

A Framework for Analyzing Hate Speech Towards Refugees and Asylum Seekers in Israeli Parliamentary Proceedings

Dorielle Lonke, Tomer Shor
November 20, 2025

Abstract

This project aims to develop an annotation framework for expressions of hate speech directed at refugees and asylum seekers in Israel. Hate speech and offensive language have been widely discussed and analyzed in many languages, but still poses a challenge to under-resourced languages such as Hebrew. The topic of refugees and asylum seekers is particularly challenging, as it is further marginalized in public and institutional discourse. As a result, there is a lack of comprehensive resources and tools to investigate these phenomena on a large scale. The proposed project aims to bridge this gap, by presenting an operative conceptual framework, and applying it to sentences from the Knesset corpus, a Hebrew dataset of sentences from the proceedings of Israeli parliamentary discussions. The expected outcome is the first annotated dataset of Hebrew hate speech targeting refugees and asylum seekers, which will contribute to the analysis and detection of hate speech in diverse linguistic and cultural settings.

Keywords: Hebrew NLP, annotation framework, hate speech detection, refugees and asylum seekers

1 Personal Information of Applicant

Name: Dorielle Lonke

email address: doriellelonke@gmail.com

Affiliation: University of Amsterdam, Master of Logic (graduated September 2025)

Residence: Amsterdam, Netherlands

2 Project Description

2.1 Objective and Motivation

Offensive language and hate speech detection have been widely studied in many languages. In low-resourced languages such as Hebrew, there is a scarcity of resources and implementations for investigating these phenomena. In the case of refugees and asylum seekers in Israel this issue is even more pronounced, as this topic is marginalized both in terms of the public interest, as well as institutional attention. This project aims at presenting an annotation framework and the first annotated dataset of hate speech towards refugees and asylum seekers in Israel. We will engage with previous work on hate speech targeting refugees (Arcila-Calderón et al. 2022), as well as Hebrew hate speech (Hamad et al. 2023; Litvak et al. 2021) which addresses the diverse challenges of annotating hate speech in low-resourced languages. In particular, we face the challenges of providing a ground truth for a subjective task – viz., determining offense – that requires also close familiarity with colloquial Hebrew as well as the Israeli political climate. This framework aims at contributing a reliable resource that could be used towards developing more elaborate tools and methodologies for analyzing and detecting Hebrew hate speech. It also aims to shed light on the current institutional attitude which fails to protect, deprives of humanitarian needs and rights, and collectively criminalizes refugees and asylum seekers in Israel (Mandil 2025).

2.2 Methodology

As an initial step, we will devise an annotation framework for capturing the particularities of Hebrew hate speech directed at refugees and asylum seekers. This framework will rely on the guidelines presented by the Hrandt Dink Foundation (2025), as part of a 16-year long media monitoring project that documented expressions of hate speech towards various identity groups in Turkey. These guidelines include an annotation framework consisting of labels for hate speech categories, target groups, and the perceived degree of hate speech. These categories will be supplemented by the framework presented by Ron et al. (2023), which provides mostly overlapping categories for hate speech classification. On the basis of the combined frameworks, we will annotate a selection of sentences pertaining to refugees and asylum seekers, taken from the recently published Knesset corpus – a corpus of over 30 million sentences from all plenary and committee protocols held in the Israeli parliament in the past three decades (Goldin et al. 2025). Using a list of Hebrew keywords we will extract relevant sentences, and provide manual annotations from multiple native Hebrew speakers. The annotations will be independent and inter-annotator agreement will be measured using commonplace metrics such as Krippendorf's Alpha or Fleiss' Kappa.

Employing the Knesset corpus for our task presents several methodological advantages. For one, it is widely encompassing, and the earliest records predate the first wave of arrival of asylum seekers, which began around 2005-2007¹. It could therefore be used towards a diachronic analysis of the political discussions on refugees and asylum seekers, which have become more extreme over the years². Second, it has been annotated with morpho-syntactic information and named entities, and contains meta-information about the speaker's demographic and political properties, which could be incorporated in future research alongside our annotations. In addition, it contains a subset of 5,000 sentences with annotations adhering to the Universal Dependency UD-V2.10 guidelines (De Marneffe et al. 2021). The dependency annotations could serve as a baseline for semantic role labeling, paving a path towards a frame-semantics analysis of the narratives concerning refugees and asylum seekers, taking a similar approach to the one presented by Ryazanov et al. (2024).

3 Relevance to UniDive

This project will benefit immensely from the contents of the training school. Israel consists of much cultural and linguistic diversity, with large groups of speakers of Hebrew, Arabic, Russian, Amharic, Tigrinya, English and Ukrainian, to name a few. This richness also translates to a very polarized and variegated political discourse, which poses challenges to tasks such as ours. Practically speaking, devising parsing and annotation procedures to handle Hebrew text requires additional attention to the syntax-semantics interface. Even simple tasks such as identifying relevant keywords may face the unique challenges posed by Hebrew – a root-based, morphologically rich language, whose vowels are not visible in standard written form (Tsarfaty et al. 2019). Additionally, as a low-resourced language, there are not many benchmarks for evaluating Hebrew large language models. It would be very useful to learn best practices for transforming our dataset into a functional benchmark, which could be used to evaluate existing approaches, and develop new LLM-based approaches for hate speech detection.

4 Project Phase

This project has not started. It is planned to be carried out in collaboration with ASSAF – Aid Organization for Refugees and Asylum Seekers in Israel. It is not affiliated with any Israeli institutions.

¹see *General info about asylum seekers in Israel* n.d.

²This is reflected in the Israeli nomenclature used by politicians to refer to refugees and asylum seekers, namely 'infiltrators', and the corresponding attitudes arising from this framing. See Hochman 2015.

References

- Arcila-Calderón, Carlos et al. (Oct. 2022). “How to Detect Online Hate towards Migrants and Refugees? Developing and Evaluating a Classifier of Racist and Xenophobic Hate Speech Using Shallow and Deep Learning”. en. In: *Sustainability* 14.20, p. 13094. ISSN: 2071-1050. DOI: 10.3390/su142013094. URL: <https://www.mdpi.com/2071-1050/14/20/13094> (visited on 09/20/2025).
- De Marneffe, Marie-Catherine et al. (May 2021). “Universal Dependencies”. en. In: *Computational Linguistics*, pp. 1–54. ISSN: 0891-2017, 1530-9312. DOI: 10.1162/coli_a_00402. URL: https://direct.mit.edu/coli/article/doi/10.1162/coli_a_00402/98516/Universal-Dependencies (visited on 09/20/2025).
- General info about asylum seekers in Israel* (n.d.). en. URL: <https://assaf.org.il/en/refugees-in-israel/>.
- Goldin, Gili et al. (Sept. 2025). “The Knesset corpus: an annotated corpus of Hebrew parliamentary proceedings”. en. In: *Language Resources and Evaluation* 59.3, pp. 2973–3004. ISSN: 1574-020X, 1574-0218. DOI: 10.1007/s10579-025-09833-4. URL: <https://link.springer.com/10.1007/s10579-025-09833-4> (visited on 09/19/2025).
- Hamad, Nagham et al. (2023). “Offensive Hebrew Corpus and Detection using BERT”. In: *2023 20th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA)*, pp. 1–8. DOI: 10.1109/AICCSA59173.2023.10479258.
- Hochman, Oshrat (Oct. 2015). “Infiltrators or Asylum Seekers? Framing and Attitudes Toward Asylum Seekers in Israel”. en. In: *Journal of Immigrant & Refugee Studies* 13.4, pp. 358–378. ISSN: 1556-2948, 1556-2956. DOI: 10.1080/15562948.2014.982779. URL: <http://www.tandfonline.com/doi/full/10.1080/15562948.2014.982779> (visited on 09/20/2025).
- Hrant Dink Foundation (Mar. 2025). *Utilizing AI Against Hate Speech: A Guide to Annotation, Classification, and Detection*. English. Tech. rep. Hrant Dink Foundation, p. 72. URL: <https://hrantdink.org/attachments/article/4413/UTILIZING%20AI%20AGAINST%20HATE%20SPEECH%20A%20guide%20to%20annotation,%20classification,%20and%20detection.pdf>.
- Litvak, Marina et al. (2021). “Offensive language detection in semitic languages”. In: *Multimodal hate speech workshop*. Vol. 2020, pp. 7–12.
- Mandil, Shahar (Apr. 2025). *Forum of Refugee and Asylum Seeker Organizations in Israel, Statement of Opinion Concerning legislative proposals that endanger the lives of asylum seekers and refugees living in Israel*. Tech. rep. Mandil. ASSAF - Aid Organization for Refugees and Asylum Seekers in Israel. URL: <https://assaf.org.il/en/forum-of-refugee-and-asylum-seeker-organizations-in-israel-statement-of-opinion-concerning-bill-for-the-immediate-deportation-of-infiltrators-who-support-the-regime-in-their-country-of-origin-2024/>.
- Ron, Gal et al. (2023). “Factoring Hate Speech: A New Annotation Framework to Study Hate Speech in Social Media”. en. In: *The 7th Workshop on Online Abuse and Harms (WOAH)*. Ron et al. Toronto, Canada: Association for Computational Linguistics, pp. 215–220. DOI: 10.18653/v1/2023.woah-1.21. URL: <https://aclanthology.org/2023.woah-1.21> (visited on 09/19/2025).
- Ryazanov, Igor, Carl Öhman, and Johanna Björklund (Nov. 2024). “How ChatGPT Changed the Media’s Narratives on AI: A Semi-automated Narrative Analysis Through Frame Semantics”. en. In: *Minds and Machines* 35.1. Ryazanov et al., p. 2. ISSN: 1572-8641. DOI: 10.1007/s11023-024-09705-w. URL: <https://doi.org/10.1007/s11023-024-09705-w> (visited on 08/01/2025).
- Tsarfaty, Reut et al. (Nov. 2019). “What’s Wrong with Hebrew NLP? And How to Make it Right”. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations*. Ed. by Sebastian Padó and Ruihong Huang. Hong Kong, China: Association for Computational Linguistics, pp. 259–264. DOI: 10.18653/v1/D19-3044. URL: <https://aclanthology.org/D19-3044/>.