A New Semantic Artifact Based Framework for Studying and Documenting Algospeak and Related Phenomena

Anas Fahad Khan[1], **Elisa Gugliotta**[4], **Elisa Squadrito**[1,2], **Maura Tarquini**[3], **Francesca Frontini**[1]

1.CNR-ILC, 2.Università degli Studi di Macerata, 3.Università degli Studi di Sassari - DiSSUF, 4.Universitè Grenoble Alpes - LIDILEM

{fahad.khan,elisa.squadrito,francesca.frontini}@ilc.cnr.it, {mtarquini,egugliotta}@uniss.it

Abstract

In this paper we present a new framework for analysis, documenting and publishing resources about the recent linguistic phenomenon of *algospeak*. This proposed framework features the use of two semantic artifacts (both of which we make available as SKOS semantic artifacts in RDF), and a cross-lingual lexicon of algospeak terms which follows a schema intended to facilitate the comparison of algospeak across languages and cultural contexts. Our article also features a discussion of the use of algospeak in two non-anglophone contexts (Italian and Arabic) which resulted from a period of data collection which the authors undertook as preparation for the creation of our framework and the categories which underlie it.

Keywords: algospeak, semantic artifacts, arabic, italian, social media, anti-language, emoji

1. Introduction

Algospeak is a form of coded language or linguistic self-censorship used to bypass content moderation algorithms on social media platforms[1]. It is "commonly understood as abbreviating, misspelling, or substituting specific words, [...] when creating a social media post with the particular goal to circumvent a platform's content moderation systems" (Steen et al., 2023). Algospeak is most commonly associated with Tiktok but has also become increasingly common on other platforms such as Reddit, Instagram and Youtube[2]. Beyond this, a number of instances of algospeak have entered into wider cultural circulation and/or have begun to be used for reasons other than circumventing social media platform content moderation systems.

Well known examples of algospeak are *p0rn* and 🍅, both stand ins for the word *porn*; the 🍉 emoji for *Palestine*; and *unalive* for *dead*. Algospeak can frequently be as simple as the substitution of numbers for letters, letters for letters, or the substitution of emojis for words or morphemes, and in this it is similar to an earlier kind of internet language, namely, *leetspeak*. However it can often be more sophisticated than that – sophisticated at least from a linguistic point of view – by playing on phonological similarities between words or making use of speakers' morphological and lexical knowledge, and combining these together with other (often obscure) socio-cultural allusions. It is no wonder, then, that despite being a relatively new phenomenon, algospeak has already been the subject of numerous articles and at least one book (Aleksic, 2025).

In the current article, we will propose an approach that uses language resources to document, analyse and compare algospeak and related linguistic phenomena across socio-cultural and linguistic contexts. More precisely: in this submission we introduce a conceptual framework that will facilitate the documentation, analysis and comparison of algospeak using a FAIR language resource

All authors contributed to the study and the linguistic resources described. The development of the two taxonomies described above was largely carried out by Khan. Material preparation and data collection were largely carried out by Khan with respect to the English dataset and by Squadrito with respect to the Italian dataset; the analysis of the Italian examples given below was carried out by Squadrito. Tarquini and Gugliotta were responsible for material prerparation and data collection of the Arabic dataset, whose empirical analysis contributed to the refinement and expansion of initial resources. Frontini provided supervision with regard to FAIR data principles, data management, and ethical considerations related to resource publication and interoperability. The structure of the manuscript was jointly developed by all of the authors. The initial draft was written collaboratively, and all of the authors contributed to reviewing and editing subsequent versions. All authors approved the final manuscript.

[1]Content Warning: This document discusses examples of harmful content (hate, abuse, misinformation and negative stereotypes). The authors do not support the use of harmful language, nor any of the harmful representations quoted below.

[2]Searches for the most common and cross-lingual algospeak examples in the CLARIN Federated Content Search engine show first hits in European reference corpora starting from 2010.

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 11896–11906
11–16 May 2026.

(Wilkinson et al., 2016). This framework is based on the use of semantic artifacts to categorise instances of algospeak according to a number of different dimensions in a multilingual lexicon describing the phenomena across a number of domains. Consultable versions of these resources are currently available in a github repository[3]: https://github.com/anasfkhan81/AlgospeakLex.

The rest of the paper is structured as follows. In Section 2 we give some background on previous work on algospeak and similar linguistic phenomena, pointing out the main inspirations for our particular approach; Section 2.1 expands on this and motivates our decision focus on reusability. In Section 3, we present the framework itself. Section 4.1 reflects on algospeak formation processes and general tendencies in Italian. In Section 4.2 we present studies on the use of algospeak in Arabic, which is described using our framework; further details can be found in the TSV version of the current draft of the lexicon[4].

2. Background

As we have already mentioned, algospeak has given rise to numerous scientific articles over the last few years; this includes, for instance, the work of Fillies and Paschke (2024), who propose a simple taxonomy of algospeak to organize ChatGPT prompts, a framework later adopted by Blažević and Žuvela Blažević (2025). It also includes Klug et al. (2023) who compiled an English algospeak dataset based on keyword searches on TikTok. Alongside these studies, several work have appeared in the popular press and one book in the genre of popular science (Aleksic, 2025). Despite this growing interest and a steady increase in research on algospeak as well as its categorization for hate speech detection purposes, most of the existing literature remains focused on anglophone contexts. As Calhoun and Fawcett (2023) demonstrate, current taxonomies/categorisation schemes are predominantly based on English-derived examples, with limited attention (beyond translingual pictorial representations such as emojis) to algospeak phenomena in other languages. In terms of the previous literature on algospeak, we have been particularly inspired in the current work by two articles which, in our opinion, stand out for their rigorous linguistic analysis of the phenomenon and for the success with which they are able to situate it in a wider social context, namely Calhoun and Fawcett (2023) and Hughes et al. (2024). The former provides a taxonomy of linguistic processes underlying algospeak, which we have adapted as the foundation for the machine-actionable taxonomy described in Section 3.1, while the latter includes a lexicographic appendix that directly informed the design of our language resource. In undertaking the current work, we also sought to make connections with previous literature on euphemism as well as forms of coded-languages associated with marginalised groups.

The necessity of camouflaging the direct meaning of an utterance and/or engaging in linguistic self-censorship has given rise to numerous different linguistic strategies over the ages; it has also given rise to a rich scholarly literature dedicated to the study of these strategies. One concept from this previous scholarship which we have found to be particularly relevant to the current discussion is that of an *anti-language*. This was first proposed by M.A.K. Halliday in the 1970s as a means of describing specialised languages associated with so-called anti-societies: that is, communities existing within some larger community or mainstream society and which are perceived as somehow standing in opposition to this latter. Halliday (1976) proposes a classification of the different linguistic processes underlying anti-languages, a classification which has proven useful to the framework which we have developed and which we describe in more detail in Section 3. We have also been inspired by previous scholarly work in the study of euphemism, another topic which is closely related to algospeak, and in particular by Warren's influential taxonomy of euphemistic processes from 1992 (Warren, 1992). The latter has a lot in common with Halliday's antilanguage taxonomy, and both share numerous similarities with the categorisation proposed in Calhoun and Fawcett (2023). We have attempted to align these different categorisations together as part of our descriptive framework for algospeak, as detailed in Section 3.1.

Additionally, studies in computer-mediated communication, such as Danesi (2016), have provided useful analytical tools for interpreting the use of emojis in algospeak. In this context, they appear to privilege the iconic dimension. Unlike the phatic or pragmatic functions that modulate interpersonal tone, emojis in algospeak rarely serve to manage interaction or relational stance. They may nonetheless retain an emotive or "softening" role, mitigating the opacity of coded language, and they can also reinforce community signaling, as shared emoji codes help insiders identify one another while evading algorithmic detection.

[3] Our intention is to deposit a first complete version of the language resource in a trusted CLARIN repository

[4] https://github.com/anasfkhan81/AlgospeakLex/blob/main/AlgospeakLexicon.tsv

2.1. The Challenges of Delimiting Algospeak

It is not always clear where the dividing line between algospeak and other kinds of internet language lies, especially when that language is used in offensive or derogatory ways; this can make it difficult to delimit exactly what it meant by algospeak. Another related issue here is that social media platforms don't tend to make their lists of 'filter words' public. Indeed, the highly opaque nature of content moderation practices in social media platforms poses numerous other difficulties in this context. Certain common internet terms such as 'chad', 'soy' or 'red pill' aren't generally regarded as offensive in and of themselves (although they can be used in offensive ways), but they can give rise to similar terms which contain deeply racist and sexist connotations, e.g., terms such as 'tyrone' and 'dog pill'. It is difficult therefore to determine what should count as algospeak and what shouldn't. This is especially the case with many of the neologisms used by the so-called *incel* or *femcel* communities, both of which fit under Halliday's description of anti-societies. Elements of our framework, especially the aligned taxonomy described in Section 3.1 are therefore also intended to be (re-)used for the broader task of studying neologisms used by internet anti-societies. This motivated our decision to make our framework modular and to incorporate FAIR resources, in particular by adopting the well known and popular semantic web language for semantic artifacts, SKOS.[5]

Given all of these issues therefore we decided to develop a framework that could be used both for the analysis, documentation and comparison of instances algospeak, but that was also intended to be highly re-usable, and also extensible, for other similar tasks and in other related contexts.

3. A New Framework for Studying and Documenting Algospeak

In this section, we will present our framework for documenting and studying algospeak via FAIR language resources. This framework is intended to offer an intuitive way of organising and sharing data on algospeak via a lexical resource by using previously existing standards and technologies; it includes the provision of a number of semantic artifacts published in SKOS that can be easily re-used both for this and other purposes. Our primary intention is to make these semantic artifacts available to facilitate the creation of lexical resources documenting algospeak and other related linguistic phenomena, but they can also be used to annotate corpora (especially the Taxonomy of Intentions described below). Currently these taxonomies are available via our github repository[6] but we also intend to make them available through an instance of the SKOSMOS platform (designed to provide a use-friendly, easily navigable interface for SKOS resources), and to publish everything under an open license – as well as depositing it in a CLARIN repository. More precisely, our framework proposes the following semantic artifacts to describe (and potentially annotate) instances of algospeak; they are both described in detail in Section 3.1:

- A taxonomy which categorises the different linguistic processes underlying algospeak and related phenomena; this was initially bootstrapped by aligning the categorisation proposed by Calhoun and Fawcett (2023) with two other taxonomies and then expanded;
- A taxonomy of intentions behind the use of algospeak which the authors of this paper have developed.

In addition, we have made available a multilingual lexicon of instances of algospeak which makes use of these two taxonomies to categorise and describe each instance. These resources were developed through an analysis of the literature, however they were also refined and updated as a result of a process of data collection and the analysis of potential examples of algospeak in both Italian (see Section 4.1) and Arabic (see Section 4.2).

In summary, the resources which we have made available with the current work, via github[7], include two semantic taxonomies (the Aligned Taxonomy and the Taxonomy of Intentions) and a machine-readable lexicon of algospeak instances. These are described in more detail in the rest of the paper. We have not yet made available an annotated corpus of algospeak and although the development of the framework was based on data which we collected via several social media platforms and internet forums, our contribution in the current work

[5] It is important to note that practices aimed at circumventing digital censorship are by no means limited to the web-based contexts which we discuss in this paper. In particular, Chinese online communities have long developed sophisticated strategies to evade platform- and state-level moderation, including homophonic substitutions, character decomposition, coded expressions, and creative lexical innovation (Hughes et al., 2024). Although our current case studies focus on Italian and Arabic, the Chinese context represents a crucial precedent for understanding algospeak-like phenomena as part of a broader global ecology of digitally mediated linguistic adaptation.

[6] https://github.com/anasfkhan81/AlgospeakLex

[7] https://github.com/anasfkhan81/AlgospeakLex

lies in providing reusable semantic artifacts rather than (for now) a fully annotated dataset.

3.1. The Aligned Taxonomy and The Taxonomy of Intentions

The purpose of the **Aligned Taxonomy (AT)** is to describe the linguistic processes which underlie to individual instances of algospeak. One of the main uses for the AT is in lexical resources which document algospeak and similar linguistic phenomena – especially those comparing instances of these phenomena across communities and across languages – and this is how we use it in the multilingual lexicon of algospeak described in Section 3.2.

The AT began as an effort at aligning the categorisation of algospeak proposed in Calhoun and Fawcett (2023) (summarised in Table 1) with Halliday's classification of the linguistic processes behind the creation of anti-language terms (Halliday, 1976) and Warren's 1992 taxonomy of euphemism (Warren, 1992), after we had realised how much these different categorisations had in common. After going through the analysis of numerous different examples of algospeak, we ended up modifying this initial taxonomy and adding new concepts, as described below.

Process	Examples
Use of non-letters	n@z1 'Nazi'
Innovative Use of Morphology	unalive 'kill'
Lexical Replacement	accounting 'sex work'
Innovative Phonological Patterns	seggsy, sessy 'sexy'
Intentional Spoonerism	woke smeed 'smoke weed'
Orthographic Reanalysis	knee grow 'negro'
Phonotactic/Prosodic Template	nip nops 'nipples'

Table 1: Calhoun and Fawcett's Classification

The AT is a multilevel SKOS taxonomy on three different levels[8]. It includes provenance information for each of the different concepts which it includes. In cases where a concept may be unclear from its name, we offer examples, where possible from algospeak. The top-level elements of AT are the following:

- **Orthography**: this element is used to categorise linguistic processes that alter the orthography of a word in order to disguise it or to create new terms. Example sub-concepts include, *initialism, use of non-letters*;
- **Morphology**: this element is used to categorise linguistic processes that involve the combination of pre-existing morphemes. Example sub-concepts include: *blending, clipping, affixing*;
- **Phonology**: This element categorises linguistic processes which generate new terms on the basis of phonological properties of words. Example sub-concepts include: *metathesis, intentional spoonerism*;
- **Lexical Semantic Innovation**: This element encompasses different lexical and semantic processes resulting in new terms/senses. Example sub-concepts include: *metaphor, metonymy, loan words, naming*.

Often we can identify more than one process behind a given instance of algospeak and in Section 3.2, we show how these concepts, either individually or in combination, can be used in the description of single instances of algospeak.

Even though algospeak was initially meant to circumvent internet filters on social media platforms there might be different motivations behind its use in particular cases; e.g., instances of algospeak that gain wider traction might end up being used for reasons of kudos or in-group signalling. In order to classify these different motivations we have created the **Taxonomy of Intentions (TI)**[9]. The TI is a two level taxonomy encoded in SKOS which describes the intentionality behind the choice of a given algospeak term. This taxonomy can be associated with a particular instance of the use of the term (in annotations for instance), or it can describe the intention or intentions, with which that term tends to be used in general[10]. The TI currently consists of the following concepts[11]:

- **Avoidance of Moderation**

[8]The current version of the AT is accessible here: https://github.com/anasfkhan81/AlgospeakLex/blob/main/aligned_taxonomy.ttl

[9]The TI can be accessed here: https://github.com/anasfkhan81/AlgospeakLex/blob/main/intent_taxonomy.ttl

[10]This is similar, in the latter case, to the assignment of sentiment and polarity in sentiment lexicons and this is how we use the TI in our lexicon of algospeak.

[11]The names of the concepts should be fairly self-explanatory, nevertheless definitions of individual concepts are given in full in the SKOS taxonomy available in our github repository.

  - Non-Offensive Avoidance (NOA)
  - Harmful / Offensive Intentions (HOI)
  - Trauma Sharing (TS)
  - Political Activism (PA)
- **Care-Oriented Intentions**
  - Trigger Mitigation (TM)
  - Community Solidarity / In-group Signal (CS)
- **Playful / Stylistic Intentions**
  - Humour / Irony / Parody (HIP)

We are currently developing these taxonomies as part of continuing collaborative work within task 2.6 of the ENEOLI European COST action[12], however we would like to eventually open both of these taxonomies to crowd-sourcing.

3.2. A Lexicon of Algospeak

As mentioned above, one of the original inspirations behind our intention to develop a lexicon for algospeak was the lexicographic supplement included with Hughes et al. (2024). Building upon this initial inspiration, we decided that we would compile a multilingual lexical resource that could facilitate the cross-linguistic comparison of algospeak across languages and (sub-)cultures and that we would make it available via commonly used standards. In what follows we will describe a first version of this resource which we have made available as a TSV file on our github repository[13], although the intention is eventually to publish this linked data, as an RDF resource[14]. Our lexical data is organised using the following categories[15]:

- **Language** (Obligatory): This can either be a specific language or languages (when a given instance has been used in more than one language); it can also be 'translingual'.
- **Form** (Obligatory): this is the equivalent to the lemma form in a more traditional lemma although it may include emojis e.g., 🍁 or *p🍁rn*.
- **Base Form** (Optional): When the algospeak instance derives from a identifiable word via some linguistic process we cite the original form (in most cases this will be a lemma form); this allows us to trace back the different variants of the same form e.g., for *p🍁rn* it would be 'porn'. The base form may be a phrase in the case of an initialism, e.g., *thot* standing for 'that ho over there'.
- **Evoked Concept** (Optional): One or more concepts which the term evokes, preferably referring to an ID in some Semantic Web ontology such as DBPedia.
- **Part of Speech** (Optional): The part of speech of the algospeak instance where relevant. For instance the use of the *(())* in antisemitic language (one of the examples in our lexicon) does not necessarily correspond to a traditional part of speech.
- **Grammatical Note** (Optional): Allows for the addition of extra grammatical information or context if needed.
- **Processes** (Obligatory): This is a description of the linguistic processes underlying each instance of algospeak using the AT, described above. In many cases there will be more than one such process involved. When there is a clear order in the application of processes we include the '>' symbol between process names; more details are given between square brackets. Otherwise we separate process names with a ',' or a '&'.
- **Example** (Optional): An example of the use of the instance of algospeak in a phrase or sentence; where available a link to a relevant wiktionary entry. At the moment we are looking into the best way to document Tiktok/Instagram/etc videos via screenshots.
- **Definition** (Optional): A definition of the instance of algospeak where this isn't clear from the word itself (in cases when it is a mere substitution of letters in a word or phrase which would clearly
- **Synset** (Optional): An existing synset (in the given language) to which the word (in the sense referred to) belongs[16].

[12] https://eneoli.eu/

[13] This is available at https://github.com/anasfkhan81/AlgospeakLex/blob/main/AlgospeakLexicon.tsv. For legal and ethical reasons, this version of lexicon contains abstracted and processed representations rather than raw social media content (see Ethics Statement).

[14] As of the time of writing, existing RDF vocabularies were not sufficient to capture all of the salient parts of our data. We plan to propose a number of new properties and classes to extend existing vocabularies including both OntoLex and lexinfo.

[15] This list of categories is also part of the overall framework which we describe in this paper.

[16] Algospeak (in common with anti-languages and other kinds of euphemism) demonstrates over-lexicalisation, meaning we have potentially very many linguistic means of referring to the same concept (often one that's considered taboo); this suggests the usefulness of the concept of a synset for organising lexical data on algospeak, Synsets can also then be used to link together different instances of algospeak via the Wordnet Interlingual Index (ILI).

- **Community Guidelines** (Optional): The relevant community guideline(s) *usually* being broken in uses of the instance of algospeak. Here we re-use the categories suggested in Calhoun and Fawcett (2023): **Hateful Speech and Ideologies** (HSI), **Suicide, self-harm, violent acts** (SSVA), **Illegal activities and regulated goods** (IARG), **Adult nudity and sexual activites** (ANSA).
- **Intention** (Optional): the usual intention behind the use of an instance of algospeak. Here we refer to the TI.

Take, for instance, the verb *rope* which is used in algospeak to mean 'to commit suicide (especially by hanging)': this has the evoked concept 'suicide'; it is a result of a metonymic shift followed by a word class shift, this is represented as '*metonymy [instrumental, use of rope for committing suicide] & word class shift*'. We describe it as *SSVA* according to the Community Guidelines classification we have adopted; and we associate it with the Trauma Sharing (TS), and Community Solidarity (CS) elements of the TI taxonomy. The lexicon is a work in progress and currently comprises 179 documented occurrences across seven languages, including but not limited to English, Italian, and Arabic (see Table 2). As part of a broader effort to document algospeak in multilingual settings, the dataset also includes Croatian and Portuguese and remains open to further additions. However, Italian and Arabic constitute the primary focus of analysis of the current article.

Language	Coverage
English	39.1%
Italian	36.3%
Croatian	11.7%
Arabic	9%
Portuguese	2.2%
Translingual	1.7%
TOTAL	100%

Table 2: Lexicon Composition Overview

which we plan to expand via crowdsourcing using github following a similar workflow to that used by the Open English Wordnet Project[17]. The lexicon and the kinds of description which it proposes is further described in Section 3.2.

4. Algospeak: Two Non-Anglophone Case Studies

Our framework has been developed with the aim of facilitating the study of algospeak as a complex linguistic and sociological phenomenon cross-linguistically. The semantic artifacts which we have developed have taken into consideration examples of algospeak (and related phenomena) in various linguistic and cultural contexts. Up until now we have been able to collect and document examples in Italian, Portuguese (Brazilian), Croatian and Arabic, along with numerous instances in both British, American English and English as a lingua franca from different Internet communities. In what follows we present a number of examples of algospeak from Italian and Arabic from two qualitative studies carried out by the authors of the current article together with some initial reflections. The examples have been analysed using our semantic artifacts and are described in full in our Algospeak lexicon[18] using the categories given above, in Section 3.2.

4.1. An Italian Case Study

The analysis conducted on Italian social media platforms, mostly TikTok, Instagram, and YouTube, confirms that the use of algospeak in this context is fairly common and that it reflects the same kinds of user perceptions with respect to the risk of being censored by content moderation filters as in anglophone contexts. Also, in common with the general tendency observed for English, most of the instances we retrieved fall under orthographic modification, specifically the use of non-letters.

The motivations behind these examples range from avoidance of moderation to care-oriented intentions. Examples extracted from the lexicon include examples of the following (Italian base lemma given in parentheses and categories taken from the Intention Taxonomy):

- Political Activism (PA): g@z@, IS R4 3LE, gen*cidio 'genocide',
- Non-offensive avoidance (NOA): p4ll3 (palle) 'balls, testicles', pis3llo (pisello) 'penis', 🎻 (violenza) 'violence', 🍆, t3tt3 (tette) 'breasts',
- Harmful/offensive intentions (HOA): stūprø (stupro) 'rape', āmm4zzø (amazzo) 'kill', Ospēdale (ospedale) 'hospital'

The examples *Ospēdale* and *āmm4zzø* are particularly interesting when the contexts in which they were employed are examined more closely. In the first case, *Ospēdale* was extracted from the sentence *Appena beçco te o [censored] finite all'Ospēdale (se siete fortunate)* which roughly translates to 'If I catch you or [censored] you will

[17] https://en-word.net/

[18] available at https://github.com/anasfkhan81/AlgospeakLex/blob/main/AlgospeakLexicon.tsv

end up in the hospital (if you are lucky)'. Here, although the message is clearly one of hate (HOI), even seemingly neutral words such as *becco* 'catch' undergo self-censorship as a form of linguistic hyper-vigilance.

A similar situation occurs with *āmm4zzø*, coupled in the original comment with *trøvø* (from *trovo*, the first-person singular of *trovare* 'to find'). Again, the verb is obfuscated through the use of *ø* despite not being problematic in itself. The same occurs with the other examples documented in the lexicon, as *MU01A [..] F4M3*, which stands for dying of hunger.

Where anglophone contexts frequently reveal greater creativity in the use of different linguistic processes to circumvent filters, our analysis suggests that in Italian social media the perception is that moderation can be bypassed simply by inserting an asterisk or numerals in place of letters.

However, our analysis did yield a few algospeak examples formed through phonetical and morphological strategies. Such cases tended to fall under the non-offensive annotation (NOA) category of intentions, and specifically involved the description of sexual body parts. In this regard, two examples featured in the lexicon are *voolva* and *nepe*. The form *voolva*, attested as an algospeak variant of *vulva*, results from a process of **vowel lengthening** that falls under *Phonology* → *Innovative Phonological Patterns* in the AT taxonomy (concept derived from Calhoun and Fawcett (2023)). The doubling of the vowel *u* into *oo* functions as a graphemic strategy of obfuscation while maintaining recognisability of the base form. In English orthography *oo* often already represents the /uː/ sound (as in *food* or *moon*), making the alteration appear natural within the language's spelling system. In Italian, by contrast, such doubling has no phonological value, so its adoption here signals a purely visual borrowing from English orthographic conventions rather than a sound-based motivation. *Nepe*, instead, is a censored version of *pene* ("penis"), obtained through the phonological process of *adjacent metathesis*, that is, the interchange of two contiguous segments or syllables within a word. The use of metathesis for cryptic purposes is well attested, a famous example being *verlan*, a French argot first documented in the 19th century among robbers, but also *trancorio*[19] and *riocontro*, slang terms used in the provinces of Brescia and Milan to soften insults and functioning as anti-languages. Such examples shows how mechanisms used for decades in real life are now being repurposed and given new life on social media platforms.[20].

A closing consideration concerns the significant influence exerted by the English language and anglophone media on user habits. Italian users not only tend to employ algospeak forms that point to English lemmas, but also adopt usages that originate from foreign contexts. One interesting example of this is the 🍊 emoji, used online to refer to Jewish people. The substitution of letters with this emoji originally relies on the phonetic and graphical similarity between the words *juice* and *Jews*. As a result, the emoji rapidly became a vehicle for anti-Semitic messages circulating in English-speaking online spaces. In the Italian context, however, it has started to be used with the same meaning even though the original phonetic motivation is no longer present in the target language and culture.

4.2. An Arabic Case Study

4.2.1. Linguistic Background

Despite the growing scholarly attention devoted to algospeak, Arabic algospeak remains a largely unexplored territory. Indeed, in Arabic-language digital contexts – and not only in these contexts – sensitive content seems to evade algorithmic moderation with disconcerting ease. Potentially sensitive materials regularly pass unnoticed, for reasons that are manifold yet closely tied to the linguistic landscape of Arabic itself. Concerning the sociolinguistic situation of contemporary Arabic, it is based on a dynamic and fluid repertoire, being a structurally diglossic language (Ferguson, 1959) or, more specifically, a multiglossic one, which challenges any fixed notion of linguistic norm or code univocity (Haeri, 2003, 58–61). It is precisely this structure, articulated across *fuṣḥā* (Modern Standard Arabic, MSA), *ᶜāmmiyya* (vernaculars), hybrid forms, and borrowings, that makes Arabic an extreme case for automated moderation systems. A theoretical framing of the concepts of diglossia, multiglossia, and code-switching in the contemporary Arab world is required, in order to clarify how Arabic linguistic practices offer a paradigmatic example of structural linguistic opacity, highly performative and resistant to computational domestication. The diglossic model proves excessively rigid and hierarchical, and is therefore insufficient to describe the complexity of actual linguistic practices (Hary, 1996; Rosenbaum, 2000). Bassiouney (2009) analysed several instances of alternation between *fuṣḥā* and *ᶜāmmiyya* in Egyptian communication, demonstrating how such practices are imbued with identity and social meaning (Bassiouney, 2006). The scholar

[19] www.odiopiccolo.com/il-trancorio (Last access: 22/10/2025)

[20] The word verlan itself is created by metathesis of "l'envers", in the same way as the words "trancorio" and "riocontro" both stand for "contrario".

introduced the concept of indexicality, namely the capacity of every linguistic choice to index relations of power, belonging, affiliation, or distance, even in seemingly neutral contexts (Bassiouney, 2009, 69–70). Further empirical studies confirming the complexity of the Arabic linguistic repertoire were conducted by Avallone (2022), who examinated the writing practices of Egyptian Arabic. The corpus-based study shows that the absence of a shared orthographic norm does not produce chaos; rather, it generates recurring strategies linked to phonological, pragmatic, and identity-related factors (Avallone, 2022). In this context, orthographic hybridity, code-mixing, and the playful alternation between Arabic and Latin scripts can be understood as usage-based mechanisms of linguistic innovation rather than deviations from a norm 4.2.2. A central aspect of the lexical modernisation of Arabic is the process of adaptation and integration of lexical borrowings from European languages. Arabic has undergone an intense process of linguistic renewal since the nineteenth century, stimulated by the translation of scientific, technical, and administrative works from Europe (Versteegh, 1997, 161–165). The new terms, often derived from languages such as French, English, and Italian, were absorbed through various strategies: semantic calques, phonological adaptations, or neologisms built upon existing Arabic roots[21]. Bassiouney (2020) explored in depth the dynamics of code-switching and code-mixing within Arabic, which challenge traditional Western categories, being characterised by a complex stratification of coexisting and fluidly interacting varieties. While code-switching refers to the alternation between linguistic varieties (*fuṣḥā* and *ᶜāmmiyya*, Arabic and English or French, or different dialects) within the same communicative act, motivated by pragmatic, identity-based, or discursive reasons Bassiouney's code-mixing implies a more stable and profound integration of foreign elements into the Arabic morphosyntactic structure, as in the case of phonologically or morphologically adapted borrowings (Bassiouney, 2013, 30). This type of hybridization has led, and continues to lead, to the creation of neologisms formed in accordance with the morphological rules of the Arabic linguistic system. An illustrative example is the verb تفرّمت 'tafarramat' (to be formatted), derived from the English verb to format and adapted to the Arabic verbal pattern taC$_1$aC$_2$C$_2$aC$_3$a, which traditionally corresponds to a reflexive-passive form. Such integration occurs through two main strategies: on the one hand, the direct adoption of the loanword in its original form, with minimal phonological or orthographic adaptation; on the other, its structural Arabisation, achieved by extracting three consonants perceived as a root and inserting them into the Arabic verbal morphological pattern. For instance, the colloquial verb دوّش 'dawwəš' (to have a shower), originating from the French noun 'douche', has been morphologically integrated into Arabic as a triradical verb on the pattern C$_1$aC$_2$C$_2$aC$_3$a[22]. Instead, the verb تلفن 'talfana' (to phone), from the Italian 'telefonare', entered in MSA as a quadriliteral verb on the pattern C$_1$aC$_2$C$_3$aC$_4$a. Borrowings, as well as alternation, are not perceived as marked by Arabic speakers, due to the normalised multiglossic context that allows them to overlook foreign origins. Whether involving fully adopted loans or root extraction, the morphosyntactic integration within Arabic grammar is complete. In this sense, lexical borrowing is not merely a sign of external contact but a linguistic operation embedded in the strategic management of the Arabic continuum.

4.2.2. A usage-based analysis

To the best of our knowledge, no annotated corpus specifically dedicated to algospeak exists in Arabic. Although several resources focus on Arabic hate speech (Haddad et al., 2019; Ahmad et al., 2024), none of them currently includes annotations for phenomena of algorithmic self-censorship or deliberate linguistic manipulation. Therefore, this section is based on a qualitative analysis of data collected from various social media platforms (primarily TikTok, Twitter/X, and Instagram). Before turning to the qualitative observations, it is worth noting that written practices in Arabic Computer-Mediated Communication are themselves highly innovative. The widespread use of Arabizi, the Latin-based transliteration system employing numerals such as 3 for *ᶜayn* and 7 for *ḥāʾ* has long illustrated how users creatively adapt writing conventions across the Arabic linguistic continuum. These same adaptive tendencies provide the sociolinguistic background against which instances of Arabic algospeak must be interpreted (Bassiouney and Walters, 2020)[23]. From the qualitative observations conducted, it emerges that although explicit content in Arabic is both frequent and easily accessible online, the development of a *fully-fledged algo-*

[21] Many borrowings first appeared in formal contexts e.g., the press, education, and bureaucracy. The coexistence of unadapted borrowings and locally adapted forms created a fertile ground for the ongoing negotiation between prescriptive norms and actual usage (Versteegh, 1997, 164).

[22] Since immediacy and recognisability prevail over purist concerns, integration processes occur both in the standard language and in vernaculars.

[23] Historically, Arabizi was also reported to function as a practical workaround during the Arab uprisings, helping users bypass keyword-based filtering and facilitating rapid cross-platform coordination.

speak system does not appear to be an urgent necessity, nor is it comparable in scale or diffusion to that observed among English-speaking communities. In many cases, the pressure exerted by automated moderation appears to be less strict. Nevertheless, the analysis reveals a number of phenomena that can be mapped onto the categories proposed by the AT, as shown in tables 3 and 4.

A group of examples concerns the transliteration of English loanwords into the Arabic script, as in سكس (səks, 'sex') or السكس (al-səks, 'the sex'), where the English lemma is morphosyntactically integrated through the affixation of the Arabic definite article (**al-*). In other cases, such as أنا اسكس (ʾanā uskus, I have sex), a deeper morphological adaptation can be observed, whereby the loanword is embedded within the MSA verbal structure following the 1st person present-tense pattern (ʔuC$_1$C$_2$uC$_3$), combined with an explicit subject pronoun. A similar case of morphological integration is found in the Arabizi verb *nrapi* (I do rap). Here, the English root 'rap' is embedded within an Arabic verbal template through the addition of the first-person prefix n- (typical of Maghrebi dialects), illustrating a hybrid formation that combines an English lexical base with an Arabic morphological pattern. The resulting form exemplifies loanword adaptation through compounding and prefixation, where the borrowed verb is fully integrated into the colloquial Arabic verbal system, maintaining its foreign lexical identity while conforming to native morphological constraints. Among the strategies for integrating new vocabulary into colloquial Arabic, we find hybrid forms such as *rappeurat*, a French–Arabic blend encoded in Arabizi, formed from the French word *rappeur* ('rapper') and the Arabic suffix -āt (feminine plural). It is an example of morphological code-mixing, in which a foreign lexical base is integrated into the Arabic morphological system. The pluralised form (-āt) is often used ironically or as a group marker, not necessarily as a grammatical plural.

Lang.	Form	POS	Processes
ar	سكس	Noun	Orthography > Adaptating loanw. > Transliteration
ar	السكس	Noun	Morphology > Adapting loanw. > Affixation; Translit.
ar	اسكس	Verb	Morphology > Adapting loanw. > Compounding; Translit.
ar-arabizi	*nrapi*	Verb	Morphology > Adapting loanw. > Compounding

Table 3: Examples of different processes involved in the adaptation of the English words into Arabic.

A set of strategies relies on the use of non-alphabetic elements to obscure graphemes or taboo words, as in كـ🙈س (kəss, vulva) or زبـ🙈ي (zəbbī, my penis). The use of emoji of edited images (with portions masked or overpainted) constitutes a form of visual 'self-censorship' that keeps the content intelligible to the intended audience while making it opaque to moderation algorithms. These practices, which echo graphic-substitution mechanisms already observed in English algospeak, function here primarily as strategies of self-moderation rather than as genuine process of linguistic innovation. As outlined in Table 4, a type of phenomenon concerns semantic and cultural re-elaboration, where substitution operates not on the surface form but on the implicit meaning. Emblematic examples include كرموسة (karmūsa, 'little fig' or lit. 'a unit of figs'), used metaphorically to refer to the female genitalia, and سنّمار (Sinnimār), used to evoke the proverb جزاء سنّمار (Sinnimār reward, ǧazāʾ) and its moral connotation of unjust punishment (Kassis, 1999, 258–259). In such cases, the implicit reference replaces an explicit term, producing a euphemistic or ironic effect analogous to certain forms of algospeak in other languages. In the medical (and specifically sexual) domain, we observe instances of syntagmatic calque and euphemistic attenuation, as in سرعة القذف (surᶜat al-qaḏf, ejaculation speed), الإمتاع الذاتي (al-imtāᶜ al-ḏātī, self-pleasure), or العادة السريّة (al-ᶜada al-sirriyya, the secret habit). These expressions often derive from international lexical models but are re-adapted to the Arabic cultural register, where the need for respect and decorum tends to prevail over the drive to evade algorithmic moderation.

Form	Example	Evoked Concept	Community Guide
سنمار	شارلي كيرك [...] جزاء سنمار فعلا	punishment	SSVA
الهربس (al-hərbəs)	سمعتوا ايه عن الهربس من الخرافات دي؟	medical problems	ANSA
العادة السريّة	العادة السريّة	self-erotism	ANSA

Table 4: Examples of different processes involved in the adaptation of the English words into Arabic. Concerning the Example column, the first sentence means: *Charlie Kirk [...] got the fate of Sinnimār*. The second one means: *What have you heard about herpes from these myths?* The third one means: *How to deal with masturbation*.

5. Conclusion and Future Work

In the current article we have presented a framework for documenting algospeak via a lexical resource. This framework consists of two semantic artifacts and a set of categories for compiling

a lexicon of instances of algospeak (and other related phenomena). In addition we have also described a multilingual lexicon of algospeak which we have presented. This lexicon contains entries in different languages including Arabic and Italian. We have described entries in both these language and given a qualitative analysis of the situation as regards algospeak in both of these languages based on a process of data collection undertaken by the authors of the paper. Our plan in future work is to expand our lexicon by including examples from other languages. A first version of the lexicon and taxonomies will be published and deposited with an open license next year. We also intend to open our resource up to crowd-sourcing. At the time of writing, we are collaborating with other researchers as part of the ENEOLI COST action to extend the current work to other languages and domains. We have already begun to compile a corpus of annotated algospeak examples using our taxonomies which we plan to publish in future while respecting the ethical principles set out in the Ethics Statement given below.

Acknowledgements

This work is based upon work from COST Action CA22126 – ENEOLI (European Network on Extreme Online Linguistic Identities), supported by COST (European Cooperation in Science and Technology). COST (European Cooperation in Science and Technology) is a funding agency for research and innovation networks.

6. Bibliographical References

Ashraf Ahmad, Mohammad Azzeh, Eman Elnagi, Qasem Abu Al-Haija, Dana Halabi, Abdullah Aref, and Yousef Abu Hour. 2024. Arabic hate speech dataset 2023.

Adam Aleksic. 2025. *Algospeak: How social media is transforming the future of language*. Knopf Publishing.

Lucia Avallone. 2022. Spelling variants in written egyptian arabic: Observations on communicative practices in informal writing. *Zeitschrift für Arabische Linguistik*, page 75.

Reem Bassiouney. 2006. *Functions of code switching in Egypt: Evidence from monologues*, volume 46. Brill.

Reem Bassiouney. 2009. *Arabic Sociolinguistics*. Edinburgh University Press.

Reem Bassiouney. 2013. The social motivation of code-switching in mosque sermons in egypt. *International journal of the sociology of language*, 220:49–66.

Reem Bassiouney. 2020. *Arabic sociolinguistics: Topics in diglossia, gender, identity, and politics*. Georgetown University Press.

Reem Bassiouney and Keith Walters. 2020. *The Routledge handbook of Arabic and identity*. Routledge London & New York.

Mara Ruža Blažević and Tania Žuvela Blažević. 2025. Insights into algospeak and sensitive language on tiktok. *Zbornik radova Međimurskog veleučilišta u Čakovcu*, 16(1):40–53.

Kendra Calhoun and Alexia Fawcett. 2023. "They Edited Out her Nip Nops": Linguistic Innovation as Textual Censorship Avoidance on TikTok. *Language@Internet*, 21:1–30.

Thierry Chanier, Céline Poudat, Benoit Sagot, Georges Antoniadis, Ciara R Wigham, Linda Hriba, Julien Longhi, and Djamé Seddah. 2014. The comere corpus for french: structuring and annotating heterogeneous cmc genres. *Journal for language technology and computational linguistics*, 29(2):1–30.

Marcel Danesi. 2016. *The Semiotics of Emoji: The Rise of Visual Language on the Internet*. Bloomsbury Academic, London.

Charles A Ferguson. 1959. Diglossia. *word*, 15(2):325–340.

Jan Fillies and Adrian Paschke. 2024. Simple llm based approach to counter algospeak. In *Proceedings of the 8th Workshop on Online Abuse and Harms (WOAH 2024)*, pages 136–145.

Hatem Haddad, Hala Mulki, and Asma Oueslati[1]. 2019. T-hsab: A tunisian hate speech. In *Arabic Language Processing: From Theory to Practice: 7th International Conference, ICALP 2019, Nancy, France, October 16–17, 2019, Proceedings*, volume 1108, page 251. Springer Nature.

Niloofar Haeri. 2003. *Sacred language, ordinary people: Dilemmas of culture and politics in Egypt*. Springer.

M. A. K. Halliday. 1976. Anti-Languages. *American Anthropologist*, 78(3):570–584.

Benjamin Hary. 1996. The importance of the language continuum in arabic multiglossia. *Understanding Arabic*, pages 69–90.

Brianne Hughes, Kendra Calhoun, Alexia Fawcett, Kelly E. Wright, Benjamin Zimmer, Emily Brewster, Jaidan McLean, and Lynn Zhang. 2024. Among the New Words. *American Speech: A Quarterly of Linguistic Usage*, 99(1):78–90.

Riad Aziz Kassis. 1999. *The book of Proverbs and Arabic proverbial works*, volume 74. Brill.

Daniel Klug, Ella Steen, and Kathryn Yurechko. 2023. How algorithm awareness impacts algospeak use on tiktok. In *Companion Proceedings of the ACM Web Conference 2023*, pages 234–237.

Gabriel M Rosenbaum. 2000. " fuṣḥāmmiyya": Alternating style in egyptian prose. *Zeitschrift für arabische Linguistik*, (38):68–87.

Ella Steen, Kathryn Yurechko, and Daniel Klug. 2023. You Can (Not) Say What You Want: Using Algospeak to Contest and Evade Algorithmic Content Moderation on TikTok. *Social Media + Society*, 9(3):20563051231194586.

Kees Versteegh. 1997. *The Arabic language*. New York: Columbia University Press.

Beatrice Warren. 1992. What euphemisms tell us about the interpretation of words. *Studia linguistica*, 46(2).

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J. G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A. C. 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3:160018.

Ethics Statement

The examples presented may reference hate, harassment, sexual content, or self-harm. We include such material only when essential to illustrate specific linguistic mechanisms. The paper contains a content warning at the beginning, following the HARMCHECK framework (Kirk et al., 2022).

All examples are drawn from publicly accessible online data. No personal identifiers, usernames, or metadata are retained. Because annotation consisted solely of conceptual categorization and masking of examples, the authors served as annotators; no third-party annotators were exposed to harmful content. The legal and ethical framework adopted follows the best practices developed for CLARIN CMC corpora (https://www.clarin.eu/resource-families/cmc-corpora), including reference models such as the CoMeRe corpus (Chanier et al., 2014), which address copyright, platform terms of service, and personal data protection through controlled research access and distribution of processed data only.

A key ethical concern in this research is dual use. While the taxonomy can support the study of online discourse and moderation bias, it might also be repurposed to improve censorship or surveillance. To reduce this risk, in our final release we will (a) publish the taxonomies alongside a minimal set of examples sufficient to demonstrate linguistic phenomena with an open license, while taking into account legal and ethical considerations in consultation with CLARIN K-centres; (b) distribute the complete lexicon under a research-only license, with controlled access granted through authenticated login and monitored usage conditions; and (c) include clear documentation and intended-use statements discouraging enforcement, surveillance, or automated moderation applications.

The analysis seeks to increase transparency about how moderation and linguistic adaptation interact, not to aid automated enforcement.