

Theory and Applications of Natural Language Processing

Series editors

Julia Hirschberg

Eduard Hovy

Mark Johnson

Aims and Scope

The field of Natural Language Processing (NLP) has expanded explosively over the past decade: growing bodies of available data, novel fields of applications, emerging areas and new connections to neighboring fields have all led to increasing output and to diversification of research.

“Theory and Applications of Natural Language Processing” is a series of volumes dedicated to selected topics in NLP and Language Technology. It focuses on the most recent advances in all areas of the computational modeling and processing of speech and text across languages and domains. Due to the rapid pace of development, the diversity of approaches and application scenarios are scattered in an ever-growing mass of conference proceedings, making entry into the field difficult for both students and potential users. Volumes in the series facilitate this first step and can be used as a teaching aid, advanced-level information resource or a point of reference.

The series encourages the submission of research monographs, contributed volumes and surveys, lecture notes and textbooks covering research frontiers on all relevant topics, offering a platform for the rapid publication of cutting-edge research as well as for comprehensive monographs that cover the full range of research on specific problem areas.

The topics include applications of NLP techniques to gain insights into the use and functioning of language, as well as the use of language technology in applications that enable communication, knowledge management and discovery such as natural language generation, information retrieval, question-answering, machine translation, localization and related fields.

The books are available in printed and electronic (e-book) form:

- * Downloadable on your PC, e-reader or iPad
- * Enhanced by Electronic Supplementary Material, such as algorithms, demonstrations, software, images and videos
- * Available online within an extensive network of academic and corporate R&D libraries worldwide
- * Never out of print thanks to innovative print-on-demand services
- * Competitively priced print editions for eBook customers thanks to MyCopy service <http://www.springer.com/librarians/e-content/mycopy>

More information about this series at <http://www.springer.com/series/8899>

Kemal Ofłazer • Murat Saraçlar
Editors

Turkish Natural Language Processing

 Springer

Editors

Kemal Oflazer
Carnegie Mellon University Qatar
Doha-Education City, Qatar

Murat Saraçlar
Electrical and Electronic Engineering
Boğaziçi University
Istanbul-Bebek, Turkey

ISSN 2192-032X ISSN 2192-0338 (electronic)
Theory and Applications of Natural Language Processing
ISBN 978-3-319-90163-3 ISBN 978-3-319-90165-7 (eBook)
<https://doi.org/10.1007/978-3-319-90165-7>

Library of Congress Control Number: 2018948167

© Springer International Publishing AG, part of Springer Nature 2018

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 the registered company Springer International Publishing AG part of Springer Nature.

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Turkish has proved to be a very interesting language for natural language processing techniques and applications. There has been a significant amount of work on Turkish since the early 1990s on introducing and/or adapting fundamental techniques, compiling resources, and developing applications.

The idea for this book came after one of us gave an invited talk at the LREC Conference held in Istanbul, Turkey, in 2012. Since then, the authors and we have worked hard to bring this effort to fruition. This book brings together most of the work done on Turkish in the last 25 years or so. After a bird’s-eye overview of relevant aspects of Turkish, it covers work on morphological processing and disambiguation, statistical language modeling, speech processing, named-entity recognition, dependency, and deep parsing. It then continues with statistical machine translation from English to Turkish and from Turkic languages to Turkish and sentiment analysis for Turkish, a topic that has recently been quite popular with the advent of social media. Finally, the book covers the most important natural language processing resources that have been developed for Turkish including the Turkish WordNet, the Turkish Treebank, Turkish National Corpus, and Turkish Discourse Bank.

We hope that this book helps other researchers in advancing the state of the art for Turkish and possibly other Turkic languages that share nontrivial similarities with Turkish.

Doha, Qatar
Istanbul, Turkey
July, 2017

Kemal Oflazer
Murat Saraçlar

Acknowledgements

The work presented in the following chapters has been supported by grants from various institutions: Early work was funded by a grant by NATO Science for Stability Program. The Turkish Scientific and Technological Research Council (TÜİTAK) has funded much of the more recent work in the last 15–20 years through many grants. Turkish Academy of Sciences (TUBA) provided support through its Outstanding Young Scientist Award program (TUBA-GEBİP). Boğaziçi University Research Fund (BU-BAP) also funded the language modeling and speech recognition research through grants. Additional funding was provided by the EU Framework 5 program.

A number of people have been very supportive of our work throughout these years: Lauri Karttunen has encouraged us and provided us with the Xerox Finite State Toolkit since the beginning, in order for us to build the very basic resources we needed. Other colleagues (then) at XRCE/PARC, notably Ken Beesley and Ron Kaplan, have supported our work by helping with the intricacies of the toolkit and by including our team in the ParGram Project. We thank them. The FSM Library and AMTools software packages developed by AT&T Labs-Research also have been essential in the early work on Turkish large vocabulary speech recognition. These were superseded by the OpenFST and Kaldi toolkits. We thank the authors of these toolkits and their organizations.

During these years, many graduate students and/or research assistants contributed to the work described in these chapters with their theses. Without their contributions, most of the work reported in the following chapters would not have been possible. We won't attempt to list them, lest we forget some. We thank them also.

Finally, we owe a lot to our families who supported us through these years. We cannot thank them enough.

Contents

1	Turkish and Its Challenges for Language and Speech Processing	1
	Kemal Oflazer and Murat Saraçlar	
2	Morphological Processing for Turkish	21
	Kemal Oflazer	
3	Morphological Disambiguation for Turkish	53
	Dilek Zeynep Hakkani-Tür, Murat Saraçlar, Gökhan Tür, Kemal Oflazer, and Deniz Yuret	
4	Language Modeling for Turkish Text and Speech Processing	69
	Ebru Arısoy and Murat Saraçlar	
5	Turkish Speech Recognition	93
	Ebru Arısoy and Murat Saraçlar	
6	Turkish Named-Entity Recognition	115
	Reyyan Yeniterzi, Gökhan Tür, and Kemal Oflazer	
7	Dependency Parsing of Turkish	133
	Gülşen Eryiğit, Joakim Nivre, and Kemal Oflazer	
8	Wide-Coverage Parsing, Semantics, and Morphology	153
	Ruket Çakıcı, Mark Steedman, and Cem Bozşahin	
9	Deep Parsing of Turkish with Lexical-Functional Grammar	175
	Özlem Çetinoğlu and Kemal Oflazer	
10	Statistical Machine Translation and Turkish	207
	Kemal Oflazer, Reyyan Yeniterzi, and İlknur Durgar-El Kahlout	
11	Machine Translation Between Turkic Languages	237
	A. Cüneyd Tantuğ and Eşref Adalı	
12	Sentiment Analysis in Turkish	255
	Gizem Gezici and Berrin Yanıkoğlu	

13	The Turkish Treebank	273
	Gülşen Eryiğit, Kemal Oflazer, and Umut Sulubacak	
14	Linguistic Corpora: A View from Turkish	291
	Mustafa Aksan and Yeşim Aksan	
15	Turkish Wordnet	317
	Özlem Çetinoğlu, Orhan Bilgin, and Kemal Oflazer	
16	Turkish Discourse Bank: Connectives and Their Configurations	337
	Deniz Zeyrek, Işın Demirşahin, and Cem Bozşahin	
	Index	357

About the Authors

Eşref Adalı received his B.Sc. and Ph.D. from Istanbul Technical University, Turkey, in 1971 and 1976, respectively. Currently, he is the Dean of the Faculty of Computer Engineering and Informatics, which he founded in 2010 after founding the Department of Computer Engineering in 1998. He was a Visiting Research Fellow at Case Western University, Cleveland, OH, USA, in 1977–1978; a Visiting Professor at Carleton University, Ottawa, Canada, in 1979; and a Visiting Professor at Akron University, Akron, OH, USA, in 1985. Adalı worked at the TÜBİTAK Marmara Research Center as the Founding Director of Informatics Group in 1990–1991. Although his more recent work has been on Turkish natural language processing, he has worked on microprocessors, system analysis, and design in the past and published two books on microcomputers and real-time systems.

Mustafa Aksan is Professor of Linguistics at Mersin University, Turkey, where he is also the Head of the Turkish Center for Corpus Studies. His main research interests are in corpus linguistics, lexical semantics, pragmatics, and morphology. He is the Principal Researcher in the construction of the *Turkish National Corpus* and also a coauthor of *A Frequency Dictionary of Turkish* (Routledge 2016). Recently, Aksan started a project on affix ordering in Turkish.

Yeşim Aksan is Professor of Linguistics at Mersin University, Turkey. Her main research interests are in corpus linguistics, lexical semantics, cognitive semantics, and pragmatics. She is the Project Director of the *Turkish National Corpus* and a coauthor of *A Frequency Dictionary of Turkish* (Routledge 2016). Previously, Aksan conducted corpus-based and corpus-driven studies on various aspects of Turkish.

Ebru Arısoy received her B.Sc., M.Sc., and Ph.D. degrees from the Electrical and Electronics Engineering Department, Boğaziçi University, Istanbul, Turkey, in 2002, 2004, and 2009, respectively. From 2010 to 2013, she worked as a postdoctoral researcher at IBM T. J. Watson Research Center, Yorktown Heights, NY, USA. She then moved to IBM Turkey, where she had a key role in developing language modeling approaches for voice search and mobile dictation applications. Since 2014, she has worked as an Assistant Professor at the Electrical and Electronics

Engineering Department of the MEF University, Istanbul, Turkey. Her research interests include automatic speech recognition, statistical language modeling, as well as speech and language processing for educational technologies.

Orhan Bilgin is a translator, editor, and a dictionary publisher based in Istanbul, Turkey. He holds a B.A. in Economics and an M.A. in Cognitive Science from the Boğaziçi University, Istanbul, Turkey. His master's thesis was on frequency effects in the processing of morphologically complex nouns in Turkish. Between 2001 and 2004, he worked as a lexicographer on the Turkish team at the Sabancı University, Istanbul, Turkey, as part of the Balkanet project: an EU-funded project that designed and developed medium-sized word nets for six Balkan languages, including Turkish. Bilgin is the founder of zargan.com, an online English-Turkish dictionary that has been active since 2001, as well as a partner of Banguoglu Ltd., a translation company specializing in the translation of legal documents.

Cem Bozşahin works on the learning and projection of structure, including argument structure and constituent structure, at the intersection of computer science, linguistics, and philosophy. He holds a Ph.D. in Computer Science from the Arizona State University, Tempe, AZ, USA. He worked at the Ohio University (1990–1992) before permanently joining the Middle East Technical University, Ankara, Turkey, in 1992. He held visiting research assignments at the University of Edinburgh, Scotland, UK (2002–2017), at the Boğaziçi University, Istanbul, Turkey (2011), and at the University of Lisbon, Portugal (2015–2016).

Ruket Çakıcı is a lecturer at the Department of Computer Engineering of the Middle East Technical University, Ankara, Turkey. She received her Ph.D. from the School of Informatics of the University of Edinburgh, Scotland, UK, after receiving her M.Sc. and B.Sc. degrees in Computer Engineering from the Middle East Technical University. Her research interests are in empirical and data-driven methods for natural language processing, in particular (multilingual) morphological, syntactic, and semantic parsing; combinatory categorical grammar; automatic description generation for images and videos; and natural language generation.

Özlem Çetinoğlu is a postdoctoral researcher at the IMS, University of Stuttgart, Germany. She received her Ph.D. in Computer Science from the Sabancı University, Istanbul, Turkey, in 2009, where she developed a large-scale Turkish LFG grammar as part of her thesis in the context of the ParGram project. From 2009 to 2011, she worked at the CNGL, Dublin City University, Ireland, on automatically labeling English with deep syntactic information and on parsing noncanonical data. Çetinoğlu's research interests include deep grammars, statistical dependency parsing, morphologically rich languages, and code switching.

Işın Demirşahin holds a Ph.D. in Cognitive Science from the Middle East Technical University, Ankara, Turkey. Her research focuses on discourse structures in general and Turkish discourse and information structure in particular. Demirşahin currently works for Google UK Ltd. as a computational linguist and focuses on internationalizing end-to-end dialogue systems.

İlknur Durgar-El Kahlout is currently Chief Researcher at the TÜBİTAK-BİLGEM Research Center in Gebze, Turkey. She received her B.Sc. degree in Computer Engineering from the Başkent University, Ankara, Turkey, in 1997 and her M.Sc. and Ph.D. degrees in Computer Science and Engineering from the Sabancı University, Istanbul, Turkey, in 2003 and 2009 respectively. She worked as a postdoctoral fellow in the Quero project at LIMSI-CNRS, Orsay, France. Durgar-El Kahlout's current research interests are in natural language processing and statistical machine translation, especially of morphologically complex languages.

Gülşen Eryiğit obtained her M.Sc. and Ph.D. degrees from the Computer Engineering Department of the Istanbul Technical University, Turkey, in 2002 and 2007, respectively, where she is currently a faculty member. In addition, she is a founding member and coordinator of the Natural Language Processing Group and a member of the Learning from Big Data Group. In 2006, she worked as a Visiting Student Researcher at Växjö University, Sweden, in the first-ranked team in CoNLL shared tasks on multilingual dependency parsing. In 2007, Eryiğit won the Siemens Excellence Award for her Ph.D. thesis. Her current research focuses on natural language processing of Turkish on which she has authored and coauthored publications in prestigious journals and conferences. She represented Turkey in the CLARIN Project (EU 7th Framework Program, Common Language Resources and Technology Infrastructure) and in PARSEME (EU Cost Action, Parsing and Multiword Expressions). Recently, Eryiğit acted as the Principal Investigator of two research projects "Parsing Web 2.0 Sentences" (funded by EU Cost Action and TÜBİTAK) and "Turkish Mobile Personal Assistant" (funded by the Turkish Ministry of Science, Industry and Technology). Moreover, she is also the NLP coordinator of the interdisciplinary research project "A Signing Avatar System for Turkish-to-Turkish Sign Language Machine Translation" (funded by TÜBİTAK). Eryiğit also used her research to consult several domestic and international IT companies on Turkish natural language processing.

Gizem Gezici received her B.Sc. and M.Sc. degrees in Computer Science and Engineering from the Sabancı University, Istanbul, Turkey, in 2011 and 2013. She is currently a fourth-year Ph.D. student at the Sabancı University and is involved in a project in the emerging research area of *bias in search*. Gezici's research interests include sentiment analysis, machine learning, and data mining.

Dilek Zeynep Hakkani-Tür is a Research Scientist at Google. Prior to joining Google, she was a researcher at Microsoft Research (2010–2016), the International Computer Science Institute (2006–2010), and AT&T Labs-Research (2001–2005). She received her B.Sc. degree in Computer Engineering from the Middle East Technical University, Ankara, Turkey, in 1994, and her M.Sc. and Ph.D. degrees in Computer Engineering from the Bilkent University, Ankara, Turkey, in 1996 and 2000, respectively. Her research interests include natural language and speech processing, spoken dialogue systems, and machine learning for language processing. She coauthored more than 200 papers in natural language and speech processing and is the recipient of three Best Paper Awards for her work on active learning

for dialogue systems from IEEE Signal Processing Society, ISCA, and EURASIP. She was an associate editor of *IEEE Transactions on Audio, Speech, and Language Processing* (2005–2008), a member of the IEEE Speech and Language Technical Committee (2009–2014), and an area editor for speech and language processing for Elsevier’s *Digital Signal Processing Journal* and *IEEE Signal Processing Letters* (2011–2013). Since 2014, she has been a Fellow of IEEE and ISCA and currently serves on the ISCA Advisory Council (2015–2019). In addition, Hakkani-Tür was granted over 50 patents for her work.

Joakim Nivre is Professor of Computational Linguistics at Uppsala University, Sweden. He holds a Ph.D. in General Linguistics from the University of Gothenburg, Sweden, and a Ph.D. in Computer Science from Växjö University, Sweden. His research focuses on data-driven methods for natural language processing, in particular for syntactic and semantic analysis. Nivre is one of the main developers of the transition-based approach to syntactic dependency parsing, described in his book *Inductive Dependency Parsing* (Springer 2006) and implemented in the widely used MaltParser system. In addition to his current position of President of the Association for Computational Linguistics, he is one of the founders of the “Universal Dependencies” project, which aims to develop cross-linguistically consistent treebank annotation for many languages and currently involves over 50 languages and over 200 researchers around the world. As of July 2017, Nivre was cited more than 11,000 times and produced over 200 scientific publications.

Kemal Oflazer received his Ph.D. in Computer Science from Carnegie Mellon University in Pittsburgh, PA, USA, and his M.Sc. in Computer Science and B.Sc. in Electrical and Electronics Engineering from the Middle East Technical University, Ankara, Turkey. He is currently a faculty member at Carnegie Mellon University in Doha, Qatar, where he is also the Associate Dean for Research. He held visiting positions at the Computing Research Laboratory of the New Mexico State University, Las Cruces, USA, and at the Language Technologies Institute, Carnegie Mellon University. Prior to joining CMU-Qatar, he worked in the faculties of the Sabanci University in Istanbul, Turkey (2000–2008), and the Bilkent University in Ankara, Turkey (1989–2000). He has worked extensively on developing natural language processing techniques and resources for Turkish. Oflazer’s current research interests are in statistical machine translation into morphologically complex languages, the use of NLP for language learning, and machine learning for computational morphology. In addition, he was a member of the editorial boards of *Computational Linguistics*, *Journal of Artificial Intelligence Research*, *Machine Translation*, and *Research on Language and Computation* and was a book review editor for *Natural Language Engineering*. Apart from having been a member of the Nomination and Advisory Boards for EACL, he served as the Program Co-chair for ACL 2005, an area chair for COLING 2000, EACL 2003, ACL 2004, ACL 2012, EMNLP 2013, and the Organization Committee Co-chair for EMNLP 2014. Currently, he is an editorial board member of both *Language Resources and Evaluation* and *Natural Language Engineering* journals and is a member of the advisory board for “SpringerBriefs in Natural Language Processing.”

Murat Saraçlar received his B.Sc. degree in 1994 from the Electrical and Electronics Engineering Department of the Bilkent University, Ankara, Turkey; his M.S.E. degree in 1997; and Ph.D. degree in 2001 from the Electrical and Computer Engineering Department of the Johns Hopkins University, Baltimore, MD, USA. From 2000 to 2005, he was with the multimedia services department of the AT&T Labs Research. In 2005, he joined the Electrical and Electronic Engineering Department of the Boğaziçi University, Istanbul, Turkey, where he is currently Full Professor. He was a Visiting Research Scientist at Google Inc., New York, NY, USA (2011–2012) and an Academic Visitor at IBM T. J. Watson Research Center (2012–2013). Saraçlar was awarded the AT&T Labs Research Excellence Award in 2002, the Turkish Academy of Sciences Young Scientist (TUBA-GEBIP) Award in 2009, and the IBM Faculty Award in 2010. He published more than 100 articles in journals and conference proceedings. Furthermore, he served as an associate editor for *IEEE Signal Processing Letters* (2009–2012) and *IEEE Transactions on Audio, Speech, and Language Processing* (2012–2016). Having been editorial board member of *Language Resources and Evaluation* from 2012 to 2016, Saraçlar is currently an editorial board member of *Computer Speech and Language* as well as a member of the IEEE Signal Processing Society Speech and Language Technical Committee (2007–2009, 2015–2018).

Mark Steedman is Professor of Cognitive Science at the School of Informatics of the University of Edinburgh, Scotland, UK. Previously, he taught at the Department of Computer and Information Science of the University of Pennsylvania, Philadelphia, PA, USA, which he joined as Associate Professor in 1988. His Ph.D. on artificial intelligence is from the University of Edinburgh. He is a Fellow of the American Association for Artificial Intelligence, the British Academy, the Royal Society of Edinburgh, the Association for Computational Linguistics, the Cognitive Science Society and a member of the European Academy. Steedman's research interests cover issues in computational linguistics, artificial intelligence, computer science and cognitive science, including syntax and semantics of natural language, wide-coverage parsing and open-domain question answering, comprehension of natural language discourse by humans and by machine, grammar-based language modeling, natural language generation, and the semantics of intonation in spoken discourse. Much of his current NLP research addresses probabilistic parsing and robust semantics for question answering using the CCG grammar formalism, including the acquisition of language from paired sentences and meanings by child and machine. Some of his research also concerns the analysis of music by humans and machines. Steedman occasionally works with colleagues in computer animation where these theories are used to guide the graphical animation of speaking virtual or simulated autonomous human agents.

Umut Sulubacak is a research and teaching assistant at the Department of Computer Engineering of the Istanbul Technical University, Turkey. As part of his B.Sc. and M.Sc. studies, his research focused on the morphological and syntactic analysis of Turkish, using both rule-based and data-driven methods and optimizing morpho-syntactic annotation processes for Turkish dependency treebanks. He was

involved in the construction of the Turkish treebank as part of the “Universal Dependencies” project and has remained an active contributor to the project ever since. In addition to his teaching responsibilities, he currently pursues his Ph.D. degree at the same institution with research in treebank linguistics and machine learning for Turkish language processing.

A. Cüneyd Tantuğ is currently Associate Professor at the Faculty of Computer and Informatics Engineering of the Istanbul Technical University, Turkey, where he completed his Ph.D. and has been a faculty member since 2009. His research areas include natural language processing, machine translation, and machine learning.

Gökhan Tür is a computer scientist focusing on human/machine conversational language understanding systems. He was awarded his Ph.D. in Computer Science from the Bilkent University, Ankara, Turkey, in 2000. Between 1997 and 1999, he was a Visiting Scholar at the Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA, USA; then at the Johns Hopkins University, Baltimore, MD, USA; and at the Speech Lab of SRI, Menlo Park, CA, USA. He was at AT&T Research (2001–2006), working on pioneering conversational systems like “How May I Help You?” Later at SRI, he worked for the DARPA GALE and CALO projects (2006–2010). He was a founding member of the Microsoft Cortana team, focusing on deep learning methods (2010–2016), and was the Conversational Understanding Architect at the Apple Siri team (2014–2015). Tür is currently with the Deep Dialogue team at Google Research. Apart from frequent presentations at conferences, he coauthored more than 150 papers in journals and books. He is also a coeditor of the book *Spoken Language Understanding: Systems for Extracting Semantic Information from Speech* (Wiley 2011) and of a special issue on Spoken Language Understanding of the journal *Speech Communication*. He is also the recipient of the Best Paper Award of *Speech Communication* by ISCA for 2004–2006 and by EURASIP for 2005–2006. Tür was the organizer of the HLT-NAACL 2007 Workshop on Spoken Dialog Technologies, the HLT-NAACL 2004, and the AAAI 2005 Workshops on Spoken Language Understanding. He also served as the area chair for spoken language processing for IEEE ICASSP conferences from 2007 to 2009 and IEEE ASRU Workshop in 2005, as the spoken dialog area chair for the HLT-NAACL conference in 2007, and as an organizer of the SLT Workshop in 2010. Having been a member of the IEEE Speech and Language Technical Committee (SLTC) (2006–2008) and the IEEE SPS Industrial Relations Committee (2013–2014) as well as an associate editor for the *IEEE Transactions on Audio, Speech, and Language Processing* (2010–2014) and *Multimedia Processing* (2014–2016) journals, Tür is currently a senior member of IEEE, ACL, and ISCA.

Berrin Yamkoğlu is Professor of Computer Science at the Sabancı University, Istanbul, Turkey. She received a double major in Computer Science and Mathematics from the Boğaziçi University, Istanbul, Turkey, in 1988 and her Ph.D. in Computer Science from the Dartmouth College, Hanover, NH, USA, in 1993. Her research interests lie in the areas of machine learning, with applications to image and language understanding, currently focusing on multimodal deception detection,

sentiment analysis, online handwriting recognition, and object recognition from photographs. Yanıkoğlu received an IBM Research Division award in 1996 and first place positions in several international signature verification competitions in collaboration with her students and colleagues.

Reyyan Yeniterzi received her B.Sc. and M.Sc. degrees in Computer Science and Engineering from the Sabancı University, Istanbul, Turkey, in 2007 and 2009, and her M.Sc. and Ph.D. degrees from the Language Technologies Institute of Carnegie Mellon University, Pittsburgh, PA, USA, in 2012 and 2015. Since 2015, she has worked as an Assistant Professor at the Computer Science Department of the Özyeğin University, Istanbul, Turkey. Previous to her various visiting research positions at the International Computer Science Institute (ICSI), Berkeley, CA, USA; at Vanderbilt University, TN, USA; and at Carnegie Mellon University in Doha, Qatar, Yeniterzi gained practical experience as an intern at Google. Her main research interests include natural language processing, text mining, social media analysis, information retrieval, search engines, and statistical machine translation.

Deniz Yuret is Associate Professor of Computer Engineering at the Koç University in Istanbul, Turkey, where he has worked at the Artificial Intelligence Laboratory since 2002. Previously, he was at the MIT AI Lab, Cambridge, MA, USA (1988–1999), and later cofounded Inquire, Inc., a company commercializing question answering technology (2000-2002). Yuret worked on supervised and unsupervised approaches to syntax, morphology, lexical semantics, and language acquisition. His most recent work is on grounded language acquisition and understanding.

Deniz Zeyrek is Professor of Cognitive Science and Director of the Informatics Institute at the Middle East University in Ankara, Turkey. She holds a Ph.D. in Linguistics from the Hacettepe University, Ankara, Turkey. Her broad research interests are in Turkish discourse and pragmatics, development of annotation schemes for recording discourse, and pragmatic phenomena on corpora. She contributed to the development of the METU Turkish Corpus, a corpus of modern written Turkish, and is the principal developer of Turkish Discourse Bank, a discourse corpus annotated in the PDTB style. Zeyrek's research mainly focuses on discourse relations and means of expressing discourse relations in Turkish and expands to discourse structure as revealed by discourse relations.