

Lecture Notes in Artificial Intelligence

10583

Subseries of Lecture Notes in Computer Science

LNAI Series Editors

Randy Goebel

University of Alberta, Edmonton, Canada

Yuzuru Tanaka

Hokkaido University, Sapporo, Japan

Wolfgang Wahlster

DFKI and Saarland University, Saarbrücken, Germany

LNAI Founding Series Editor

Joerg Siekmann

DFKI and Saarland University, Saarbrücken, Germany


More information about this series at <http://www.springer.com/series/1244>


Nathalie Camelin · Yannick Estève
Carlos Martín-Vide (Eds.)

Statistical Language and Speech Processing

5th International Conference, SLSP 2017
Le Mans, France, October 23–25, 2017
Proceedings

Editors
Nathalie Camelin
University of Le Mans
Le Mans
France

Carlos Martín-Vide 
Rovira i Virgili University
Tarragona
Spain

Yannick Estève 
University of Le Mans
Le Mans
France

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Artificial Intelligence
ISBN 978-3-319-68455-0 ISBN 978-3-319-68456-7 (eBook)
DOI 10.1007/978-3-319-68456-7

Library of Congress Control Number: 2017954906

LNCS Sublibrary: SL7 – Artificial Intelligence

© Springer International Publishing AG 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

These proceedings contain the papers that were presented at the 5th International Conference on Statistical Language and Speech Processing (SLSP 2017), held in Le Mans, France, during October 23–25, 2017.

The scope of SLSP deals with topics of either theoretical or applied interest discussing the employment of statistical models (including machine learning) in language and speech processing, namely:

- Anaphora and coreference resolution
- Authorship identification, plagiarism, and spam filtering
- Computer-aided translation
- Corpora and language resources
- Data mining and Semantic Web
- Information extraction
- Information retrieval
- Knowledge representation and ontologies
- Lexicons and dictionaries
- Machine translation
- Multimodal technologies
- Natural language understanding
- Neural representation of speech and language
- Opinion mining and sentiment analysis
- Parsing
- Part-of-speech tagging
- Question-answering systems
- Semantic role labeling
- Speaker identification and verification
- Speech and language generation
- Speech recognition
- Speech synthesis
- Speech transcription
- Spelling correction
- Spoken dialogue systems
- Term extraction
- Text categorization
- Text summarization
- User modeling

SLSP 2017 received 39 submissions. Every paper was reviewed by three Programme Committee members. There were also a few external experts consulted. After a thorough and vivid discussion phase, the committee decided to accept 21 papers (which

represents an acceptance rate of about 54%). The conference program included three invited talks and presentations of work in progress as well.

The excellent facilities provided by the EasyChair conference management system allowed us to deal with the submissions successfully and handle the preparation of these proceedings in time.

We would like to thank all invited speakers and authors for their contributions, the Program Committee and the external reviewers for their diligent cooperation, and Springer for its very professional publishing work.

July 2017

Nathalie Camelin
Yannick Estève
Carlos Martín-Vide

Organization

Program Committee

Jon Barker	University of Sheffield, UK
Laurent Besacier	Grenoble Informatics Laboratory, France
Paul Buitelaar	National University of Ireland, Galway, Ireland
Felix Burkhardt	Telekom Innovation Laboratories, Germany
Xavier Carreras	Xerox Research Centre Europe, France
Francisco Casacuberta	Technical University of Valencia, Spain
Ciprian Chelba	Google, USA
Eng Siong Chng	Nanyang Technological University, Singapore
Jennifer Chu-Carroll	Elemental Cognition, USA
Doug Downey	Northwestern University, USA
Robert Gaizauskas	University of Sheffield, UK
Julio Gonzalo	National University of Distance Education, Spain
Keikichi Hirose	University of Tokyo, Japan
Gerhard Jäger	University of Tübingen, Germany
Gareth Jones	Dublin City University, Ireland
Joseph Keshet	Bar-Ilan University, Israel
Tomi Kinnunen	University of Eastern Finland, Finland
Lun-Wei Ku	Academia Sinica, Taiwan
Kong Aik Lee	Institute for Infocomm Research, Singapore
Elizabeth D. Liddy	Syracuse University, USA
Xunying Liu	Chinese University of Hong Kong, Hong Kong
Suresh Manandhar	University of York, UK
Carlos Martín-Vide (Chair)	Rovira i Virgili University, Spain
Yuji Matsumoto	Nara Institute of Science and Technology, Japan
Ruslan Mitkov	University of Wolverhampton, UK
Marie-Francine Moens	KU Leuven, Belgium
Seiichi Nakagawa	Toyohashi University of Technology, Japan
Preslav Nakov	Qatar Computing Research Institute, Qatar
Hermann Ney	RWTH Aachen University, Germany
Cécile Paris	CSIRO, Australia
Fuchun Peng	AI Sense Inc., USA
Pascal Perrier	Grenoble Institute of Technology, France
Leon Rothkrantz	Delft University of Technology, Netherlands
Horacio Saggion	Pompeu Fabra University, Spain
Murat Saraçlar	Bogaziçi University, Turkey
Holger Schwenk	Facebook, France
Brad Story	University of Arizona, USA
Karin Verspoor	University of Melbourne, Australia

Stephan Vogel	Qatar Computing Research Institute, Qatar
Xiaojun Wan	Peking University, China
Phil Woodland	University of Cambridge, UK
Chuck Wooters	Semantic Machines, USA
François Yvon	LIMSI-CNRS, France

Additional Reviewers

Evans, Richard
Makary, Mireille
Rohanian, Omid
Taslimipoor, Shiva

Contents

Invited Paper

- Author Profiling in Social Media: The Impact of Emotions
on Discourse Analysis 3
Paolo Rosso and Francisco Rangel

Language and Information Extraction

- Neural Machine Translation by Generating Multiple Linguistic Factors 21
Mercedes García-Martínez, Loïc Barrault, and Fethi Bougares
- Analysis and Automatic Classification of Some Discourse Particles
on a Large Set of French Spoken Corpora 32
Denis Jovet, Katarina Bartkova, Mathilde Dargnat, and Lou Lee
- Learning Morphology of Natural Language as a Finite-State Grammar. 44
Javad Nouri and Roman Yangarber
- Incorporating Coreference to Automatic Evaluation
of Coherence in Essays 58
Michal Novák, Kateřina Rysová, Magdaléna Rysová, and Jiří Mírovský
- Graph-Based Features for Automatic Online Abuse Detection. 70
*Etienne Papegnies, Vincent Labatut, Richard Dufour,
and Georges Linarès*
- Exploring Temporal Analysis of Tweet Content from Cultural Events 82
*Mathias Quillot, Cassandre Ollivier, Richard Dufour,
and Vincent Labatut*
- Towards a Relation-Based Argument Extraction Model
for Argumentation Mining 94
Gil Rocha and Henrique Lopes Cardoso

Post-processing and Applications of Automatic Transcriptions

- Three Experiments on the Application of Automatic Speech Recognition
in Industrial Environments 109
Ferdinand Fuhrmann, Anna Maly, Christina Leitner, and Franz Graf
- Enriching Confusion Networks for Post-processing 119
Sahar Ghannay, Yannick Estève, and Nathalie Camelin

Attentional Parallel RNNs for Generating Punctuation in Transcribed Speech	131
<i>Alp Öktem, Mireia Farrús, and Leo Wanner</i>	
Lightweight Spoken Utterance Classification with CFG, tf-idf and Dynamic Programming	143
<i>Manny Rayner, Nikos Tsourakis, and Johanna Gerlach</i>	
Low Latency MaxEnt- and RNN-Based Word Sequence Models for Punctuation Restoration of Closed Caption Data	155
<i>Máté Ákos Tündik, Balázs Tarján, and György Szaszák</i>	
Speech: Paralinguistics and Synthesis	
Unsupervised Speech Unit Discovery Using K-means and Neural Networks	169
<i>Céline Manenti, Thomas Pellegrini, and Julien Pinquier</i>	
Noise and Speech Estimation as Auxiliary Tasks for Robust Speech Recognition	181
<i>Gueorgui Pironkov, Stéphane Dupont, Sean U.N. Wood, and Thierry Dutoit</i>	
Unified Approach to Development of ASR Systems for East Slavic Languages	193
<i>Radek Safarik and Jan Nouza</i>	
A Regularization Post Layer: An Additional Way How to Make Deep Neural Networks Robust	204
<i>Jan Vaněk, Jan Zelinka, Daniel Soutner, and Josef Psutka</i>	
Speech Recognition: Modeling and Resources	
Detecting Stuttering Events in Transcripts of Children’s Speech	217
<i>Sadeen Alharbi, Madina Hasan, Anthony J.H. Simons, Shelagh Brumfitt, and Phil Green</i>	
Introducing AmuS: The Amused Speech Database	229
<i>Kevin El Haddad, Ilaria Torre, Emer Gilmartin, Hüseyin Çakmak, Stéphane Dupont, Thierry Dutoit, and Nick Campbell</i>	
Lexical Emphasis Detection in Spoken French Using F-BANKS and Neural Networks	241
<i>Abdelwahab Heba, Thomas Pellegrini, Tom Jorquera, Régine André-Obrecht, and Jean-Pierre Lorré</i>	

Speaker Change Detection Using Binary Key Modelling
with Contextual Information 250
Jose Patino, Héctor Delgado, and Nicholas Evans

Perception of Expressivity in TTS: Linguistics, Phonetics or Prosody?. 262
Marie Tahon, Gwéno!é Lecorvé, Damien Lolive, and Raheel Qader

Author Index 275