

LREC 2020 WORKSHOP
Language Resources and Evaluation Conference
11–16 May 2020

**International FrameNet Workshop 2020
Towards a Global, Multilingual FrameNet**

PROCEEDINGS

Edited by:
Tiago T. Torrent, Collin F. Baker, Oliver Czulo, Kyoko Ohara and
Miriam R. L. Petruck

**Proceedings of the LREC
International FrameNet Workshop 2020:
Towards a Global, Multilingual FrameNet**

Edited by: Tiago T. Torrent, Collin F. Baker, Oliver Czulo, Kyoko Ohara and Miriam R. L. Petruck

ISBN: 979-10-95546-58-0
EAN: 9791095546580

For more information:

European Language Resources Association (ELRA)
9 rue des Cordelières
75013, Paris
France
<http://www.elra.info>
Email: lrec@elda.org

© European Language Resources Association (ELRA)

These Workshop Proceedings are licensed under a Creative Commons
Attribution-NonCommercial 4.0 International License

Introduction

This workshop continues the series of International FrameNet Workshops, based on frame semantics and FrameNets around the world, including meetings in Berkeley, California in 2013 (sponsored by the Swedish FrameNet group), in Juiz de Fora, Brazil in 2016 (sponsored by FrameNet Brasil), and in Miyazaki, Japan in 2018 (in conjunction with the LREC conference there). The last of these was specifically designed to bring together two sometimes separate groups of researchers, those concentrating on frame semantics and those concentrating on construction grammar, which Charles J. Fillmore and many of the same colleagues developed in parallel over a period of several decades. The call for papers of the current conference emphasized that the workshop would welcome studies of both theoretical and practical issues, and we are fortunate to have strong contributions of both types, sometimes within a single paper. A number of the papers arise out of the parallel annotation of Ken Robinson's TED talk "Do schools kill creativity?". This very popular talk has been translated (or at least subtitled) in a wide variety of languages. The attendees at the International FrameNet Workshop in 2016 decided that different groups would undertake Frame Semantic annotation for it, based on Berkeley FrameNet Release 1.7 as shared standard. These efforts have raised a number of important theoretical issues while also helping new FrameNet projects to get underway in several languages.

Three papers deal with expansion of FrameNets to new languages and content areas:

- Gargett and Leung describe the building of a new FrameNet for Emirati Arabic, and how they were able to use novel combinations of existing databases and tools to create an Emirati Arabic corpus and begin annotating it semiautomatically with frames and frame elements derived mainly from Berkeley FrameNet. They report encouraging numbers for inter-annotator agreement and good prospects for further automatic processing.
- Giouli et al. describe participating in the shared annotation task on the TED talk and the issues they encountered in trying to use the Berkeley FrameNet frames for Modern Greek. Some of these are simply gaps in the BFN lexical coverage, others are interesting issues related to differences between English and Greek.
- L'Homme et al. are building a multilingual frame semantic resource for the specialized terminology relating to the environment. They have created more than 100 new frames for this purpose and "slightly modified" 40 more in the process; they began with French and expanded to English, Spanish, Portuguese, and Chinese recently. They propose a general technique for FrameNet expansion in technical domains.

Two papers deal with how frame semantics applies across languages and what relation it has to differences in constructions across languages:

- Ohara concentrates on constructions, comparing constructions in English and in Japanese, also based on examples from the shared annotation task. In this study, she expands on earlier work comparing frames and constructions between English and Japanese children's books, and points out that the Japanese version of the TED talk is not exactly intended as a complete translation of the English version.
- Baker and Lorenzi describe a variety of techniques for aligning frames between FrameNets in different languages. For this purpose, the authors have created a tool to visualize the alignments and compare the techniques used, an effort that should aid in testing different combinations of alignment methods.

Three papers deal with practical application of frame semantics to other NLP tasks:

- Marzinotto projects FrameNet annotation from English to French using two cutting-edge NLP methods, attention-based machine translation and parsing based on BERT embeddings, using data from two French frame semantic resources, ASFALDA and CALOR. The study found that projecting annotation in this way slightly improves the performance of the parser.
- Lee et al. report on the creation of a database of Mandarin Chinese verbs, and how it was used to analyze the occurrences of verbs in Chinese language textbooks used in China across grades 1-12. They found that the distinction between core and non-core frame elements helps to predict the reading difficulty of the texts.
- Gruzitis et al., in a surprising change of direction, start from the Latvian FrameNet corpus, which also has a universal dependency layer and semi-automatically create a complete set of Propbank annotation for the same corpus; this process was found to be much easier than a mapping from Propbank to FrameNet would have been.

Four papers deal with connecting frame semantics to other types of meaning:

- Czulo et al. remind us that from early in the development of frame semantics, Fillmore conceived of the theory as encompassing both lexical frames where meanings are tied to specific lexical items and pragmatic frames (called "scenes" at that point) tied to the situational context of usage, rather than lexical items. The paper offers an update on the concept, with examples from the shared annotation task.
- Postma et al. describe a new approach to annotating both semantic frames and real-world references (based on Wikidata) on the same texts, creating rich possibilities for future work analyzing text coherence and variations in framing of news events. The authors have developed a tool which both assists and constrains the annotator in this process.
- Remijnse and Minnema also deal with real-world references, but for the problem of constructing narratives from news reports. The authors found that pragmatically inferred frames were needed to achieve coherence across documents.
- Belcavello et al. discuss a novel tool they have created for annotating multimodal data with semantic frames. The paper provides an example worked in detail about how the text, audio, and video can all be annotated somewhat differently, with attention to how they are actually temporally related in the data – here, a TV series about travel to different countries. The paper also illustrates how a highly elaborated set of ternary qualia relations can help to connect the annotations of the different modalities.

The Organizing Committee would like to thank the authors for submitting so many substantial papers and the reviewers for helping to select and improve those included here. We would also like to thank LREC for scheduling a full day for this workshop, so that we could cover the full range of relevant topics. As this preface is being written, the pandemic caused by COVID-19 makes it uncertain whether the workshop originally planned for May 16, 2020 in Marseilles in conjunction with the LREC conference will simply be postponed or will be converted into an on-line meeting. Still, we are eager to press ahead with the publication of the proceedings, so that these exciting papers can receive the attention they deserve.

Collin Baker, for the Organizing Committee

Organizers:

Tiago Timponi Torrent, FrameNet Brasil, Federal University of Juiz de Fora, Brazil
Collin Baker, FrameNet, International Computer Science Institute, USA
Oliver Czulo, Department of Translation Studies, Universität Leipzig, Germany
Kyoko Ohara, Japanese FrameNet, Keio University, Japan
Miriam R. L. Petruck, FrameNet, International Computer Science Institute, USA

Program Committee:

Hans Boas, University of Texas, Austin
Dana Dannells, Gothenburg University
Gerard de Melo, Rutgers University
Ellen Dodge, Google Inc.
Jerome Feldman, International Computer Science Institute
Maucha Gamonal, University of Leipzig
Laurent Gautier, University of Bourgogne
Normunds Gruzitis, University of Latvia
Karin Friberg Heppin, Apple Inc.
Richard Johansson, Gothenburg University
Dimitrios Kokkinakis, Gothenburg University
Marie-Claude L'Homme, University of Montréal
Russell Lee-Goldman, Google Inc.
Ely Matos, Federal University of Juiz de Fora
Sebastian Padó, University of Stuttgart
Michael Roth, University of Stuttgart
Josef Ruppenhofer, Institute for the German Language (IDS)
Nathan Schneider, Georgetown University
Simon Varga, University of Mainz
Alexander Ziem, Heinrich-Heine University Düsseldorf

Table of Contents

<i>Beyond Lexical Semantics: Notes on Pragmatic Frames</i> Oliver Czulo, Alexander Ziem and Tiago Timponi Torrent	1
<i>Finding Corresponding Constructions in English and Japanese in a TED Talk Parallel Corpus using Frames-and-Constructions Analysis</i> Kyoko Ohara	8
<i>Towards Reference-Aware FrameNet Annotation</i> Levi Remijnse and Gosse Minnema	13
<i>Frame-Based Annotation of Multimodal Corpora: Tracking (A)Synchronies in Meaning Construction</i> Frederico Belcavello, Marcelo Viridiano, Alexandre Diniz da Costa, Ely Matos and Tiago Timponi Torrent	23
<i>Combining Conceptual and Referential Annotation to Study Variation in Framing</i> Marten Postma, Levi Remijnse, Filip Ilievski, Antske Fokkens, Sam Titarsolej and Piek Vossen	31
<i>FrameNet Annotations Alignment using Attention-based Machine Translation</i> Gabriel Marzinotto	41
<i>Greek within the Global FrameNet Initiative: Challenges and Conclusions so far</i> Voula Giouli, Vera Pilitsidou and Hephæstion Christopoulos	48
<i>Using Verb Frames for Text Difficulty Assessment</i> John Lee, Meichun Liu and Tianyuan Cai	56
<i>Deriving a PropBank Corpus from Parallel FrameNet and UD Corpora</i> Normunds Gruzitis, Roberts Dargis, Laura Rituma, Gunta Nešpore-Bērzkalne and Baiba Saulite	63
<i>Building the Emirati Arabic FrameNet</i> Andrew Gargett and Tommi Leung	70
<i>Exploring Crosslinguistic Frame Alignment</i> Collin Baker and Arthur Lorenzi	77
<i>Building Multilingual Specialized Resources Based on FrameNet: Application to the Field of the Environment</i> Marie-Claude L' Homme, Benoît Robichaud and Carlos Subirats	85

Conference Program

Saturday, May 16, 2020

09:00–09:20 Opening Session

09:00–09:20 *Welcome and introduction to the workshop*
Collin F. Baker

09:20–11:00 Oral Session 1: Theoretical Papers

09:20–09:45 *Beyond Lexical Semantics: Notes on Pragmatic Frames*
Oliver Czulo, Alexander Ziem and Tiago Timponi Torrent

09:45–10:10 *Finding Corresponding Constructions in English and Japanese in a TED Talk Parallel Corpus using Frames-and-Constructions Analysis*
Kyoko Ohara

10:10–10:35 *Towards Reference-Aware FrameNet Annotation*
Levi Remijnse and Gosse Minnema

10:35–11:00 *Frame-Based Annotation of Multimodal Corpora: Tracking (A)Synchronies in Meaning Construction*
Frederico Belcavello, Marcelo Viridiano, Alexandre Diniz da Costa, Ely Matos and Tiago Timponi Torrent

11:00–11:30 Coffee Break

Saturday, May 16, 2020 (continued)

11:30–13:00 Poster Session

11:30–13:00 *Combining Conceptual and Referential Annotation to Study Variation in Framing*
Marten Postma, Levi Remijnse, Filip Ilievski, Antske Fokkens, Sam Titarsolej and Piek Vossen

11:30–13:00 *FrameNet Annotations Alignment using Attention-based Machine Translation*
Gabriel Marzinotto

11:30–13:00 *Greek within the Global FrameNet Initiative: Challenges and Conclusions so far*
Voula Giouli, Vera Pilitsidou and Hephaestion Christopoulos

11:30–13:00 *Using Verb Frames for Text Difficulty Assessment*
John Lee, Meichun Liu and Tianyuan Cai

11:30–13:00 *Deriving a PropBank Corpus from Parallel FrameNet and UD Corpora*
Normunds Gruzitis, Roberts Dargis, Laura Rituma, Gunta Nešpore-Bērzkalne and Baiba Saulite

11:30–13:00 *Building the Emirati Arabic FrameNet*
Andrew Gargett and Tommi Leung

13:00–14:00 Lunch Break

Saturday, May 16, 2020 (continued)

14:00–14:50 Oral Session 2: Practical Papers

14:00–14:25 *Exploring Crosslinguistic Frame Alignment*

Collin Baker and Arthur Lorenzi

14:25–14:50 *Building Multilingual Specialized Resources Based on FrameNet: Application to the Field of the Environment*

Marie-Claude L'Homme, Benoît Robichaud and Carlos Subirats

14:50–15:50 Group Discussion Session

14:50–15:50 *Group Discussion Session 1: Fostering Collaboration in Global FrameNet*

Tiago Torrent

14:50–15:50 *Group Discussion Session 2: Alignment Issues in FrameNet*

Collin F. Baker

15:50–16:10 Coffee Break

16:10–17:00 Closing Session

16:10–17:00 *Planning Meeting*