Introduction

The 2nd Workshop on Stylistic Variation (StyVa) at NAACL 2018 follows up the successful first iteration of the workshop at EMNLP 2017. The goal of the workshop is to offer a venue for bringing together a large but previously underserved and splintered community within computational linguistics, attracting 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. This spirit of diversity is reflected this year particularly in our first invited speaker, James Pennebaker, whose main research has been published in psychology (though we note that his LIWC lexicon is a standard tool for NLP researchers).

In this iteration, we received 11 submissions, of which we accepted 6 as talks (54.5%); 1 paper was later withdrawn due to acceptance elsewhere and instead of the paper we will have a third invited talk. Even with a smaller set of papers than last year, the topics covered are diverse, including stylistic difference due to social variables, language background, and genre. Perhaps the most striking difference compared to our last iteration was a relative lack of natural language generation papers (we did have one, though!).

We thank the authors for choosing StyVa as a venue for their excellent work, all of our invited speakers (James Pennebaker, Rada Mihalcea, and Barbara Plank) 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 their support.

We would also like to thank Thomson Reuters for their generous sponsorship of this workshop.

We look forward to a great workshop in New Orleans!

Julian Brooke
Lucie Flekova
Moshe Koppel
Thamar Solorio
Organizers:

Julian Brooke, Thomson Reuters
Lucie Flekova, Amazon Research
Moshe Koppel, Bar-Ilan University
Thamar Solorio, University of Houston

Program Committee:

Nikolaos Aletras (University of Sheffield)
Yves Bestgen (Université catholique de Louvain)
Alberto Barrón-Cedeño (Qatar Computing Research Institute)
Walter Daelemans (University of Antwerp)
Jacob Eisenstein (Georgia Tech)
Roger Evans (University of Brighton)
Alexander Gelbukh (Instituto Politécnico Nacional)
Adam Hammond (San Diego State University)
Graeme Hirst (University of Toronto)
Ekaterina Kochmar (Cambridge University)
Vasileios Lampos (University College London)
Dominique Legallois (Université Paris 3, Sorbonne-Nouvelle)
Manuel Montes-y-Gomez (Instituto Nacional de Astrofísica, Optica y Electronica)
Dong Nguyen (Alan Turing Institute)
Umashanthi Pavalanathan (Georgia Institute of Technology)
Ellie Pavlick (University of Pennsylvania)
Barbara Plank (University of Groningen)
Martin Potthast (Leipzig University)
Vinod Prabhakaran (Computer Science, Stanford)
Daniel Preotiuc-Pietro (University of Pennsylvania)
Emily Prud’hommeaux (Rochester Institute of Technology)
Sudha Rao (University of Maryland)
Maarten Sap (University of Washington)
Andrew H. Schwartz (Stony Brook University)
Anders Soegaard (University of Copenhagen)
Benno Stein (Bauhaus-Universität Weimar)
Joel Tetreault (Grammarly)
Sandra Uitdenbogerd (RMIT University)
Sowmya Vajjala (Iowa State University)
Svitlana Volkova (Pacific Northwest National Laboratory)
Wei Xu (Ohio State University)
Marcos Zampeiri (University of Wolverhampton)

Invited Speakers:

Prof. Rada Mihalcea, University of Michigan
Prof. James W. Pennebaker, University of Texas at Austin
Prof. Barbara Plank, University of Groningen
Stylistic variation over 200 years of court proceedings according to gender and social class
Stefania Degaetano-Ortlieb ................................................................. 1

Stylistic Variation in Social Media Part-of-Speech Tagging
Murali Raghu Babu Balusu, Taha Merghani and Jacob Eisenstein ................. 11

Detecting Syntactic Features of Translated Chinese
Hai Hu, Wen Li and Sandra Kübler ....................................................... 20

Evaluating Creative Language Generation: The Case of Rap Lyric Ghostwriting
Peter Potash, Alexey Romanov and Anna Rumshisky ................................. 29

Cross-corpus Native Language Identification via Statistical Embedding
Francisco Rangel, Paolo Rosso, Julian Brooke and Alexandra Uitdenbogerd ........ 39
Conference Program

Tuesday, June 5, 2018

9:15–9:30  Opening Remarks

9:30–10:30  Invited Talk by James W. Pennebaker: Measuring Linguistic Variation with Function Words

10:30–11:00  Break

11:00–11:30  Stylistic variation over 200 years of court proceedings according to gender and social class
  Stefania Degaetano-Ortlieb

11:30–12:00  Stylistic Variation in Social Media Part-of-Speech Tagging
  Murali Raghu Babu Balusu, Taha Merghani and Jacob Eisenstein

12:00–12:30  Detecting Syntactic Features of Translated Chinese
  Hai Hu, Wen Li and Sandra Kübler

12:30–14:00  Lunch

14:00–15:00  Invited Talk by Rada Mihalcea: What Does Language Tell us about the People Behind it

15:00–15:30  Evaluating Creative Language Generation: The Case of Rap Lyric Ghostwriting
  Peter Potash, Alexey Romanov and Anna Rumshisky
Tuesday, June 5, 2018 (continued)

15:30–16:00 Break

16:00–17:00 Invited Talk by Barbara Plank: Author Profiling from Text and Beyond

17:00–17:30 Cross-corpus Native Language Identification via Statistical Embedding
Francisco Rangel, Paolo Rosso, Julian Brooke and Alexandra Uitdenbogerd