

Coling 2008

**22nd International Conference on
Computational Linguistics**

**Proceedings of the workshop on
Human Judgements in Computational
Linguistics**

Workshop chairs:
Ron Artstein, Gemma Boleda,
Frank Keller, and Sabine Schulte im Walde

23 August 2008
Manchester, UK

©2008 The Coling 2008 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

Order copies of this and other Coling 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 978-1-905593-49-1

Design by Chimney Design, Brighton, UK
Production and manufacture by One Digital, Brighton, UK

Introduction

The papers in these proceedings were presented at the *Coling 2008 workshop on human judgements in Computational Linguistics*, held in Manchester on 23 August 2008.

Human judgements play a key role in the development and the assessment of linguistic resources and methods in Computational Linguistics. They are commonly used in the creation of lexical resources and corpus annotation, and also in the evaluation of automatic approaches to linguistic tasks: In the developmental phase, human judgements help to define an inventory of categories as well as robust annotation criteria, and in the assessment phase they are used to evaluate the results of automatic systems against existing linguistic standards. Furthermore, systematically collected human judgements provide clues for research on linguistic issues that underlie the judgement task, providing insights complementary to introspective analysis or evidence gathered from corpora.

The workshop included papers about experiments that collect human judgements for Computational Linguistic purposes. A particular focus of the workshop was concerned with human judgements on ‘controversial’ linguistic tasks (those that are not clear from a theoretical point of view, such as many tasks having to do with semantics or pragmatics). Such experimental tasks are usually difficult to design and interpret, and they typically result in low agreement scores. They are often poorly documented in the literature; however, they are especially well-suited as a basis for a fruitful discussion.

We were pleased to receive 22 submissions, of which we accepted eight papers for presentation at the workshop. We would like to thank the program committee for the excellent job they did in reviewing the submissions. As well as the paper presentations, the workshop also included a keynote talk by Martha Palmer of the University of Colorado at Boulder. These proceedings include an abstract of her talk.

The workshop was supported by the Spanish Education and Science Ministry via the KNOW project (TIN2006-15049-C03-03) and by Sonderforschungsbereich 732, Universität Stuttgart.

Ron Artstein
Gemma Boleda
Frank Keller
Sabine Schulte im Walde

Organizers:

Ron Artstein, University of Southern California
Gemma Boleda, Universitat Politècnica de Catalunya
Frank Keller, University of Edinburgh
Sabine Schulte im Walde, Universität Stuttgart

Programme Committee:

Toni Badia, Universitat Pompeu Fabra
Marco Baroni, University of Trento
Beata Beigman Klebanov, Northwestern University
André Blessing, Universität Stuttgart
Chris Brew, Ohio State University
Kevin Cohen, University of Colorado Health Sciences Center
Barbara Di Eugenio, University of Illinois at Chicago
Katrín Erk, University of Texas at Austin
Stefan Evert, University of Osnabrück
Afsaneh Fazly, University of Toronto
Alex Fraser, Universität Stuttgart
Jesus Gimenez, Universitat Politècnica de Catalunya
Roxana Girju, University of Illinois at Urbana-Champaign
Ed Hovy, University of Southern California
Nancy Ide, Vassar College
Adam Kilgarriff, University of Brighton
Alexander Koller, University of Edinburgh
Anna Korhonen, University of Cambridge
Mirella Lapata, University of Edinburgh
Diana McCarthy, University of Sussex
Alissa Melinger, University of Dundee
Paola Merlo, University of Geneva
Sebastian Padó, Stanford University
Martha Palmer, University of Colorado
Rebecca Passonneau, Columbia University
Massimo Poesio, University of Trento
Sameer Pradhan, BBN Technologies
Horacio Rodriguez, Universitat Politècnica de Catalunya
Bettina Schrader, Universität Potsdam
Suzanne Stevenson, University of Toronto

Invited Speaker:

Martha Palmer, University of Colorado at Boulder

Table of Contents

<i>Invited Talk: The Relevance of a Cognitive Model of the Mental Lexicon to Automatic Word Sense Disambiguation</i>	
Martha Palmer and Susan Brown	1
<i>Analyzing Disagreements</i>	
Beata Beigman Klebanov, Eyal Beigman and Daniel Diermeier	2
<i>Exploiting ‘Subjective’ Annotations</i>	
Dennis Reidsma and Rieks Op den Akker	8
<i>Human Judgement as a Parameter in Evaluation Campaigns</i>	
Jean-Baptiste Berthelin, Cyril Grouin, Martine Hurault-Plantet and Patrick Paroubek	17
<i>Native Judgments of Non-Native Usage: Experiments in Preposition Error Detection</i>	
Joel Tetreault and Martin Chodorow	24
<i>Polysemy in Verbs: Systematic Relations between Senses and their Effect on Annotation</i>	
Anna Rumshisky and Olga Batiukova	33
<i>Eliciting Subjectivity and Polarity Judgements on Word Senses</i>	
Fangzhong Su and Katja Markert	42
<i>Human Judgements in Parallel Treebank Alignment</i>	
Martin Volk, Torsten Marek and Yvonne Samuelsson	51
<i>An Agreement Measure for Determining Inter-Annotator Reliability of Human Judgements on Affective Text</i>	
Plaban Kumar Bhowmick, Anupam Basu and Pabitra Mitra	58

Workshop Programme

Saturday, 23 August 2008

- 9:20–9:30 Introductory remarks
- 9:30–10:30 *Invited Talk: The Relevance of a Cognitive Model of the Mental Lexicon to Automatic Word Sense Disambiguation*
Martha Palmer and Susan Brown
- 10:30–11:00 Break
- 11:00–11:30 *Analyzing Disagreements*
Beata Beigman Klebanov, Eyal Beigman and Daniel Diermeier
- 11:30–12:00 *Exploiting ‘Subjective’ Annotations*
Dennis Reidsma and Rieks Op den Akker
- 12:00–12:30 *Human Judgement as a Parameter in Evaluation Campaigns*
Jean-Baptiste Berthelin, Cyril Grouin, Martine Hurault-Plantet and Patrick Paroubek
- 12:30–14:00 Lunch
- 14:00–14:30 *Native Judgments of Non-Native Usage: Experiments in Preposition Error Detection*
Joel Tetreault and Martin Chodorow
- 14:30–15:00 *Polysemy in Verbs: Systematic Relations between Senses and their Effect on Annotation*
Anna Rumshisky and Olga Batiukova
- 15:00–15:30 *Eliciting Subjectivity and Polarity Judgements on Word Senses*
Fangzhong Su and Katja Markert
- 15:30–16:00 Break
- 16:00–16:30 *Human Judgements in Parallel Treebank Alignment*
Martin Volk, Torsten Marek and Yvonne Samuelsson
- 16:30–17:00 *An Agreement Measure for Determining Inter-Annotator Reliability of Human Judgements on Affective Text*
Plaban Kumar Bhowmick, Anupam Basu and Pabitra Mitra
- 17:00–17:30 Discussion

