

A Report

from the North American Research Network in Historical Sociolinguistics

Kelly E. Wright



On Deck

- ★ Background on NARNiHS' history, goals, and happenings
- ★ A report from the LSA
 - Summary of works presented
 - Questions
- ★ My work in Historical Sociolinguistics
 - If there is time

NARNiHS

Who are we?

- ★ History of NARNiHS
 - Brainchild of collaborative socializing following Mobility, variability and changing literacies in Modern Times (2014) in Utrecht
 - Organized at UW Madison October 2016
- ★ Incorporated Steering Group
 - November 2017
 - Current Membership

NARNiHS

What we've been up to



First Events

- LSA Summer Institute 2017
 - Course Guides, Conference, Workshop (by yours truly)
- NWAV 46



LSA Newest Sister Society

- January 2018



New Annual Events

- LSA Annual Meeting (cooked)
- NARNiHS @ the KFLC (raw)
 - Savi Namboodiripad and Ariana Bancu
- Branded Panels

Steering Group

Rights and Privileges

- ★ Define bylaws and organizing principles
- ★ Head an (inter)national organization
- ★ Review abstracts
- ★ Chair sessions and lead discussions and workshops at high-level conferences
- ★ Shape the field!

Historical Sociolinguistics

narnihs.org

- ★ The study of the interaction of language and society **in historical periods**.
- ★ The application of **sociolinguistic theories** and **methods** to the study of historical **language variation and change** over time, or more broadly.

Historical Sociolinguistics

narnihs.org

- ★ Fossil Record problem (Chomsky)
- ★ Bad Data problem (Labov)

How does one study historical states of language or society?

How does one study historical states of language or society?

Interface with other disciplines!

Other Disciplines

narnihs.org

- ★ Social Historians,
 - ★ Historical Sociologists,
 - ★ Computer Scientists, Systems Analysts,
 - ★ Social and Cultural Anthropologists, Political Scientists,
 - ★ Civil Engineers, Geographers, Economists,
 - ★ Art Historians, Business Historians,
 - ★ Archivists, Curators, and Library Scientists,
 - ★ Evolutionary Biologists, Botanists, and Medical Historians
-

NARNiHS' Goals

Historical Sociolinguistics...

- ★ Relatively unknown field in North America.
- ★ Researchers come from diverse disciplinary backgrounds
 - We are not readily identifiable to each other
- ★ Researchers don't readily identify with the field.
- ★ A young field, especially in the North American context.
 - HiSoN

NARNiHS' Goals

Historical Sociolinguistics...

- ★ Field in North America needs active professional dialog
 - colleagues working in related areas across humanities and social science disciplines.
- ★ Field in North America needs collective ties to colleagues and sister organizations across the globe
 - to harness the collaborative energy of the discipline.

Report from the LSA

(So you can see what we do)

Overview

- ★ Researchers from 7 countries
- ★ First (!) General Membership Meeting
 - 27 in attendance!
- ★ Poster Session
 - Seven Posters in coordinated session with LSA general poster session
 - Our very own Carrie Ann Morgan presented *Inflammatory infinitives and light dialects: enregistering Central Albanian*

Four Sessions

- ★ Language on the Move
- ★ Pathways of Change
- ★ Language and Officialdom
- ★ Language in Contact

Language on the Move

- ★ Separated at birth?
Scandinavian-American English
and its representation over time.
 - Joe Salmons and Laura Moquin (UW Madison)
- ★ ‘I wonder does he know you are in America’: Exploring the development of embedded inversion in Irish English
 - Carolina P. Amador-Moreno (Univ. of Extremadura)

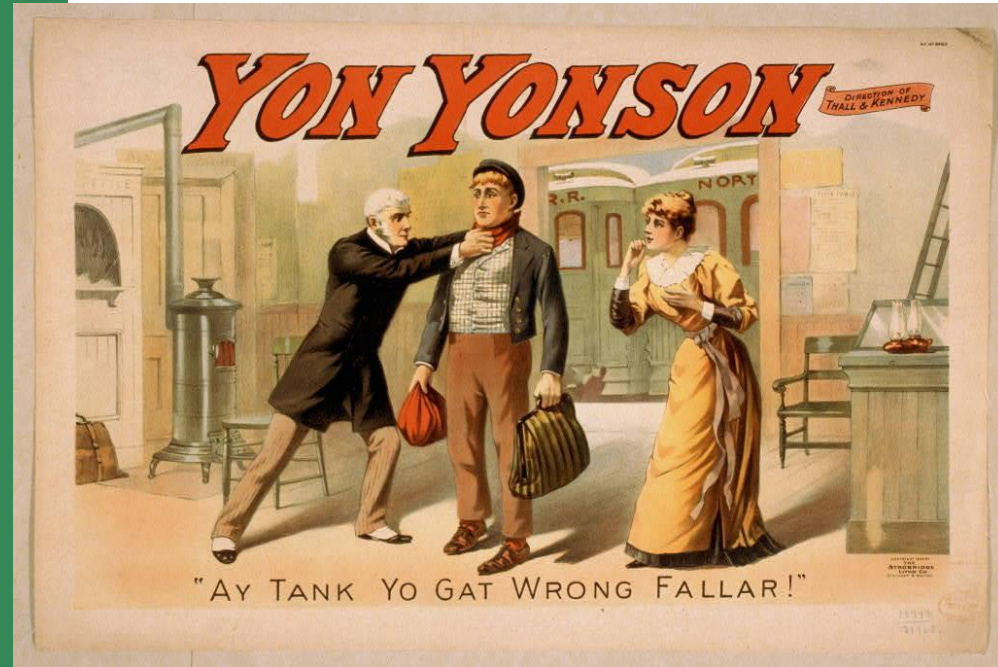
Scandinavian-American English over time

Salmons and Moquin

- ★ Portrayals of variety by looking at 19th C dialect comedians
 - Showing us systematic differences between varieties that Haugen or Moen didn't mention
 - ★ Norwegian Variety solidified really early
 - Evidence of German contact in the US context
 - ★ State varieties based on dialect patterns
-

Scandinavian- American English over time

Salmons and Moquin



Scandinavian-American English over time

Salmons and Moquin

- ★ Later in early 20th C
 - hyperforeignisms
 - Stereotype representations follow similar patterns across all varieties
 - Portrayals include generalized, over-realizations
 - affricate /dʒ/ as a glide [j] and the often-omitted neutralization of /s ~ z/.

Embedded Inversion in Irish English

Carolina P. Amador-Moreno

- ★ CORIECOR
 - (which I helped make, y'all!)
 - ★ Systematic?
 - Distinctly different pattern from Irish Gaelic AND from other colonial period varieties of British English
 - ★ Stigmatized?
 - Irish Immigrants
 - Irish Authors
 - ★ Wonder, ask, know, see
 - Wonder not so much: why?
-

Pathways of Change

- ★ Change from above: benne vs
sijn ‘are (pl)’ variation in Early
Modern Dutch
 - Christina Ilse Terpstra (Univ. of
Calgary)
 - ★ The Success of French
Orthographic reform among the
religious elite in 17th century
New France: the case of the
apostrophe
 - Sylvie Dubois and Malcolm
Richardson (Louisiana State Univ.)
-

Pathways of Change

- ★ Interpreting a written change: the usefulness of scripta theory
 - William Johnson (OSU)
- ★ Koineization and culture: the early modern generalization of Spanish vuestra/merced/usted
 - Don Tuten (Emory)

Change from above

Christina Ilse Terpstra

- ★ *benne* vs *sijn* ‘are (pl)’ variation in Early Modern Dutch
- ★ Nationalism, Printing Press, Religion
 - Centre -> Periphery
- ★ Private letters
- ★ *Benne* fell into disuse as speakers opted for the more prestigious *sijn*
 - Older speakers felt more pressure to conform

French Orthographic Reform

Sylvie Dubois

- ★ Quebec and New Orleans
 - 1685-1727
 - Immediately preceding typographical standardization
 - ★ Ursulines' handwritten eulogies
 - High class, educated, shared ideology
 - Writing (le belle main) highly prized and widely distributed
 - ★ Graphic fusions, elision of final vowels
 - Their strategies became that of secular women
-

French Orthographic Reform

Sylvie Dubois

- ★ Perfect HS project
 - Understudied Genre
 - Influence of individuals on the group
 - Mixture of elite and marginalized forms
 - Effects of private use on the public (prescriptive) perceptions

The Usefulness of Scripta Theory

William Johnson

- ★ Scripta: A Koine used for writing
- ★ *Pas* and *Mie* negative particles in Old French
 - The Jespersen Cycle
 - Proposing a scripta emerged only after 1300
- ★ Various written forms from 2 dialect regions
- ★ *Mie* becomes “fossilized” used only at the end of a line of poetry
- ★ Parisian Linguistic Gravity and hyperlocal varieties in competition
 - ○ Levelled to a scripta

Generalization of Spanish vuestra merced/ usted

Don Tuten

- ★ 16th C demographic expansion and dialect mixing in the Spanish Empire
 - New Titled Nobility
 - Constant renegotiation of social status; mobility fervently discouraged
- ★ Nevertheless, social climbers persisted
 - Plays out in address forms

Generalization of Spanish *vuestra* *merced*/ *usted*

Don Tuten

- ★ *Vuestra merced* (VM) was for a “middling sort”
- ★ Stereotype of VM associated with rurality
- ★ *Merced* was used for lowest nobility
 - Becomes used as performance of urbanity
- ★ Illustrates language users conscious manipulation of forms.
 - Social actuation of LVC

Language and Officialdom

- ★ Interpretation and linguistic malpractice in the 18th century Louisiana courtroom
 - Janelle Thomas (Oxford)

Interpretation and Linguistic Malpractice

Janelle Thomas

- ★ Pointe Coupee, LA trials of slave conspiracy in 1790s
- ★ Trials in French, recorded in Spanish by a French speaker
 - Defendants spoke neither language well (if at all)
 - “Creating a separateness”
- ★ Layers of quotedness in trial transcripts
- ★ Tracing scribal fingerprints
 - Machine learning

Interpretation and Linguistic Malpractice

Janelle Thomas

- ★ Perfect HS project
 - Understudied Genre
 - Influence of individuals on the group (and vice versa)
 - Fluency of dominant users affecting marginalized users
 - Capsule of above and below influence
 - Consequences social perceptions on doctrine and practice (linguistic or otherwise)

Language in Contact

- ★ Contact-induced lenition in Arabic ‘emphatic’ consonants: a sociohistorical perspective
 - Uri Horesh (Univ of Essex)
 - ★ Communication contexts, sound change dating, and historical sociolinguistics: Colonial Spanish loanwords in Indigenous American languages
 - Israel Sáenz-Sanchez (West Chester Univ.)
-

Contact-induced lenition in Arabic 'emphatic' consonants

Uri Horesh

- ★ Arabic-Hebrew in Jaffa
 - Jaffa is a high contact speech community that is somewhat isolated
- ★ Doesn't correspond to what Hebrew used to be and how Arabic is now.
 - Loss of pharyngealization
- ★ Affricate merger with non-emphatic counterparts NOT historical antecedents
 - Change in progress in Arabic is leading to a near-merger
 - $CI^{\text{f}} \rightarrow CI$, and the lenited segments are [d, s, t]

Colonial Spanish loanwords in Indigenous American language

Israel Sáenz-Sanchez

- ★ Adaptation of Spanish sibilants in a corpus of Colonial Nahuatl
 - Alphabetic and pictographic writing
- ★ L1 influence of Nahuatl phonology in spelling trends
 - Fricatives and interdentals
 - Seseo accent (And)
- ★ L2 Colonial Spanish Pronunciation Patterns
- ★ Noisy Results
 - x/j variations correspond to unmerged or premerged varieties
 - But x/j for /s/ cannot be explained by L1 influence or L2 speech patterns

Questions?

JOIN NARNiHS

Every network needs more nodes!

- ★ Join the ListServ!
 - Send an email to listserv@lsv.uky.edu with an empty subject line and in the body of the email the simple message: subscribe NARNiHS yourfirstname yourlastname "
 - ★ Uploading your bio!
 - ★ Get our quarterly announcements!
 - ★ Organize a branded panel!
 - ★ Come to a conference!
-

My HS work!!!

(which I did not present, but you still get to hear because you gave me a platform)

Lexical Racialization examined through Machine Learning

NWAV 47

- ★ Semantic Field Theory
 - Polysemy problem
 - How is meaning organized cognitively and applied materially?
- ★ 8.5 Million-word Racialized Semantics in Athletics Corpus (RSEAC)
 - 120 athletes
 - 60 Black, 60 White
 - 60 Males, 60 Females
 - 15,500 lexemes

Lexical Racialization examined through Machine Learning

NWAV 47

- ★ Measuring Implicit human biases through the statistical properties of language
 - Caliskan et al (2017)
- ★ Black athletes are described **predictably** different than White athletes
 - Wright (2017)

Previous Methodologies in Combination

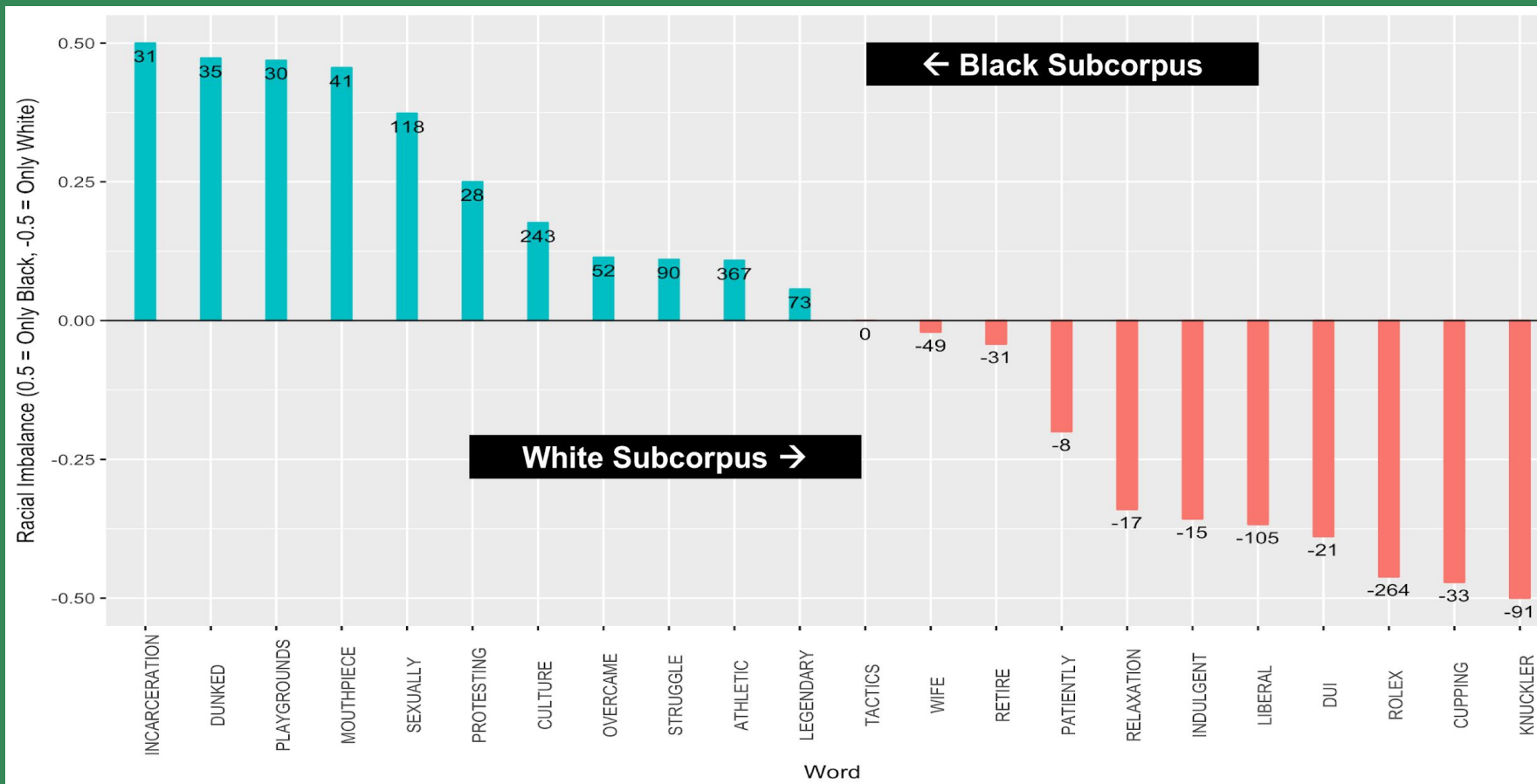
Or why RSEAC is boss

- ★ Metadata behind Sports Journalism
- ★ Massive Text Corpus, highly controlled for stylistic variation
- ★ Longform and event pieces describing individual athletes, not teams
- ★ Variation based on patterns of actual disparities are baked into the data, making racialization isolable.

Machine Learning

- ★ A Support Vector Machine is a learning algorithm
 - I trained my SVM to predict athlete race with lexical token counts for each athlete as input.
- ★ A Random Forest is a learning algorithm building multiple decisions trees.
 - Modeling based on the most accurate.
 - Trained on the same dataset.

The SVM returned a two-dimensional representation of the **RSEAC**'s lexical semantic field.



Machine Learning

Support Vector Machine

**The SVM sorted Racial
Probability By
Athlete Subcorpus**

The most impressive result
here is the *lack of a gradient*
in the probabilities of
category membership.

But, how?



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%

Machine Learning

Random Forest Modeling

But, how?



The RandomForest lets us crack into Lexical Importance to Categorization

The model output predicts which individual words were most useful to the SVM's categorization task.

Lexical Type	Importance to Race Sorting	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

Implications

But wait, aren't we talking about
Historical Sociolinguistics?

- ★ Caliskan et al. (2017) argue that algorithms are super racist.
 - They are, because they are trained on text
 - Text is made by fallible, implicitly biased humans
- ★ My machine learning + corpus method allows us to observe semantic change in apparent time **in any dataset, ever!**
 - And we really should because algorithms are super racist
- ★ Our knowledge is hobbled with half our data out of theoretical
—— consideration

JOIN NARNiHS

Every network needs more nodes!

- ★ Join the ListServ!
 - Send an email to listserv@lsv.uky.edu with an empty subject line and in the body of the email the simple message: subscribe NARNiHS yourfirstname yourlastname "
 - ★ Uploading your bio!
 - ★ Get our quarterly announcements!
 - ★ Organize a branded panel!
 - ★ Come to a conference!
-