Quasy 2021

**Second Workshop on
Quantitative Syntax
(Quasy, SyntaxFest 2021)**

Proceedings

To be held as part of SyntaxFest 2021

21–25 March, 2022

Sofia, Bulgaria

Preface

Following its first edition that was held in Paris in 2019, the Second Workshop on Quantitative Syntax (Quasy 2021) takes place again as part of SyntaxFest, which co-locates four related but independent events:

- The Sixth International Conference on Dependency Linguistics (Depling 2021)

- The Second Workshop on Quantitative Syntax (Quasy 2021)

- The 20th International Workshop on Treebanks and Linguistic Theories (TLT 2021)

- The Fifth Workshop on Universal Dependencies (UDW 2021)

The reasons that suggested bringing these four events together in 2019 still hold in 2021. There is a continuing, strong interest in corpora and dependency treebanks for empirically validating syntactic theories, studying syntax from quantitative and theoretical points of view, and for training machine learning models for natural language processing. Much of this research is increasingly multilingual and cross-lingual, made in no small part possible by the Universal Dependencies project, which continues to grow at currently nearly 200 treebanks in over 100 languages.

For these reasons and encouraged by the success of the first SyntaxFest, which was held in 2019 in Paris, we – the chairs of the four events – decided to bring them together again in 2021. Due to the vagaries of the COVID-19 pandemic, it was eventually decided to push the actual SyntaxFest 2021 back to March 2022. In order not to delay the publication of new research and not to conflict with other events, we decided however to publish the proceedings that you are now reading in advance, in December 2021.

As in 2019, we organized a single reviewing process for the whole SyntaxFest, with identical paper formats for all four events. Authors could indicate (multiple) venue preferences, but the assignment of papers to events for accepted papers was made by the program chairs.

38 long papers were submitted, 25 to Depling, 11 to Quasy, 17 to TLT, and 24 to UDW. The program chairs accepted 30 (79%) and assigned 8 to Depling, 5 to Quasy, 7 to TLT, and 10 to UDW. 22 short papers were submitted, 6 to Depling, 7 to Quasy, 9 to TLT, and 9 to UDW. The program chairs accepted 14 (64%) and assigned 3 to Depling, 3 to Quasy, 3 to TLT, and 5 to UDW.

At the time of this writing, we do not yet know whether SyntaxFest will be a hybrid or purely online event. We regret this uncertainty but are nevertheless looking forward to it very much. Our sincere thanks go to everyone who is making this event possible, including everybody who submitted their papers, and of course the reviewers for their time and their valuable comments and suggestions. We would like to thank Djamé Seddah, whose assistance and expertise in organizing SyntaxFests was invaluable. Finally, we would also like to thank ACL SIGPARSE for its endorsement and the ACL Anthology for publishing the proceedings.

Radek Čech, Xinying Chen, Daniel Dakota, Miryam de Lhoneux, Kilian Evang, Sandra Kübler, Nicolas Mazziotta, Simon Mille, Reut Tsarfaty (co-chairs)

Petya Osenova, Kiril Simov (local organizers and co-chairs)

December 2021

Program co-chairs

The chairs for each event (and co-chairs for the single SyntaxFest reviewing process) are:

- Depling:

  - Nicolas Mazziotta (Université de Liège)
  - Simon Mille (Universitat Pompeu Fabra)

- Quasy:

  - Radek Čech (University of Ostrava)
  - Xinying Chen (Xi'an Jiaotong University)

- TLT:

  - Daniel Dakota (Indiana University)
  - Kilian Evang (Heinrich Heine University Düsseldorf)
  - Sandra Kübler (Indiana University)

- UDW:

  - Miryam de Lhoneux (Uppsala University / KU Leuven / University of Copenhagen)
  - Reut Tsarfaty (Bar-Ilan University / AI2)

Local Organizing Committee of the SyntaxFest

- Petya Osenova (Bulgarian Academy of Sciences)
- Kiril Simov (Bulgarian Academy of Sciences)

Program Committee for the Whole SyntaxFest

Patricia Amaral (Indiana University Bloomington)
Valerio Basile (University of Turin)
David Beck (University of Alberta)
Ann Bies (Linguistic Data Consortium, University of Pennsylvania)
Xavier Blanco (UAB)
Igor Boguslavsky (Universidad Politécnica de Madrid)
Bernd Bohnet (Google)
Cristina Bosco (University of Turin)
Gosse Bouma (Rijksuniversiteit Groningen)
Miriam Butt (Universität Konstanz)
Marie Candito (Universtité Paris 7 / INRIA)
Radek Cech (University of Ostrava)
Giuseppe Giovanni Antonio Celano (University of Pavia)
Xinying Chen (Xi'an Jiaotong University)
Silvie Cinková (Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied
Linguistics)
Cagri Coltekin (University of Tuebingen)
Benoit Crabbé (Université Paris 7 / Institut national de recherche en informatique et en automatique, Paris)
Daniel Dakota (Indiana University)
Eric De La Clergerie (Institut national de recherche en informatique et en automatique, Paris)
Felice Dell'Orletta (Institute for Computational Linguistics, National Research Council, Pisa)
Kaja Dobrovoljc (Jožef Stefan Institute)
Kilian Evang (Heinrich Heine University Düsseldorf)
Thiago Ferreira (University of São Paulo)
Ramon Ferrer-I-Cancho (Universitat Politècnica de Catalunya)
Kim Gerdes (Laboratoire Interdisciplinaire des Sciences du Numérique, Université Paris-Saclay)
Carlos Gómez-Rodríguez (Universidade da Coruña)
Jan Hajic (Institute of Formal and Applied Linguistics, Charles University, Prague)
Eva Hajicova (Institute of Formal and Applied Linguistics, Charles University, Prague)
Dag Haug (University of Oslo)
Richard Hudson (University College London)
András Imrényi (Eszterházy Károly Egyetem)
Leonid Iomdin (Institute for Information Transmission Problems, Russian Academy of Sciences)
Jingyang Jiang (Zhejiang University)
Sylvain Kahane (Modyco, Université Paris Ouest Nanterre / CNRS)
Vaclava Kettnerova (Institute of Formal and Applied Linguistics)
Sandra Kübler (Indiana University Bloomington)
Guy Lapalme (University of Montreal)
François Lareau (Observatoire de linguistique Sens-Texte, Université de Montréal)
Alessandro Lenci (University of Pisa)
Nicholas Lester (University of Zurich)
Lori Levin (Carnegie Mellon University)
Haitao Liu (Zhejiang University)

Marketa Lopatkova (Institute of Formal and Applied Linguistics, Charles University, Prague)

Olga Lyashevskaya (National Research University Higher School of Economics)

Teresa Lynn (Dublin City University)

Jan Macutek (Mathematical Institute of the Slovak Academy of Sciences / Constantine the Philosopher University in Nitra)

Robert Malouf (San Diego State University)

Alessandro Mazzei (Dipartimento di Informatica, Università di Torino)

Nicolas Mazziotta (Université de Liège)

Alexander Mehler (Text Technology Group, Goethe-University Frankfurt am Main)

Wolfgang Menzel (Department of Informatics, Hamburg University)

Jasmina Milicevic (Dalhousie University)

Simon Mille (Pompeu Fabra University)

Yusuke Miyao (The University of Tokyo)

Simonetta Montemagni (Institute for Computational Linguistics, National Research Council, Pisa)

Kaili Müürisep (University of Tartu)

Alexis Nasr (Laboratoire d'Informatique Fondamentale, Université de la Méditerranée, Aix-Marseille II)

Sven Naumann (University of Trier)

Anat Ninio (The Hebrew University of Jerusalem)

Joakim Nivre (Uppsala University)

Pierre Nugues (Lund University, Department of Computer Science Lund, Sweden)

Kemal Oflazer (Carnegie Mellon University-Qatar)

Timothy Osborne (Zhejiang University)

Petya Osenova (Sofia University / Institute of Information and Communication Technologies, Sofia)

Robert Östling (Department of Linguistics, Stockholm University)

Agnieszka Patejuk (Polish Academy of Sciences / University of Oxford)

Alain Polguère (Université de Lorraine)

Prokopis Prokopidis (Institute for Language and Speech Processing/Athena RC)

Laura Pérez Mayos (Pompeu Fabra University)

Owen Rambow (Stony Brook University)

Rudolf Rosa (Institute of Formal and Applied Linguistics, Charles University, Prague)

Tanja Samardzic (University of Zurich)

Giorgio Satta (University of Padua)

Nathan Schneider (Georgetown University)

Olga Scrivner (Indiana University Bloomington)

Djamé Seddah (Alpage, Université Paris la Sorbonne)

Alexander Shvets (Institute for Systems Analysis of Russian Academy of Sciences)

Maria Simi (Università di Pisa)

Achim Stein (University of Stuttgart)

Reut Tsarfaty (Faculty of Mathematics and Computer Science, The Weizmann Institute of Science, Rehovot)

Francis M. Tyers (Indiana University Bloomington)

Zdenka Uresova (Institute of Formal and Applied Linguistics, Charles University, Prague)

Gertjan Van Noord (University of Groningen)

Giulia Venturi (Institute for Computational Linguistics, National Research Council, Pisa)

Veronika Vincze (Hungarian Academy of Sciences, Research Group on Articial Intelligence)

Relja Vulanovic (Kent State University at Stark)

Chunshan Xu (anhui jianzhu university)
Xiang Yu (University of Stuttgart)
Zdenek Zabokrtsky (Institute of Formal and Applied Linguistics, Charles University, Prague)
Amir Zeldes (Georgetown University)
Daniel Zeman (Institute of Formal and Applied Linguistics, Charles University, Prague)
Hongxin Zhang (Zhejiang University)
Yiyi Zhao (Institute of Applied Linguistics, Communication University of China, Beijing)
Heike Zinsmeister (University of Hamburg)
Miryam de Lhoneux (University of Copenhagen)
Marie-Catherine de Marneffe (The Ohio State University)
Valeria de Paiva (Samsung Research America and University of Birmingham)

Additional Reviewers

Chiara Alzetta (National Research Council of Italy)
Aditya Bhargava (University of Toronto)
Lauren Cassidy (Dublin City University)
Simon Petitjean (Heinrich Heine University Düsseldorf)
Xenia Petukhova (National Research University Higher School of Economics)
Daniel Swanson (Indiana University)
He Zhou (Indiana University)
Yulia Zinova (Heinrich Heine University Düsseldorf)

Table of Contents