Can political dogwhistles be predicted by distributional methods for analysis of lexical semantic change?

Max Boholm, Björn Rönnerstrand, Ellen Breitholtz, Robin Cooper, Elina Lindgren, Gregor Rettenegger, and Asad Sayeed
University of Gothenburg

Marianne och Marcus Wallenbergs Stiftelse
WASP—HS

Goal: identify semantic change of dogwhistles

We combine methods of **lexical semantic change (LSC) detection** and **survey data from lexical replacement tests** to model the temporal dynamics of dogwhistle meaning.

- Computational measures of LSC $\leftrightarrow$ shifts in "hidden" (in-group) and "public" (out-group) meanings?
- Compare distributional methods w.r.t. the modeling of semantic change of dogwhistles.

Background

What is a political dogwhistle?

$\rightarrow$ An expression used to broadcast a benign-sounding message to the wider public while expressing a potentially unacceptable "payload" message to a target affinity group.

Example in Swedish: *återvandring* (re-migration)

- Part of immigration/refugee debate.
- "Public" meaning: voluntary return to home country.
- "Hidden" meaning: forced deportation.

Data 1. Diachronic Corpus (2000–2022)

- Discussion forum *Flashback* (1.5M users, 80M posts)
- Controversial topics and opinions, incl. discrimination and racism (anonymous users)

Swedish dogwhistles in study

Swedish	English	Corpus frq.	Mean	S.D.
berika	enrich	20,936	27.92	12.18
förortsgäng	suburban gang	227	0.23	0.26
globalist	globalist	31,156	32.07	39.62
återvandring	re-migration	12,999	13.19	22.20

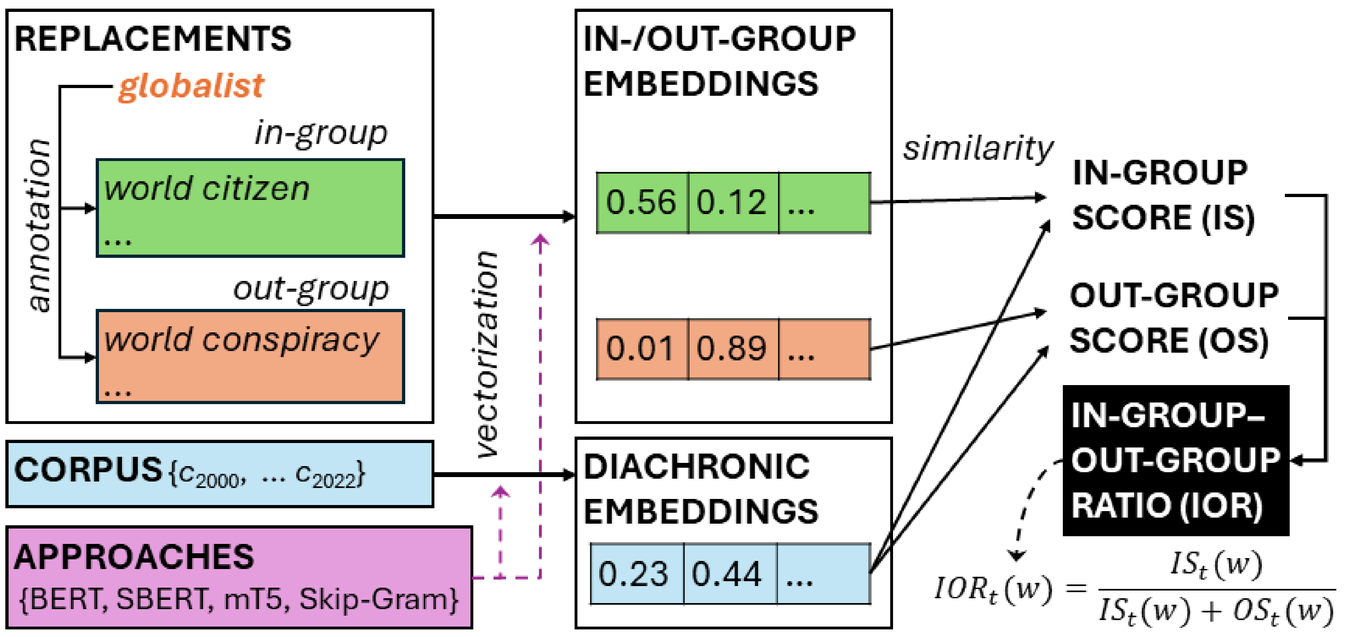
Data 2. Word replacement task

Survey conducted via *Swedish Citizens' Panel*.

- 1,780 respondents were asked to replace the dogwhistle word in a hand-constructed sentence context.
- Example: "The Swedish unions are controlled by **globalists**."
- *In-group replacements*: anti-Semitic slurs, etc.
- *Out-group replacements*: "benign" mentions of trade policy, etc.
- Inter-annotator agreement: Krippendorff's $\alpha > 0.6$

Ethics: IRB approval ; anonymous

Modeling in-group and out-group meanings



Regression: Predicting LSC from IOR

- **LSC:** Angular distance of time-specific word embeddings.
- **IOR:** normalized measure of *w*'s in-group meaning relative to its out-group meaning.

$$\Delta_{t_i,t_j}(w) = \frac{\arccos(cossim(\overrightarrow{w_{t_i}}, \overrightarrow{w_{t_j}}))}{\pi}$$

IOR *difference*:

$$\Delta^{IOR}_{t_i,t_j}(w) = abs(IOR_{t_j}(w) - IOR_{t_i}(w))$$

Four approaches for modeling meaning:

- BERT
- Sentence-BERT (SBERT)
- Multilingual Text-to-Text Transfer Transformer (mT5)
- Skip-Gram with negative sampling (SGNS)

Results

- The methods for detecting LSC **are** sensitive to the dynamic meaning of dogwhistles: observed meaning shifts for dogwhistles using distributional methods are explained by their in-group and out-group dimensions.
- Suggests that LSC measures could be used to detect dogwhistles online.
- The LLMs explain more variability of the data and have larger coefficients for $\Delta^{IOR}_{t_i,t_j}$, than the SGNS models.

By the numbers

$$\Delta_{t_i,t_j}(w) = \beta_0 + \beta_1 \times \Delta^{IOR}_{t_i,t_j}(w) + \beta_2 \times log_2(FPM_{t_i}(w)) + \beta_3 \times \Delta^{FPM}_{t_i,t_j}(w)$$

	SBERT	BERT	mT5-XL	SGNS *w*5,*d*100	SGNS *w*10,*d*100	SGNS *w*15,*d*100	SGNS *w*5,*d*200	SGNS *w*10,*d*200	SGNS *w*15,*d*200
$\Delta^{IOR}_{t_i,t_j}$	**0.79*****	**0.55*****	**0.56*****	**0.25***	0.13	**0.25***	**0.27***	0.18	**0.36****
$\Delta^{FPM}_{t_i,t_j}$	-0.02	-0.08	-0.22*	0.26*	0.24	0.20	0.27*	0.20	0.21
FPM, lg	-0.27***	-0.24*	-0.45***	0.05	-0.03	-0.08	0.35**	0.26*	0.22
Const.	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
Adj. R^2	0.80	0.40	0.40	0.09	0.02	0.07	0.21	0.11	0.20

Note: $N = 64$; *p<0.05; **p<0.01; ***p<0.001; standardized coefficients; *FPM* = frq. per million

{asad.sayeed@gu.se}