TLT 2020

Proceedings of the 19th International Workshop on Treebanks and Linguistic Theories

27–28 October, 2020
University of Düsseldorf
Düsseldorf, Germany

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-952148-01-9

Introduction

Welcome to the 19th edition of the International Workshop on Treebanks and Linguistic Theories! It was meant to take place in Düsseldorf, but like so many events in 2020 had to pivot to online-only. We hope it will be a great experience for everyone all the same!

TLT's aim is to bring together developers and users of linguistically annotated natural language corpora. It addresses all aspects of treebank design, development, and use. By "treebank" we mean any pairing of natural language data (spoken, signed, or written) with annotations of linguistic structure at various levels of analysis, including e.g., morpho-phonology, syntax, semantics, and discourse. Annotations can take any form, including trees and general graphs.

The program of TLT 2020 reflects this broad view of work on treebanks. It includes papers on the construction of annotated resources, parsing, typology and universals, under-resourced and historical languages, and new tools for processing and querying. The program shows that an increasing amount of work in the treebank space – as well as in computational linguistics and natural language processing in general – is multilingual, looking at multiple languages from the start instead of just one. Another increasingly important theme is semantic annotation. Both themes are reflected in our invited talks. Miryam de Lhoneux will talk about parsing multiple languages, especially truly low-resource ones, about when cross-lingual learning helps and where more work is needed. And Johan Bos will talk about possibilities and difficulties in large-scale deep semantic annotation, and present a new method that may make it easier. We are delighted they accepted our invitations and we include their abstracts in this volume.

We received a total of 4 short paper submissions, of which 3 were accepted (75%), and a total of 13 long paper submissions (not counting one submission that was withdrawn), of which 11 were accepted (85%) following the reviews by our program committee. We are very grateful for the hard work of the reviewers, as well as for that of the authors, especially seeing as moving the event online resulted in a tight schedule and a video requirement.

This will be the first TLT that is held online. We opted for a setup where a regular two-day schedule of sessions takes place via video chat with talks and Q&A, but talks are pre-recorded to minimize the impact of any technical difficulties, and to make them more accessible, e.g., to participants in different timezones. A social event will also take place via video chat on the eve of the workshop. In parallel, we use text chat for asynchronous communication between participants before, during and after the workshop. We hope that it will work out well and inspire people to come back for many more TLTs, online and offline!

Kilian Evang, Laura Kallmeyer, Rafael Ehren, Simon Petitjean, Esther Seyffarth, and Djamé Seddah

Düsseldorf & Paris

October 2020

Organisers:

Kilian Evang
Laura Kallmeyer
Rafael Ehren
Simon Petitjean
Esther Seyffarth
Djamé Seddah

Program Committee:

Lasha Abzianidze, Patricia Amaral, Emily M. Bender, Johan Bos, Cristina Bosco, Giuseppe Giovanni Antonio Celano, Silvie Cinková, Daniel Dakota, Miryam de Lhoneux, Jennifer Foster, Carlos Gómez-Rodríguez, Daniel Hershcovich, Sandra Kübler, François Lareau, Nicholas Lester, Haitao Liu, Nicolas Mazziotta, Alexander Mehler, Yusuke Miyao, Jiří Mírovský, Sven Naumann, Joakim Nivre, Pierre Nugues, Stephan Oepen, Alain Polguère, Rudolf Rosa, Rik van Noord, Amir Zeldes

Invited Speakers:

Johan Bos, University of Groningen
Miryam de Lhoneux, University of Copenhagen

Invited Talks

Johan Bos: Grammar, Meaning & Annotation

What is the role of computational grammars in semantic annotation? In the Parallel Meaning Bank, grammar plays a pivotal role. This has good sides, and bad sides. It is good, because annotation is ensured to be carried out in a systematic, consistent and efficient way. But it can also be counterproductive, as linguistic input can be full of surprises. In such cases the grammar is a showstopper. Well, you might say, why not bypass the grammar in such cases? Sure, but annotating meanings from scratch is not straightforward when the targets are expressive semantic representations, such as the Discourse Representation Structures from Discourse Representation Theory used in the Parallel Meaning Bank. I present a new notation for these meaning representations: without variables, without explicit recursion, and without reliance on grammar.

Miryam de Lhoneux: Parsing Typologically Diverse Languages

This talk is about parsing typologically diverse languages. I first argue that the Universal Dependencies (UD) dataset is the best multilingual dataset that we currently have and allows us to ask general questions that are relevant for multilingual NLP. I then ask the question of how well our current parsers generalize across languages and the question of how we evaluate that.

I subsequently ask the question of how accurate our parsers currently are for truly low-resource languages. I explain recent developments in cross-lingual learning that are great at leveraging data from related languages and that improve parsing accuracy for low-resource languages. I show that for low-resource languages for which we do not have a high-resource related language, our parsers are currently highly inaccurate. Since such cases represent the majority of world languages, we might want to shift our focus on these. I finally suggest that we may find answers in the use of typological information, discuss work that has tried to do that and highlight what more can be done.

Table of Contents