

NAACL HLT 2010

**Workshop on Computational
Linguistics and Writing:
Writing Processes and
Authoring Aids
(CL&W 2010)**

Proceedings of the Workshop

June 6, 2010
Los Angeles, California

USB memory sticks produced by
Omnipress Inc.
2600 Anderson Street
Madison, WI 53707
USA

©2010 The Association for Computational Linguistics

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

Introduction

Writing today, whether professional, academic, or private, relies heavily on computers. Most texts composed in the 21st century are probably written on computers or other electronic devices, such as mobile phones. People compose texts in word processors, text editors, content management systems, blogs, wikis, e-mail clients, and instant messaging applications. Each of these tools supports authors in different ways and to different degrees.

Writing research has been concerned with word processing since the 1970s. Writing researchers today investigate specific characteristics of writing with computers and the effect of tools on writing processes. The current rise of new writing environments and genres (e.g., blogging) has prompted new studies in this area of research.

During the last few decades, computational linguistics has mostly been concerned with static or finished texts. We believe there is now a growing need to explore how computational linguistics can support human text production and word processing. However, there are still very few projects where computational linguists and writing researchers work together here.

The Workshop on Computational Linguistics and Writing (CL&W 2010) provides an overview of current developments in the area of computational linguistics for authoring aids, and an overview of recent advances in writing research. CL&W 2010 continues and builds on the workshops on authoring aids at the Sixth International Conference on Language Resources and Evaluation (LREC 2008) and the 2008 Swedish Language Technology Conference (SLTC 2008). The papers included here present research that explores writing processes and text production, as well as actual systems that support writers. In both areas, research on all languages is relevant, including less-resourced languages. CL&W 2010 brings together researchers from both communities, to identify areas where computational linguistics and writing research can benefit from each other and to stimulate discussion and interdisciplinary cooperation between these two areas of research.

In our call for papers we posed some questions:

- How can writing be supported by methods, resources, and tools from computational linguistics? This includes NLP tools and techniques that can be used or have been used to support writing (e.g., grammar and style checking, document structuring, thematic segmentation, and editing and revision aids).
- How can we gain a better understanding of writing processes, strategies, and needs? How can techniques from HCI research and psychology help us to gain new insights into the composition and writing processes, and to improve writing tools?
- Which methods, resources, and tools from computational linguistics might support research in this area?
- How do high-level writing processes and the mechanics of writing relate to each other?
- How does the tool used influence composition (including editing and revising)? Are writers aware of the possibilities and limitations of their writing tools?

- Is there a need for the development of new writing tools? What can we learn from earlier approaches and tools like RUSKIN, Writer’s Workbench, or Augment, or from source code editors for programming languages?
- How can insights from writing research and methods from computational linguistics help to support the needs of particular user groups (e.g., foreign language learners, children, persons with disabilities)?

We received 15 submissions from both computational linguists and writing researchers. After a rigorous review process we selected 9 papers for the workshop. We would like to thank the members of the Program Committee for their excellent work—the reviews were all very thorough, carefully written, detailed, and helped the authors to improve their papers.

The papers cover a variety of topics, ranging from actually working—and freely available!—systems to support novice and expert authors, to more general thoughts on the intersection of writing research and computational linguistics. We are pleased to present these papers in this volume.

We hope the work presented here will trigger discussion and collaboration between researchers, bringing together expertise and interest from writing research and computational linguistics.

Michael Piotrowski, Cerstin Mahlow, and Robert Dale

Organizers:

Michael Piotrowski, University of Zurich (Switzerland)
Cerstin Mahlow, University of Zurich (Switzerland)
Robert Dale, Macquarie University (Australia)

Program Committee:

Gerd Bräuer, Zurich University of Applied Sciences (Switzerland)
Jill Burstein, ETS (USA)
Rickard Domeij, The Language Council of Sweden (Sweden)
Kevin Egan, University of Southern California (USA)
Caroline Hagège, Xerox Research Centre Europe (France)
Soe Johansson Kokkinakis, University of Gothenburg (Sweden)
Ola Karlsson, The Language Council of Sweden (Sweden)
Ola Knutsson, KTH (Sweden)
Sabine Lehmann, acrolinx GmbH (Switzerland)
Eva Lindgren, Umeå University (Sweden)
Aurlien Max, LIMSI (France)
Guido Nottbusch, University of Bielefeld (Germany)
Daniel Perrin, Zurich University of Applied Sciences (Switzerland)
Martin Reynaert, Tilburg University (The Netherlands)
Gert Rijlaarsdam, University of Amsterdam (The Netherlands)
Dietmar Rösner, Otto-von-Guericke University Magdeburg (Germany)
Koenraad de Smedt, University of Bergen (Norway)
Sylvana Sofkova Hashemi, University of Gothenburg (Sweden)
Scott Warnock, Drexel University (USA)
Eric Wehrli, University of Geneva (Switzerland)
Carl Whithaus, UC Davis (USA)
Michael Zock, CNRS (France)

Table of Contents

<i>Computational Linguistics in the Translator’s Workflow—Combining Authoring Tools and Translation Memory Systems</i>	
Christoph Rösener	1
<i>Scientific Authoring Support: A Tool to Navigate in Typed Citation Graphs</i>	
Ulrich Schäfer and Uwe Kasterka	7
<i>Grammaticality Judgement in a Word Completion Task</i>	
Alfred Renaud, Fraser Shein and Vivian Tsang	15
<i>The Design of a Proofreading Software Service</i>	
Raphael Mudge	24
<i>A Toolkit to Assist L2 Learners Become Independent Writers</i>	
John Milton and Vivying S.Y. Cheng	33
<i>Learning Simple Wikipedia: A Cogitation in Ascertaining Abecedarian Language</i>	
Courtney Napoles and Mark Dredze	42
<i>Questions Worth Asking: Intersections between Writing Research and Computational Linguistics</i>	
Anne Ruggles Gere and Laura Aull	51
<i>Exploring Individual Differences in Student Writing with a Narrative Composition Support Environment</i>	
Julius Goth, Alok Baikadi, Eun Young Ha, Jonathan Rowe, Bradford Mott and James Lester . .	56
<i>The Linguistics of Readability: The Next Step for Word Processing</i>	
Neil Newbold and Lee Gillam	65

Workshop Program

Sunday, June 6, 2010

09:00 Opening

Session 1

09:15–09:40 *Computational Linguistics in the Translator’s Workflow—Combining Authoring Tools and Translation Memory Systems*
Christoph Rösener

09:40–10:05 *Scientific Authoring Support: A Tool to Navigate in Typed Citation Graphs*
Ulrich Schäfer and Uwe Kasterka

10:05–10:30 *Grammaticality Judgement in a Word Completion Task*
Alfred Renaud, Fraser Shein and Vivian Tsang

10:30 Coffee Break

Session 2

11:00–11:25 *The Design of a Proofreading Software Service*
Raphael Mudge

11:25–11:50 *A Toolkit to Assist L2 Learners Become Independent Writers*
John Milton and Vivying S.Y. Cheng

11:50–12:15 *Learning Simple Wikipedia: A Cogitation in Ascertaining Abecedarian Language*
Courtney Napoles and Mark Dredze

12:15 Lunch Break

Sunday, June 6, 2010 (continued)

Session 3

- 1:40–2:05 *Questions Worth Asking: Intersections between Writing Research and Computational Linguistics*
Anne Ruggles Gere and Laura Aull
- 2:05–2:30 *Exploring Individual Differences in Student Writing with a Narrative Composition Support Environment*
Julius Goth, Alok Baikadi, Eun Young Ha, Jonathan Rowe, Bradford Mott and James Lester
- 2:30–2:55 *The Linguistics of Readability: The Next Step for Word Processing*
Neil Newbold and Lee Gillam
- 3:00 Coffee Break
- Discussion**
- 3:30 Discussion
- 5:00 Closing