Communications in Computer and Information Science 380

Cerstin Mahlow Michael Piotrowski (Eds.)

Systems and Frameworks for Computational Morphology

Third International Workshop, SFCM 2013
Berlin, Germany, September 6, 2013
Proceedings

Springer

Volume Editors

Cerstin Mahlow
University of Konstanz
78457 Konstanz, Germany
E-mail: cerstin.mahlow@uni-konstanz.de

Michael Piotrowski
Leibniz Institute of European History
Alte Universitätsstr. 19
55116 Mainz, Germany
E-mail: piotrowski@ieg-mainz.de

ISSN 1865-0929 e-ISSN 1865-0937
ISBN 978-3-642-40485-6 e-ISBN 978-3-642-40486-3
DOI 10.1007/978-3-642-40486-3
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013945728

CR Subject Classification (1998): I.2.7, J.5

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This volume contains the papers presented at SFCM 2013: The Third International Workshop on Systems and Frameworks for Computational Morphology, held on September 6, 2013, at the Humboldt-Universität zu Berlin.

Morphological resources are the basis for all higher-level natural language processing applications. Morphology components should thus be capable of analyzing single word forms as well as whole corpora. For many practical applications, not only morphological analysis, but also generation is required, i.e., the production of surfaces corresponding to specific categories.

Apart from uses in computational linguistics, there are numerous practical applications that either require morphological analysis and generation, or that can greatly benefit from it, for example, in text processing, user interfaces, or information retrieval. These applications have specific requirements for morphological components, including requirements from software engineering, such as programming interfaces or robustness.

With the workshop on Systems and Frameworks for Computational Morphology (SFCM) we have established a place for presenting and discussing recent advances in the field of computational morphology. In 2013 the workshop took place for the third time. SFCM focuses on actual working systems and frameworks that are based on linguistic principles and that provide linguistically motivated analyses and/or generation on the basis of linguistic categories. The main theme for SFCM 2009 was systems for a specific language, namely, German; SFCM 2011 looked at phenomena at the interface between morphology and syntax in various languages. SFCM 2013 had three main goals:

- To discuss the role of morphological analysis and generation to improve the rather disappointing situation with respect to language technology for languages other than English, as described in the recently published White Paper Series by META-NET.
- To stimulate discussion among researchers and developers and to offer an up-to-date overview of available morphological systems for specific purposes.
- To stimulate discussion among developers of general frameworks that can be used to implement morphological components for several languages.

On the basis of the number of submissions and the number of participants at the workshop we can definitely state that the topic of the workshop was met with great interest from the community, both from academia and industry. We received 15 submissions, of which seven were accepted after a thorough review by the members of the Program Committee and additional reviewers. The peer review process was double-blind, and each paper received four reviews.

In addition to the regular papers, we had the pleasure of Georg Rehm giving an invited talk on the role of morphology systems in the META-NET Strategic Research Agenda.

The discussions after the talks and during the demo sessions, as well as the final plenum, showed the interest in and the need and the requirements for further efforts in the field of computational morphology. We will maintain the website for the workshop series at `http://www.sfcm.eu`.

This book starts with the invited paper by Georg Rehm ("The State of Computational Morphology for Europe's Languages and the META-NET Strategic Research Agenda"), emphasizing that computational morphology is all but a solved problem. Only for a few European languages appropriate resources and tools are available. He argues that a joint effort of the members of the European research community is needed to create "adequate, precise, robust, scalable and freely available morphology components" for all European languages.

The following paper "A Case Study in Tagging Case in German: An Assessment of Statistical Approaches" by Simon Clematide presents a study that assesses the performance of purely statistical approaches using supervised machine learning for predicting case in German nouns. The study evaluates different approaches—Hidden Markov Models, Decision Trees, and Conditional Random Fields—on two treebanks. The author shows that his CRF-based approach outperforms all other approaches and results in an improvement of 11% compared to an HMM trigram tagger.

In their paper "Jabalín: A Comprehensive Computational Model of Modern Standard Arabic Verbal Morphology Based on Traditional Arabic Prosody," Alicia González Martínez, Susana López Hervás, Doaa Samy, Carlos G. Arques, and Antonio Moreno Sandoval note that—despite its richness—the Arabic morphological system is in fact highly regular. By taking inspiration from the traditional description of Arabic prosody, the authors' Jabalín system implements a compact and simple morphological description for Modern Standard Arabic, which takes advantage of the regularities of Arabic morphology.

Both SFCM 2009 and SFCM 2011 featured papers on HFST, and we are happy to see this tradition continue: The paper "HFST—A System for Creating NLP Tools" by Krister Lindén, Erik Axelson, Senka Drobac, Sam Hardwick, Juha Kuokkala, Jyrki Niemi, Tommi Pirinen, and Miikka Silfverberg presents and evaluates various NLP tools that have been created using HFST. What makes this paper particularly interesting, however, is that the authors describe an implementation and application of `pmatch` finite-state pattern matching algorithm presented by Lauri Karttunen at SFCM 2011.

The next paper, "A System for Archivable Grammar Documentation" by Michael Maxwell, first describes a number of criteria for archivable documentation of grammars for natural languages and then presents a system for writing and testing morphological and phonological grammars, which aims to satisfy these criteria. The paper explains some of the decisions that went into the design of the formalism and describes experiences gained from its use with grammars for a variety of languages.

Fiammetta Namer's paper "A Rule-Based Morphosemantic Parser for French for a Fine-Grained Semantic Annotation of Texts" describes the DériF system. Unlike existing word segmentation tools, DériF annotates derived and compound words with semantic information, namely a definition, lexical-semantic features, and lexical relations.

Next, in their paper "Implementing a Formal Model of Inflectional Morphology," Benoît Sagot and Géraldine Walther describe the implementation of a formal model of inflectional morphology that aims to capture typological generalizations. The authors show that the availability of such a model—and an implementation thereof—is beneficial for studies in descriptive and formal morphology, as well as for the development of NLP tools and resources.

Finally, the paper "Verbal Morphosyntactic Disambiguation through Topological Field Recognition in German-Language Law Texts" by Kyoko Sugisaki and Stefan Höfler introduces an incremental system of verbal morphosyntactic disambiguation that exploits the concept of topological fields, and demonstrates that this approach is able to significantly reduce the error rate in POS tagging.

The contributions show that high-quality research is being conducted in the area of computational morphology: Mature systems are further developed and new systems and applications are emerging. Other languages than English are becoming more important. The papers in this book come from six countries and two continents, discuss a wide variety of languages from many different language families, and illustrate that, in fact, a rich morphology is better described as the norm rather than the exception—proving that for most languages, as we have stated above, morphological resources are indeed the basis for all higher-level natural language processing applications.

The trend toward open-source developments still goes on and evaluation is considered an important issue. Making high-quality morphological resources freely available will help to advance the state of the art and allow the development of high-quality real-world applications. Useful applications with carefully conducted evaluation will demonstrate to a broad audience that computational morphology is an actual science with tangible benefits for society.

We would like to thank the authors for their contributions to the workshop and to this book. We also thank the reviewers for their effort and for their constructive feedback, encouraging and helping the authors to improve their papers. The submission and reviewing process and the compilation of the proceedings was supported by the EasyChair system. We thank Aliaksandr Birukou, the editor of the series *Communications in Computer and Information Science* (CCIS), and the Springer staff for publishing the proceedings of SFCM 2013. We are grateful for the financial support given by the German Society for Computational Linguistics and Language Technology (GSCL). We thank Anke Lüdeling and Carolin Odebrecht and the staff from the Corpus Linguistics and Morphology Group at the Department of German Language and Linguistics at the Humboldt-Universität zu Berlin for the local organization.

June 2013

Cerstin Mahlow
Michael Piotrowski

Organization

The Third International Workshop on Systems and Frameworks for Computational Morphology (SFCM 2013) was organized and chaired by Cerstin Mahlow and Michael Piotrowski. The workshop was held at Humboldt-Universität zu Berlin.

Program Chairs

Cerstin Mahlow	University of Konstanz, Germany
Michael Piotrowski	Leibniz Institute of European History, Germany

Program Committee

Bruno Cartoni	University of Geneva, Switzerland
Simon Clematide	University of Zurich, Switzerland
Piotr Fuglewicz	TiP Sp. z o. o., Katowice, Poland
Thomas Hanneforth	University of Potsdam, Germany
Kimmo Koskenniemi	University of Helsinki, Finland
Winfried Lenders	University of Bonn, Germany
Krister Lindén	University of Helsinki, Finland
Anke Lüdeling	Humboldt-Universität Berlin, Germany
Cerstin Mahlow	University of Konstanz, Germany
Günter Neumann	DFKI Saarbrücken, Germany
Michael Piotrowski	Leibniz Institute of European History, Germany
Benoît Sagot	INRIA/Université Paris 7, France
Helmut Schmid	University of Stuttgart, Germany
Angelika Storrer	University of Dortmund, Germany
Pius ten Hacken	Swansea University, UK
Andrea Zielinski	Fraunhofer IOSB, Germany

Additional Reviewers

Lenz Furrer	University of Zurich, Switzerland

Local Organization

Anke Lüdeling	Humboldt-Universität Berlin, Germany
Carolin Odebrecht	Humboldt-Universität Berlin, Germany

Sponsoring Institutions

German Society for Computational Linguistics and Language Technology (GSCL)
Humboldt-Universität Berlin, Germany

Table of Contents