

# **Fourth Widening Natural Language Processing Workshop 2020**

Seattle, Washington, USA  
5 July 2020

ISBN: 978-1-7138-1380-4

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2020) by the Association for Computational Linguistics  
All rights reserved.

Printed with permission by Curran Associates, Inc. (2020)

For permission requests, please contact the Association for Computational Linguistics  
at the address below.

Association for Computational Linguistics  
209 N. Eighth Street  
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006

Fax: 1-570-476-0860

[acl@aclweb.org](mailto:acl@aclweb.org)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

# TABLE OF CONTENTS

CORPUS BASED AMHARIC SENTIMENT LEXICON GENERATION .....	1
<i>Girma Neshir Alemneh ; Andreas Rauber ; Solomon Atnafu</i>	
NEGATION HANDLING FOR AMHARIC SENTIMENT CLASSIFICATION .....	4
<i>Girma Neshir Alemneh ; Andreas Rauber ; Solomon Atnafu</i>	
EMBEDDING ORIENTED ADAPTABLE SEMANTIC ANNOTATION FRAMEWORK FOR AMHARIC WEB DOCUMENTS.....	7
<i>Kidane Woldemariam ; Fekade Getahun</i>	
SIMILARITY AND FARNNESS BASED BIDIRECTIONAL NEURAL CO-ATTENTION FOR AMHARIC NATURAL LANGUAGE INFERENCE .....	8
<i>Abebawu Eshetu ; Getenesh Teshome ; Ribka Alemayehu</i>	
LARGE VOCABULARY READ SPEECH CORPORA FOR FOUR ETHIOPIAN LANGUAGES: AMHARIC, TIGRIGNA, OROMO, AND WOLAYTTA .....	13
<i>Solomon Teferra Abate ; Martha Yifiru Tachbelie ; Michael Melese ; Hafte Abera ; Tewodros Gebreselassie ; Wondwossen Mulugeta ; Yaregal Assabie ; Million Meshesha Beyene ; Solomon Atnafu ; Binyam Ephrem Seyoum</i>	
SIMPLEX-PB 2.0: A RELIABLE DATASET FOR LEXICAL SIMPLIFICATION IN BRAZILIAN PORTUGUESE.....	18
<i>Nathan Hartmann ; Gustavo Henrique Paetzold ; Sandra Aluísio</i>	
TOKEN LEVEL IDENTIFICATION OF MULTIWORD EXPRESSIONS USING CONTEXTUAL INFORMATION .....	23
<i>Reyhaneh Hashempour ; Aline Villavicencio</i>	
BI-DIRECTIONAL ANSWER-TO-ANSWER CO-ATTENTION FOR SHORT ANSWER GRADING USING DEEP LEARNING .....	26
<i>Abebawu Eshetu ; Getenesh Teshome ; Ribka Alemahu</i>	
EFFECTIVE QUESTIONS IN REFERENTIAL VISUAL DIALOGUE.....	31
<i>Mauricio Mazuecos ; Alberto Testoni ; Raffaella Bernardi ; Luciana Benotti</i>	
A TRANSLATION-BASED APPROACH TO MORPHOLOGY LEARNING FOR LOW RESOURCE LANGUAGES.....	36
<i>Tewodros Gebreselassie ; Amanuel Mersha ; Michael Gasser</i>	
MONOLINGUAL CORPUS CREATION AND EVALUATION OF TRULY LOW-RESOURCE LANGUAGES FROM PERU .....	41
<i>Gina Bustamante ; Arturo Oncevay</i>	
TIGRINYA AUTOMATIC SPEECH RECOGNITION WITH MORPHEME BASED RECOGNITION UNITS.....	46
<i>Hafte Abera ; Sebsibe Hailemariam</i>	
VARIANTS OF VECTOR SPACE REDUCTIONS FOR PREDICTING THE COMPOSITIONALITY OF ENGLISH NOUN COMPOUNDS .....	51
<i>Pegah Alipoormolabashi ; Sabine Schulte Im Walde</i>	
SENTITEL: TABSA FOR TWITTER REVIEWS ON UGANDA TELECOMS.....	55
<i>David Kabiito ; Joyce Nakatumba Nabende</i>	

AN ASSESSMENT OF LANGUAGE IDENTIFICATION METHODS ON TWEETS AND WIKIPEDIA ARTICLES .....	58
<i>Pedro Verneti ; Larissa Freitas</i>	
A COMPARISON OF IDENTIFICATION METHODS OF BRAZILIAN MUSIC STYLES BY LYRICS .....	61
<i>Patrick Guimarães ; Jader Froes ; Douglas Costa ; Larissa Freitas</i>	
ENABLING FAST AND CORRECT TYPING IN 'LEICHTE SPRACHE' (EASY LANGUAGE) .....	64
<i>Ina Steinmetz ; Karin Harbusch</i>	
AI4D - AFRICAN LANGUAGE DATASET CHALLENGE .....	68
<i>Kathleen Siminyu ; Sackey Freshia</i>	
CAN WIKIPEDIA CATEGORIES IMPROVE MASKED LANGUAGE MODEL PRETRAINING?.....	78
<i>Diksha Meghwal ; Katharina Kann ; Iacer Calixto ; Stanislaw Jastrzebski</i>	
ADVERSARIAL EVALUATION OF BERT FOR BIOMEDICAL NAMED ENTITY RECOGNITION.....	79
<i>Vladimir Araujo ; Andrés Carvallo ; Denis Parra</i>	
FFR V1.1: FON-FRENCH NEURAL MACHINE TRANSLATION.....	83
<i>Chris Chinenye Emezue ; Femi Pancrace Bonaventure Dossou</i>	
CLASSIFICATION AND ANALYSIS OF NEOLOGISMS PRODUCED BY LEARNERS OF SPANISH: EFFECTS OF PROFICIENCY AND TASK.....	88
<i>Shira Wein</i>	
DEVELOPING A MONOLINGUAL SENTENCE SIMPLIFICATION CORPUS FOR URDU .....	92
<i>Yusra Anees ; Sadaf Abdul Rauf ; Nauman Iqbal ; Abdul Basit Siddiqi</i>	
TRANSLATING NATURAL LANGUAGE INSTRUCTIONS FOR BEHAVIORAL ROBOT NAVIGATION WITH A MULTI-HEAD ATTENTION MECHANISM.....	96
<i>Patricio Cerda-Mardini ; Vladimir Araujo ; Álvaro Soto</i>	
TOWARDS MITIGATING GENDER BIAS IN A DECODER-BASED NEURAL MACHINE TRANSLATION MODEL BY ADDING CONTEXTUAL INFORMATION .....	99
<i>Christine Basta ; Marta R. Costa-Jussà ; José A. R. Fonollosa</i>	
PREDICTING AND ANALYZING LAW-MAKING IN KENYA.....	103
<i>Oyinlola Babafemi ; Adewale Akinfaderin</i>	
DEFINING AND EVALUATING FAIR NATURAL LANGUAGE GENERATION .....	107
<i>Catherine Yeo ; Alyssa Chen</i>	
POLITICAL ADVERTISING DATASET: THE USE CASE OF THE POLISH 2020 PRESIDENTIAL ELECTIONS .....	110
<i>Lukasz Augustyniak ; Krzysztof Rajda ; Tomasz Kajdanowicz ; Micha Bernaczyk</i>	
THE HUMAN UNLIKENESS OF NEURAL LANGUAGE MODELS IN NEXT-WORD PREDICTION .....	115
<i>Cassandra L. Jacobs ; Arya D. McCarthy</i>	
A STUDY ON THE INFLUENCE OF ARCHITECTURE COMPLEXITY OF RNNs FOR INTENT CLASSIFICATION IN E-COMMERCE CHATS IN BAHASA INDONESIA.....	116
<i>Renny Pradina Kusumawardani ; Muhammad Azzam</i>	

LONG-TAIL PREDICTIONS WITH CONTINUOUS-OUTPUT LANGUAGE MODELS .....	119
<i>Shiran Dudy ; Steven Bedrick</i>	
ANALYZING THE FRAMING OF 2020 PRESIDENTIAL CANDIDATES IN THE NEWS .....	123
<i>Audrey Acken ; Dorottya Demszky</i>	
UNDERSTANDING THE IMPACT OF EXPERIMENT DESIGN FOR EVALUATING DIALOGUE SYSTEM OUTPUT .....	124
<i>Sashank Santhanam ; Samira Shaikh</i>	
STUDYING THE EFFECT OF EMOTIONAL AND MORAL LANGUAGE ON INFORMATION CONTAGION DURING THE CHARLOTTESVILLE EVENT .....	128
<i>Khyati Mahajan ; Samira Shaikh</i>	
MAPPING OF NARRATIVE TEXT FIELDS TO ICD-10 CODES USING NATURAL LANGUAGE PROCESSING AND MACHINE LEARNING .....	131
<i>Risuna Nkolele</i>	
MULTITASK MODELS FOR CONTROLLING THE COMPLEXITY OF NEURAL MACHINE TRANSLATION .....	136
<i>Sweta Agrawal ; Marine Carpuat</i>	
USING SOCIAL MEDIA FOR BITCOIN DAY TRADING BEHAVIOR PREDICTION .....	140
<i>Anna Paula Pawlicka Maule ; Kristen Johnson</i>	
HAUSAMT V1.0: TOWARDS ENGLISH-HAUSA NEURAL MACHINE TRANSLATION.....	144
<i>Adewale Akinfaderin</i>	
OUTCOMES OF COMING OUT: ANALYZING STORIES OF LGBTQ+.....	148
<i>Krithika Ramesh ; Tanvi Anand</i>	
AN EVALUATION OF SUBWORD SEGMENTATION STRATEGIES FOR NEURAL MACHINE TRANSLATION OF MORPHOLOGICALLY RICH LANGUAGES .....	151
<i>Aquia Richburg ; Ramy Eskander ; Smaranda Muresan ; Marine Carpuat</i>	
ENHANCED URDU WORD SEGMENTATION USING CONDITIONAL RANDOM FIELDS AND MORPHOLOGICAL CONTEXT FEATURES .....	156
<i>Aamir Farhan ; Mashrukh Islam ; Dipti Misra Sharma</i>	
FLEXIBLE NON-AUTOREGRESSIVE NEURAL MACHINE TRANSLATION VIA REPOSITIONING EDIT OPERATIONS.....	160
<i>Weijia Xu ; Marine Carpuat</i>	