Linguistic Issues in Language Technology – LiLT Submitted, January 2010

Special Volume Introduction

D. Terence Langendoen and Emily M. Bender

> Submitted, January 2010, Published by CSLI Publications

February 2010

LiLT volume 3, issue 1

Special Volume Introduction

D. TERENCE LANGENDOEN, University of Arizona AND EMILY M. BENDER, University of Washington

This volume grows out of a special session that we organized at the January 2009 Annual Meeting of the Linguistic Society of America entitled "Computational Linguistics: Implementation of Analyses against Data". It had two parts: an hour-long plenary presentation by us on Friday evening, and a set of five half-hour-long papers and general discussion on Saturday morning. Our goal was to present computational linguistics not as a subfield of linguistics but as a resource for accessing all sorts of machine-readable language and linguistic data and tools to record and analyze those data. We asked our presenters to address three questions:

- 1. How does your work facilitate the extension of analyses to a wide range of naturally occurring data and/or the extraction of analyses from such data?
- 2. Does such a quantitative change in the amount of data countenanced lead to qualitative changes in analysis/theory compared to smaller-scale methods? If so, what kind of changes?
- 3. How does this work inform your perspective on the relationship between data and analyses in linguistics?

In our plenary presentation, we looked broadly across the field of linguistics and asked how computational techniques can contribute to its development, as well as what we need to know and do in order to maximize the potential of these techniques to advance our research,

LiLT Volume 3, Issue 1, February 2010. Special Volume Introduction. Copyright © 2010, CSLI Publications.

teaching and outreach. In addition, we assessed the current state of the infrastructure that is available to support the use of the resources that computational linguistics provides. We went on to ask how we can all work together to improve and extend that infrastructure so as to make massive amounts of high quality linguistic data and resources readily accessible and usable by researchers, teachers, and the interested public worldwide.

The first three of the papers in this volume are revisions of papers presented in the special session. The fourth is on a topic similar to the one that was presented by the same authors.¹ The first paper (by Bender and Langendoen) deals with all the issues raised in the special session, while the final three examine specific questions in depth that relate computational linguistics to syntax (Bender, on grammar engineering for hypothesis testing in syntax), morphology (Palmer et al, on building an environment for morphological annotation that can be cost-effectively used in the field, particularly for working with endangered languages), and pragmatics (Potts and Schwarz, on analyzing large web text corpora for the expression of affect and solidarity). These papers describe a few of the steps that are now being taken to build the "linguistic research environment of the future" envisioned in section 4 of the first paper.

Finally, we would like to note that the papers in the present volume are situated in the context of a broader discussion in both the linguistics and computational linguistics/language technology communities about the relationship between the two fields. In partcular, the April 2009 meeting of the European Association for Computational Linguistics in Athens, Greece included a workshop entitled "The Interaction between Linguistics and Computational Linguistics: Virtuous, Vicious or Vacuous?",² and another special issue of LiLT (edited by Tim Baldwin and Valia Kordoni) containing papers from that workshop and others on the same theme is in preparation. Then in July, another workshop entitled "Cyberling 2009: Toward a Cyberinfrastructure for Linguistics" was held at the LSA Linguistic Institute in Berkeley, CA, to provide specific directions for the field to explore in building and sustaining its research, teaching and outreach infrastructure. The working group reports from that workshop are available on the web.³ In addition, the organizers of that workshop have set up a blog⁴ to keep the conversation

¹Potts and Schwarz's special session paper can be found at

 $http://semanticsarchive.net/Archive/jFjNGNjZ/potts-schwarz-exclamatives08.pdf \ ^{2}http://www.csse.unimelb.edu.au/~tim/events/eacl2009/$

³http://cyberling.elanguage.net/

⁴http://blog.cyberling.org/

going on how to move forward. The 2010 meeting of the Association for Computational Linguistics in Uppsala, Sweden will include a workshop entitled "Natural Language Processing and Linguistics: Finding the Common Ground".⁵ We are encouraged by this evidence of interest from general and theoretical linguists as well as computational linguists and language technologists in exploring ways in which we can work together to advance our understanding of language in all its aspects.

 $^{^{5}} http://faculty.washington.edu/fxia/nlpling2010/$