Measuring Implicit Human Biases Through the Statistical Properties of Language

Kelly E. Wright

University of Michigan

LSA Department of Linguistics

HiSoN Summer School

Speaker Series

17 July 2019

Racialization

- Happens when a word with **no preexisting racial connotation** comes to describe people of color asymmetrically.
  - *Thug*
- A new Polysemous Sense
  - One word, Multiple meanings

Articulate

More than a microaggression

(See Alim and Smitherman 2012; Hill 2008)

- In the US, standard linguistic varieties are ideologically associated with **Whiteness** and formal **education**
  - Calling a minority articulate notes surprisal at standard language usage in the absence of these traits.

Racism

- How do we see social ideology reflected in text?
  - (aka language use)
- Wright (2017) hypothesis:
  - at the lexical level.

Lee Atwater

You start out in 1954 by saying, “[N-Word, N-word, N-Word].” By 1968 you can't say “[N-Word]” — that hurts you. Backfires. So you say stuff like **forced busing**, **states' rights** and all that stuff. You're getting so abstract now [that] you're talking about cutting taxes, and all these things you're talking about are totally economic things and a byproduct of them is [that] blacks get hurt worse than whites… I'm saying that if it **is getting that abstract, and that coded, that we are doing away with the racial problem** one way or the other. You follow me — because obviously sitting around saying, “We want to cut this,” is much more abstract than even the busing thing, and a hell of a lot more abstract than “[N-word, N-word].”

Covert Racism

Bussing in our above example

- Racialization happens
  - Regular words, with no racial connotation, become asymmetrically applied to the descriptions of actions or events of actors of color.

The Myth of Unbeatable Black Athleticism

- **Folk ideology**, mistakenly applied throughout US history
- Black athletes are *not* exceptional or naturally suited for physical activity, or for violent displays or prowess in team or individual sport.

Wright (2017)

Covert Racialization in Sports Journalism

- Black athletes are described predictably differently
  - Exceptionality and Animalistic traits
- White athletes are described
  - Leadership or Skill-based terms

What are there ways in which lexical semantic space is renegotiated as new (racialized) meaning is bound to old (non-racialized) words?

Formal Semantics

The Polysemy Problem

(See Trier 1932)

- The simplest processes of semantic change have proven reliably difficult to model.
- Spam: potted meat → Spam: junk email
  - Both meanings active, how do we choose?

Dataset

RSEAC

- **R**acialized **SE**mantics in **A**thletics **C**orpus
  - 120 Athletes
  - 60 White; 60 Black
  - 30 Male; 30 Female
  - 15,500 lexemes
  - 8.5 million total words

Latent Semantic Analysis

(See Caliskan et al. 2017)

- Using Word Vector Models trained on the Google corpus, the authors recreate human participant’s results from Implicit Association Tasks.

Caliskan et al. (2017)

N= Number of Subjects N_T= Target Words N_A= Attribute Words   D = Effects Sizes   P = Pvalues

Target words	Attribute words	Original finding				Our finding			
		Ref.	N	d	P	N_T	N_A	d	P
Flowers vs. insects	Pleasant vs. unpleasant	(5)	32	1.35	10^{-8}	25 × 2	25 × 2	1.50	10^{-7}
Instruments vs. weapons	Pleasant vs. unpleasant	(5)	32	1.66	10^{-10}	25 × 2	25 × 2	1.53	10^{-7}
European-American vs. African-American names	Pleasant vs. unpleasant	(5)	26	1.17	10^{-5}	32 × 2	25 × 2	1.41	10^{-8}

❖ Word Vectors created from Target word stimuli and Attribute word stimuli

Machine Learning

- Support Vector Machine
  - Counting Stuff
- Random Forest Modeling Task
  - Analyzing Counted Stuff

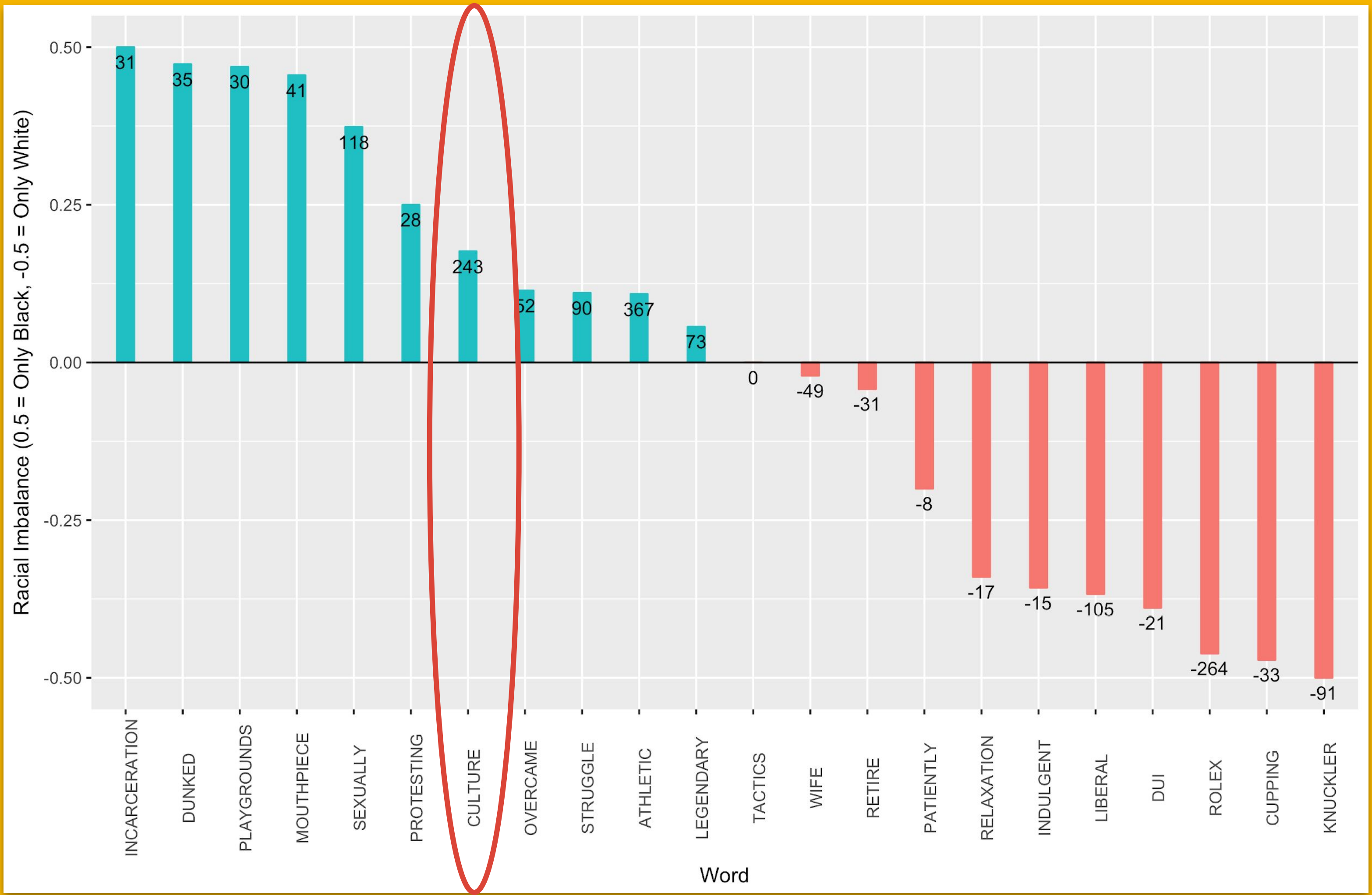
Support Vector Machine

- Learning Algorithm, trained to predict athlete race from lexical token counts
  - Subcorpora organized by athlete

Culture
2:1 ratio

Black Subcorpus →

White Subcorpus →



Athlete	Race Probability
Eric Berry	3%
Chamique Holdsclaw	3%
Brittney Griner	3%
Anthony Davis	3%
Jackie Joyner Kersee	3%
Alia Atkinson	3%
Hope Solo	96%
Phil Mickelson	96%
Andrew Luck	96%
Ronda Rousey	96%
Drew Brees	96%
Katie Ledecky	96%

- The SVM sorted Racial Probability by Athlete Subcorpus
- The most impressive result here is a lack of gradience in the probabilities of category membership.

Random Forest Modeling

- A learning algorithm building multiple decision trees and modeling based on the most accurate.
- Trained on the same dataset

Random Forest

- Predict class membership by building decision trees
- See a portion of the data, use it to predict class membership on the rest of the data.
- 90/10 training ratio.

Lexical Type	Importance	Black Sum	White Sum
BLOCKS	0.2092	520	102
NIKE	0.1515	660	217
COAST	0.1511	132	214
EFFORTS	0.1451	310	205
AVERAGED	0.1421	573	109
ATHLETIC	0.1130	1029	662
WONDERFUL	0.1075	204	379
UNDEFEATED	0.1064	241	74
WHOM	0.1052	363	227
APPEARED	0.0958	545	333
SOCIAL	0.0904	872	560
LLC	0.0871	104	222

- RandomForest lets us crack into Lexical Importance to Categorization
- Model outputs predict which words were most useful in the racial categorization task.

Implications

- Supports Caliskan et al. (2017) and Garg et al (2017)
- Supports the Distributional Hypothesis
- Supports Trier’s assertion that meaning is epiphenomenal
- Algorithms are super racist (Speer 2017)

Future Directions

- BETTER VISUALIZATION!!!
  - (I have ideas; I need help)
- Extend time depth and sport balance in RSEAC
  - Apparent time studies
- Expand and SVM test on new datasets
- Henderson & McCready (2018)
  - Dogwhistles: Where world knowledge and self knowledge mesh to control semantic judgements

Thanks!

Kelly E. Wright

@raciolinguistic

kellywri@umich.edu