Proceedings of the NoDaLiDa 2017 Workshop on Processing Historical Language

edited by
Gerlof Bouma and Yvonne Adesam

22 May 2017
Gothenburg

Proceedings of the NoDaLiDa 2017 Workshop on Processing Historical Language
edited by Gerlof Bouma and Yvonne Adesam
NEALT Proceedings Series 32
ISBN 978-91-7685-503-4

Linköping Electronic Conference Proceedings 133
ISSN 1650-3686 eISSN 1650-3740

ACL Anthology W17-05

Photo front cover: Kjell Holmner/Göteborg & Co

Preface

The papers in this volume are presented at the Workshop on Processing Historical Language, held in conjunction with the 40-year anniversary of NoDaLiDa, 22 May 2017 in Gothenburg.

While historical texts have long attracted interest from language historians and historical linguists, we have seen an increased attention to the problems particular to processing historical data from a computational perspective in the last decade or so. 'Processing' here entails a wide range of text processing tasks, such as creating electronic transcriptions and editions of manuscripts, constructing lexica, tagging, and parsing, as well as content-oriented processing such as semantic parsing and information extraction. The aim of the workshop is to bring together researchers working on processing historical materials with a particular focus on work that investigates the combination of data-driven and knowledge-driven modelling.

We received 16 submissions, which were each reviewed (double blind) by three programme committee members. Because of the amount and quality of the submissions, the workshop, initially planned as a half-day workshop, was prolonged to accommodate 9 oral presentations.

The authors come from eight different European countries. The research presented at the workshop covers a range of topics related to historical materials, including spelling standardization, linguistic analysis, identification of text re-use, and data visualization. Featured languages are Dutch, English, Finnish, German, Icelandic, Latin, and Spanish, at varying historical stages. The programme also includes an invited talk by Stefanie Dipper, titled *Variance in historical data: how bad is it and how can we profit from it for historical linguistics?*

We are excited to have such a varied and inspiring programme and would like to thank the invited speaker, authors, and reviewers for their valuable contributions.

May 1, 2017 — Gerlof Bouma
Gothenburg — Yvonne Adesam

The workshop is organized as part of project MAÞiR – Methods for the automatic Analysis of Text in digital Historical Resources – funded by Marcus and Amalia Wallenberg Foundation, grant MAW 2012.0146.

Workshop Organization / Programme Chairs

Yvonne Adesam, University of Gothenburg
Gerlof Bouma, University of Gothenburg

Programme Committee

David Alfter, University of Gothenburg
Marcel Bollmann, Ruhr-Universität Bochum
Lars Borin, University of Gothenburg
Gosse Bouma, University of Groningen
Hanne Martine Eckhoff, University of Tromsø
Markus Forsberg, University of Gothenburg
Iris Hendrickx, Radboud University Nijmegen
Richard Johansson, University of Gothenburg
Alex Speed Kjeldsen, University of Copenhagen
Beáta Megyesi, Uppsala University
Eva Pettersson, Uppsala University
Nina Tahmasebi, University of Gothenburg
Erik Tjong Kim Sang, Meertens Institute
Marjo van Koppen, Utrecht University
Shafqat Virk, University of Gothenburg

Contents