Linguistic Creativity in the Age of Content Moderation and Filter Word Lists

Fahad Khan[1] & Elisa Squadrito[2]

[1] Consiglio Nazionale delle Ricerche, Istituto di Linguistica Computazionale 'A. Zampolli', Italy

[2]Università di Macerata, Italy

The need to limit offensive speech and content in digital forums has prompted social media platforms to adopt increasingly stringent content moderation policies. These policies are largely enforced via automated detection methods, which do not always take into consideration the context of use of many words, thereby penalising users from marginalised communities or those who wish to discuss topics, e.g., related to sexual health or suicidal thoughts. In response to such policies, internet users have developed numerous linguistic strategies to circumvent automated filters and maintain the visibility of their posts – with many of these being described with the label of *algospeak*. In practice algospeak consists of the use of seemingly innocuous words, phrases, or symbols as neologisms to discuss restricted topics (Isam, 2024; Steen et al., 2023). Equiparable to a cryptolect (Cavaglià, 2023), algospeak has gained the attention of sociolinguists, computational linguistics, and lexicographers, who have proposed typologies to describe its use (Hughes et al., 2024). However, much existing research on algospeak focuses on English, the web's dominant lingua franca, without looking at the phenomena in a wider historical or global context. In order to contribute to the current literature on algospeak the present study concerns itself with the following questions:

1. How do current categorisations of algospeak capture the phenomenon in non-anglophone cultural contexts?
2. How do modern day lexical creation processes used to represent offensive or taboo content online vary from one linguistic/cultural system to another?
3. How does algospeak relate to other linguistic phenomena, e.g., anti-languages (Halliday, 1974), aesopian language (Parry, 1950), described in previous work in historical linguistics and in particular on semantic shift and neologism creation?

To answer these questions, this study will examine the lexical creation processes underlying algospeak across multiple cultural and linguistic contexts, and seek to describe and classify them using theoretical approaches proposed in previous work in historical linguistics.

This study is being carried out both within the ambit of ENEOLI[16] an EU COST action focusing on neology and lexical innovation, from which among other things it draws theoretical and methodological inspiration, as well as the H2IOSC[17] Italian national infrastructural project, which will provide tools, resources and workflows. In particular, as part of ENEOLI, a comparative analysis of convergent and divergent practices involving algospeak and similar practices across several European languages (e.g., Italian, Croatian, and Portuguese) will be carried out focusing on social media platforms such as TikTok, Instagram and Reddit. In addition, we will work on a number of developing resources which will allow for the documentation as well as facilitating cross-comparison of the use of algospeak across languages and cultural contexts as well as with other previously studied historical linguistic phenomena. One instance of this is the use of a pre-existing SKOS-based taxonomy of metaphoric and metonymic semantic shifts[18], produced as part of the H2IOSC project, to organise instances of algospeak (where relevant), in turn this work will be used to enrich the resource with new algospeak based examples.

References

adFrej. (2021, October 20). *Steam lets you check all their words used to filter profanities. (And it's a hilarious lecture)* [Reddit post]. r/Steam. https://www.reddit.com/r/Steam/comments/qccuu8/steam_lets_you_check_all_their_words_used_to/

Cavaglià, R. (2023, February 14). *Algospeak e censura: Il linguaggio online segreto | Siamomine Mag*. Siamomine. https://siamomine.com/algospeak-censura-attivismo/

Halliday, M. A. K. (1976). Anti-languages. *American Anthropologist*, *78*(3), 570–584. https://doi.org/10.1525/aa.1976.78.3.02a00150

Hughes, B., Calhoun, K., Fawcett, A., Wright, K. E., Zimmer, B., Brewster, E., McLean, J., & Zhang, L. (2024). Among the new words. *American Speech*, *99*(1), 78–90. https://doi.org/10.1215/00031283-11186920

Isam, H. (2024). Algospeak and digital culture: Navigating social media challenges. In Z. Rarastesa et al. (Eds.), *Proceedings of the Third International Conference on Communication, Language, Literature, and Culture (ICCoLLiC 2024)* (Vol. 883, pp. 656–669). Atlantis Press SARL. https://doi.org/10.2991/978-2-38476-321-4_50

[16] https://eneoli.eu/
[17] https://www.h2iosc.cnr.it/
[18] https://vocabs.ilc4clarin.ilc.cnr.it/chroskos/figss/en/

Parry, A. (1950). On “Aesopian” language and borrowings from Russian. *American Speech*, *25*(3), 190–196. https://doi.org/10.2307/454317

Steen, E., Yang, L., Feng, P., & Agapie, E. (2023). You can (not) say what you want: Using algospeak to contest and evade algorithmic content moderation on TikTok. *Social Media + Society*, *9*(3), 20563051231194584. https://doi.org/10.1177/20563051231194586