

The take of linguistics on (religion-related) hate speech

Bernd Kortmann

FRIAS, University of Freiburg, Germany

Overall spirit

- showcasing concepts, theories, methods & approaches that linguistics provides for detecting and classifying hate speech and related types of speech which are hurtful and downright dangerous for individuals and entire groups, and which ultimately may lead to societal disruptions
- implementation of linguistic approaches and methodologies may contribute to a better understanding of the phenomenon

Structure

1. Hate speech and related terms / concepts
2. Relevant fields and approaches in linguistics
3. Relevant methods
4. Some major results and lessons, specifically concerning religion-related hate speech

1. Related terms and concepts

- **hate speech** (-> at the top end of a continuum of offensive and potentially personally & societally dangerous speech)
- **cyberhate** (= online variety of hate speech; targeting certain groups on the basis of a common characteristic)
- **cyberbullying** (targeting individuals usually in the setting of a particular community like school, workplace)
- **toxic speech** (Tirrell)
- **dogwhistles** (-> *doublespeak*; roots in Orwell's *Newspeak* and *Doublethink*)
- **fake news** (panel abstract: „open or covert means to disseminate hate speech“)

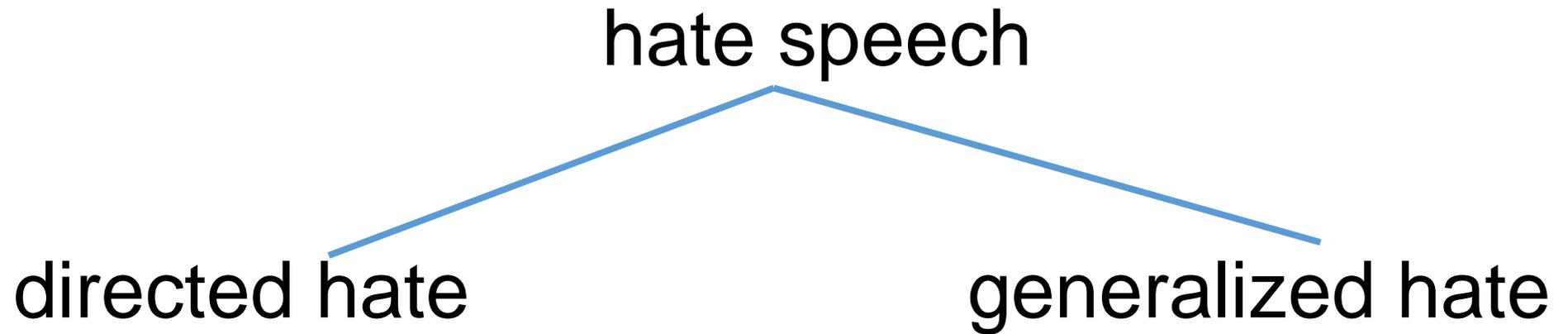
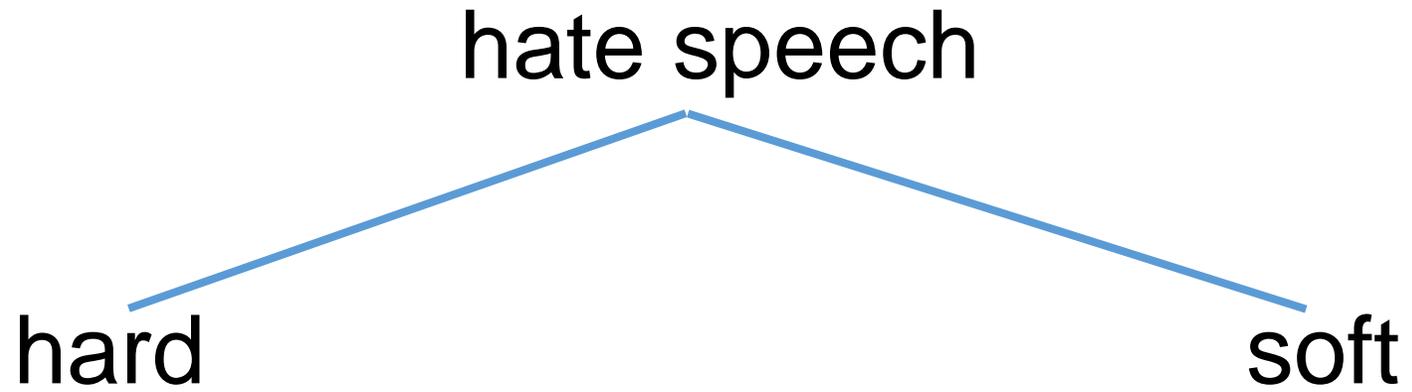
Definitions of hate speech

- no universally accepted definition
- Hate speech ultimately amounts to an “advocacy of discriminatory hatred which constitutes incitement to hostility, discrimination or violence. (UN General Assembly 1966)
- Council of Europe’s Committee of Ministers definition (1997): “all forms of expressions which spread, incite, promote or justify racial hatred, xenophobia, antisemitism or other forms of hatred based on intolerance, including: intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin”
- EU definition (2008): “all conduct publicly inciting to violence or hatred directed against a group of persons or a member of such a group defined by reference to race, colour, religion, descent or national or ethnic origin” (Council of EU)

Definitions of hate speech

- “Hate speech is language that attacks or diminishes, that incites violence or hate against groups, based on specific characteristics, such as physical appearance, religion, descent, national or ethnic origin, sexual orientation, gender identity or other,…” (Fortuna/Nunes 2018: 5 in Vrysis et al. 2021)
- We adopt the definition of hate speech along the same lines of prior literature... **and inspired by social networking community standards and hateful conduct policy (Facebook 2016; Twitter 2016) as**

“direct and serious attacks on any protected category of people based on their race, ethnicity, national origin, religion, sex, gender, sexual orientation, disability or disease”.



hard hate speech

- prosecutable forms prohibited by law

soft hate speech

- lawful but raising serious concerns in terms of intolerance and discrimination (e.g. the “playful” dimension of hate speech, such as codified language in online exchanges among extreme-right groups: *juices* for ‘Jews’, *Jewrope* for *Europe*)
- threshold for distinguishing between hard and soft hate speech differs from country to country = one of the thorniest issues about hate speech
- different legal systems may have altogether different approaches towards regulating and combating hate speech (e.g. USA vs. EU, and even within the EU).

directed hate

- hate language towards a **specific individual** or entity (-> cyberbullying):
 - example: “@usr4your a f*cking queer f*gg*t b*tch”.

generalized hate

- hate language towards a **general group** of individuals who share a common protected characteristic, e.g., ethnicity or sexual orientation:
 - example: “— was born a racist and — will die a racist! — will not rest until every worthless n*gger is rounded up and hung, n*ggers are the scum of the earth!! wPww WHITE America”

- Our analysis reveals that Directed hate speech, in addition to being more personal and directed, is more informal, angrier, and often explicitly attacks the target (via name calling) with fewer analytic words and more words suggesting authority and influence. Generalized hate speech, on the other hand, is dominated by religious hate, is characterized by the use of lethal words such as *murder*, *exterminate*, and *kill*; and quantity words such as *million* and *many*.

Toxic speech (Tirrell 2017, 2018, 2021)

- From an interview with Lynne Tirrell (2017): “Toxicity is the degree in which a substance, often a poison, can harm people. That’s my basic conception, [...] Toxic speech, like any toxin, is a threat to the well-being or even the very lives of those against whom it’s deployed ... The damage can be local or systemic, but toxicity damages all it touches.”
- toxic speech may explicitly use offensive slurs, but more commonly it does not explicitly use derogatory terms and still has toxic effects
- different aspects of toxic speech to be kept in mind:
nature of toxin, dose, shape/way & frequency in which it is fed, susceptibility of different individuals to it

Toxic speech (Tirrell 2017, 2018, 2021)

- Cf. Viktor Klemperer (in *LTI -- Lingua Tertii Imperii*, 1947):
“Nazism permeated the flesh and blood of the people through single words, idioms and sentence structures, [...] What happens if the cultivated language is made up of poisonous elements? Words can be like tiny doses of arsenic: they are swallowed unnoticed, appear to have little effect, and then after a little time the toxic reaction sets in after all.”
- data from e.g. Nazi speeches/propaganda, Rwandan genocide, Trump’s presidency (e.g. on Mexicans)

Dogwhistles

- **definition:** „an utterance whose full meaning is only understood by – at most – a subset of audience members“
- known from political advertising: „implicit political messaging“ with the aim of manipulation

dogwhistles

overt

covert (= default)

use:

+ intentional

use of *Google* to mean 'black person' by a member of far-right groups

- intentional

**amplifier
dogwhistles**

use of *Google* to mean the search engine by someone unaware of the far-right meaning, misinterpreted in an online forum

use of the term *welfare* without intending to activate racial associations [...] These uses are unintentional because the people uttering the term do not intend it to function this way.

+ intentional

deliberate campaign in US to associate the term *welfare* with Black people. [...] This dogwhistle functions without the conscious awareness of those whose racial attitudes are activated by the term.

Note: not only linguistic dogwhistles, also visual dogwhistles (→ semiotics)

Tirrell (personal communication, 2021) on abortion as a dogwhistle for signalling religious stance

“In the US, a lot of the manipulative discourse that triggers religion uses **abortion as a dogwhistle** to rouse up the religious folks against women. Then this circles back to get liberals to think evangelicals are neanderthals. It has become intractable. So it might be useful to look at the way that discourse about abortion rights (well, they discuss it as evil and not a right) is a dogwhistle for signalling a religious stance.”

1. Related terms and concepts

- **hate speech** (-> at the top end of a continuum of offensive and potentially personally & societally dangerous speech)
- cyberhate (= online variety of hate speech; targeting certain groups on the basis of a common characteristic)
- cyberbullying (targeting individuals usually in the setting of a particular community like school, workplace)
- **toxic speech** (Tirrell)
- **dogwhistles** (-> *doublespeak*; roots in Orwell's *Newspeak* and *Doublethink*)
- fake news (panel abstract: „open or covert means to disseminate hate speech“)
- **focus here on studies published in and on the situation of hate speech in the European Union, the US, SE Asia, and in the Middle East**

Stavros Assimakopoulos
Fabienne H. Baider
Sharon Millar

Online Hate Speech in the European Union

A Discourse-
Analytic
Perspective

- The C.O.N.T.A.C.T. project (2015–2017):
- **C**reating an **O**n-line **N**etwork, monitoring **T**eam and phone **A**pp to **C**ounter hate crime **T**actics <https://reportinghate.eu/en/>
- major aim: the identification of hate crime
- in light of
 - the increasing use of the internet as a tool of hate and propaganda,
 - the under-reporting of hate crime,
 - the rise of extremist groups and political parties in the EU
- experts from & studies on 10 EU member states

PHARM: another recent EU hate speech project

- **P**reventing **H**ate against **R**efugees and **M**igrants (2019-)

<https://pharmproject.usal.es/>

- major aim:

monitor and model hate speech in Greece, Spain, and Italy to predict and combat hate crime and also counter ist effects using cutting-edge algorithms

- use of natural language processing mechanisms that identify the textual hate and sentiment load, along with related metadata,...

2. Relevant fields & approaches in linguistics

- pragmatics, meaning in context (key notion of **(speaker)** **intention**, which is also key for identifying hate speech)
said vs meant, literal vs. non-literal meaning, direct vs. indirect speech acts, inferencing mechanisms
- language philosophy
- **cognitive linguistics** (categorization, **metaphors** like the PARASITES metaphor in Nazi propaganda applied to Jews, metonymies, **frames** / Frame Theory)
- **corpus linguistics**
- discourse analysis (especially: **Critical Discourse Analysis**)
- forensic linguistics
- sociolinguistics (language and macro-/microsocial variables, stereotyping)

Frames/frame theory

- Frames are grounded in our cognitive and epistemological knowledge (Busse 2012). They show us how this knowledge has been structured by our previous (linguistic and nonlinguistic) experiences (Barsalou 1992).
- Frequent patterns of usage “represent speaker’s knowledge of their language, including the conceptual structures that motivate language” (Glynn 2010: 89). Therefore, we can hypothesize that concepts such as ISLAMIC are organized in frames that “govern our thought” and “our everyday functioning” (Lakoff and Johnson 1980: 3), and that **the frequency patterns we find of lexical co-occurrences** provide us with data to identify and describe those frames ... (Kopytowska 2009: 4).

Critical Discourse Analysis

- a study of discourse that views language as a form of social practice. CDA takes (non-linguistic) social practice and linguistic practice as constituting one another
- special focus: investigating how societal power relations are established and reinforced through language use

Critical Discourse Analysis

As argued by social constructivists and critical discourse analysis scholars alike (cf. Berger and Luckmann 1966/1991; Fairclough and Wodak 1997; Searle 1995, 2010), **discursive representations** of individuals, groups, events, issues, phenomena and relations **are both constituted by and constitutive of the socio-political status-quo of these entities.** Since contemporary public discourse abounds in messages of hate, and research findings demonstrate that there exists a link between verbal and physical aggression, it seems vital to explore the dynamics of hate speech production and reception in the public sphere in its current mediatised form.

Critical Discourse Analysis

- a study of discourse that views language as a form of social practice. CDA takes (non-linguistic) social practice and linguistic practice as constituting one another
- special focus: investigating how societal power relations are established and reinforced through language use
- CDA explores the linguistic features & organization of concrete instances of discourse, such as choices & patterns in vocabulary, rhetorical figures (e.g. metaphors), grammar, cohesion, etc.

3. Relevant methods (quantitative & qualitative)

- Critical Discourse Analysis (-> triangulation of both qualitative & quantitative methods)

largely or purely quantitative:

- corpus analysis (e.g. n-grams, collocations)
- sentiment analysis & emotion analysis
- algorithms and web interfaces for automatic detection of hate speech

Critical Discourse Analysis -- method

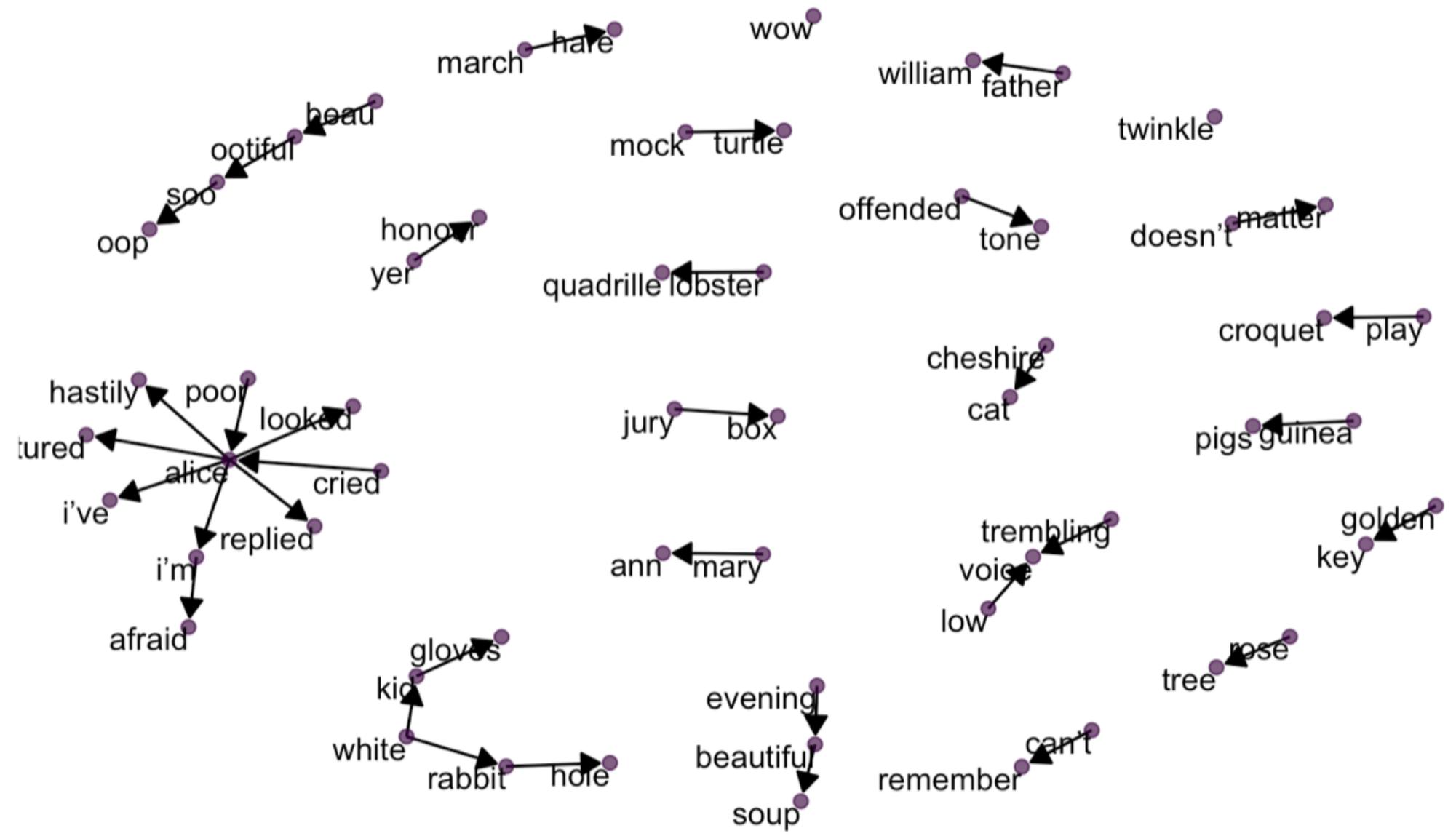
- CDA explores the linguistic features & organization of concrete instances of discourse, such as choices & patterns in vocabulary, rhetorical figures (e.g. metaphors), grammar, cohesion, etc.
- it does so by combining quantitative (corpus-linguistic) methods with qualitative textual and interview analysis techniques
- Assimakopoulos et al. (2017) triangulate the following sources in exploring (primarily soft) hate speech: online comments on news reporting, interviews (of young people) and questionnaires on the perception of hate speech

Quantitative approaches: Sentiment & emotion analysis

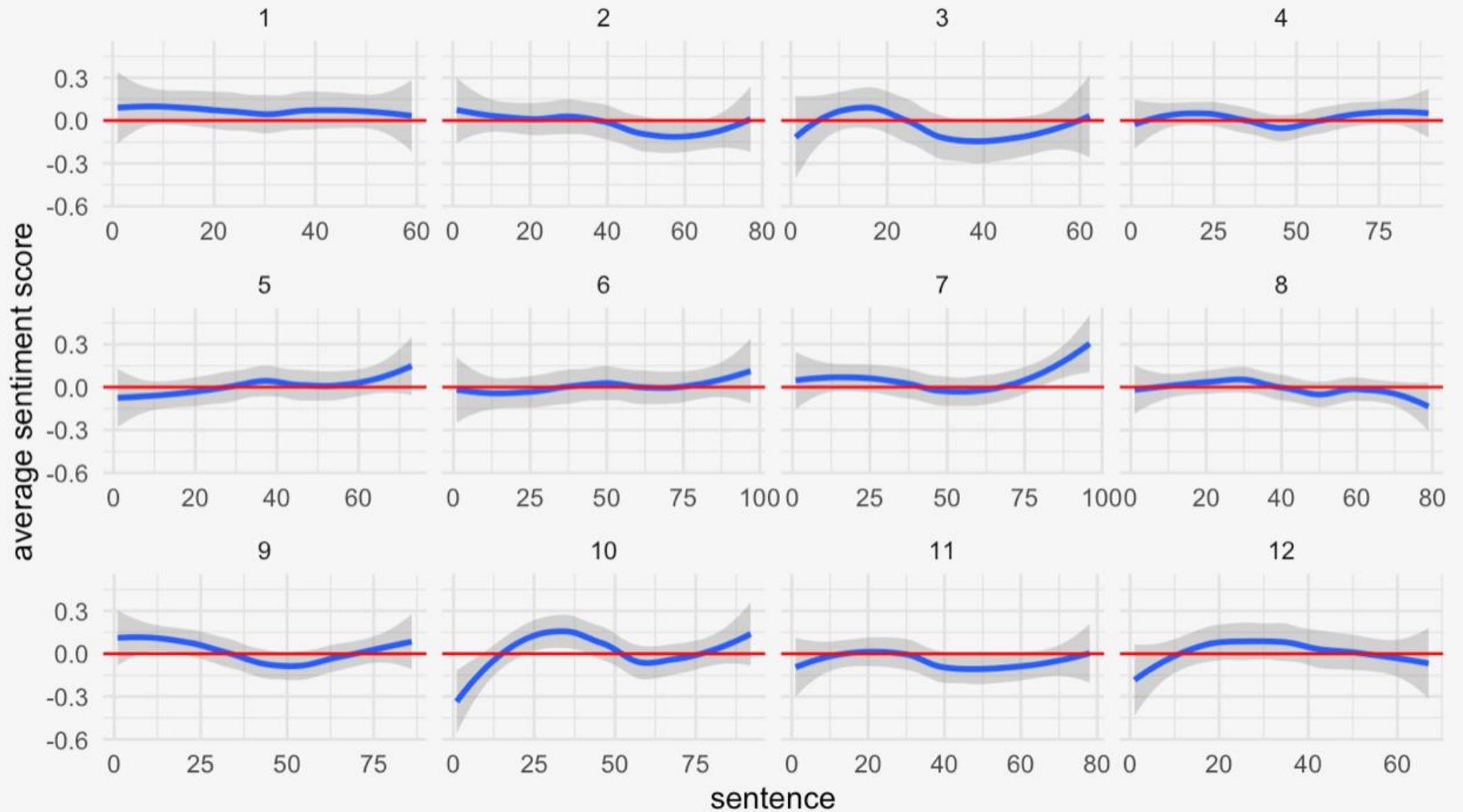
- sentiment analysis:

words or word combinations (n-grams) are looked up in an available, pre-compiled sentiment “dictionary” (often crowd-sourced). Some dictionaries will rate words on a numeric scale (-5 to 5) on an axis of negative/positive sentiment, others have various emotions like joy, fear, disgust, etc. Relative to the same sentiment / emotions dictionary, different corpora, texts, or even book chapters can be compared (e.g. sentiment scores across each chapter of *Alice in Wonderland*).

Bigrams (two-word combinations) in "Alice's Adventures in Wonderland"



Sentiment scores in Alice in Wonderland chapters



Quantitative approaches: Sentiment & emotion analysis

- sentiment analysis
- source data: Big data (typically social media)
- hate speech corpora (e.g. based on Twitter alone or in combination with other social media; for different languages)
- lexical hate speech databases (for different languages)
- emotion analysis: via specialized algorithms and learning mechanisms applied to natural language processing (-> based on dictionaries organized by entries arranged by emotions like joy, fear, disgust, etc.)

4. Some major results and lessons to be learnt

- Note, first of all:

preferred targets of speech in consulted literature

- religion: here primarily Islam / Muslims
- refugees / migrants
- racism / ethnic minorities
- sexual orientation (LGBTIQ)

Assimakopoulos et al. 2017

Metaphors

- Another way to promote anxiety and panic is the use of **metaphors conceptualizing immigration as an invasion and as flooding the country** ... Perceived in this way, migrants and refugees inevitably constitute a threat to the collective Self and the survival of a community as a cohesive unit (cf. Buzan et al. 1998): in the particular setting of the current migration crisis, coming from a predominantly Muslim background, they are likely to bring in beliefs and traditions incompatible with the European Christian worldview.
- This stance gains even more relevance in the case of **ethnically and religiously homogenous societies**, such as the Polish one. (-> especially in Poland: “refugee-related hate speech with ‘patriotic’ undertones”)

- **Metaphors can reveal the underlying conceptual frame** of their producer and give access to a set of assumptions made by competent members of a discourse community about **the ‘typical’ aspects** of a member of a minority or any person belonging (or appearing as belonging) to that group. This then leads to **the conceptualisation of metaphors as creating or confirming stereotypes** (Zinken 2003).
- Cyprus: xenophobic metaphors are used to construct the social Other in social media.

Table 3.4 Recurrent metaphors used for Othering immigrants in the Cypriot C.O.N.T.A.C.T. corpus

Metaphor of	Example
DISEASE	Refugees have not done ‘ <i>medical tests.</i> ’
DIRT	Zero policy migration is ‘ <i>a global clean up.</i> ’
AMORALITY	Female foreigners being referred to as ‘ <i>prostitutes.</i> ’
SUBHUMAN/ ALIEN	Immigrants being referred to using animal categories, such as ‘ <i>mice</i> ’, ‘ <i>worms</i> ’ and ‘ <i>monkeys.</i> ’
OUTLAW	‘ <i>Migrants do everything illegally.</i> ’
BURDEN	‘ <i>Migrants expect to be taken care of.</i> ’
DANGER/ THREAT	Migrants have ‘ <i>dangerous relationships</i> ’ with Islam, ‘ <i>foreigners spread the terror.</i> ’

- ... the category 'refugee(s)' ... comprises mainly metaphors characterising the relevant individuals as disgusting animals (worms, mice)
- the category 'migrant(s)' seems to be more commonly attributed metaphors of OUTLAW, VIOLENCE and DISEASE
- This study confirms that previously identified metaphors used to Othering migrants and refugees in other languages are also found in Cypriot discourse.
- **Cyprus-specific:** the category 'foreigner(s)' is more typically approached using the metaphor of AMORALITY (especially for female foreigners)

Conceptual contiguity of race & religion in Cyprus

- Cyprus: religion is a central part of the collective identity
- In the C.O.N.T.A.C.T. Interviews conducted in Cyprus, Islam is predominantly confounded with a particular ethnicity: Turks
- In most cases where the word/topic *religion* was mentioned, reference was also made to *race* in the same participant turn.
- Race and religion among the most frequent collocates; this lexical „company“ runs parallel to the lexical „company“ of Turks with fanaticism and Muslims with terrorism.

Alcántara-Plá/Ruiz-Sánchez (2017)

Framing of Muslims on the Spanish Internet

Frames / Frame theory

- We can identify frames through the most frequent collocations obtained with corpus linguistics tools.
- Example: **the adjective “Islamic” is frequently used with violent concepts such as “terrorism”**. This indicates a conceptual contiguity between terrorism and Islam, i.e. that **terrorism is part of the framing conveyed by the word islámico** (‘Islamic’).

Representation of Muslims on the Spanish Internet

Two main results:

- First, there is a negative framing that links Muslims to terrorism
- Secondly, this negative framing is referred to with the adjective *islámico*. Though *musulmán* could be understood as a synonym, ***musulmán* is used in neutral or positive frames** related to culture and contemporary politics/society in all the corpora.
 - > In fact, there is a positive semantic shift when *islámico* appears together with *musulmán* or *árabe*: the framing becomes neutral.

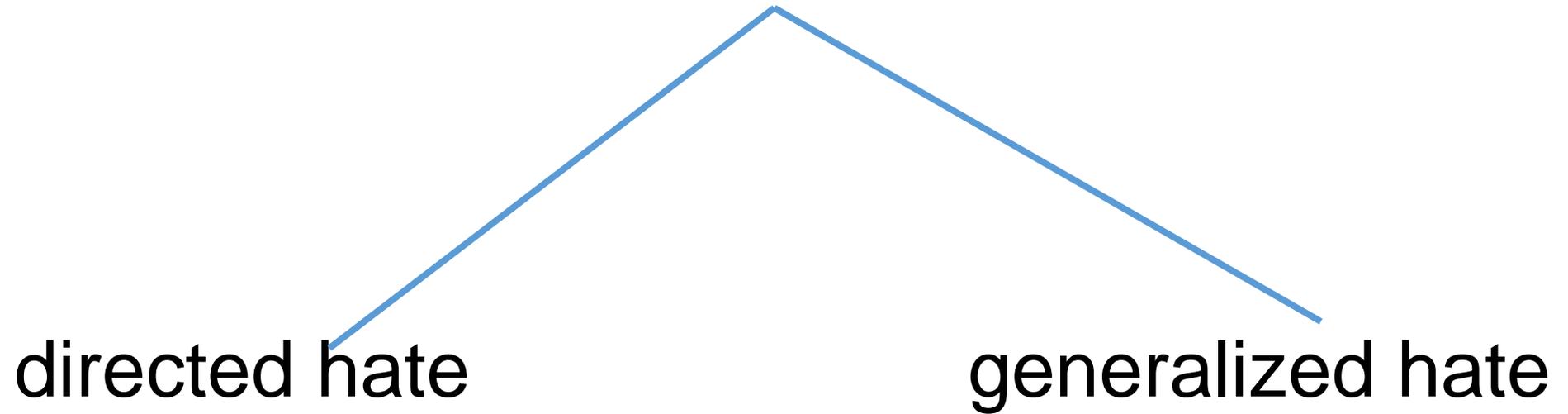
- “As result of our research, we can confirm the stigmatization of this minority [Muslims in Spain] in the digital discourse. [...]

If the detection of **stigmatizations** is, as pointed out by experts, **the first step in escalating into hate speech and hate crime**, online discourse about Muslims in Spain should be considered as worrying.”

ElSherief et al. (2018)

Hate speech analysis in social media

hate speech



Directed hate

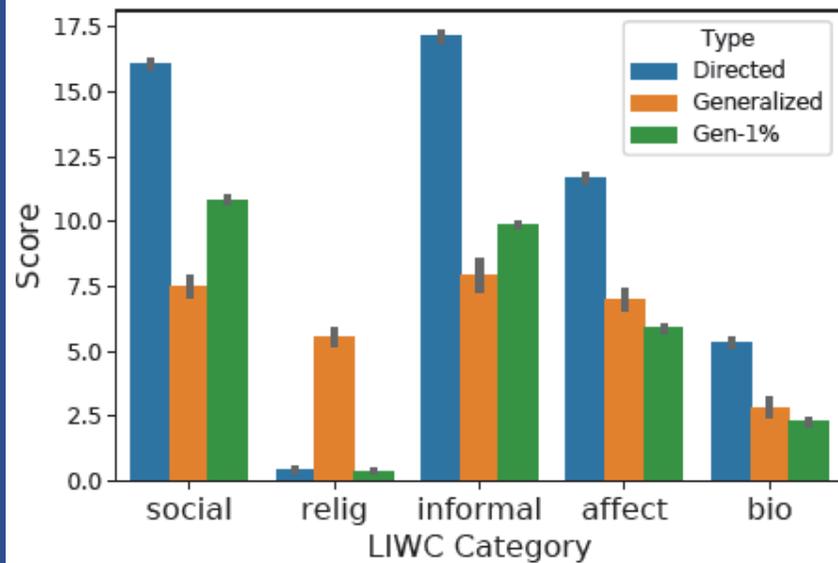
- hate language towards a specific individual or entity.
 - An example is: “@usr4your a f*cking queerf*gg*t b*tch”.

Generalized hate

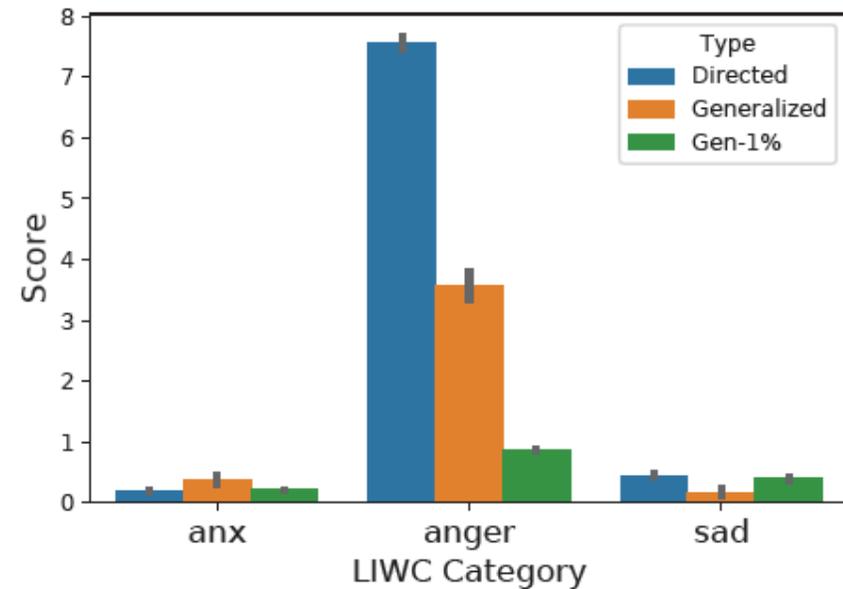
- hate language towards a general group of individuals who share a common protected characteristic, e.g., ethnicity or sexual orientation.
 - An example is: “— was born a racist and — will die a racist! — will not rest until every worthless n*gger is rounded up and hung, n*ggers are the scum of the earth!! wPww WHITE America”

Archaic Generalized	Archaic Directed	Class Generalized	Class Directed
Anti wigger	hillbilly	Catholics	Rube
hillbilly	chinaman	hollering	#redneck
bitch	verbally	#racist	ALABAMA
	prostitute	Cracker	batshit
Disability Generalized	Disability Directed	Ethnicity Generalized	Ethnicity Directed
retards	#Retard	Anglo	coons
legit	sniping	spics	Redskins
Only	#retarded	breeds	Rhodes
yo	Asshole	hollering	#wifebeater
Phone	upbringing	actin	plantation
Gender Generalized	Gender Directed	Nationality Generalized	Nationality Directed
dyke(s)	#CUNT	Anti	chinaman
chick	judgemental	wigger	Zionazi(s)
cunts	aitercation	bitch	#BoycottIsrael
hoes	Scouse	white	prostitute
bitches	traitorous		#BDS
Religion Generalized	Religion Directed	SexOrient Generalized	SexOrient Directed
Algebra	catapults	meh	pansy
Israelis	Muzzie	#faggot(s)	Cuck
extermination	Zionazi	queers	CHILDREN
Jihadi	#BoycottIsrael	hipster	FOH
lunatics	rationalize	NFL	wrists

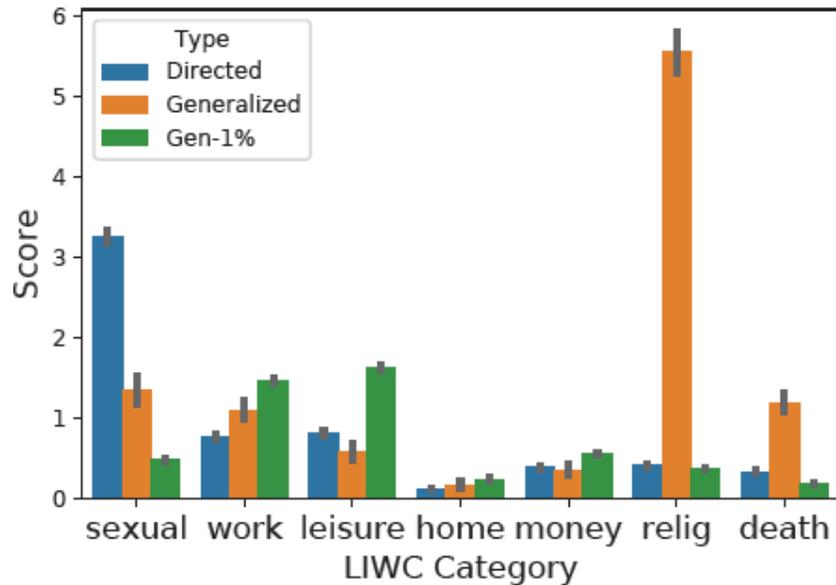
Table 2: Top five keywords learned by SAGE for each **hate speech class**. Note the presence of distinctive words related to each class (both for Generalized and Directed hate).



(b) Psychological processes



(d) Negative emotions



(f) Personal concerns

Figure 5: Mean scores for LIWC categories. Several differences exist between Directed hate speech and Generalized hate speech. For example, **Directed hate speech exhibits more anger** than Generalized hate speech, and **Generalized hate speech is primarily associated with religion**. Error bars show 95% confidence intervals of the mean.

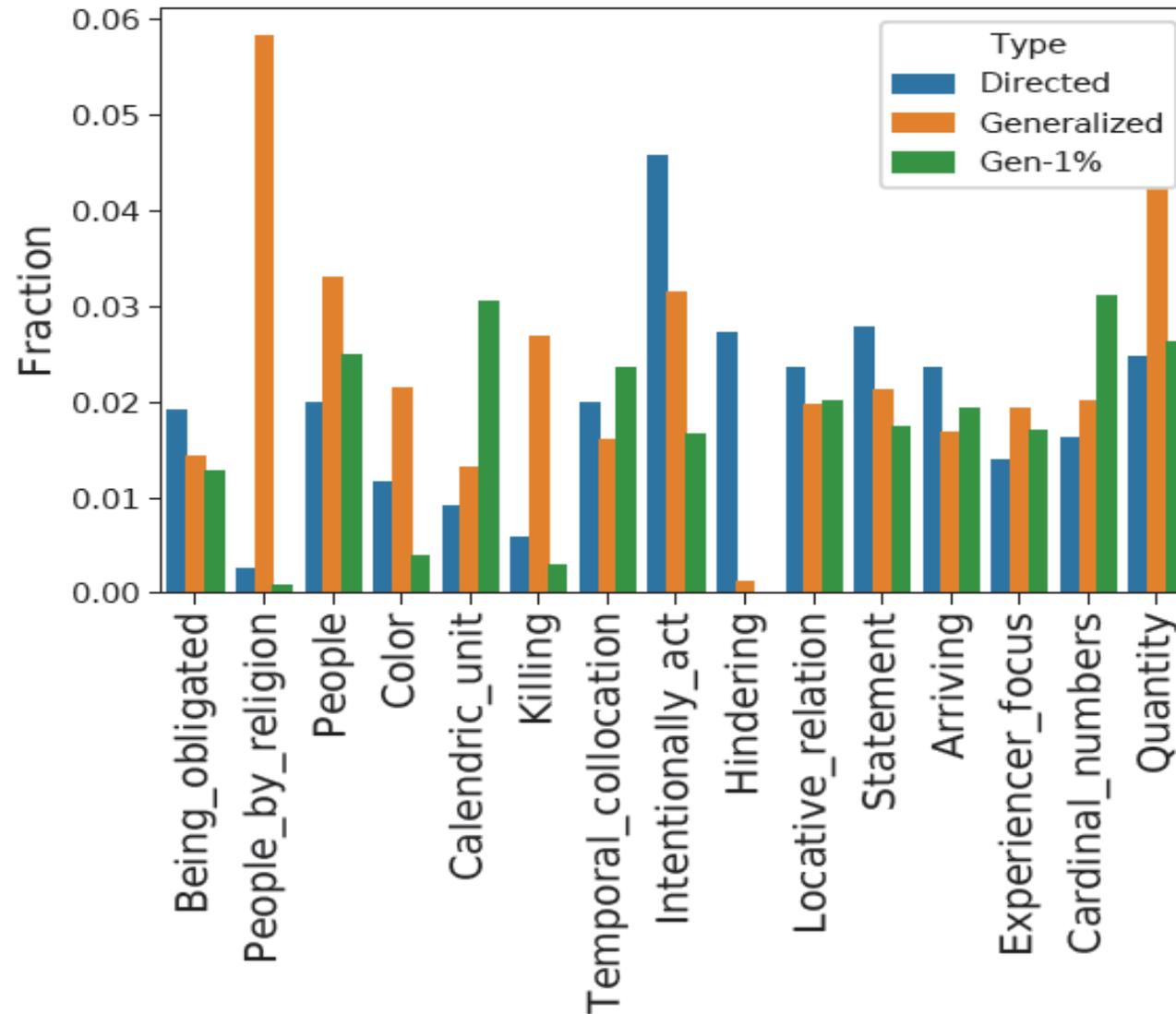


Figure 6: Proportion of frames in different types. **Note the much higher proportion of PEOPLE BY RELIGION frame mentions in Generalized hate speech.** In contrast, Directed hate speech evokes frames such as INTENTIONALLY ACT and HINDERING.

Albadi et al (2018)

Religious hate speech in the Arabic Twittersphere

- based on 6,000 tweets (6 x 1,000 tweets for the six religious groups)
- nearly 50% of the discussions about religion in Arabic Twittersphere is about hate towards religious groups
- especially towards Jews (60%), Atheists (56%) and Shia

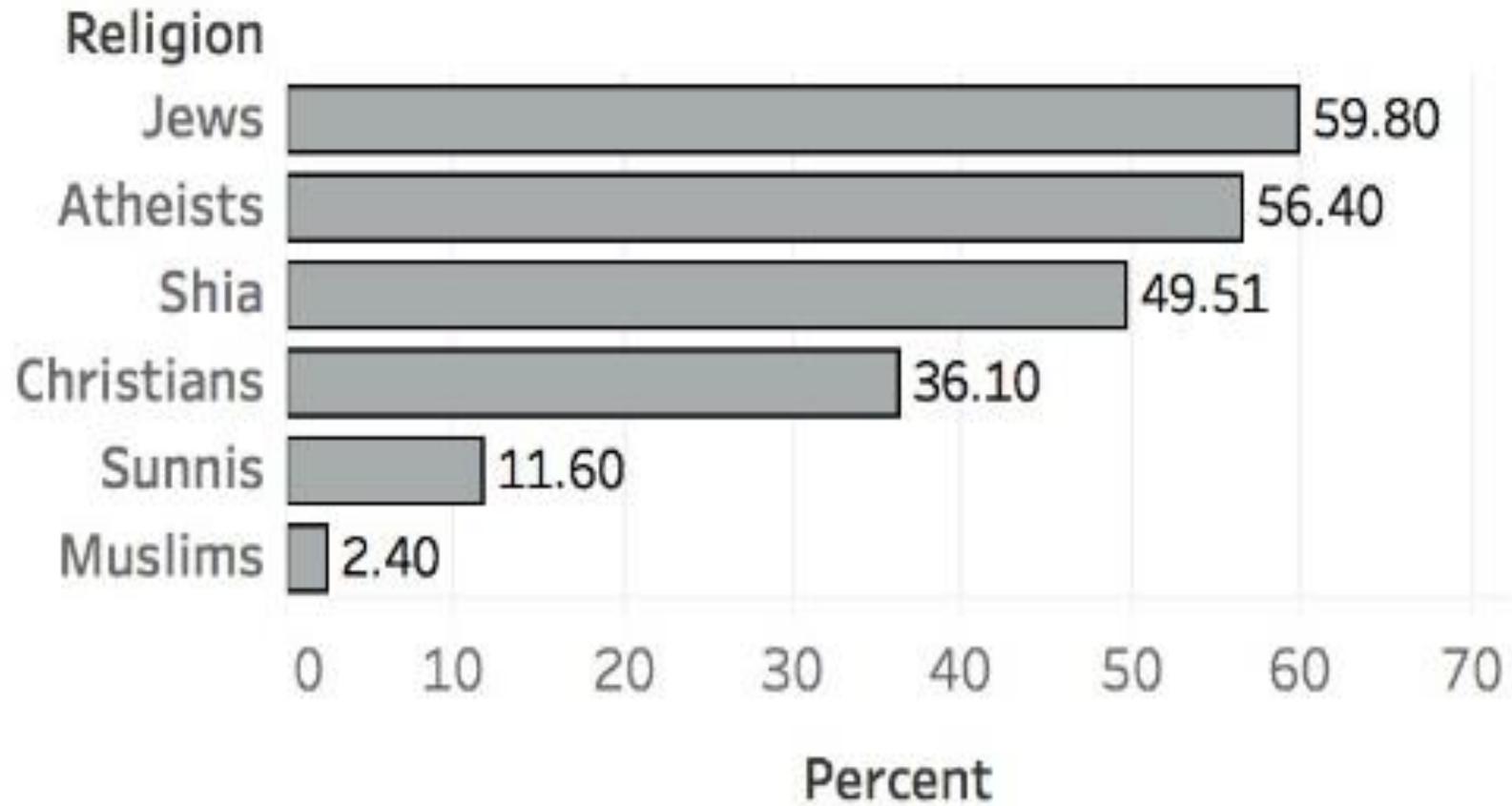


Fig. 2. Percentage of tweets labeled as *hate* against each of the religious groups when considering individually each of the 1000 tweets collected for each of the religious groups.

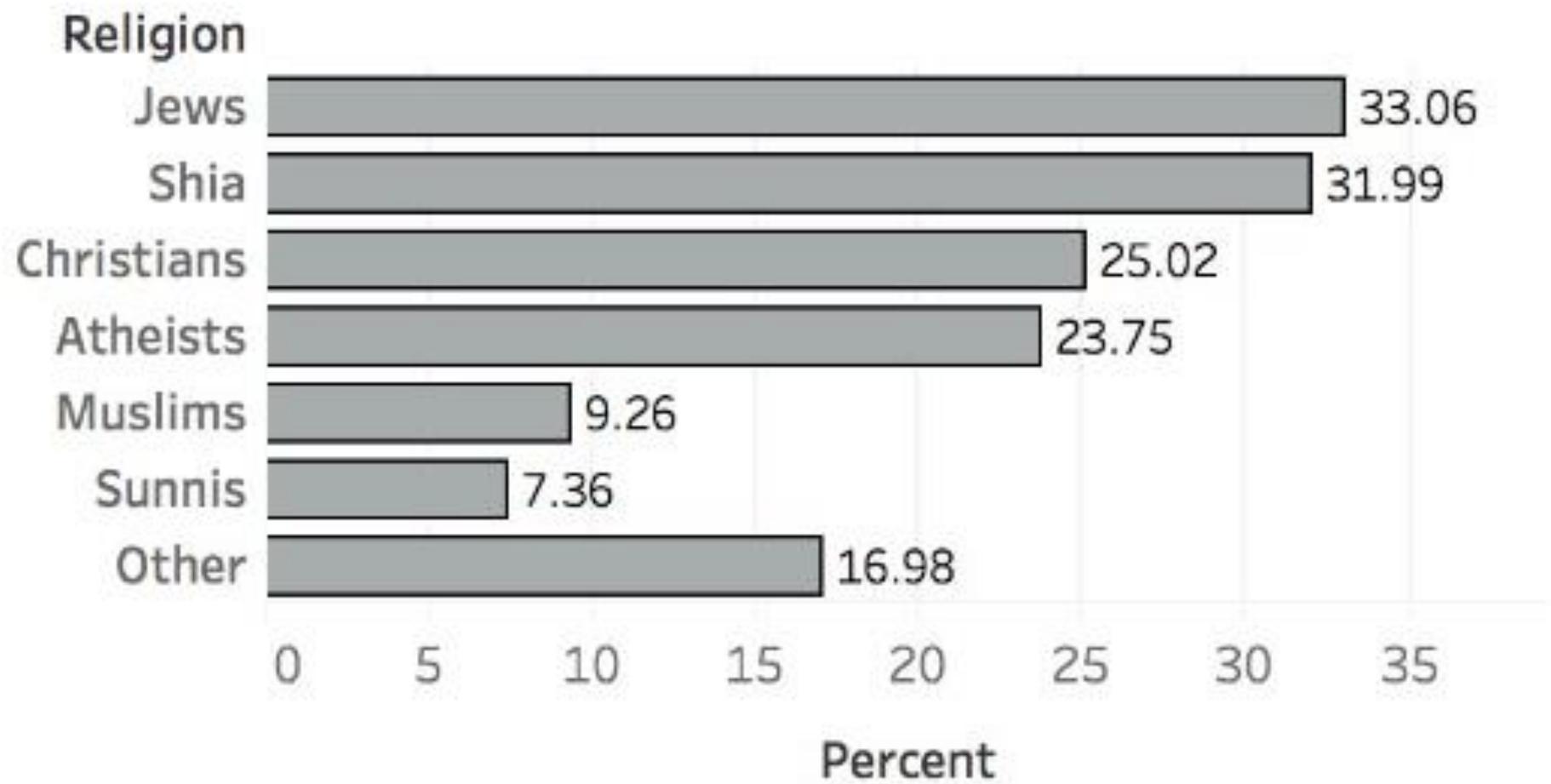


Fig. 3. Percentage of hateful tweets received by each of the religious groups among all hateful tweets.

5. Conclusion

- from a linguistic point of view: no specific mechanisms distinguishing religious-relevant hate speech from hate speech in general
- no new breakthrough insights by linguist(ic)s, BUT:
- linguistics offers concepts, theories, methods, entire frameworks for spotting and seeing through hate speech **AND** the linguistic, communicative, social, and cognitive mechanisms giving rise to it

- in line with the overall spirit of my intervention, full agreement with Assimakopoulos et al. 2017:

In closing, we hope to have shown that linguists have an important role to play in this picture [...]. Since it is intention that lies at the very core of most legal definitions of hate speech, contextualising and qualitatively analysing such speech seems central to not only tackling this complex phenomenon but also to safeguarding freedom of expression on the many platforms that the internet offers. We therefore believe that this is an endeavour that can only be accomplished by encouraging **collaboration and constructive dialogue** between policy makers, legal practitioners, linguists and computer scientists specialising in the automatic detection of hate speech, [...]

References

- Agustina, Nurizzati, Ainim, S., Adek, M. & Rahmi, A. 2020. Light and shadow in hate-speech cases: A forensic linguistics. *Proceedings of the 3rd International Conference on Language, Literature, and Education (ICLLE 2020)*, 15-22. Heidelberg: Atlantis Press. <https://www.atlantis-press.com/article/125945979.pdf>
- Alcántara-Plá, M. & Ruiz-Sánchez, A. 2017. The framing of Muslims on the Spanish Internet. *Lodz Papers in Pragmatics* 13/2: 261-283. <https://doi.org/10.1515/lpp-2017-0013>
- Assimakopoulos, S., Baider, F. H., & Millar, S. 2017. *Online hate speech in the European Union. A discourse-analytic perspective*. Heidelberg: Springer. <https://link.springer.com/content/pdf/10.1007%2F978-3-319-72604-5.pdf>
- Albadi, N., Kurdi, M., & Mishra, S. 2018. Are they our brothers? Analysis and detection of religious hate speech in the Arabic Twittersphere. *Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. IEEE Press, 69–76. <https://nuhaalbadi.com/assets/papers/AreThey.pdf>
- Baker, P., Gabrielatos, C., & McEnery, T. 2013. Sketching muslims: A corpus driven analysis of representations around the word ‘muslim’ in the British press 1998–2009. *Applied Linguistics* 34: 255-278. <https://doi.org/10.1093/applin/ams048>
- Cheng, J.E. 2015. Islamophobia, muslimophobia or racism? Parliamentary discourses on Islam and Muslims in debates on the minaret ban in Switzerland. *Discourse & Society* 26(5): 562-586. <https://www.jstor.org/stable/26376401>

- Chilton, P. & Kopytowska, M. (eds.). 2018. *Religion, Language, and the Human Mind*. Oxford: OUP.
- Clarke, I. & Grieve, J. 2017. Dimensions of abusive language on Twitter. In *Proceedings of the first workshop on abusive language online*, 1-10. <https://aclanthology.org/W17-3001.pdf>
- Darquennes, J. & Vandenbussche, W. 2011. Language and religion as a sociolinguistic field of study: some introductory notes. *Sociolinguistica* 25: 1-11. <https://doi.org/10.1515/9783110236262.1>.
- Drainville, R. & Saul, J. 2020. Visual and linguistic dogwhistles. Retrieved from https://www.researchgate.net/profile/Jennifer-Saul/publication/344441861_Visual_and_Linguistic_Dogwhistles/links/5f8ee80892851c14bcd58a6b/Visual-and-Linguistic-Dogwhistles.pdf
- ElSherief, M., Kulkarni, V., Nguyen, D., Yang Wang, W., & Belding, E. 2018. Hate lingo: a target-based linguistic analysis of hate speech in social media. *Proceedings of the International AAAI Conference on Web and Social Media*, 12(1). Retrieved from <https://ojs.aaai.org/index.php/ICWSM/article/view/15041>
- ElSherief, M., Ziems, C., Muchlinski, D., Anupindi, V., Seybolt, J., De Choudhury, M. & Yang, D. 2021. Latent Hatred: A Benchmark for Understanding Implicit Hate Speech. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 345–363, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics. <https://aclanthology.org/2021.emnlp-main.29>

- Leezenberg, M. 2017. Free speech, hate speech, and hate beards. Language ideologies of Dutch populism. In: Silva, N. D. ed. *Language and violence: Pragmatic perspectives*. Amsterdam/Philadelphia: Benjamins. 141–168. <https://doi.org/10.1075/pbns.279.07lee>
- Lumsden, K. & Harmer, E. (eds.). 2019. *Online Othering: Exploring Digital Violence and Discrimination on the Web*. Palgrave Macmillan.
- Pihlaja, S. 2014. *Antagonism on YouTube: Metaphor in Online Discourse*. London: Bloomsbury.
- Sanders, J. 2018. Introduction to the Topical Issue “Cognitive Linguistics and Theology”. *Open Theology* 4/1: 541-544. <https://doi.org/10.1515/opth-2018-0042>
- Tirrell, L. 2021. Toxic speech and damaged bodies. <https://www.youtube.com/watch?v=cWuNqktX7dU>
- Tirrell, L. 2017. Toxic Speech: Toward an epidemiology of discursive Harm. *Philosophical Topics* 45 (2): 139–161. <https://doi.org/10.5840/philtopics201745217>
- Tirrell, L. 2018. Toxic speech: Inoculations and antidotes. *The Southern Journal of Philosophy* 56(S1):116-144. <https://doi.org/10.1111/sjp.12297>
- Vrysis, L., Vryzas, N., Kotsakis, R., Saridou, T., Matsiola, M., Veglis, A., Arcila-Calderón, C. & Dimoulas, C. A. 2021. A web interface for analyzing hate speech. *Future Internet* 13(3): 80-. <https://doi.org/10.3390/fi13030080>
- Wikipedia contributors. (2021, August 23). Dog whistle (politics). In *Wikipedia, The Free Encyclopedia*. Retrieved August 30, 2021, from [https://en.wikipedia.org/w/index.php?title=Dog_whistle_\(politics\)&oldid=1040195852](https://en.wikipedia.org/w/index.php?title=Dog_whistle_(politics)&oldid=1040195852)

- **Special journal** since 2013: *Journal of Language Aggression and Conflict* (Benjamins)