

COLING 2012

**24th International Conference on
Computational Linguistics**

**Proceedings of the
First Workshop on Eye-tracking and
Natural Language Processing**

**Workshop chairs:
Michael Carl, Pushpak Bhattacharya and
Kamal Kumar Choudhary**

**15 December 2012
Mumbai, India**

Diamond sponsors

Tata Consultancy Services
Linguistic Data Consortium for Indian Languages (LDC-IL)

Gold Sponsors

Microsoft Research
Beijing Baidu Netcon Science Technology Co. Ltd.

Silver sponsors

IBM, India Private Limited
Crimson Interactive Pvt. Ltd.
Yahoo
Easy Transcription & Software Pvt. Ltd.

Proceedings of the First Workshop on Eye-tracking and Natural Language Processing

Michael Carl, Pushpak Bhattacharya and Kamal Kumar Choudhary (eds.)
Revised preprint edition, 2012

Published by The COLING 2012 Organizing Committee
Indian Institute of Technology Bombay,
Powai,
Mumbai-400076
India
Phone: 91-22-25764729
Fax: 91-22-2572 0022
Email: pb@cse.iitb.ac.in

This volume © 2012 The COLING 2012 Organizing Committee.
Licensed under the *Creative Commons Attribution-Noncommercial-Share Alike 3.0 Nonported* license.
<http://creativecommons.org/licenses/by-nc-sa/3.0/>
Some rights reserved.

Contributed content copyright the contributing authors.
Used with permission.

Also available online in the ACL Anthology at <http://aclweb.org>

Preface

The workshop “Eye-tracking and Natural Language Processing” (ETNLP) is an exploratory ½ satellite workshop in the context of Coling 2012, Mumbai. Over the last decades, eye-tracking systems have been used for reading research, human-computer interaction, user modeling and usability studies, system evaluation and feature extraction, as well as for on-line applications such as real-time human-machine interaction, intelligent user adaptation and gaze-based control.

The ETNLP workshop sets out to explore the state of the art in eye-tracking techniques and Natural Language Processing. Given the interdisciplinary nature of the workshop, we invited papers on all topics related to gaze-based language and/or reading research, gaze-based computational psycholinguistics, eye movements and attention in natural language understanding, corpora of gaze data in language processing and other, so as to foster and explore eye-tracking methods in modeling natural language processing.

We received eight submissions relevant to the theme of the workshop from different parts of the world. All submissions were reviewed in a double-blind process by three reviewers and finally five papers were accepted for presentation. In addition to the five papers, we invited Matthew Crocker, Saarland University, to talk about his eye-tracking research on spoken interaction. These presentations cover a broad area of topics, including processing effort in reading, parsing and translation, models on linguistic complexity and surprisal, investigation into classification of scanpaths and systematic error correction, as well as gaze behavior in dialog and spoken interaction. A plethora of experimental and computational methods are used and introduced to analyze the gaze data, such as a morpho-syntactic surprisal index, a seasim measure, the dirac delta and conditional random fields. Many of the contributions show a tendency to enlarge the scope of the processing units, moving from single word fixations to scan paths, syntactic chunks, the reading line, usage of micro and macro units, etc. so as to capture more general gaze-movement and text processing strategies.

The workshop provides an opportunity for researchers to present their work and to interact with peers around the world in this emerging field of research. It is an opportunity for people working in different fields such as Computational Linguistics, Psycholinguistics, and Computational Psycholinguistics to come together and share their views on Eye-tracking and Natural Language Processing.

The workshop organizers would like to thank the authors for their contributions and the programme committee for their review work. Special thanks goes to the Coling organizers, who made this event possible.

We look forward to seeing you all at ETNLP 2012

Michael Carl, Pushpak Bhattacharyya, Kamal Kumar Choudhary

ETNLP 2012 Workshop Organizers

Organizers:

Michael Carl (Copenhagen Business School, Denmark)
Pushpak Bhattacharya (IIT Mumbai, India)
Kamal Kumar Choudhary (IIT Ropar, India)

Invited Speaker:

Matthew Crocker (Saarland University, Germany)

Programme Committee:

Pushpak Bhattacharyya (IIT Mumbai, India)
Ralf Biedert (DFKI, Germany)
Sharon O'Brien (DCU, Ireland)
Michael Carl (Copenhagen Business School, Denmark)
Kamal Kumar Choudhary (IIT Ropar, India)
Matthew Crocker (Saarland University, Germany)
Silvia Hansen-Schirra (University of Mainz, Germany)
Kenneth Holmqvist (Humanities Lab, Sweden)
Sam Hutton (SR-research, Canada)
Frank Keller (University of Edinburgh)
Pascual Martinez (University of Tokyo, Japan)
Ricardo Matos (Tobii, Sweden)
Mattias Nilsson (Uppsala University, Sweden)
Kari-Jouko Räihä (University of Tampere, Finland)
RMK Sinha (JSS Academy of Technical Education, India)
Oleg Spakov (University of Tampere, Finland)
Sara Stymne (Linköping University, Sweden)
Per Henning Uppstad (Lesecenteret, Norway)

Table of Contents

<i>Grounding spoken interaction with real-time gaze in dynamic virtual environments</i> Matthew Crocker	1
<i>Identifying instances of processing effort in translation through heat maps: an eye-tracking study using multiple input sources</i> Fabio Alves, José Luiz Gonçalves and Karina Szpak	5
<i>Robustness and processing difficulty models. A pilot study for eye-tracking data on the French Treebank</i> Stéphane Rauzy and Philippe Blache	21
<i>Scanpaths in reading are informative about sentence processing</i> Titus von der Malsburg, Shravan Vasishth and Reinhold Kliegl	37
<i>Predicting Word Fixations in Text with a CRF Model for Capturing General Reading Strategies among Readers</i> Tadayoshi Hara, Daichi Mochihashi, Yoshinobu Kano and Akiko Aizawa	55
<i>A heuristic-based approach for systematic error correction of gaze data for reading</i> Abhijit Mishra, Michael Carl and Pushpak Bhattacharyya	71

First Workshop on Eye-tracking and Natural Language Processing

Program

Saturday, 15 December 2012

- 09:30–10:20 **Invited talk:** *Grounding spoken interaction with real-time gaze in dynamic virtual environments*
Matthew Crocker
- 10:20–10:45 *Identifying instances of processing effort in translation through heat maps: an eye-tracking study using multiple input sources*
Fabio Alves, José Luiz Gonçalves and Karina Szpak
- 10:45–11:10 *Robustness and processing difficulty models. A pilot study for eye-tracking data on the French Treebank*
Stéphane Rauzy and Philippe Blache
- 11:10–12:00 Tea break
- 12:00–12:25 *Scanpaths in reading are informative about sentence processing*
Titus von der Malsburg, Shravan Vasishth and Reinhold Kliegl
- 12:25–12:50 *Predicting Word Fixations in Text with a CRF Model for Capturing General Reading Strategies among Readers*
Tadayoshi Hara, Daichi Mochihashi, Yoshinobu Kano and Akiko Aizawa
- 12:50–13:15 *A heuristic-based approach for systematic error correction of gaze data for reading*
Abhijit Mishra, Michael Carl and Pushpak Bhattacharyya

