

EMNLP 2017

Workshop on Stylistic Variation

Proceedings of the Workshop

September 8, 2017
Copenhagen, Denmark

©2017 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-945626-99-9

Preface

The Workshop on Stylistic Variation (StyVa) at EMNLP 2017 is the first of its kind, offering a new venue for bringing together a large but previously underserved and splintered community within computational linguistics. Our goal in creating this workshop was to attract a variety of perspectives on style from traditional areas within NLP, including authorship attribution, author profiling, genre studies, natural language generation, distributional lexicography, and literary and educational applications; to this end we have defined stylistic variation as broadly as possible, to include any variation in phonological, lexical, syntactic, or discourse realization of particular semantic content, due to differences in extralinguistic variables such as individual speaker, speaker demographics, target audience, genre, etc.

We received 22 submissions, of which we accepted 14 (64%), seven as talks and seven as poster presentations. Though there was indeed a great deal of diversity in the submissions, including at least one submission in several of the major topic areas discussed above, we also noted a clear trend: we received several papers on style-sensitive language generation, particularly using neural network models. This clearly reflects a more general interest in the field, and one we would expect to continue. More generally, we are pleased that this workshop has served as a venue for both traditional and cutting-edge approaches to style.

We'd like to thank the authors for choosing StyVa as a venue for their excellent work, our invited speakers (Ani Nenkova and Walter Daelemans) for their invaluable contribution, and of course the reviews provided by our esteemed Program Committee. We'd also want to thank the ACL workshop organizing committee for giving us this opportunity to bring together the NLP stylistic community.

We look forward to a great workshop in Copenhagen!

Julian Brooke, Moshe Koppel, and Thamar Solorio

Organizers:

Julian Brooke, University of Melbourne
Moshe Koppel, Bar-Ilan University
Thamar Solorio, University of Houston

Program Committee:

Shlomo Argamon, Illinois Institute of Technology
Tim Baldwin, University of Melbourne
Ritwik Banerjee, Stony Brook University
Alberto Barrón-Cedeño, Qatar Computing Research Institute
Yves Bestgen, Université catholique de Louvain
Dasha Bogdanova, Dublin City University
Walter Daelemans, University of Antwerp
Jacob Eisenstein, Georgia Tech
Roger Evans, University of Brighton
Lucie Flekova, Technische Universität Darmstadt
Alexander Gelbukh, Instituto Politécnico Nacional
Rachel Greenstadt, Drexel University
Adam Hammond, San Diego State University
Graeme Hirst, University of Toronto
Dirk Hovy, University of Copenhagen
Eduard Hovy, Carnegie Mellon University
Diana Inkpen, University of Ottawa
Mike Kestemont, University of Antwerp
Ekaterina Kochmar, Cambridge University
Shibamouli Lahiri, University of Michigan
Annie Louis, University of Edinburgh
Suraj Maharjan, University of Houston
Manuel Montes-y-Gomez, Instituto Nacional de Astrofísica, Óptica y Electrónica
Francisco Rangel, Autoritas Consulting, S.A.
Ellie Pavlick, University of Pennsylvania
Barbara Plank, University of Groningen
Martin Potthast, Bauhaus-Universität Weimar
Paolo Rosso, Universitat Politècnica de València
Jan Rybicki, Jagiellonian University
Horacio Saggion, Universitat Pompeu Fabra
Prasha Shrestha, University of Houston
Anders Søgaard, University of Copenhagen
Lucia Specia, University of Sheffield
Efsthios Stamatatos, University of the Aegean
Benno Stein, Bauhaus-Universität Weimar
Joel Tetreault, Grammarly
Oren Tsur, Harvard University
Sandra Uitdenbogerd, RMIT University
Sowmya Vajjala, Iowa State University

Marilyn Walker, UC Santa Cruz
Wei Xu, Ohio State University
Marcos Zampeiri, University of Cologne

Invited Speakers:

Walter Daelemans, University of Antwerp
Ani Nenkova, University of Pennsylvania

Table of Contents

| | |
|--|-----|
| <i>From Shakespeare to Twitter: What are Language Styles all about?</i> Wei Xu | 1 |
| <i>Shakespearizing Modern Language Using Copy-Enriched Sequence to Sequence Models</i> Harsh Jhamtani, Varun Gangal, Eduard Hovy and Eric Nyberg | 10 |
| <i>Discovering Stylistic Variations in Distributional Vector Space Models via Lexical Paraphrases</i> Xing Niu and Marine Carpuat | 20 |
| <i>Harvesting Creative Templates for Generating Stylistically Varied Restaurant Reviews</i> Shereen Oraby, Sheideh Homayon and Marilyn Walker | 28 |
| <i>Is writing style predictive of scientific fraud?</i> Chloé Braud and Anders Søgaard | 37 |
| <i>"Deep" Learning : Detecting Metaphoricity in Adjective-Noun Pairs</i> Yuri Bizzoni, Stergios Chatzikiyriakidis and Mehdi Ghanimifard | 43 |
| <i>Authorship Attribution with Convolutional Neural Networks and POS-Eliding</i> Julian Hitschler, Esther van den Berg and Ines Rehbein | 53 |
| <i>Topic and audience effects on distinctively Scottish vocabulary usage in Twitter data</i> Philippa Shoemark, James Kirby and Sharon Goldwater | 59 |
| <i>Differences in type-token ratio and part-of-speech frequencies in male and female Russian written texts</i> Tatiana Litvinova, Pavel Seredin, Olga Litvinova and Olga Zagorovskaya | 69 |
| <i>Modeling Communicative Purpose with Functional Style: Corpus and Features for German Genre and Register Analysis</i> Thomas Haider and Alexis Palmer | 74 |
| <i>Stylistic Variation in Television Dialogue for Natural Language Generation</i> Grace Lin and Marilyn Walker | 85 |
| <i>Controlling Linguistic Style Aspects in Neural Language Generation</i> Jessica Fidler and Yoav Goldberg | 94 |
| <i>Approximating Style by N-gram-based Annotation</i> Melanie Andresen and Heike Zinsmeister | 105 |
| <i>Assessing the Stylistic Properties of Neurally Generated Text in Authorship Attribution</i> Enrique Manjavacas, Jeroen De Gussem, Walter Daelemans and Mike Kestemont | 116 |

Conference Program

Friday, September 8, 2017

09:00–10:30 Opening Session

09:00–09:10 *Opening remarks*

Julian Brooke, Tamar Solorio, and Moshe Koppel

09:10–10:00 *Invited Talk: Style Analysis for Practical Semantic Interpretation of Text*

Ani Nenkova

10:00–10:30 *From Shakespeare to Twitter: What are Language Styles all about?*

Wei Xu

10:30–11:00 Coffee Break

11:00–12:30 Technical Papers I

11:00–11:30 *Shakespearizing Modern Language Using Copy-Enriched Sequence to Sequence Models*

Harsh Jhamtani, Varun Gangal, Eduard Hovy and Eric Nyberg

11:30–12:00 *Discovering Stylistic Variations in Distributional Vector Space Models via Lexical Paraphrases*

Xing Niu and Marine Carpuat

12:00–12:30 *Harvesting Creative Templates for Generating Stylistically Varied Restaurant Reviews*

Shereen Oraby, Sheideh Homayon and Marilyn Walker

Friday, September 8, 2017 (continued)

12:30–14:00 Lunch

14:00–14:50 *Invited Talk: Problems in Personality Profiling*
Walter Daelemans

14:50–15:30 Poster Session

Is writing style predictive of scientific fraud?

Chloé Braud and Anders Søggaard

"Deep" Learning : Detecting Metaphoricity in Adjective-Noun Pairs

Yuri Bizzoni, Stergios Chatzikyriakidis and Mehdi Ghanimifard

Authorship Attribution with Convolutional Neural Networks and POS-Eliding

Julian Hitschler, Esther van den Berg and Ines Rehbein

Topic and audience effects on distinctively Scottish vocabulary usage in Twitter data

Philippa Shoemark, James Kirby and Sharon Goldwater

Differences in type-token ratio and part-of-speech frequencies in male and female Russian written texts

Tatiana Litvinova, Pavel Seredin, Olga Litvinova and Olga Zagorovskaya

Modeling Communicative Purpose with Functional Style: Corpus and Features for German Genre and Register Analysis

Thomas Haider and Alexis Palmer

Stylistic Variation in Television Dialogue for Natural Language Generation

Grace Lin and Marilyn Walker

Friday, September 8, 2017 (continued)

15:30–16:00 Coffee Break

16:00–17:30 Technical Papers II

16:00–16:30 *Controlling Linguistic Style Aspects in Neural Language Generation*
Jessica Fidler and Yoav Goldberg

16:30–17:00 *Approximating Style by N-gram-based Annotation*
Melanie Andresen and Heike Zinsmeister

17:00–17:30 *Assessing the Stylistic Properties of Neurally Generated Text in Authorship Attribution*
Enrique Manjavacas, Jeroen De Gussem, Walter Daelemans and Mike Kestemont

17:30–17:35 *Closing Remarks*
Julian Brooke, Tamar Solorio, and Moshe Koppel

