

WISDOM FROM WORDS: INSIGHT FROM LANGUAGE AND TEXTUAL ANALYSIS

11th Triennial Invitational Choice Symposium
Hyatt Regency Chesapeake Bay Golf and Spa
May 29 – June 1, 2019

Hello again! We're looking forward to seeing everyone at the Symposium. We hope you'll enjoy this opportunity for in-depth interaction among scholars deeply interested in using text analysis to understand psychological and social phenomena. And that you'll also take advantage of the chance the symposium offers to mingle with a diverse range of elite scholars¹ during meal and break sessions.

-Jonah and Grant

WORKSHOP SUMMARY

Our session brings together researchers from a variety of disciplines (psychology, sociology, computer science, information science, and consumer behavior) to share insights and challenges on research that uses text to examine social and psychological phenomenon, and to explore new domains and methods to gain insight from text. While natural language processing work in computer science has created innovative tools, much of this work has focused on prediction, rather than understanding. Similarly, while social scientists have long been interested in language, with few exceptions (e.g., Pennebaker et al's Linguistic Inquiry and Word Count) this work has been hampered by the ability to accurately extract and quantify behavioral insight from text.

This workshop hopes to help bridge these divides. We raise a number of key questions, and by sharing new perspectives on language and methodological advances in textual analysis, begin to address them. Among the questions we will explore:

1. What does language *reveal* about attitudes and preferences? About the choices people will make? How can we extract and understand this information from texts?
2. How does language shape important choices and behaviors?
3. What are the methodological challenges in extracting meaning from words? As we move beyond the goal of simply measuring sentiment, how might we extract text measures that better align with human meaning, emotions, and judgments?

By analyzing the language people use, researchers can not only gain deeper insight into what choices and decisions people make, but *why* they made them.

¹ You can see who else will be at the Choice Symposium by clicking on any of the 25 workshop titles here--
<http://www.choicesymposium.org/accepted.html>.

WORKSHOP SCHEDULE

While we hope to keep things loose depending on how the conversation goes, we've come up with a tentative schedule we hope will provoke interesting, productive, and fun conversation.

Our main workshop days, Thursday and Friday (May 30-31), consist of--

1. Five minute "speed date" introductions, where each of us introduces themselves to the group in five minutes or less (and five slides or less). Have some fun here! Possible topics could include--
 - What academic graveyard tourists will see as your epitaph.
 - Your favorite or most important research discovery.
 - The project you wish you could find time for.
 - Something you wish you knew how to do (in work or otherwise).
 - The most interesting or confounding text you've ever seen.

To do: Prepare Powerpoint slides within range of [0,5]. Email these to gpackard@yorku.ca by Wednesday, May 29. If you're doing them on the plane or a napkin the night before, bring a thumb drive to our workshop room 15 minutes early (8:45am) Thursday morning.

2. 45 minute talk time for each participant. Ideally, presenters use this time not just to present, but to discuss. Lots of freedom beyond that... you can use part or all of this time to get some useful comments, suggestions, and ideas about work in progress. Additionally (or alternatively) you can share research or methods from somebody else (!) that you think is the best thing ever. We hope participants will be exposed to working theories, methods, and problems we can collaboratively learn from and help improve upon across radically different domains.

To do: We've put together a tentative schedule on the pages that follow based on the information you provided about your talk interests. If you'd like us to change or update anything, email gpackard@yorku.ca and he'll send around an updated schedule.

3. Three "Big Question" workshop sessions of 1-1.5 hours each. In these sessions, we'll discuss topics submitted by workshop participants such as (a) methodological opportunities and challenges (e.g., moving beyond dictionaries, sourcing, validation), (b) identifying important new domains that would benefit from exploration using text analysis, and (c) talking about ideas and challenges emerging from the workshop over drinks on a sunny patio overlooking the Chesapeake. These sessions may include breakouts into smaller groups.

The workshop organizers will collect notes from the "Big Question" sessions and share these with participants after the event in the hopes they might encourage future action or collaboration on the ideas developed.

SCHEDULE DETAIL (V3)

DAY	TIME	TOPIC
Wednesday, May 29	6-8:30pm	Welcome Reception / Dinner
Thursday, June 1	7:30-9:00am	Breakfast (hosted by the event)
	9:00-10:30am	1. Speed Date Intros (5 minute "getting to know you" presentations; see prior page for instructions!). 2. Narrative Development (Jonah)
	10:30-11:00am	Morning Coffee / Snack Break (hosted)
	11:00-12:30pm	1. Understanding Empathy (Lyle) 2. Some features I like, and some questions I have (Lillian)
	12:30-2:00pm	Lunch (hosted)
	2:00-3:30pm	1. Demographic-Aware NLP (Rada) 2. Legitimizing Language (Ashlee)
	3:30-4:00pm	Afternoon Coffee / Snack Break (hosted)
	4:00-5:30pm	Big Questions A: Beyond the Bag of Words. Inference Making from Language. Methods. Validation. (Jonah facilitates)
	6:00pm -	Dinner (hosted)
Friday, June 2	7:30-9:00am	Breakfast (hosted)
	9:00-10:30am	1. Conversational Markers of Anti-Social Dynamics (Cristian) 2. Exploring the Language of Care Conversations (Grant)
	10:30-11:00am	Morning Coffee / Snack Break (hosted)
	11:00-12:30pm	Big Questions B: Where is Language Going? And Where can we Take Text Analysis Next? (Grant facilitates)
	12:30-2:00pm	Lunch (on our own)
	2:00-3:30pm	1. Evolution of Crowdsourced Ideas (Wendy) 2. Identity and Upheaval in Social Media (Jamie)
	3:30-4:00pm	Afternoon Coffee / Snack Break (on our own)
	4:00-5:30pm	Big Questions C: Opportunities & Challenges over drinks on the patio at Blue Point Provision Company (10 minute walk)
	6:00pm -	Dinner with our group at Blue Point or Bistro Poplar (we have reservations at both) or on your own.
<i>Rather than offering extensive free time on Friday, we have compressed activity into the first two days of the event. This gives you freedom Saturday to enjoy networking or personal activities around the hotel, or to head home early. The event includes all meals on Saturday if you want to partake.</i>		
Saturday, June 3	7:30-9:00am	Breakfast (hosted)
	9:00-12:30am	<i>Free time (travel departures or enjoy the resort and area)</i>
	12:30-2:00pm	Lunch (hosted)
	2:00-6:00pm	<i>Free time (travel departures or enjoy the resort and area)</i>
	6:00pm -	Final Reception / Dinner / Party (hosted)

WORKSHOP PARTICIPANTS



Jonah Berger

Professor Jonah Berger is a world-renowned expert on word of mouth, social influence, consumer behavior, and how products, ideas, and behaviors catch on. He has published dozens of articles in top-tier academic journals, teaches Wharton's highest rated online course, and popular accounts of his work often appear in places like *The New York Times*, *Wall Street Journal*, and *Harvard Business Review*. His most recent work uses automated textual analysis and natural language processing to pull behavioral insights from text data (e.g., predicting song success from lyrics, movie success from scripts, and customer satisfaction from service calls).



Cristian Danescu-Niculescu-Mizil

Cristian is an assistant professor in the information science department at Cornell University. His research aims at developing computational frameworks that can lead to a better understanding of human social behavior, by unlocking the unprecedented potential of the large amounts of conversational data generated online. He is the recipient of several awards--including an NSF CAREER Award, the WWW 2013 Best Paper Award, a CSCW 2017 Best Paper Award, and two Google Faculty Research Awards--and his work has been featured in popular-media outlets such as the Wall Street Journal, NBC's The Today Show, NPR and the New York Times.



Ashlee Humphreys

Ashlee is an associate professor at the Medill School of Communications at Northwestern University. Her research investigates the role of legal and cultural institutions in creating markets, the influence of language on consumer judgments of legitimacy, and the process of consumer co-creation. Humphreys is the author of *Social Media: Enduring Principles*, and her work has been published in the Journal of Marketing, the Journal of Consumer Research and Sociology Compass. She also serves as an associate editor at the Journal of Consumer Research.



Lillian Lee

Lillian Lee is a professor of computer science and of information science at Cornell University, and co-editor-in-chief of the Transactions of the ACL (TACL). Her research interests include natural language processing and computational social science. She is a AAAI Fellow, an ACL Fellow, and an ACM Fellow. She received one of three inaugural awards for the Test of Time (2002-2012) Paper on Computational Linguistics (joint with Bo Pang), and best paper awards at NAACL 2004 (joint with Regina Barzilay) and the IJCAI 2016 Natural Language Processing meets Journalism workshop (joint with Liye Fu and Cristian Danescu-Niculescu-Mizil). She earned a citation in "Top Picks: Technology Research Advances of 2004" by Technology Research News (also joint with Regina Barzilay). Her work has received several mentions in the popular press, including The New York Times, NPR's All Things Considered, and NBC's The Today Show, and one of her co-authored papers on the memorability of movie quotes was publicly called "boring" by Youtubers Rhett and Link in a video viewed 2.6 million times.



Rada Mihalcea

Rada is a professor in the Language and Information Technologies (LIT) group at the University of Michigan and is director of the Michigan Artificial Intelligence Lab. She works on computational linguistics, with a focus on lexical semantics, multilingual natural language processing, and computational social sciences. She has held leadership roles in her research community, currently serving as Vice-President elect for the Association for Computational Linguistics. She is the recipient of a Presidential Early Career Award for Scientists and Engineers, awarded in 2010 by President Barack Obama and the 2019 Sarah Goddard Power Award for her work introducing young women and girls to the field of CS.



Wendy Moe

Wendy is the Associate Dean of Master's Programs, Dean's Professor of Marketing, and Co-Director of the Smith Analytics Consortium at the University of Maryland's Robert H. Smith School of Business. She is an expert in online and social media marketing with a focus on analytics. Wendy's research appears in numerous leading business journals. She is also the author of *Social Media Intelligence*. Professor Moe has been recognized by as a leading scholar in her field (Howard Award, MSI Young Scholar Award, Erin Anderson Award, Buzzell Award). She is co-editor of *Journal of Interactive Marketing*, serves on the Board of Trustees for the *Marketing Science Institute*, and is on several editorial boards.



Grant Packard

Grant is an associate professor at the Schulich School of Business at York University in Toronto. He studies the production and consumption of language in the marketplace-- from how and why people talk about products and services in online reviews and communities, to how customer service and sales people speak to customers, and how language itself is consumed as a product (e.g. song lyrics). Grant has been recognized for his research through the MSI Young Scholar Award, the Lazaridis Junior Research Excellence Award, the Leo Burnett Scholars Grant, and the University of Michigan Kendrick Award. Grant's research has been covered by media such as CBC, Globe and Mail, Harvard Business Review, NPR, and The Wall Street Journal.



James Pennebaker

James is the Regents Centennial Professor of Liberal Arts and Professor of Psychology at the University of Texas at Austin. He and his students are exploring natural language use, group dynamics, and personality in both laboratory and real world settings. His earlier work on expressive writing found that physical health and work performance can improve by simple writing and/or talking exercises. His cross-disciplinary research is related to linguistics, clinical and cognitive psychology, communications, medicine, and computer science. Author or editor of 12 books and over 300 articles, Pennebaker has received numerous research and teaching awards and honors.



Lyle Ungar

Lyle is a Professor in the Computer and Information Science department at the University of Pennsylvania. His research group develops scalable machine learning and text mining methods, including clustering, feature selection, and semi-supervised and multi-task learning for natural language, psychology, and medical research. He has published over 250 articles and holds ten patents. His current research focuses on statistical natural language processing, deep learning, and the use of social media to understand the psychology of individuals and communities. Example projects include spectral learning of language models, multi-view learning for gene expression and MRI data, and mining social media to better understand personality and well-being.