for *suicide*. Finally, a *produser* asks for information on the signified behind the signifier *herbacide*, which clearly indicates that a reliable group norm for referring to suicide has not been established.

While much humorous self-censorship takes place on the level of individual lexemes or phrases, it may also affect larger units and dynamic discourse processes. A case in point is pseudo-corrections in the written mode, which use the different options for correction in the spoken and the written modes. Corrections in spoken language necessarily leave traces: the correctible cannot be undone but is replaced *post hoc* or midway by an alternative formulation, i.e., a different word, phrase of grammatical construction (see Section 4.2 for how this can result in involuntary humor in the context of oral censorship signs on CL 1). In written language, correctibles can usually be deleted without a trace (real-time processing in DMC excepted). Example 7, taken from the written comments (CL 3), plays exactly with this difference:

---

[3] Evidence from the corpus suggests that the skull emojis in the first reply function as a humor indicator (along the lines of "I'm dead from laughing", cf. Example 9 below), similar to other laughter-indicating emojis.

Example 7: Comment on "The Worst Religious Channel"

> My sister watched this channel, like what the fu- I mean frick

In the written mode within YouTube's communication form, such corrections are obviously unnecessary, as *produsers* may correct their message before posting it (Meredith and Stokoe 2014). However, the corpus contains instances of such corrections in the written mode where they are explicitly performed in constructing the message, therefore taking on a pseudo-like character. The effect is that the taboo lexeme is almost called up but ultimately avoided. Therefore, this kind of pseudo-correction in the written mode is a deliberately transparent self-censorship technique. If the censorship could be hidden in the editing process, but is not, there must be reasons for not doing so. We see these potentially in identity construction and/or humor. What is common to all these forms is that they may trigger humor reactions in the corpus and fulfil the function of undermining social norms (Zijderveld 1976: 178). While paying lip service to these norms, they clearly indicate their transgression.

### 4.2. Unintentional (?) sexual self-censorship

In the following example, the commentary YouTuber Repzion discusses another content creator (Eugenia Cooney), who allegedly suffers from an eating disorder.

Example 8: Video excerpt from "We Need To Talk About This Woman" (01:31-01:45)

> 3    **Repzion:** because if you go to her comment section you'll even find people who are recovering from ED
> <draws letters in the air with fingers>
> It's an abbreviation for eating uh D hah
> I just realized that that sounds like uh when I say ED ugh
>                                         <moves hand in front of face>

Repzion uses the initialism *ED* as a substitute for the compound noun *eating disorders*. Afterwards, he attempts to explain what the abbreviation stands for but seems to realize midway that he wants to avoid the term *disorder*, hence shortening it to [di:]. This is indicated by the hesitancy marker *uh* (Pomerantz 1975), and results in the formulation *eating d*. He then laughs awkwardly, seemingly realizing that the formulation has now taken up an additional meaning. Then he *meta*-metacommunicates, i.e., he metacommunicates about his own metacommunicative comment, before ultimately stopping to speak. He then indicates embarrassment through 'face palming'.

The unintended meaning generated in his first metacommunicative utterance is a sexual one, as the initialism *D* may also be used as a substitute for the word *dick*. *Eating d* thus has taken on a sexual meaning, which is incongruous with the topic currently discussed. The absurdity of the potential formulation "people who are recovering from eating dick" makes the sequence humorous, as well as the fact of an unintended additional censorable meaning generated in the censorship process (Brock 2016). Rather than creating plausible deniability with a censorship sign (*ED*), the result is the creation of a *more* rather than a *less* offensive alternative interpretation.

This segment is metacommunicatively discussed in the video's comment section (CL 3). We were able to identify 120 comments reacting to this incident. The blunder is perceived as humorous, with commenters identifying the incongruity of the sexual meaning of Repzion's own metacommunication as the source of involuntary humor. At the same time, his need for self-censoring is ridiculed through the further uptake of the ill-conceived form in the comments.

Example 9: Comment on "We Need to Talk about This Woman"

> *Eating D*, I'm dying.
> I shouldn't be this entertained by the video, but man that made me laugh

Since Repzion has used the initialism *ED* in the video before to describe Cooney's condition, the commenters also associate the sexual meaning with her person. They then use this interpretation for further humorous comments, using Repzion's involuntary humor as a springboard. This results in a game of one-upmanship, where *produsers* make new jokes using the sexual meaning of the ill-conceived censorship sign as the source of humor.

Example 10: Comment on "We Need to Talk about This Woman"

> I'm not guilty of ever eating D, but I've fed D to people before.

Example 11: Comment on "We Need to Talk about This Woman"

> Hey she's got to be eating something right lol just kidding.... Sorry if that was in poor taste (I hate my phone's voice to text) but that's the first thing that popped up when I read this comment.

Such games can be considered affiliative behavior in the service of social cohesion (Marone 2015:74). In terms of function, this humorous passage shows strong signs of building group solidarity.

Further, some commenters offer improvement on the self-censorship measure taken by Repzion to avoid such involuntarily humorous instances in the future, as in Example 12.

Example 12: Comment on "We Need to Talk about This Woman"

> @Repzion "disordered eating" gets the same point across without it being so…. amusing to my immature brain. lmao (I did get a couple of giggles out of it and I use humor to cope so it helped me chuckle through the content a little easier. Thank you for the accidental black humor.)

This shows that his attempts at self-censorship are accepted by the viewers, affirming the taboo character of the term *eating disorder*.

In Example 8, the censorship conventions attached to a specific form (abbreviation) lead to involuntary humorous self-censorship. This is possible because the censorship potential of abbreviations lies in their potential polysemy. The censorship sign *ED* has competing interpretations, not only *eating disorder*, as intended by Repzion, but also *erectile dysfunction*, which is also referred to in the humorous comments on the video (Example 13). This shows that the sign *ED* is not fully conventionalized and its polysemy has humorous potential.

Example 13: Comment on "We Need to Talk about This Woman"

> Oh my gawd Eugenia haz Erectile Disfunction 😱

Lastly, the involuntariness of humor in Example 8 is ultimately questionable. While the accidental sexual formulation seems unintentional at first, its inclusion in the video is clearly intentional, as it was not deleted in the editing process. This fundamentally differs from how accidental humor shows up in live video formats, as the participants in such frameworks do not have *post hoc* control over what is included in the broadcast (see Chovanec 2021). In other words, the humorous utterance on CL 1 at the time the video was recorded may have been involuntary. Still, the collective sender on CL 2 approved of its presence in the video, making the humor intentional for the audience on this level. This demonstrates that the multi-level environment on YouTube leads to specific options for creating humor.

4.3. **Name censorship**

Another instance of humorous censorship could be identified through indicators of humorous reception in the comments. Here, humor originates in an incongruous censorship practice of names. Names are usually censored – online and offline – for privacy concerns (e.g., Bös and Kleinke 2017:90-92, 107). The censorship in this case can be achieved through various means, such as complete omission or via censorship signs like abbreviations, blurring, and blackout. The goal is usually to obscure the respective name as an identifier completely. In the material considered here, name censorship not only occurs in less obscure forms, but recipients also interpret it as humorous, as seen in the following examples. Example 14 shows a comment thread (CL 3) posted under one of the videos.

Example 14: Comments on "Shane Dawson"

| *Comment* | sh*ne used to be one of my biggest role models and to realize that your role model is an awful human being hurts so much | commenter1 |
|---|---|---|
| *Reply 1* | Did you just censor "shane" 😂 | commenter2 |
| *Reply 2* | Don't think you need to censor his name 🤣 […] | commenter3 |
| *Reply 3* | @commenter2 its a bad word man | commenter1 |

The crying-laughing emojis in Reply 1 and 2 show a humorous reception of the previous comment, in which the name Shane has been censored via typographic concealment in the form of *sh*ne*.[4] This indicates an incongruity present in the original comment. It stems from a break in conventions of name censorship, as evident by the question in Reply 1 and the correction undertaken in Reply 2: names of public YouTubers are not normally censored in the *commentary sphere*. In a response (Reply 3), the original poster of the comment provides an account for the censorship sign, explaining that the name is "a bad word". This evaluative function of the censorship sign seems to be dominant. An avoidance of the name due to privacy concerns is unlikely here, because the censored name can be found under a video explicitly discussing Dawson, which means that the name is accessible to all *produsers* anyway.

The name is also featured as a censorship sign (*Sh/ne D/wson*) in a different video's content warning,[5] where it occurs alongside other censorship signs like *s/icide* and *s/xual* (Hotbox, "The Haunting Return of Shane Dawson", 00:00-00:13). Even though the name occurs in this censored form in the beginning of the video (CL 2), its title contains

---

[4] The name refers to Shane Dawson, a well-known YouTuber who was involved in many controversies over the years. A sub-section of the corpus considered here consists of commentary videos discussing him.
[5] Trigger or content warnings warn recipients of potentially triggering (i.e. retraumatizing) content contained in a succeeding text.

the name in an uncensored form (CL 4). As with Example 14, censorship in the pursuit of privacy does not seem to be the primary concern here. In terms of humorous incongruities, not only is the name censorship interpreted as humorous in the comments but also its inclusion in the content warning in the first place. An incongruity can be assumed to lie in the inclusion of a name in addition to potentially triggering topics, as the metacommunicative comment in Example 15 suggests:

Example 15:  Comment on "The Haunting Return of Shane Dawson"

> I love how you censored his name in the CW list LOLOL

Example 14 also allows for an interpretation according to Bergson's theory: the *produsers* who censor Shane's name pretend to indiscriminately and mechanically censor any potentially harmful linguistic expression, thereby creating the character of the mechanical censor. As trigger warnings or censorship signs are usually related to taboo topics or words only, the name censorship or inclusion in trigger warnings puts the referent of the name on the same level. It thereby marks them as taboo or potentially offensive. In this way, the censorship of the name achieves various functions: It generates humor for the audience and allows the sign maker to take a negative stance on Dawson (Willenberg 2024a:110-111).

Another such instance in the material confirms the incongruous nature of name inclusion in trigger warnings (TWs). In a video about Dawson by greenistnotnick, the name is included in an uncensored form (CL 2). The inclusion of the name is considered humorous in the same video's comments (CL 3), too. This is indicated again by the use of emojis, digital laughter, expressions of incredulity, or metacommunication.

Example 16:  Comment on "I Watched Shane Dawson's New Video"

> "Tw: Shane Dawson" LMFAO [6]

Example 17:  Comment on "I Watched Shane Dawson's New Video"

> I love how Shane Dawson was the TW... I mean... It's true... But it made me laugh

What these examples have in common is that the incongruous censorship of the name *Shane Dawson* and its incongruous inclusion in the censorship-adjacent trigger warnings seem to break with community practices related to both.[7] The incongruity is thus situated on the macro level of social semiotics, reflecting a break in the expectation concerning TWs in the *commentary sphere*.

---

[6] This initialism stands for *laughing my fucking ass off*.
[7] For more on the relationship between trigger warnings and censorship in terms of YouTube's participation framework, see Willenberg (2024b:266-270).

# 5. Conclusion

The analysis of the material shows that online content regulation often triggers both self-censorship and humor. Our presentation of the examples has demonstrated that what is common to all instances of humorous self-censorship is the function to undermine communicative norms about sayability. This affects both micro-semiotic dimensions of individual signifiers and macro-level semiotic structures such as taboo topics, trigger warnings, etc. Sign-makers use the semiotic affordances of YouTube's communication form in these endeavors. More specifically, it has been demonstrated how manipulations on the micro-semiotic levels affect the macro level of social semiotics in that they fulfil different functions and contribute to shaping language ideologies and linguistic norms on YouTube. What can and cannot be said on the platform is negotiated – among other things – via humorous censorship signs. While self-censorship more or less directly reflects the norms of sayability, humor needs these norms to determine an expectation which is then humorously transgressed and undermined. While it is impossible to predict whether self-censorship will stabilize the norms of communicative conduct or humorous transgression will gradually establish altered or new norms of sayability and face-work, it remains to be seen how these opposing tendencies will play out in the long run.

The material considered here allows us to identify incongruities within and in the use of censorship signs. Some cases, such as Example 6, show that censorship conventions in the *commentary sphere* are dynamic and a matter of negotiation. Acceptance or rejection of humorous censorship signs may contribute to the renegotiation or emergence of new language norms within the *commentary sphere*. This process allows sign-makers to engage in playful and non-serious self-censorship, particularly in the context of otherwise serious topics that may be subject to platform moderation, such as suicide or eating disorders.

Another significant function potentially fulfilled by humorous censorship signs is the establishment of group solidarity. Humor can function as a social glue, especially in the face of socially charged domains such as censorship. Members of the *commentary sphere* engage in humorous semiotic practices to signal group membership through shared communicative norms concerning self-censorship and the creation of an in-group identity tied to specific knowledge, which is obscured through self-censorship and accessible only to those 'in-the-know'.

The present discussion attempted to show that self-censorship techniques and conventions continue to evolve in the context of constantly changing content regulation methods on online platforms like YouTube, generating potential for new humor signs and shaping social-semiotic patterns.

# References

Androutsopoulos, Jannis and Jana Tereick 2015. YouTube: Language and discourse practices in participatory culture. In: Alexandra Georgakopoulou and Tereza Spilioti (eds.) *The Routledge Handbook of Language and Digital Communication*. Milton Park: Routledge, 354-370.

Allan, Keith and Kate Burridge 2006. *Forbidden Words. Taboo and the Censoring of Language*. Cambridge: Cambridge University Press.

Brock, Alexander 2004. *Blackadder, Monty Python und Red Dwarf – eine linguistische Untersuchung britischer Fernsehkomödien*. Tübingen: Stauffenburg.

Brock, Alexander 2016. The borders of humorous intent. The case of TV Comedies. *Journal of Pragmatics* 95: 58-66.

Brock, Alexander and Peter Schildhauer 2017. Communication form: A concept revisited. In: Alexander Brock and Peter Schildhauer (eds.) *Communication Forms and Communicative Practices. New Perspectives on Communicative Forms, Affordances and What Users Make of Them*. Frankfurt am Main: Peter Lang, 13-44.

Bateman, John A. 2016. Methodological and theoretical issues in multimodality. In: Nina-Maria Klug and Hartmut Stöckl (eds.) *Handbuch Sprache im multimodalen Kontext*. Berlin: De Gruyter. 36-74.

Bergson, Henri 2013[1900]. *Laughter. An Essay on the Meaning of the Comic*. Hamburg: Tredition.

Bös, Birte and Sonja Kleinke 2017. Publicness and privateness. In: Christian Hoffmann and Wolfram Bublitz (eds.) *Pragmatics of Social Media*. Berlin, Boston: De Gruyter Mouton, 83-121.

Bruns, Axel 2008. Blogs, *Wikipedia, Second Life, and Beyond. From Production to Produsage*. New York, NY: Lang.

Calhoun, Kendra and Alexia Fawcett 2023. 'They edited out her nip nops': Linguistic innovation as textual censorship avoidance on TikTok. *Language@Internet* 21: 1-30.

Chamizo-Domínguez, Pedro J. 2018. Ambiguity and vagueness as cognitive tools for euphemistic and politically correct speech. In: Andrea Pizarro Pedraza (ed.) *Linguistic Taboo Revisited. Novel Insights from Cognitive Perspectives*. Berlin, Boston: De Gruyter Mouton, 79-96.

Chovanec, Jan 2021. Saving one's face from unintended humor: Impression management in follow-up sports interviews. *Journal of Pragmatics* 176: 198-212.

Chovanec, Jan 2022. 'Nice try, loser': participation and embedded frames of interaction in online sports commentaries. In: Alexander Brock, Janet Russell, Peter Schildhauer and Merle Willenberg (eds.) *Participation and Identity. Empirical Investigations of States and Dynamics*. Berlin et al.: Peter Lang, 65-84.

Dynel, Marta 2009. *Humorous Garden-Paths. A Pragmatic-Cognitive Study*. Newcastle: Cambridge: Scholars Publishing.

Dynel, Marta 2014. Participation framework underlying YouTube interaction. *Journal of Pragmatics* 73: 37-52.

Fry Jr., William F. 1977. The appeasement function of mirthful laughter. In: Antony J. Chapman and Hugh C. Foot (eds.) *It's a Funny Thing, Humor*. Oxford: Pergamon Press, 23-26.

Goffman, Erving 1981. *Forms of Talk*. Philadelphia: University of Pennsylvania Press.

Haugh, Michael 2011. Humor, face and im/politeness in getting acquainted. In: Bethan L. Davis, Michael Haugh, and Andrew John Merrison (eds.) *Situated Politeness*. London: Continuum International Publishing Group, 165-184.

Hay, Jennifer 2000. Functions of humor in the conversations of men and women. *Journal of Pragmatics* 32: 709-742.

Holmes, Janet 2000. Politeness, power and provocation: How humor functions in the workplace. *Discourse Studies* 2(2): 159-185.

Husband, Charles 1977. The mass media and the functions of ethnic humor in a racist society. In: Antony J. Chapman and Hugh C. Foot (eds.) *It's a Funny Thing, Humor*. Oxford: Pergamon Press, 267-272.

Jewitt, Carey 2015. Multimodal analysis. In: Alexandra Georgakopoulou and Tereza Spilioti (eds.) *The Routledge Handbook of Language and Digital Communication*. Milton Park et al.: Routledge, 69-84.

Levinson, Stephen C. 1988. Putting linguistics on a proper footing: Explorations in Goffman's concept of participation. In: Paul Drew and Anthony Wootton (eds.) *Erving Goffman. Exploring the Interaction Order*. Oxford/Cambridge: Polity Press, 161-227.

Long, Debra L. and Arthur C. Graesser 1988. Wit and humor in discourse processing. *Discourse Processes* 11: 35-60.

Marone, Vittorio 2015. Online humor as a community-building cushioning glue. *European Journal of Humor Research* 3 (1): 61-83.

Martin, Mike W. 1987. Humor and aesthetic enjoyment of incongruities. In: John Morreall, (ed.) *The Philosophy of Laughter and Humor*. Albany: State University of New York Press, 172-186.

Martineau, William H. 1972. A model of the social functions of humor. In: Jeffrey H. Goldstein and Paul E. McGhee (eds.) *The Psychology of Humor. Theoretical Perspectives and Empirical Issues*. New York: Academic Press, 101-125.

Martínez, Javier 2023. A general mechanism of humor: Reformulating the semantic overlap. *Humor* 36 (4): 529-565.

Marwick, Alice E. and danah boyd 2011. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society 13* (1): 114-133.

Meredith, Joanne and Elizabeth Stokoe 2014. Repair: Comparing Facebook 'chat' with spoken interaction. *Discourse & Communication* 8 (2): 181-207.

Nevins, Andrew and Bert Vaux 2003. Metalinguistic, shmetalinguistic: The phonology of shmreduplication. *Proceedings from the Annual Meeting of the Chicago Linguistic Society* 1: 702-721.

Nunez-Ramos, Rafael and Guillermo Lorenzo 1997. On the aesthetic dimension of humor. *Humor* 10 (1): 105-116.

Pomerantz, Anita 1975. *Second Assessments. A Study of Some Features of Agreements/Disagreements*. Ph.D. thesis. Irvine: University of California.

Ritchie, Graeme 2004. *The Linguistic Analysis of Jokes*. London and New York: Routledge.

Schmidt, Axel and Konstanze Marx 2019. Multimodality as challenge: YouTube data in linguistic corpora. In: Janina Wildfeuer, Jana Pflaeging, John A. Bateman, Ognyan Seizov and Chiao-I Tseng (eds.) *Multimodality. Disciplinary Thoughts and the Challenge of Diversity*. Berlin, Boston: De Gruyter Mouton, 115-144.

Stöckl, Hartmut 2019. Linguistic multimodality - Multimodal linguistics: A state-of-the-art sketch. In: Janina Wildfeuer, Jana Pflaeging, John A. Bateman, Ognyan Seizov and Chiao-I Tseng (eds.) *Multimodality. Disciplinary Thoughts and the Challenge of Diversity*. Berlin, Boston: De Gruyter Mouton, 41-69.

Suls, Jerry M. 1972. A two-stage model for the appreciation of jokes and cartoons: An information-processing analysis. In: Jeffrey H. Goldstein and Paul E. McGhee (eds.) *The Psychology of Humor. Theoretical Perspectives and Empirical Issues*. New York: Academic Press, 81-100.

Suls, Jerry M. 1983. Cognitive processes in humor appreciation. In: Paul E. McGhee and Jeffrey H. Goldstein (eds.) *Handbook of Humor Research*. Volume 1, Basic Issues, New York: Springer Verlag, 39-57.

Tsakona, Villy 2017. Humor research and humor reception: Far away, so close. In: Wladyslaw Chlopicki and Dorota Brzozowska (eds.) *Humorous Discourse*. Berlin, Boston: De Gruyter Mouton, 179-202.

Tsur, Reuven 1994. Droodles and cognitive poetics: Contribution to an aesthetics of disorientation. *Humor* 7 (1): 55-70.

Willenberg, Merle 2023. Talking about bodies on YouTube - Linguistic and multimodal avoidance strategies in the context of body shaming and eating disorders. In: Daniela Pietrini (ed.) *Lingua e discriminazione. Studi diacronici, lessicali e discorsivi*. Berlin: Peter Lang, 149-166.

Willenberg, Merle 2024a. TW: su1(1d3 - Multimodal self-censorship on YouTube. *JournaLIPP 9: Sprache in der digitalen Welt – Vermittlung, Variation, Politik*: 99-123, https://lipp.ub.uni-muenchen.de/lipp/issue/view/293/31

Willenberg, Merle 2024b. *What the eff. A Linguistic Investigation of Censorship and Self-censorship on YouTube Commentary Channels*. Unpublished Ph.D. thesis, Philosophical Faculty II, Martin-Luther-University Halle-Wittenberg.

Zijderveld, Anton C. 1976. *Humor und Gesellschaft. Eine Soziologie des Humors und des Lachens*. Graz, Wien and Köln: Styria.

## Videos References

Drama Investigator 2021. "I watched the haunting of Shane Dawson so you didn't have to…" *YouTube*. https://www.youtube.com/watch?v=d3wb5yDyzNA (8 Oct.).

greenisnotnick 2021. "I Watched Shane Dawson's New Video So You Don't Have To." *YouTube*. https://www.youtube.com/watch?v=wHAEtvRiwjM (11 Oct.).

Hotbox 2021. "The Haunting Return of Shane Dawson." *YouTube*. https://www.youtube.com/watch?v=dCGhDym0LDA (10. Oct.).

ImAllexx 2019. "The Worst Religious Family Channel." *YouTube*. https://www.youtube.com/watch?v=ISDCkG9mwHY (31 Oct.).

ImAllexx 2021. "Shane Dawson." *YouTube*. https://www.youtube.com/watch?v=OW763MJRzxI (10 Oct.).

Kuncan Dastner 2021. "shane's comeback video proves he's learned nothing." *YouTube*. https://www.youtube.com/watch?v=z5C6SnZ2bXs (14 Oct.).

penguinz0 2022. "Family Vlog Channels Are Sad." *YouTube*. https://www.youtube.com/watch?v=BgK0DtbwMTc. (13 Jan.).

Repzion 2020. "We Need To Talk About This Woman." *YouTube*. https://www.youtube.com/watch?v=FJteipLDngw (13 Dec.).

sixteenleo 2022. "Nikocado Avocado Is Eating Himself To the Grave… (It's Sad)." *YouTube*. https://www.youtube.com/watch?v=0OkU1xghOXE (18 Jan.).

TehMimi 2022. "These family vloggers just made the worst video on YouTube." *YouTube*. https://www.youtube.com/watch?v=aFukgbuaQd0 (16 Apr.).

AUTHOR

**Alexander Brock** is Professor of English Linguistics at Martin-Luther-University, Halle-Wittenberg, Germany.

AUTHOR

**Merle Willenberg** is a research assistant and doctoral candidate at Martin-Luther-University, Halle-Wittenberg, Germany.