

Algospeak and Platform-Driven Language Change : A Small-Scale Corpus-Based Study

Dr. NASRI Chadia*

Echahid Cheikh Larbi Tebessi University -Tebessa-, n.chadia@yahoo.com

Received: 18/05/2025

Accepted: 30/06/2026

Published: 01/07/2026

Abstract

This demonstration article presents a small-scale, corpus-based qualitative approach to the study of algospeak, the coded vocabulary that social media users develop to navigate or evade automated content moderation. A corpus of twenty documented terms is compiled from secondary, publicly available sources, coded along two dimensions—formation mechanism and topic domain—and quantified descriptively. The resulting patterns are interpreted through established sociolinguistic concepts, including the euphemism treadmill, x-phemism, and algorithmic folk theories, to illustrate how a small dataset can support a theoretically grounded analysis without being treated as statistically representative.

Key Words: Algospeak ; Algorithmic Content ; Moderation ; Corpus Linguistics ; Euphemism Treadmill ; Digital Discourse

Introduction

The rapid expansion of social media platforms has created new conditions for linguistic innovation. Among the most visible responses to automated content moderation is algospeak, a term used to describe the alternative words, spellings, expressions, and symbols that users adopt when referring to content they perceive as sensitive or potentially subject to moderation (Lorenz, 2022). Rather than constituting a single linguistic strategy, algospeak encompasses a range of formal, semantic, and visual modifications. Users may slightly alter an existing word, replace it with another expression, shorten it, or represent it through an emoji or other symbol. The resulting forms can remain highly transparent to human readers while becoming less similar to the conventional form of the referent.

This phenomenon is particularly interesting from a sociolinguistic perspective because it brings together processes of language change and users' understandings of technological systems. In conventional lexical change, new expressions emerge through processes such as semantic shift, borrowing, clipping, and euphemistic replacement. Algospeak draws on many of these same resources, but its development takes place within an environment in which users believe that particular words or forms may affect the visibility, reach, or moderation status of their content. Consequently, linguistic choices can be shaped not only by interpersonal or stylistic considerations but also by users' perceptions of how platform moderation operates.

Although the term is relatively recent, the linguistic practices it describes are connected to longstanding processes of lexical innovation, euphemistic replacement, and indirect reference. The present study therefore draws on three complementary concepts—the euphemism treadmill, x-phemism, and algorithmic folk theories—to interpret the forms identified in the corpus. These concepts address different aspects of the phenomenon: the first helps explain lexical replacement over time, the second provides a framework for interpreting the social and evaluative meanings of alternative expressions, and the third connects linguistic innovation to users’ understandings of platform moderation.

The euphemism treadmill describes the process through which a substitute expression gradually acquires some of the negative associations of the expression it replaces, encouraging speakers to seek new alternatives (Pinker, 2002, as cited in Steen et al., 2023). In the context of algospeak, this idea is especially relevant because the pressure for lexical innovation may come not only from social attitudes toward taboo subjects but also from perceived technological surveillance. Once an alternative expression becomes recognizable and potentially detectable, users may develop another form. Algospeak can therefore be viewed as a particularly dynamic environment for lexical replacement, in which linguistic innovation may be influenced by the changing relationship between users and moderation systems.

The second perspective is the concept of x-phemism, which encompasses euphemism, dysphemism, and orthophemism as different ways of referring to a socially sensitive or evaluatively marked referent. Moskal and Supernak (2026) apply this perspective to algospeak and show that alternative platform expressions can be understood as part of broader practices of negotiating taboo and social meaning. The value of the x-phemism framework for the present study lies in its ability to distinguish the linguistic form of an expression from the function it may perform. An algospeak item may involve formal alteration while simultaneously softening, disguising, or otherwise reframing the referent. Accordingly, the present analysis does not assume that every algospeak form is necessarily a euphemism; rather, x-phemism provides a broader interpretive framework for considering the relationship between alternative forms and their socially sensitive referents.

The third concept, algorithmic folk theories, concerns the informal understandings users develop about how platforms and their algorithms operate. Because moderation systems are generally opaque and their precise detection criteria are not fully available to users, people may construct working assumptions about which words, topics, or forms are likely to attract moderation. These assumptions do not have to be technically accurate to influence behaviour. As Moskal and Supernak (2026) note, users can adapt their linguistic practices in response to their perceptions of algorithmic control. This perspective is particularly important for the present study because the existence of an alternative form does not, by itself, demonstrate that its creator intentionally sought to evade moderation. Instead, the form can be examined as part of a broader

communicative environment in which users negotiate what they believe platforms will detect or restrict.

Taken together, these perspectives allow algospeak to be approached as both a linguistic and a platform-mediated phenomenon. The euphemism treadmill draws attention to processes of lexical replacement; x-phemism highlights the social and evaluative relationship between a form and its referent; and algorithmic folk theories explain how users' perceptions of opaque moderation systems may influence linguistic choices. The three perspectives are therefore complementary rather than interchangeable. In the present analysis, they are used to interpret the selected items without assuming that a single motivation or theoretical category can account for every form. This approach also makes it possible to consider algospeak as a continuum of modification, ranging from relatively minor formal changes to more substantial lexical, semantic, or visual substitutions.

Methodology

Corpus Selection

The present study draws on a small corpus of twenty documented algospeak items selected from the examples reported by Steen, Yurechko, and Klug (2023). The items were not intended to constitute a statistically representative sample of algospeak across social media platforms. Rather, they were selected purposively to provide sufficient variation for an exploratory analysis of how sensitive referents can be encoded through different linguistic and semiotic mechanisms. Three selection criteria guided the filtering of the larger source set down to twenty items: first, that the item's formation mechanism be clearly identifiable rather than ambiguous between two or more processes; second, that the resulting subset include at least one example of every mechanism and topic domain attested in the source material, so that no single strategy or subject area would be over- or under-represented by chance; and third, that near-duplicate variants encoding the same referent (e.g., multiple respellings of the same word) be consolidated into a single corpus entry rather than counted separately. In particular, the selection was guided by the presence of different formation strategies, including character substitution, phonetic modification, semantic substitution, clipping, acronym formation, emoji substitution, and cultural reference. Attention was also given to variation in the topic domains represented in the corpus, such as health, mental health, politics, violence, substance use, and reproductive health. This combination of mechanisms and domains makes it possible to examine algospeak not as a single, uniform process, but as a set of alternative encoding practices that differ in both linguistic form and degree of modification. The twenty items therefore serve primarily as an analytically diverse corpus through which recurring patterns and illustrative cases can be identified and discussed, rather than as a basis for frequency claims about algospeak as it is used across social media generally.

Coding Procedure

Each of the twenty algospeak items was coded according to two primary analytical dimensions: formation mechanism and topic domain. The first dimension captures the linguistic or semiotic process through which the conventional form of the referent is modified or replaced,

whereas the second identifies the sensitive subject area to which the referent belongs. The two dimensions were coded independently so that the linguistic structure of an item would not be conflated with the topic it encodes.

For the formation mechanism, each item was examined in relation to its conventional referent and assigned to the category that best described the principal process involved in its construction. The coding categories identified in the corpus were character substitution, phonetic strategy, semantic strategy, clipping, emoji substitution, acronym formation, and cultural reference. Character substitution was used when letters or characters were replaced by visually similar symbols or alternative characters. Phonetic strategy referred to forms that relied on pronunciation, phonetic resemblance, or sound-based alteration. Semantic strategy covered cases in which the original referent was expressed through a different lexical or descriptive expression. Clipping was assigned to shortened forms of existing expressions, while acronym formation covered forms created from the initial letters of a longer expression. Emoji substitution referred to cases in which a visual symbol or emoji was used in place of a conventional lexical item. Cultural reference was used where an established cultural association provided an indirect way of referring to the original concept. Where an item could plausibly be associated with more than one process, the category representing its most salient formation mechanism was retained for the descriptive coding, while secondary interpretive possibilities were addressed in the qualitative analysis.

The second dimension, topic domain, was used to classify the referent according to the sensitive or potentially moderated subject matter it represents. The categories emerging from the corpus included reproductive health/politics, disability/identity, health, platform mechanics, race, violence, politics/conspiracy, mental health, substance use, body image, and emotion/politics. Topic-domain coding was based on the referent rather than on the surface form of the algospeak item. Thus, an item could belong to a health-related domain even when its linguistic form contained no overt indication of that topic.

In addition to these two primary coding dimensions, the subsequent qualitative analysis considers the degree and type of modification separating each algospeak form from its conventional referent. This dimension is treated as an interpretive continuum rather than as a numerical variable. At one end are relatively minor modifications that preserve much of the original form, such as character substitution or clipping. At the other end are more substantial forms of lexical, semantic, or visual replacement in which the conventional expression is considerably altered or absent altogether. This additional distinction allows the analysis to examine not only which mechanisms are used, but also how far an algospeak form may depart from its referent while remaining interpretable within a particular communicative context.

Table 1. Filtered corpus ($n = 20$), adapted from Steen, Yurechko, & Klug (2023, Table 1)

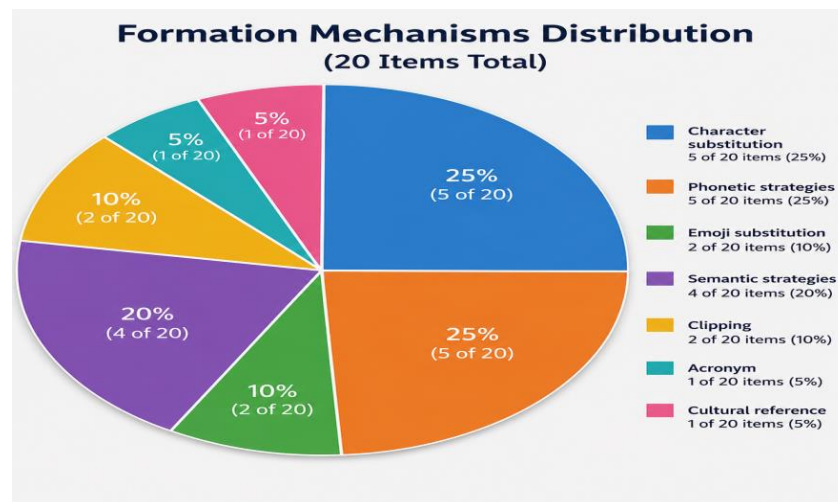
Algospeak term	Referent	Domain	Formation mechanism
@b0rt!0n	abortion	Reproductive health/politics	Character substitution

Algospeak term	Referent	Domain	Formation mechanism
auti\$m / tism	autism	Disability/identity	Character substitution
Backstreet Boys reunion tour / panda express / Panini / panorama	COVID-19 pandemic	Health	Cultural reference
the vid	COVID-19 pandemic	Health	Clipping
blink in lio / 👉 in bio	link in bio	Platform mechanics	Phonetic strategy
blk / wh1te / yt / ○ / ●	Black / White	Race/identity	Clipping
bl00d	blood	Violence (mild)	Character substitution
clock app	TikTok	Platform mechanics	Semantic strategy
cue anon	QAnon	Politics/conspiracy	Phonetic strategy
depressi0n	depression	Mental health	Character substitution
ED	eating disorder	Mental health	Acronym formation
fake body	N/A	Body image (ambiguous referent)	Semantic strategy
kerm1t sewer slide / sewer slide / unalive	suicide	Mental health/self-harm	Phonetic strategy
kl1l	kill	Violence	Character substitution
not see	Nazi	Politics/hate	Phonetic strategy
opposite of love	hate	Emotion/politics	Semantic strategy
Ouid	weed	Substance use	Phonetic strategy
swimmers	vaccinated people	Health	Semantic strategy
UA (flag emoji)	Ukraine	Politics/current events	Emoji substitution
❄️ (snowflake emoji)	cocaine	Substance use	Emoji substitution

Comparison of the algospeak items with their referents reveals variation in the linguistic mechanisms used to encode them.

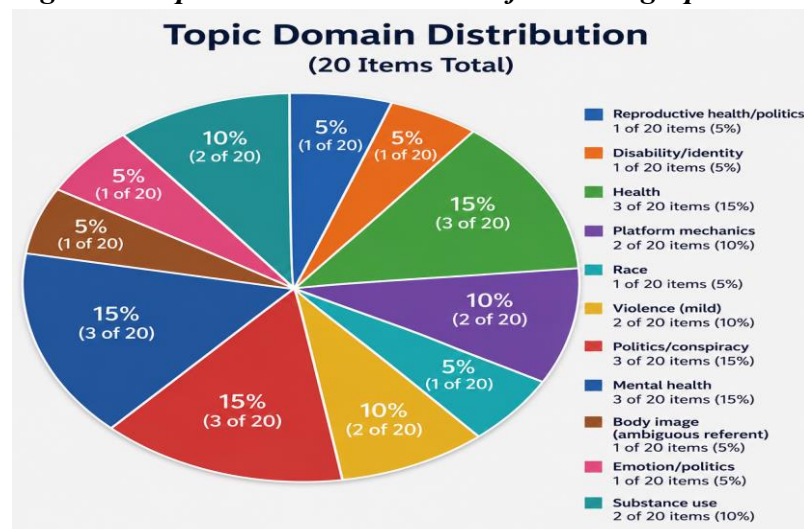
As Table 1 and Figure 1 show, character substitution and phonetic strategies were the most frequent formation mechanisms in the corpus, with five items each (25%). Together they account for half of the corpus, suggesting that many algospeak forms preserve substantial aspects of the original expression while introducing sufficient formal variation to distinguish them from their conventional spellings. Semantic strategies accounted for four items (20%), while emoji substitution and clipping each accounted for two items (10%). Acronym formation and cultural reference were the least frequent mechanisms, with one item each (5%). These figures describe the composition of this small corpus and should not be interpreted as estimates of the prevalence of these mechanisms across social media more generally.

Figure 1. Formation mechanisms distribution of the 20 algospeak items



As Figure 2 shows, mental health, politics/conspiracy, and health each account for 15% of the corpus, with the health category largely reflecting COVID-19-related terms. This suggests that especially sensitive topic areas may be associated with a higher likelihood of automated moderation, which could help explain why these domains attracted a larger number of coded items in the corpus. Platform mechanics, violence, and substance use each accounted for 10% of the corpus, while reproductive health/politics, disability/identity, race, body image, and emotion/politics each comprised a single item (5%). The distribution suggests that algospeak is employed across a diverse range of socially sensitive domains rather than being restricted to a single topic. Although the present corpus cannot establish platform-wide tendencies, it illustrates that linguistic innovation occurs wherever users perceive a need for indirect or alternative forms of reference.

Figure 2. Topic domain distribution of the 20 algospeak items



Notably, several referents (e.g., COVID-19 and autism) are represented by more than one formation mechanism. This suggests that a particular topic does not determine a single linguistic

strategy; instead, users may draw on different forms depending on community conventions, communicative purpose, or the platform context.

Table 2

Algospeak	@b0rt!0n
Referent	abortion
Formation mechanism	Character substitution (creative respelling)
Domain	Reproductive health/politics
Degree of modification	Moderate modification
Communicative function	Discussing a sensitive topic without triggering moderation
Interpretation	Letters are replaced by visually similar symbols or numbers. The result remains interpretable to human readers while differing from the conventional spelling.

In the domain of reproductive health, @b0rt!0n illustrates how algospeak can encode a politically sensitive referent through orthographic manipulation. Letters are replaced with visually similar symbols and numbers, producing a form that remains readily interpretable to human readers while differing formally from the conventional spelling. Such creative respelling implies an indirect reference that may be intended to reduce the likelihood of automated detection, although the present dataset does not allow that intention to be established directly. The example is therefore better understood in relation to algorithmic folk theories: users may develop assumptions about what forms are likely to attract moderation and modify their language preemptively. Whether or not those assumptions accurately reflect the platform’s actual moderation procedures, they can shape users’ linguistic choices and help maintain communication within communities discussing abortion and related reproductive-health issues.

Table 3

Algospeak	cue anon
Referent	QAnon
Formation mechanism	Phonetic strategy
Domain	Politics/conspiracy
Degree of modification	Moderate modification
Communicative function	Maintaining discussion of the topic while avoiding explicit mention of the group’s name
Interpretation	The initial letter in “QAnon” is represented through its pronunciation, while the remainder of the expression is retained. The form is orthographically altered but remains recognizable to users familiar with the reference.

The item “cue anon” illustrates a different strategy within the politics/conspiracy domain. The initial letter in “QAnon” is rendered through its pronunciation, while the remainder of the

expression is retained. The resulting moderate modification is orthographically different but remains interpretable to users who know the reference; its recognizability therefore depends partly on shared community knowledge. Rather than demonstrating that the term is objectively subject to moderation, the example can be interpreted as evidence of an algorithmic folk theory: users may believe that altering the visible form of a politically sensitive name reduces the risk of automated detection. The form thus combines linguistic creativity with an assumed understanding of platform moderation.

Table 4

Algospeak	kermit sewer slide / sewer slide / unalive
Referent	suicide
Formation mechanism	Phonetic strategy
Domain	Mental health
Degree of modification	Total replacement
Communicative function	Softening the directness of the expression by avoiding the exact concept of suicide
Interpretation	These expressions replace the conventional wording with indirect forms. They may soften the directness of the reference and may also be used where users perceive a risk of automated moderation.

In the mental-health domain, the forms “kermit sewer slide,” “sewer slide,” and “unalive” illustrate the interaction between euphemistic softening and perceived moderation risk. These expressions avoid the conventional wording of suicide and may therefore reduce the directness of the reference for some audiences. At the same time, their use has been widely associated with attempts to navigate platform moderation. The present corpus, however, provides no direct evidence about the intentions of the users who originally produced the forms. It is therefore more cautious to describe them as potentially serving both interpersonal and platform-oriented functions. In this respect, the examples can be discussed through the euphemism framework as well as through algorithmic folk theories, particularly where users believe that alternative wording may reduce the likelihood of moderation.

Table 5

Algospeak	clock app
Referent	TikTok
Formation mechanism	Semantic strategy
Domain	Platform mechanics
Degree of modification	Total replacement
Communicative function	Avoiding explicit mention of the app’s name, potentially to protect the reach of a video
Interpretation	The term draws on an auditory or conceptual association between a ticking clock and the sound of the word “TikTok.”

In this example, “the clock app” illustrates a semantic strategy based on an association between the referent and the sound or name of the platform. The algospeak item is entirely different in form from its referent: instead of altering characters in “TikTok,” users refer to the platform indirectly through a descriptive expression. The form can therefore function as a way of avoiding explicit mention of the platform name, although the available data do not establish whether moderation avoidance was always the motivating factor. The example demonstrates how semantic substitution can produce an indirect but recognizable reference within a community.

Table 6

Algospeak	❄️ (snowflake emoji)
Referent	cocaine
Formation mechanism	Emoji substitution
Domain	Substance use
Degree of modification	Total replacement
Communicative function	Referencing the substance visually rather than lexically
Interpretation	The resemblance is visual: the white, powder-like colour of the emoji parallels the appearance of the substance.

The use of an emoji instead of a word to encode meaning illustrates a visual strategy. Representing “cocaine” with a snowflake emoji creates an association between the white appearance of the substance and the colour of the snowflake. The two do not otherwise share obvious semantic properties, so the emoji functions primarily as a visual symbol rather than as a phonetic or orthographic substitution.

Table 7

Algospeak	ED
Referent	eating disorder
Formation mechanism	Acronym formation
Domain	Mental health
Degree of modification	Substantial modification
Communicative function	Discussing the topic more discreetly
Interpretation	The acronym reduces the likelihood of triggering automated moderation while remaining recognizable within relevant communities.

The algospeak term “ED” refers to “eating disorder,” a topic classified here within the mental-health domain. The acronym provides a shorter and less explicit form than the full expression and may also be perceived as less likely to trigger automated detection. However, because the corpus does not contain user-level evidence, the analysis cannot establish that moderation avoidance was the sole or primary motivation. The example can instead be situated within the broader concept of

algorithmic folk theories, in which users develop assumptions about platform sensitivity and adapt their linguistic choices accordingly.

Table 8

Algospeak	the vid
Referent	COVID-19 / the virus
Formation mechanism	Clipping
Domain	Health
Degree of modification	Substantial modification
Communicative function	Shortening the expression for casual, everyday discussion
Interpretation	The clipped form suggests conversational economy rather than a specifically moderation-oriented motivation.

In the health domain, “the vid” is a clipping associated with “the virus” or COVID-19. Unlike several other examples in the corpus, the form does not necessarily signal an attempt to evade moderation; it may simply reflect conversational economy, informality, or a preference for a shorter expression. This distinction is important because not every abbreviated or indirect form should automatically be classified as an evasive strategy. On the evidence available here, “the vid” is better treated as a case of lexical shortening whose communicative function may overlap with, but cannot be reduced to, euphemistic softening.

Collectively, the selected cases demonstrate that algospeak does not rely on a single linguistic process. Rather, users employ a range of orthographic, phonetic, lexical, semantic, and visual resources to encode sensitive referents. These mechanisms differ not only in their linguistic form but also in the degree of modification separating the algospeak item from its conventional referent. While some forms remain highly transparent, others require greater contextual or community knowledge for successful interpretation.

Discussion

Rather than constituting a single strategy of lexical avoidance, algospeak operates along a continuum of linguistic and semiotic modification, ranging from minimal orthographic alteration to complete visual replacement. This continuum is reflected in the diversity of formation mechanisms identified in the present corpus. Character substitution and phonetic strategies are the most frequent mechanisms, each accounting for 25% of the dataset, followed by semantic strategies (20%). Together, these findings suggest that algospeak users tend to preserve sufficient elements of the original referent to maintain intelligibility while simultaneously introducing enough formal or semantic variation to distinguish the alternative form from its conventional counterpart. The comparative rarity of full semantic or emoji replacement, by contrast, may reflect a trade-off: forms that depart further from the conventional referent are less likely to be flagged by keyword-based moderation, but they also place a greater interpretive burden on the reader, since recognition depends on shared community or contextual knowledge rather than surface similarity

to the original word. Character substitution and phonetic strategies may therefore be favoured precisely because they sit at a point on this continuum where detectability is reduced without intelligibility being seriously compromised — a balance that is consistent with users acting on algorithmic folk theories about what moderation systems are likely to catch, rather than simply following the most linguistically economical option. From an x-phemistic perspective, these forms illustrate different ways of negotiating socially sensitive meanings, without implying that every algospeak item should be classified exclusively as a euphemism, dysphemism, or orthophemism.

The distribution of topic domains further demonstrates that algospeak extends across a wide range of sensitive subjects rather than being confined to a single area of discourse. Mental health, health, and politics/conspiracy each account for 15% of the corpus, while the remaining items are distributed across domains such as violence, platform mechanics, substance use, reproductive health, disability, race, body image, and emotion. These proportions should be interpreted as descriptive characteristics of the selected corpus rather than as evidence of platform-wide moderation practices or the relative prevalence of particular topics across social media.

Finally, the qualitative analysis indicates that the relationship between an algospeak form and its referent is not determined solely by the formation mechanism employed, but also by the degree and type of modification involved. While some items remain highly transparent through minor orthographic or phonetic alterations, others achieve greater linguistic distance through lexical substitution, semantic reformulation, or visual representation. The emoji-substitution cases are particularly noteworthy because they replace conventional lexical forms with visual symbols, demonstrating that algospeak may extend beyond lexical innovation into broader semiotic practices. These findings reinforce the view that algospeak is a flexible communicative resource whose functions may include indirect reference, community recognition, stylistic creativity, and perceived moderation avoidance, depending on the communicative context.

Taken together, the formation-mechanism and topic-domain findings point to the same underlying dynamic from two different angles. The mechanism distribution shows how users modify a form; the domain distribution shows where that pressure is concentrated. Read side by side, they suggest that algospeak is not evenly distributed noise across social media language but a response that clusters where users perceive moderation risk to be highest — mental health, politics/conspiracy, and health chief among them in this corpus — and that, within those domains, users gravitate toward mechanisms offering the best available compromise between evading detection and remaining understood. This is consistent with algorithmic folk theories functioning as the connective mechanism between the two dimensions: users' beliefs about what topics attract moderation shape which referents get encoded at all, while their beliefs about what forms attract moderation shape how those referents are encoded. Testing this connection directly would require data this corpus cannot provide — in particular, evidence of users' stated motivations and of actual moderation outcomes for the forms used — but it offers a plausible account of why the same small set of mechanisms recurs across such a varied set of topics.

Limitations

- Sample size: 20 terms demonstrate the method but cannot support statistical generalization — a full-scale study would need 100+.
- Secondary-source corpus: terms are catalogued after they became widely known, so very recent or fast-dying coinages are underrepresented.
- No frequency-of-use data: this design shows what forms exist, not how often each is actually used; a platform-data or survey layer would be needed for that.

References

- Lorenz, T. (2022). Internet ‘algospeak’ is changing our language in real time. *The Washington Post*.
- Moskal, M., & Supernak, K. (2026). Censorship in the age of social media and algorithms: A linguistic study of TikTok algospeak in light of computer-mediated communication, x-phemisms, and taboos. *Studia Neophilologica*.
- Steen, E., Yurechko, K., & Klug, D. (2023). You can (not) say what you want: Using algospeak to contest and evade algorithmic content moderation on TikTok. *Social Media + Society*.