COLING 2018

**The 27th International Conference
on Computational Linguistics**

**Proceedings of the Workshop on
Computational Modeling of Polysynthetic Languages**

Saturday, August 25, 2018
Santa Fe, New Mexico, USA

Introduction

The workshop on Computational Modeling of Polysynthetic Languages addresses the needs for documentation, archiving, creation of corpora and teaching materials that are specific to polysynthetic languages. Documentation and corpus-building challenges arise for many languages, but the complex morphological makeup of polysynthetic languages makes consistent documentation particularly difficult. This workshop is the first ever meeting where researchers and practitioners working on polysynthetic languages discuss common problems and difficulties, and it is intended as the capstone to establishing possible collaborations and ongoing partnerships of the relevant issues. One of our intentions is to discuss the possibility of creating a shared task in order to increase opportunities to share challenges and solutions.

One of the goals of the workshop is to create dialog between those language professionals who collect and annotate language data for polysynthetic languages, those who are committed to linguistic analysis, those who develop and apply computational methods to these languages, and those who are dedicated to revitalization through policy, education and community activism. Too often, these communities do not interact enough to benefit each other, so there is a lost opportunity cost all around. This lost opportunity, especially in the case of endangered languages, is one that cannot be recuperated. Thus, it is urgent to work towards the goal of leveraging each other's efforts.

On a lighter note, the Workshop has had an informal pre-title "All Together Now" to reflect the linguistic and typological nature of polysynthesis.

Organizers

Workshop Chair:

Judith L. Klavans, Senior Scientist, Army Research Laboratory and University of Maryland
Ph.D. Linguistics. University College, University of London, England
Judith.l.klavans.civ@mail.mil judith.klavans@gmail.com

Organizing Committee:

Anna Kazantseva, Research Officer - National Research Council of Canada
Ph.D. Computer Science, University of Ottawa
Anna.Kazantseva@nrc-cnrc.gc.ca

Roland Kuhn, Senior Research Officer at National Research Council of Canada
Ph.D. Computer Science, McGill University
Roland.Kuhn@cnrc-nrc.gc.ca

Steve LaRocca, Team Lead, Multilingual Computing, Army Research Laboratory
Ph.D. Linguistics, Georgetown University
stephen.a.larocca.civ@mail.mil

Jeffrey Micher, Researcher, Multilingual Computing, Army Research Laboratory
M.A. Linguistics, University of Pittsburgh,
M.S. Language Technologies Carnegie Mellon University
jeffrey.c.micher.civ@mail.mil

Clare Voss, Team Lead, Multilingual Computing, Army Research Laboratory
Ph.D. Computer Science, University of Maryland
clare.r.voss.civ@mail.mil

Consulting Co-organizers:

Maria Polinsky, Professor and Associate Director of the Language Science Center, University of Maryland. Ph.D. Linguistics, Institute for Linguistics of the Russian Academy of Sciences, Moscow.
polinsky@umd.edu

Omer Preminger, Associate Professor of Linguistics, University of Maryland. Ph.D. Linguistics, The Massachusetts Institute of Technology.
omerp@umd.edu

Program Committee:

Since the goal of this workshop has been to explore practical applications of recent developments in linguistics and computational linguistics to the preservation and revitalization of North American indigenous languages, and to build on the long history of research on polysynthesis combined with the more current computational interest in processing morphologically complex languages, the Program Committee consists of theoretical linguists, computational linguists, anthropological linguists, experts in language revitalization. Thus we have established a valuable balance in expertise.

Antti Arppe, University of Alberta
Mark Baker, Rutgers University
Steven Bird, Charles Darwin University
Aaron Broadwell, University of Florida
Lauren Clemens, State University of New York - Albany
Christopher Cox, Carleton University
Amy Dahlstrom, University of Chicago
Henry Davis, University of British Columbia
Jeff Good, State University of New York - Buffalo
Jeremy Green, Six Nations Polytechnic
David Harrison, Swarthmore College
Gary Holton, University of Hawaii - Mānoa
Marianne Ignace, Simon Fraser University
Judith Klavans, Army Research Laboratory
Roland Kuhn, National Research Council of Canada
Stephen LaRocca, Army Research Laboratory
Lori Levin, Carnegie Mellon University, National Research Council of Canada
Gina-Anne Levow, University Washington
Patrick Littell, Carnegie Mellon University
Alexa Little, 7000 Languages project
Jordan Lachler, University of Alberta
Mitch Marcus, University of Pennsylvania
Michael Maxwell, Center for Advanced Study of Language, University of Maryland
Jeffrey Micher, Army Research Laboratory
Marianne Mithun, University California - Santa Barbara
Timothy Montler, University of North Texas
Rachel Nordlinger, University of Melbourne
Boyan Onyshkevich, US Government
Carl Rubino – US Government
Lane Schwartz, University Illinois - Urbana-Champaign
Richard Sproat, Google Research
Clare Voss, Army Research Laboratory
Michelle Yuan, Massachusetts Institute of Technology

Invited Speaker: Brian Maracle (Owennatekha), Kanyen'kéha (Mohawk) First Nation[1]

Brian Maracle, also known as, author, journalist and radio host (born in 1947 in Detroit, Michigan). Brian Maracle is a member of the Mohawk First Nation and a passionate advocate for the preservation of the Kanyen'kehaka (Mohawk) language.

Brian Maracle was raised on the Six Nations Grand River Territory Reserve in Ohsweken, Ontario, until he was five when his family moved to the state of New York. He earned a BA from Dartmouth College in 1969 before going to work for Indigenous organizations in Vancouver during the 1970s. In 1980 he relocated to Ottawa to study journalism at Carleton University, graduating in 1982. Maracle was a journalist in the 1980s, writing for the *Globe and Mail* and covering Indigenous issues for mainstream and Indigenous media. He also hosted the CBC Radio program *Our Native Land*, which began in 1965 and ran for 21 years, the first national radio program focused on Indigenous culture and issues.

In 1993, Maracle's first book, *Crazywater: Native Voices on Addiction and Recovery*, was published. The book is the compilation of 200 interviews, channelling traditional oral history, conducted across the country by Maracle during three years of research. The interview subjects are exemplified by the book's 75 people who represent a cross-section of society. The importance of the book is in the voice it gives to the Indigenous perspective on alcohol and drugs, which comes long after the dominant white culture—academics, social scientists, government authorities and medical experts—has expounded on the issue.

After the book came out, Maracle left Ottawa and moved back to his reserve. His experiences in returning led Maracle to write *Back on the Rez, finding the way home* which was published in 1996. He recounts the struggles of an urban dweller assimilating to life in the country while he struggles with understanding the Mohawk language. He points out how the dominant white culture has played a significant role in destroying Indigenous identity and culture, but notes also the flaws in Indigenous society and how the community is torn between white culture and tradition.

After Maracle moved back to his reserve, he began to learn the Mohawk language and, with his wife Audrey, established a full-time adult immersion language school, Onkwawenna Kentyohkwa, in 1998. Maracle eventually all but abandoned his writing career to devote himself to revitalizing the Mohawk language, employing his skills to host a radio program called *Tewatonhwehsen!* (Let's have a good time!) and writing a blog in Mohawk. In 2012, Maracle's new media collaboration with his daughter Zoe Leigh Hopkins was presented at the 13th Annual imagineNATIVE Film and Media Arts Festival. Their sound art piece, *Karenniyohston – Old Songs Made Good*, blends oral language and sound art in a 30-minute adaptation of five national and cultural anthems from Canada, the UK and the US.

---

[1]Taken from https://www.thecanadianencyclopedia.ca/en/article/brian-maracle/, author Laura Nielson Bonikowdsky (2013), Revised by Michele Felice (2017)

Panel: "Revitalization: Bridging Technology and Language Education"

The purpose of this panel is to address and debate some of the controversial issues that arise in the process of establishing better communication and building a way to bridge the research and applications gaps between and among researchers and practitioners. Some of these issues include such provocative and often divisive points.

Invited Panel Speakers

Inée Yang Slaughter, Executive Director, Indigenous Language Institute

Inée Slaughter has been with the Indigenous Language Institute since 1995. ILI was founded as the Institute for the Preservation of the Original Languages of the Americas (IPOLA) by Joanna Hess in 1992. In 1997, IPOLA extended its scope and in 2000, IPOLA was changed to Indigenous Language Institute (ILI) to reflect working relations with all indigenous communities, nationally and internationally. ILI provides the tools and training to help First Nation language teachers and learners help themselves in their efforts to bring language back into everyday lives of the People. ILI runs local and national workshops and teacher training sessions as well as the annual Indigenous Language Institute Symposium (ILIS) in New Mexico that invites presenters to address topics for language practitioners. Inée is of Korean heritage, born and raised in Japan, and is fluent in Japanese, Korean, and English.

Zoe Leigh Hopkins, Independent Filmaker and Director, Educator, Activist

Zoe Hopkins grew up in a Heiltsuk fishing village on the coast of British Columbia, home to her mother and maternal family, in the heart of the Great Bear Rainforest. She is an alumna of the Sundance Institute's Feature Film Program and her short films have premiered at the Sundance Film Festival and the Worldwide Short Film Festival. Her 2013 short, *Mohawk Midnight Runners* was named Best Canadian Short Drama at the 2013 imagineNATIVE Film + Media Arts Festival in Toronto, Ontario. In 2017, her feature film *Kayak to Klemtu* won the Air Canada Audience Choice Award. She now lives in her father's community of Six Nations, where she teaches the Mohawk language online to students across Turtle Island (North America). (from http://www.northernstars.ca/zoe-hopkins/)

Lori Levin, Research Professor, Language Technologies Institute, Carnegie Mellon University

Lori Levin holds a B.A. Linguistics, University of Pennsylvania, 1979 and a Ph.D. Linguistics, MIT, 1986. Her research interests range from theoretical to computational linguistics. In the language documentation area, she was funded by the Chilean Ministry of Education under the Language Technologies Institute's AVENUE project to collect data and produce language technologies that support bilingual education. The main resource that came out of this partnership is a Mapudungun-Spanish parallel corpus consisting of approximately 200,000 words of text and 120 hours of transcribed speech. In the education domain, Lori has been a leader and organizer of the North American Computational Linguistics Olympiad (NACLO), a contest in which high-school students solve linguistic puzzles thereby learning about the diversity and consistency of language.

Table of Contents

Conference Program

08:45–10:30   **Session 1**

08:45–09:15   *Computational Challenges for Polysynthetic Languages*
Judith L. Klavans

09:15–09:40   *A Neural Morphological Analyzer for Arapaho Verbs Learned from a Finite State Transducer*
Sarah Moeller, Ghazaleh Kazeminejad, Andrew Cowell and Mans Hulden

09:40–10:05   *Finite-state morphology for Kwak'wala: A phonological approach*
Patrick Littell

10:05–10:30   *A prototype finite-state morphological analyser for Chukchi*
Vasilisa Andriyanets and Francis Tyers

10:30–11:00   **Coffee Break**

11:00–12:30   **Session 2**

11:00–12:00   *Invited Talk*
Brian Maracle (Owennatekha), Mohawk First Nation

12:00–12:30   ***Discussion: Opportunities for Collaboration***

12:30–13:50   **Lunch**

13:50–15:50   **Session 3**

13:50–14:15   *Natural Language Generation for Polysynthetic Languages: Language Teaching and Learning Software for Kanyen'kéha (Mohawk)*
Greg Lessard, Nathan Brinklow and Michael Levison

14:15–14:40   *Kawennón:nis: the Wordmaker for Kanyen'kéha*
Anna Kazantseva, Owennatekha Brian Maracle, Ronkwe'tiyóhstha Josiah Maracle and Aidan Pine

14:40–15:30   **Session 4**

14:40–15:30   *Panel: "Revitalization: Bridging Technology and Language Education"*
Judith L. Klavans, Chair; Inee Slaughter, Executive Director of Indigenous Language Institute; Zoe Lee Hopkins, Kanien'kehá:ka/Heiltsuk, Film Director and Educator; Dr. Lori Levin, Research Professor, Language Technologies Institute, CMU

15:30–15:50   *Corpora and Shared Task*
Judith L. Klavans, Chair

15:50–16:20   **Coffee Break**

16:20–17:35   **Session 5**

16:20–16:45   *Using the Nunavut Hansard Data for Experiments in Morphological Analysis and Machine Translation*
Jeffrey Micher

16:45–17:10   *Lost in Translation: Analysis of Information Loss During Machine Translation Between Polysynthetic and Fusional Languages*
Manuel Mager, Elisabeth Mager, Alfonso Medina-Urrea, Ivan Vladimir Meza Ruiz and Katharina Kann

17:10–17:35   *Automatic Glossing in a Low-Resource Setting for Language Documentation*
Sarah Moeller and Mans Hulden

17:35–18:00   **Way Forward**