Citizen Linguistics in Language Resource Development  
(CLLRD 2020)

PROCEEDINGS

James Fiumara, Christopher Cieri, Mark Liberman,  
Chris Callison-Burch (eds.)
Introduction

Welcome to the LREC2020 Workshop on Citizen Linguistics in Language Resource Development.

Notwithstanding advances in data collection and processing, language related research, education and technology development continue to suffer from inadequate supply of Language Resources. To supplement traditional LR development, which typically relies upon top down support from some government or private foundation, Citizen Linguistics (the Citizen Science of Language) changes the incentive model to attract a new workforce which in turn requires a different kind of workflow. Incentives to Citizen Linguists may include the opportunities to learn and develop new skills; to socialize, compete and earn status or recognition; to document their language and promote their culture and, most importantly, to contribute directly to research and indirectly to a greater cause or social good. By offering human contributors sustained access to appropriate opportunities, activities, and incentives, we can enhance LR development well beyond what traditional direct funding alone can produce. However, along with these new incentives and workflows come new challenges whose solutions are relevant even to expert (paid) annotation.

The goal of this hybrid workshop/tutorial is two-fold. First is to provide a forum for researchers and practitioners to explore and discuss the issues, advantages and challenges of using Citizen Linguistics as a method for the creation of language resources. Second is to introduce LanguageARC, a new Citizen Linguistics web portal for collecting language data and judgements.
Organizers:

Chris Callison-Burch, University of Pennsylvania, USA
Christopher Cieri, Linguistic Data Consortium, University of Pennsylvania, USA
James Fiumara, Linguistic Data Consortium, University of Pennsylvania, USA
Mark Liberman, Linguistic Data Consortium, University of Pennsylvania, USA

Program Committee:

Sonja Bosch, University of South Africa, South Africa
Chris Callison-Burch, University of Pennsylvania, USA
Nicoletta Calzolari, Institute for Computational Linguistics, Italy
Khalid Choukri, ELRA/ELDA, France
Christopher Cieri, Linguistic Data Consortium, University of Pennsylvania, USA
John Coleman, Oxford University, UK
Maxine Eskenazi, Carnegie Mellon University, USA
Karën Fort, Sorbonne Université, France
James Fiumara, Linguistic Data Consortium, University of Pennsylvania, USA
Mark Liberman, Linguistic Data Consortium, University of Pennsylvania, USA
Peter Patrick, University of Essex, UK
Massimo Poesio, Queen Mary University of London, UK
Stephanie Strassel, Linguistic Data Consortium, University of Pennsylvania, UK
Jennifer Tracey, Linguistic Data Consortium, University of Pennsylvania, UK

Invited Speaker:

John Coleman, Oxford University, UK
Table of Contents

LanguageARC: Developing Language Resources Through Citizen Linguistics
James Fiumara, Christopher Cieri, Jonathan Wright and Mark Liberman ......................... 1

Developing Language Resources with Citizen Linguistics in Austria – A Case Study
Barbara Heinisch ................................................................. 7

Objective Assessment of Subjective Tasks in Crowdsourcing Applications
Giannis Haralabopoulos, Myron Tsikandilakis, Mercedes Torres Torres and Derek McAuley ... 15

Speaking Outside the Box: Exploring the Benefits of Unconstrained Input in Crowdsourcing and Citizen Science Platforms
Jon Chamberlain, Udo Kruschwitz and Massimo Poesio ........................................... 26

Leveraging Non-Specialists for Accurate and Time Efficient AMR Annotation
Mary Martin, Cecilia Mauceri, Martha Palmer and Christoffer Heckman ......................... 35

The INCOMSLAV Platform: Experimental Website with Integrated Methods for Measuring Linguistic Distances and Asymmetries in Receptive Multilingualism
Irina Stenger, Klara Jagrova and Tania Avgustinova ................................................. 40

Identifications of Speaker Ethnicity in South-East England: Multicultural London English as a Divisible Perceptual Variety
Amanda Cole ................................................................. 49

LanguageARC - a tutorial
Christopher Cieri and James Fiumara ................................................................. 58