

NAACL HLT 2010

**Workshop on  
Creating Speech and  
Language Data with  
Amazon's Mechanical Turk**

**Proceedings of the Workshop**

June 6, 2010  
Los Angeles, California

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

Generously Sponsored By:

**amazon**mechanical turk  
beta  Artificial Artificial Intelligence

and

CrowdFlower 

©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](mailto:acl@aclweb.org)

## Preface

The NAACL-2010 Workshop on Creating Speech and Language Data With Amazon’s Mechanical Turk explores applications of crowdsourcing technologies for the creation and study of language data. Recent work has evaluated the effectiveness of using crowdsourcing platforms, such as Amazon’s Mechanical Turk, to create annotated data for natural language processing applications. This workshop further explores this area and these proceedings contain 34 papers and an overview paper that each experiment with applications of Mechanical Turk. The diversity of applications showcases the new possibilities for annotating speech and text, and has the potential to dramatically change how we create data for human language technologies.

Papers in the workshop also looked at best practices in creating data using Mechanical Turk. Experiments evaluated how to design Human Intelligence Tasks (HITs), how to attract users to the task, how to price annotation tasks, and how to ensure data quality. Applications include the creation of data sets for standard NLP tasks, developing entirely new tasks, and investigating new ways of integrating user feedback in the learning process.

The workshop featured an open-ended shared task in which 35 teams were awarded \$100 of credit on Amazon Mechanical Turk to spend on an annotation task of their choosing. Results of the shared task are described in short papers and all collected data is publicly available. Shared task participants focused on data collection questions, such as how to convey complex tasks to non-experts, how to evaluate and ensure quality and annotation cost and speed.

The organizers thank the workshop participants who contributed to an incredibly strong workshop program. We also thank the program committee for quickly reviewing the large number of submissions. Special thanks go to Sharon Chiarella, vice president of Amazon Mechanical Turk, for funding the shared task, Ted Sandler of Amazon for assistance in organizing the shared task, Stephanie Geerlings and Lukas Biewald of CrowdFlower for making their service available to shared task participants, and to Jonny Weese for editing and compiling the final proceedings.



**Organizers:**

Chris Callison-Burch, Johns Hopkins University  
Mark Dredze, Johns Hopkins University

**Program Committee:**

Breck Baldwin, Alias-i, Inc.  
Jordan Boyd-Graber, University of Maryland  
Michael Bloodgood, Johns Hopkins University  
Bob Carpenter, Alias-i, Inc.  
David Chen, University of Texas, Austin  
Maxine Eskenazi, Carnegie Mellon University  
Nikesh Garera, Kosmix Corporation  
Jim Glass, Massachusetts Institute of Technology  
Alex Gruenstein, Google, Inc.  
Janna Hamaker, Amazon.com, Inc.  
Jon Hamaker, Microsoft Corporation  
Samer Hassan, University of North Texas  
Alexandre Klementiev, Johns Hopkins University  
Benjamin Lambert, Carnegie Mellon University  
Ben Leong, University of North Texas  
Ian McGraw, Massachusetts Institute of Technology  
Scott Novotney, Johns Hopkins University  
Brendan O'Connor, Carnegie Mellon University  
Gabriel Parent, Carnegie Mellon University  
Massimo Poesio, University of Essex  
Joe Polifroni, Nokia Research Labs  
Joseph Reisinger, University of Texas, Austin  
Ted Sandler, Amazon.com, Inc.  
Stephanie Seneff, Massachusetts Institute of Technology  
Kevin Small, Tufts University  
Rion Snow, Stanford University / Twitter, Inc.



## Table of Contents

<i>Creating Speech and Language Data With Amazon’s Mechanical Turk</i> Chris Callison-Burch and Mark Dredze .....	1
<i>Corpus Creation for New Genres: A Crowdsourced Approach to PP Attachment</i> Mukund Jha, Jacob Andreas, Kapil Thadani, Sara Rosenthal and Kathleen McKeown .....	13
<i>Clustering dictionary definitions using Amazon Mechanical Turk</i> Gabriel Parent and Maxine Eskenazi .....	21
<i>Semi-supervised Word Alignment with Mechanical Turk</i> Qin Gao and Stephan Vogel .....	30
<i>Rating Computer-Generated Questions with Mechanical Turk</i> Michael Heilman and Noah A. Smith .....	35
<i>Crowdsourced Accessibility: Elicitation of Wikipedia Articles</i> Scott Novotney and Chris Callison-Burch .....	41
<i>Document Image Collection Using Amazon’s Mechanical Turk</i> Audrey Le, Jerome Ajoy, Mark Przybocki and Stephanie Strassel .....	45
<i>Using Amazon Mechanical Turk for Transcription of Non-Native Speech</i> Keelan Evanini, Derrick Higgins and Klaus Zechner .....	53
<i>Exploring Normalization Techniques for Human Judgments of Machine Translation Adequacy Collected Using Amazon Mechanical Turk</i> Michael Denkowski and Alon Lavie .....	57
<i>Can Crowds Build parallel corpora for Machine Translation Systems?</i> Vamshi Ambati and Stephan Vogel .....	62
<i>Turker-Assisted Paraphrasing for English-Arabic Machine Translation</i> Michael Denkowski, Hassan Al-Haj and Alon Lavie .....	66
<i>Annotating Large Email Datasets for Named Entity Recognition with Mechanical Turk</i> Nolan Lawson, Kevin Eustice, Mike Perkowitz and Meliha Yetisgen-Yildiz .....	71
<i>Annotating Named Entities in Twitter Data with Crowdsourcing</i> Tim Finin, William Murnane, Anand Karandikar, Nicholas Keller, Justin Martineau and Mark Dredze .....	80
<i>MTurk Crowdsourcing: A Viable Method for Rapid Discovery of Arabic Nicknames?</i> Chiara Higgins, Elizabeth McGrath and Laila Moretto .....	89
<i>An Enriched MT Grammar for Under \$100</i> Omar F. Zaidan and Juri Ganitkevitch .....	93

<i>Using the Amazon Mechanical Turk to Transcribe and Annotate Meeting Speech for Extractive Summarization</i>	
Matthew Marge, Satanjeev Banerjee and Alexander Rudnicky .....	99
<i>Using Mechanical Turk to Annotate Lexicons for Less Commonly Used Languages</i>	
Ann Irvine and Alexandre Klementiev .....	108
<i>Opinion Mining of Spanish Customer Comments with Non-Expert Annotations on Mechanical Turk</i>	
Bart Mellebeek, Francesc Benavent, Jens Grivolla, Joan Codina, Marta R. Costa-Jussà and Rafael Banchs .....	114
<i>Crowdsourcing and language studies: the new generation of linguistic data</i>	
Robert Munro, Steven Bethard, Victor Kuperman, Vicky Tzuyin Lai, Robin Melnick, Christopher Potts, Tyler Schnoebelen and Harry Tily .....	122
<i>Not-So-Latent Dirichlet Allocation: Collapsed Gibbs Sampling Using Human Judgments</i>	
Jonathan Chang .....	131
<i>Collecting Image Annotations Using Amazon’s Mechanical Turk</i>	
Cyrus Rashtchian, Peter Young, Micah Hodosh and Julia Hockenmaier .....	139
<i>Non-Expert Evaluation of Summarization Systems is Risky</i>	
Dan Gillick and Yang Liu .....	148
<i>Shedding (a Thousand Points of) Light on Biased Language</i>	
Tae Yano, Philip Resnik and Noah A. Smith .....	152
<i>Evaluation of Commonsense Knowledge with Mechanical Turk</i>	
Jonathan Gordon, Benjamin Van Durme and Lenhart Schubert .....	159
<i>Cheap Facts and Counter-Facts</i>	
Rui Wang and Chris Callison-Burch .....	163
<i>The Wisdom of the Crowds Ear: Speech Accent Rating and Annotation with Amazon Mechanical Turk</i>	
Stephen Kunath and Steven Weinberger .....	168
<i>Crowdsourcing Document Relevance Assessment with Mechanical Turk</i>	
Catherine Grady and Matthew Lease .....	172
<i>Preliminary Experiments with Amazon’s Mechanical Turk for Annotating Medical Named Entities</i>	
Meliha Yetisgen-Yildiz, Imre Solti, Fei Xia and Scott Halgrim .....	180
<i>Tools for Collecting Speech Corpora via Mechanical-Turk</i>	
Ian Lane, Matthias Eck, Kay Rottmann and Alex Waibel .....	184
<i>Measuring Transitivity Using Untrained Annotators</i>	
Nitin Madnani, Jordan Boyd-Graber and Philip Resnik .....	188
<i>Amazon Mechanical Turk for Subjectivity Word Sense Disambiguation</i>	
Cem Akkaya, Alexander Conrad, Janyce Wiebe and Rada Mihalcea .....	195



<i>Non-Expert Correction of Automatically Generated Relation Annotations</i>	
Matthew R. Gormley, Adam Gerber, Mary Harper and Mark Dredze .....	204
<i>Using Mechanical Turk to Build Machine Translation Evaluation Sets</i>	
Michael Bloodgood and Chris Callison-Burch .....	208
<i>Creating a Bi-lingual Entailment Corpus through Translations with Mechanical Turk: \$100 for a 10-day Rush</i>	
Matteo Negri and Yashar Mehdad .....	212
<i>Error Driven Paraphrase Annotation using Mechanical Turk</i>	
Olivia Buzek, Philip Resnik and Ben Bederson .....	217



# Workshop Program

**Sunday, June 6, 2010**

9:00            **Morning Session**

*Creating Speech and Language Data With Amazon's Mechanical Turk*

Chris Callison-Burch and Mark Dredze

9:10            Invited Talk

10:10          *Corpus Creation for New Genres: A Crowdsourced Approach to PP Attachment*

Mukund Jha, Jacob Andreas, Kapil Thadani, Sara Rosenthal and Kathleen McKeown

10:30          **Coffee Break**

11:00          **Poster Session 1**

*Clustering dictionary definitions using Amazon Mechanical Turk*

Gabriel Parent and Maxine Eskenazi

*Semi-supervised Word Alignment with Mechanical Turk*

Qin Gao and Stephan Vogel

*Rating Computer-Generated Questions with Mechanical Turk*

Michael Heilman and Noah A. Smith

*Crowdsourced Accessibility: Elicitation of Wikipedia Articles*

Scott Novotney and Chris Callison-Burch

*Document Image Collection Using Amazon's Mechanical Turk*

Audrey Le, Jerome Ajot, Mark Przybocki and Stephanie Strassel

*Using Amazon Mechanical Turk for Transcription of Non-Native Speech*

Keelan Evanini, Derrick Higgins and Klaus Zechner

*Exploring Normalization Techniques for Human Judgments of Machine Translation*

*Adequacy Collected Using Amazon Mechanical Turk*

Michael Denkowski and Alon Lavie

**Sunday, June 6, 2010 (continued)**

*Can Crowds Build parallel corpora for Machine Translation Systems?*

Vamshi Ambati and Stephan Vogel

*Turker-Assisted Paraphrasing for English-Arabic Machine Translation*

Michael Denkowski, Hassan Al-Haj and Alon Lavie

*Annotating Large Email Datasets for Named Entity Recognition with Mechanical Turk*

Nolan Lawson, Kevin Eustice, Mike Perkowitz and Meliha Yetisgen-Yildiz

*Annotating Named Entities in Twitter Data with Crowdsourcing*

Tim Finin, William Murnane, Anand Karandikar, Nicholas Keller, Justin Martineau and Mark Dredze

*MTurk Crowdsourcing: A Viable Method for Rapid Discovery of Arabic Nicknames?*

Chiara Higgins, Elizabeth McGrath and Laila Moretto

*An Enriched MT Grammar for Under \$100*

Omar F. Zaidan and Juri Ganitkevitch

12:30

**Lunch**

1:30

**Afternoon Session 1**

*Using the Amazon Mechanical Turk to Transcribe and Annotate Meeting Speech for Extractive Summarization*

Matthew Marge, Satanjeev Banerjee and Alexander Rudnicky

*Using Mechanical Turk to Annotate Lexicons for Less Commonly Used Languages*

Ann Irvine and Alexandre Klementiev

*Opinion Mining of Spanish Customer Comments with Non-Expert Annotations on Mechanical Turk*

Bart Mellebeek, Francesc Benavent, Jens Grivolla, Joan Codina, Marta R. Costa-Jussà and Rafael Banchs

*Crowdsourcing and language studies: the new generation of linguistic data*

Robert Munro, Steven Bethard, Victor Kuperman, Vicky Tzuyin Lai, Robin Melnick, Christopher Potts, Tyler Schnoebelen and Harry Tily

*Not-So-Latent Dirichlet Allocation: Collapsed Gibbs Sampling Using Human Judgments*

Jonathan Chang

**Sunday, June 6, 2010 (continued)**

3:10           **Coffee Break**

3:30           **Afternoon Session 2**

*Collecting Image Annotations Using Amazon's Mechanical Turk*

Cyrus Rashtchian, Peter Young, Micah Hodosh and Julia Hockenmaier

*Non-Expert Evaluation of Summarization Systems is Risky*

Dan Gillick and Yang Liu

*Shedding (a Thousand Points of) Light on Biased Language*

Tae Yano, Philip Resnik and Noah A. Smith

4:30           **Poster Session 2**

*Evaluation of Commonsense Knowledge with Mechanical Turk*

Jonathan Gordon, Benjamin Van Durme and Lenhart Schubert

*Cheap Facts and Counter-Facts*

Rui Wang and Chris Callison-Burch

*The Wisdom of the Crowds Ear: Speech Accent Rating and Annotation with Amazon Mechanical Turk*

Stephen Kunath and Steven Weinberger

*Crowdsourcing Document Relevance Assessment with Mechanical Turk*

Catherine Grady and Matthew Lease

*Preliminary Experiments with Amazon's Mechanical Turk for Annotating Medical Named Entities*

Meliha Yetisgen-Yildiz, Imre Solti, Fei Xia and Scott Halgrim

*Tools for Collecting Speech Corpora via Mechanical-Turk*

Ian Lane, Matthias Eck, Kay Rottmann and Alex Waibel

*Measuring Transitivity Using Untrained Annotators*

Nitin Madnani, Jordan Boyd-Graber and Philip Resnik

**Sunday, June 6, 2010 (continued)**

*Amazon Mechanical Turk for Subjectivity Word Sense Disambiguation*

Cem Akkaya, Alexander Conrad, Janyce Wiebe and Rada Mihalcea

*Non-Expert Correction of Automatically Generated Relation Annotations*

Matthew R. Gormley, Adam Gerber, Mary Harper and Mark Dredze

*Using Mechanical Turk to Build Machine Translation Evaluation Sets*

Michael Bloodgood and Chris Callison-Burch

*Creating a Bi-lingual Entailment Corpus through Translations with Mechanical Turk:  
\$100 for a 10-day Rush*

Matteo Negri and Yashar Mehdad

*Error Driven Paraphrase Annotation using Mechanical Turk*

Olivia Buzek, Philip Resnik and Ben Bederson