Computational Analysis of Storylines

Event structures are central in linguistics and artificial intelligence research: People can easily refer to changes in the world, identify their participants, distinguish relevant information, and have expectations of what can happen next. Part of this process is based on mechanisms similar to narratives, which are at the heart of information sharing. But it remains difficult to automatically detect events or automatically construct stories from such event representations. This book explores how to handle today's massive news streams and provides multidimensional, multimodal, and distributed approaches, like automated deep learning, to capture events and narrative structures involved in a "story." This overview of the current state-of-the-art on event extraction, temporal and casual relations, and storyline extraction aims to establish a new multidisciplinary research community with a common terminology and research agenda. Graduate students and researchers in natural language processing, computational linguistics, and media studies will benefit from this book.

TOMMASO CASELLI is an Assistant Professor in Computational Semantics at the University of Groningen. He received his PhD in computational linguistics on temporal processing of texts from the University of Pisa. His main research areas are in discourse processing, event extraction, and (event) sentiment analysis. He is one of the founders of the "Event and Stories in the News" workshop series and is currently working on developing computational models and natural language processing tools to extract plot structures from news. He took part in organizing semantic evaluation campaigns in natural language processing for English and Italian.

EDUARD HOVY is a Research Professor at the Language Technology Institute at Carnegie Mellon University. He was awarded honorary doctorates from the National Distance Education University (UNED) in Madrid in 2013 and the University of Antwerp in 2015. He is one of the initial 17 Fellows of the Association for Computational Linguistics (ACL). His research contributions include the co-development of the ROUGE text summarization evaluation method, the BLANC coreference evaluation method, the Omega ontology, the Webclopedia QA Typology, the FEMTI machine translation evaluation classification, the DAP text harvesting method, the OntoNotes corpus, and a model of structured distributional semantics.

MARTHA PALMER is a Professor at the University of Colorado in Linguistics, Computer Science, and Cognitive Science. She is a AAAI Fellow and an ACL Fellow. She works on trying to capture elements of the meanings of words that can comprise automatic representations of complex sentences and documents. She is a co-editor of *Linguistic Issues in Language Technology* and has been on the CLJ Editorial Board and a co-editor of JNLE. She is a past President of the Association for Computational Linguistics, past Chair of SIGLEX and SIGHAN, and was the Director of the 2011 Linguistics Institute held in Boulder, Colorado.

PIEK VOSSEN is Professor at Vrije Universiteit Amsterdam. He is the co-founder and co-president of the Global Wordnet Association, organizing the international Wordnet conferences since 2002. In 2013, he received the Dutch Spinoza prize for his research. He used this prize to launch a series of projects that included the structuring of news streams using storylines and reader/writer perspectives. Vossen's current main research focuses on cross-document event co-reference and perspective modeling of multiple sources with respect to event data and modeling event implications, as well as event timelines and storylines.

Studies in Natural Language Processing

This series offers widely accessible accounts of the state of the art in natural language processing. Established on the foundations of formal language theory and statistical learning, natural language processing is burgeoning with the widespread use of large annotated corpora, rich models of linguistic structure, and rigorous evaluation methods. New multilingual and multimodal language technologies have been stimulated by the growth of the web and pervasive computing devices. The series strikes a balance between statistical versus symbolic methods; deep versus shallow processing; rationalism versus empiricism; and fundamental science versus engineering. Each volume sheds light on these pervasive themes, delving into theoretical foundations and current applications. The series is aimed at a broad audience who are directly or indirectly involved in natural language processing, from fields including corpus linguistics, psycholinguistics, information retrieval, machine learning, spoken language, human-computer interaction, robotics, language learning, ontologies, and databases.

Also in the series

Douglas E. Appelt, *Planning English Sentences*
Madeleine Bates and Ralph M. Weischedel (eds.), *Challenges in Natural Language Processing*
Steven Bird, *Computational Phonology*
Peter Bosch and Rob van der Sandt, *Focus*
The Language of Word Meaning: (editors) *Federica Busa and Pierrette Bouillon*
Inheritance, Defaults and the Lexicon: (editors) *Ted Briscoe, Valeria de Paiva, and Ann Copestake*

Ronald Cole, Joseph Mariani, Hans Uszkoreit, Giovanni Varile, Annie Zaenen, Antonio Zampolli, and Victor Zue (eds.), *Survey of the State of the Art in Human Language Technology*
David R. Dowty, Lauri Karttunen, and Arnold M. Zwicky (eds.), *Natural Language Parsing*
Ralph Grishman, *Computational Linguistics*
Graeme Hirst, *Semantic Interpretation and the Resolution of Ambiguity*
András Kornai, *Extended Finite State Models of Language*
Kathleen R. McKeown, *Text Generation*
Martha Stone Palmer, *Semantic Processing for Finite Domains*
Terry Patten, *Systemic Text Generation as Problem Solving*
Ehud Reiter and Robert Dale, *Building Natural Language Generation Systems*
Manny Rayner, David Carter, Pierette Bouillon, Vassilis Digalakis, and Matis Wiren (eds.), *The Spoken Language Translator*
Michael Rosner and Roderick Johnson (eds.), *Computational Lexical Semantics*
Richard Sproat, *A Computational Theory of Writing Systems*
George Anton Kiraz, *Computational Nonlinear Morphology*
Nicholas Asher and Alex Lascarides, *Logics of Conversation*
Margaret Masterman (edited by Yorick Wilks), *Language, Cohesion and Form*
Walter Daelemans and Antal van den Bosch, *Memory-Based Language Processing*
Chu-Ren Huang, Nicoletta Calzolari, Aldo Gangemi, Alessandro Lenci, Alessandro Oltramari, and Laurent Prévot (eds.), *Ontology and the Lexicon: A Natural Language Processing Perspective*
Thierry Poibeau and Aline Villavicencio (eds.), *Language, Cognition, and Computational Models*
Bind Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Second Edition*
Marcos Zampieri and Preslav Nakov, *Similar Languages, Varieties, and Dialects: A Computational Perspective*

Computational Analysis of Storylines

Making Sense of Events

Edited by

TOMMASO CASELLI
University of Groningen

EDUARD HOVY
Carnegie Mellon University, Pennsylvania

MARTHA PALMER
University of Colorado at Boulder

PIEK VOSSEN
Vrije Universiteit, Amsterdam

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108490573
DOI: 10.1017/9781108854221

First published 2021

Printed in the United Kingdom by TJ Books Limited, Padstow Cornwall

A catalogue record for this publication is available from the British Library.

Library of Congress Cataloging-in-Publication Data
Names: Caselli, Tommaso, 1980– editor. | Palmer, Martha Stone, editor. | Hovy, Eduard H., editor. | Vossen, Piek, editor.
Title: Computational analysis of storylines : making sense of events / edited by Tommaso Caselli, Martha Palmer, Eduard Hovy, Piek Vossen.
Description: New York : Cambridge University Press, 2021. | Series: Studies in natural language processing | Includes bibliographical references.
Identifiers: LCCN 2021024558 (print) | LCCN 2021024559 (ebook) | ISBN 9781108490573 (hardback) | ISBN 9781108854221 (epub)
Subjects: LCSH: Discourse analysis, Narrative. | Narration (Rhetoric)–Data processing. | Natural language processing (Computer science) |
BISAC: COMPUTERS / Artificial Intelligence / Natural Language Processing |
LCGFT: Essays.
Classification: LCC P302.7 .C656 2021 (print) | LCC P302.7 (ebook) | DDC 401/.41–dc23
LC record available at https://lccn.loc.gov/2021024558
LC ebook record available at https://lccn.loc.gov/2021024559

ISBN 978-1-108-49057-3 Hardback