Introduction to the Special Issue: Selected papers from LREC 2018

Nancy Ide[1] · Nicoletta Calzolari[2]

The Eleventh Language Resources and Evaluation Conference (LREC) was held in Miyazaki, Japan in May 2018. This special issue collects selected papers from the conference, based on recommendations provided by reviewers in the LREC review form. Thirty-nine papers were invited to the special issue; of those, nine full-length journal papers and three project notes are included here.

The papers included in this special issue demonstrate that the types of resources of interest in the field as well as methods for their construction continue to evolve. Three papers deal with the construction of corpora supporting specific applications, including a corpus of live blogs for summarization (Avinesh et al.), a corpus containing low-frequency English constructions for predicting human processing difficulty (the Natural Stories Corpus, Futrell et al.), and a corpus of speech disorder productions to assess intelligibility and validate Disorder Severity Indexes (C2SI, Woisard et al.). Increasingly, developing resources include multi-modal data, such as the multimodal corpus of the indigenous American Choctaw language (ChoCo) discussed in Brixey et al. and the previously mentioned Natural Stories Corpus.

Other resource types are also represented in this collection. Moran et al. describe BDPROTO, a database comprising phonological inventory data from 257 ancient and reconstructed languages, and Kulahcioglu et al. overview FontLex, a typographical lexicon providing associations between emotions and fonts. Methods for constructing and enhancing resources reflect the general movement toward state-of-the-art approaches, including the use of convolutional neural networks for multi-

✉ Nancy Ide
   ide@cs.vassar.edu

1 Department of Computer Science, Vassar College, Poughkeepsie, USA

2 Istituto di Linguistica Computazionale "A. Zampolli", Consiglio Nazionale delle Ricerche Area della Ricerca di Pisa, Pisa, Italy

modal page stream and document flow segmentation (Wiedemann et al.), construction of word-formation networks using a machine-learned ranking model (Lango et al.), and the use of Convolutional Neural Network embeddings in the development of FontLex.

Evaluation and resource infrastructures have always played a role in LREC presentations. Broux et al. focus on evaluation, evaluating human corrections in a computer-assisted speaker diarization system and introduce a metric to evaluate correction cost, and Woliński et al. discuss how using three interrelated linguistic resources (parser, treebank, valency dictionary) for Polish enabled corrections in each. On infrastructure, Trippel et al. discuss the challenges of migrating major resources from one site to another–a problem little discussed but of critical importance for sustaining high-cost resources over time. Finally, Del Gratta et al. provide a global overview of the field by describing different trends provided by taking diachronic perspective of the LRE Map, which provides an ideal complement to the survey of the field provided in this issue.

We take this opportunity to recognize the members of the community who have contributed to the journal by reviewing papers, as well as those who have submitted to the journal over the past two years and the past and current members of our Editorial Board. Their contributions are essential to maintaining the high quality of materials published in LRE and the vibrancy of the field it represents.

Nancy Ide and Nicoletta Calzolari

LRE Co-editors-in-chief