NAACL HLT 2016

The 2016 Conference of the
North American Chapter of the
Association for Computational Linguistics:
Human Language Technologies

Proceedings of the Fifth Workshop on Computational Linguistics for Literature

June 16, 2016
San Diego, California, USA

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 978-1-941643-80-8

Preface

Welcome to the fifth (yes!) edition of the Workshop on Computational Linguistics for Literature, aptly nicknamed CLfL. We started our workshops as a venue for computational linguists who had an interest in processing literary texts. That somewhat underspecified, and a slightly narrow, mandate has evolved over the next four years. A rather special research community has emerged, with NLP people alongside researchers in Digital Humanities, literary scholars, poets and more. You can trace a brief history of CLfL at its latest Web site.[1]

As a token of growing maturity of our workshop series, in October 2015 a special journal issue appeared, volume 12 of *Linguistic Issues in Language Technology*,[2] with expanded versions of several papers from the CLfL workshops held in 2012-2014.

Before we tell you about the program, a big thank-you. As usual, we owe a debt of gratitude to our excellent program committee, some of whom have been with us all these years. Their hard work ensures the high quality of papers. Their dedication, passion and insights are greatly appreciated!

This year the participants will be entertained and maybe intrigued by a nicely varied program.

The workshop exists in the "twilight zone" between computation and art. Our distinguished invited speakers highlight this duality. The morning speaker is Professor Patrick Winston, one of the central figures in Artificial Intelligence. He will talk about MIT's Genesis story-understanding system and its many complex, varied aspects. Genesis reads *Macbeth*... This kind of project is what working with literature should really be about! ⌣̈

The afternoon invited speaker is a poet and a scholar, a strong voice in Digital Poetics, Loss Pequeño Glazier from SUNY Buffalo. He will tell us about poetic expression in array spaces of computational constellation, and illustrate it with his own take on Robin Blaser's influential "The Moth Poem".

There are seven regular papers on a variety of exciting topics, some of which have never been discussed at this workshop. Moshe Koppel, Moty Michaely and Alex Tall tell us about reconstructing the most original form of ancient manuscripts when multiple copies are available. Angel Daza, Hiram Calvo and Jesús Figueroa-Nazuno present their work on generating stories that are characterized by literary style. An intriguing paper by Xanda Schofield and Leo Mehr analyzes stereotypical gender roles using a corpus of film scripts.

Poetry, as always, figures strongly at the workshop. Arya Rahgozar and Diana Inkpen discuss ways in which one can distinguish automatically the periods in the Persian poet Hafez's life when he wrote his ghazals. Alex Estes and Christopher Hench take us on a brief tour of Middle High German epic poetry: how a Conditional Random Field model can automatically annotate meter in such poems.

Andrea Gagliano, Emily Paul, Kyle Booten and Marti Hearst look at figurative speech, a frequent guest in poetry. They show how continuous word vectors, a new promising technique, help generate interesting figure-of-speech relationships.

[1] https://sites.google.com/site/clfl2016/ – there are links to the Web sites for the previous workshops.

[2] http://csli-lilt.stanford.edu/

Marie Dubremetz and Joakim Nivre present another chapter of their work on detecting the rhetorical phenomenon of *chiasmus* in text.

All papers will be presented orally. Some of the authors will also put up posters. You can look at them during the day, and perhaps talk to the authors during the breaks.

There will be one new element at this year's workshop: a software demo. Olga Scrivner from Indiana University will show her team's Interactive Text Mining Suite, meant to visualize data for literary studies. She will also present a brief teaser right after lunch.

Well, that is it, more or less. Without further ado, let us invite you to enjoy this volume. We expect to see you in San Diego! ☺

Anna Feldman, Anna Kazantseva and Stan Szpakowicz

Patrick Henry Winston (http://people.csail.mit.edu/phw/)

Short bio

A graduate of MIT, Patrick Winston is Ford Professor of Artificial Intelligence and MacVicar Faculty Fellow at the Massachusetts Institute of Technology. His Genesis research group focuses on developing a computational account of human intelligence and how it differs from that of other species, with special attention to modeling human story telling and comprehension. Author and editor of numerous books, including *Artificial Intelligence*, he served as Director of the MIT Artificial intelligence Laboratory (now part of MIT's CSAIL Laboratory) for 25 years. He is now Research Coordinator for the multi-university, multi-disciplinary Center for Brains, Minds, and Machines centered at MIT.

Invited talk: *Genesis Reads Macbeth: The Role of Stories in Human Intelligence*

Abstract

I believe that human story competence—understanding, telling, authoring—lies at the center of human intelligence. To better understand that competence, my students and I are building the Genesis story-understanding system, a system that reads 100-sentence stories adapted from sources such as *Hansel and Gretel*, Crow creation myths, and Shakespeare's plays. I explain how our work has been guided by computational imperatives and challenged by our determination to model aspects of conceptual understanding, cultural bias, hypothetical reflection, personality modeling, listener-aware telling, on demand authoring, mental illness, and concept-based summary and search. I conclude with speculations on whether Genesis can really understand literature without a body, relating the question to the metaphor of the cave in Plato's Republic.

Loss Pequeño Glazier (http://epc.buffalo.edu/authors/glazier/)

Short bio

Dr. Loss Pequeño Glazier is Director of the Electronic Poetry Center/E-Poetry Festivals and Professor, Department of Media Study, SUNY Buffalo. The EPC is the world's largest digital resource for innovative and digital poetry. Glazier is the author of two books-in-progress as well as Digital Poetics: The Making of E-Poetries (Univ. of Alabama Press, 2002), *Anatman, Pumpkin Seed, Algorithm* (Salt Publishing, 2003), *Small Press: An Annotated Guide* (Greenwood, 1992), and hundreds of poems, essays, film, visual art, sound and digital works, as well as projects for dance, music, installations, and performance, including at the Neuberger Museum (SUNY Purchase), Royal Festival Hall (London), Instituto del Libro (La Habana), Guggenheim Museum (New York), UCLA Hammer Museum, Kulturforum Potsdamer Platz (Berlin), University of London, Le Divan du Monde (Paris), Bowery Poetry Club (New York), Brown University, and the Palazzo delle Arti Napoli. Glazier's work in digital writing focuses on code and its discontents, whether in natural language permutation, translation, computer programming, computational linguistics and aesthetic, spatial, and poetics. His author page contains numerous examples of his work.

Invited talk: *"The Not-Moth": Poetic Expression in Array Spaces of Computational Constellation*

Abstract

My talk provides an introduction to what I call "array poetics": using computer-generated groupings of natural language strings to explore new resonances of poetic space. To explore this space, I describe "The Not-Moth", my own digital poem written as a response to, and reflection on, Robin Blaser's "The Moth Poem", an early, influential poem in the San Francisco Renaissance of the 1960s. I investigate the dynamics, edition particulars, and the poetics of Blaser's original poem sequence. Then I think about how the qualities of this influential early work might be cast using the tools of today's technology – not to re-write the poem, but to respond to it through a computer-media composition reflecting some of its forms and language framings, while adding my own digital insights to the undertaking. To this end, the techniques he explored, innovative at the time, are viewed through my lens of "array poetics", the extension of nuances of textual variance through the computer manipulation of poetic language. In my talk, I give an overview of array poetics, including a non-technical explanation of "strings" (individual lines of natural language poetry), the "array" (groupings of such lines), and their arrangement ("coding") in exploring the possibilities of variant phrasing in expressive electronic language. Looking at the "array", I also suggest the implications of the "space between" strings, in the same way one might glimpse the "space between" lines of printed poetry, but here expanded in nuance through the basic digital frameworks (HTML, CSS, and select open-source tools) available in New Media writing. Ultimately the aim is, through technology – as well despite technology – to produce literary language in New Media in its full richness and expressive depth.

Program Committee

Apoorv Agarwal (Columbia University)
Cecilia Ovesdotter Alm (Rochester Institute of Technology)
Sandy Baldwin (Rochester Institute of Technology)
David Bamman (Carnegie Melon University)
Peter Boot (Huygens Institute for Netherlands History)
Chris Brew (Thomson-Reuters)
Julian Brooke (University of Toronto)
Micha Elsner (Ohio State University)
David Elson (Google)
Mark Finlayson (MIT)
Pablo Gervás Gómez-Navarro (Universidad Complutense de Madrid)
Roxana Girju (University of Illinois at Urbana-Champaign)
Amit Goyal (University of Maryland)
Graeme Hirst (University of Toronto)
Matthew Jockers (University of Nebraska–Lincoln)
Justine Kao (Stanford University)
Mike Kestemont (University of Antwerp)
Rada Mihalcea (University of Michigan)
David Mimno (Cornell University)
Saif Mohammad (National Research Council Canada)
Nick Montfort (Massachusetts Institute of Technology)
Sebastian Padó (Universität Stuttgart)
Livia Polanyi (LDM Associates)
Owen Rambow (Columbia University)
Caroline Sporleder (Georg-August-Universität Göttingen)
Reid Swanson (University of California, Santa Cruz)
Rob Voigt (Stanford University)
Marilyn Walker (University of California, Santa Cruz)
Bei Yu (Syracuse University)

Invited Speakers

Patrick Henry Winston, Massachusetts Institute of Technology
Loss Pequeño Glazier (State University of New York at Buffalo)

Organizers

Anna Feldman (Montclair State University)
Anna Kazantseva (National Research Council of Canada)
Stan Szpakowicz (University of Ottawa)

Table of Contents

Conference Program

Thursday, June 16, 2016

Session I

9:00–9:05 Welcome

9:05–10:00 *Genesis reads Macbeth: The role of stories in human intelligence* (invited talk)
Patrick Henry Winston

10:00–10:30 *Supervised Machine Learning for Hybrid Meter*
Alex Estes and Christopher Hench

Coffee break

Session II

11:00–11:30 *Automatic Text Generation by Learning from Literary Structures*
Angel Daza, Hiram Calvo and Jesús Figueroa-Nazuno

11:30–12:00 *Intersecting Word Vectors to Take Figurative Language to New Heights*
Andrea Gagliano, Emily Paul, Kyle Booten and Marti A. Hearst

12:00–12:30 *Gender-Distinguishing Features in Film Dialogue*
Alexandra Schofield and Leo Mehr

Thursday, June 16, 2016 (continued)

Lunch break

Session III

14:00–14:05 *Interactive Text Mining Suite, a teaser to accompany a demo*
Olga Scrivner

14:05–15:00 *"The Not-Moth": Poetic Expression in Array Spaces of Computational Constellation*
Loss Pequeño Glazier

15:00–15:30 *Reconstructing Ancient Literary Texts from Noisy Manuscripts*
Moshe Koppel, Moty Michaely and Alex Tal

Coffee break

Session IV

16:00–16:30 *Syntax Matters for Rhetorical Structure: The Case of Chiasmus*
Marie Dubremetz and Joakim Nivre

16:30–17:00 *Bilingual Chronological Classification of Hafez's Poems*
Arya Rahgozar and Diana Inkpen

17:00–17:15 Wrap-up