Contents

Preface by the Workshop Organizers

This volume contains the proceedings of the Second International Workshop on Automatic Translation for Sign and Spoken Languages (AT4SSL 2023)[1], hosted by the 24th Annual Conference of The European Association for Machine Translation (EAMT 2023)[2]. This workshop is a venue for presenting and discussing (complete, ongoing or future) research on automatic translation between sign and spoken languages.

AT4SSL 2021 The first edition of the AT4SSL workshop[3] was co-located with the AMTA conference in 2021. The workshop was conducted online, featured eight long papers, presenting completed work, and three short papers, presenting ongoing work were accepted for presentation, and was attended by approximately 35 participants.

AT4SSL 2023 scope and theme The main theme of the 2023 edition of the AT4SSL workshop is *Sign language parallel data – challenges, solutions and resolutions*: Data is one of the key factors for the success of today's AI, including language and translation models for sign and spoken languages. However, when it comes to processing sign language and training machine-learning systems we face the problems of small volumes of (parallel) data, large veracity in terms of origin of annotations (deaf or hearing interpreters), non-standardized annotations (e.g. glosses differ across corpora), video quality or recording setting, and others. In this edition of the workshop we focus on the discussion of data quantity, data quality, (re)sources, ethical and ownership concerns.

Submissions and programme The workshop welcomed two types of contributions: long and short research papers. AT4SSL 2023 received a total of 9 new submissions (4 long and 5 short papers). Following the peer-review process, 6 submissions were accepted (3 long and 3 short papers), resulting in an acceptance rate of 67% that highlights the quality of the submissions received.

The accepted papers cover a diverse range of topics related to automatic translation between signed and spoken languages, and focus on data resources, linguistics and machine translation

---

[1] https://sites.google.com/tilburguniversity.edu/at4ssl2023/home

[2] https://events.tuni.fi/eamt23/

[3] Dimitar Shterionov, ed. *Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL)*. Virtual: Association for Machine Translation in the Americas, Aug. 2021. URL: https://aclanthology.org/2021.mtsummit-at4ssl.0.

(MT) systems. The papers by McGill, E. and Saggion, H., and by De Sisto, M. et al. present new sign language corpora for BSL, VGT and NGT; the paper by Declerck, T. et al. present a unified RDF-based representation for various type of data aiming to facilitate a common signed and spoken language repository; Moryoseff, A. et al. present in their paper a new baseline MT based on a transformation from text to glosses to poses and to video in the context of translation into SL for DGS. The work presented by the paper of Hollain, N. et al. investigates the use of approximative linguistic features for sign language processing, seeking improvements over landmark-based features; the paper by Mohammed, Z. and Murtagh, I. presents their work on integrating gerunds of Irish SL in a computational lexicon framework.

In this works we feature two keynote speakers: Vincent Vandeghinste (INT, The Netherlands) and Mathias Müller (UZH, Switzerland) who will talk about general challenges related to sign language data and its use within current NLP tools, and the processing of the JWSigning corpus, respectively. In addition, a round table discussion will facilitate an open discussion between the attendees of the workshop centred around the topic: *The gap between MT for spoken and MT for signed languages: data and technology challenges.*

We hope you enjoy reading the papers and we are looking forward to a fruitful and enriching workshop!

June 2023,
D. Shterionov, M. De Sisto, M. Müller, D. Van Landuyt, R. Omardeen, S. Oboyle, A. Braffort, F. Roelofsen, F. Blain, B. Vanroy, E. Avramidis

AT4SSL 2023 Committees

Organising Committee & Workshop Chairs

Dimitar Shterionov, Department Cognitive Science and Artificial Intelligence, School of Humanities and Digital Sciences, Tilburg University, The Netherlands
Mirella De Sisto, Department Cognitive Science and Artificial Intelligence, School of Humanities and Digital Sciences, Tilburg University, The Netherlands
Mathias Müller, Department of Computational Linguistics, University of Zurich, Switzerland
Davy Van Landuyt, European Union of the Deaf, Belgium
Rehana Omardeen, European Union of the Deaf, Belgium
Shaun O'Boyle, Dublin City University, Ireland
Annelies Braffort, LISN, CNRS, Université Paris-Saclay, France
Floris Roelofsen, Institute for Logic, Language, and Computation, University of Amsterdam, The Netherlands
Frédéric Blain, Department Cognitive Science and Artificial Intelligence, School of Humanities and Digital Sciences, Tilburg University, The Netherlands
Bram Vanroy, Faculty of Arts and Philosophy, Ghent University, Faculty of Arts, KU Leuven, Belgium
Eleftherios Avramidis, German Research Center for Artificial Intelligence (DFKI), Germany

Programme Committee

Ioannis Tsochantaridis, Google Research, Switzerland
Amit Moryossef, Department of Computer ScienceBar-Ilan University, Israel
Lyke Esselink, Faculty of ScienceUniversity of Amsterdam, The Netherlands
Jampierre Rocha, Lenovo, Brazil
Mathieu De Coster, IDLab-AIROGhent University, Belgium
Myriam Vermeerbergen, Faculty of Arts, Katholieke Universiteit Leuven, Belgium
Amanda Duarte, Image Processing Group, Signal Theory and Communications DepartmentUniversitat Politècnica de Catalunya, Spain
Silvia Rodríguez Vázquez, Department of Translation Technology, Faculty of Translation and InterpretingUniversity of Geneva, Switzerland
Giacomo Inches, Martel Innovate, Switzerland
Cristina España-Bonet, Universität de Saarlandes, German Research Center for Artificial Intelligence (DFKI), Germany
Ahmet Alp Kindiroglu, Perceptual Intelligence Laboratory (PILAB), Department of Computer Engineering, Bogazici Universitesi, Turkey
Sarah Ebling, Department of Computational Linguistics, University of Zurich, Switzerland

Workshop Program

Time	Activity
09:00	**Start of the workshop**
09:00 – 09:15	**Opening notes**
09:15 – 10:30	**Keynote 1:** *Challenges with Sign Language Datasets* Vincent Vandeghinste (INT) **Keynote 2:** *JWSign: A Highly Multilingual Corpus of Bible Translations for Sign Language Processing* Mathias Müller (UZH)
10:30 – 11:00	Coffee break
11:00 – 12:30	**Presentation Session 1** *Analyzing the Potential of Linguistic Features for Sign Spotting: A Look at Approximative Features* Natalie Hollain, Martha Larson and Floris Roelofsen *A Linked Data Approach for linking and aligning Sign Language and Spoken Language Data* Thierry Declerck, Sam Bigeard, Fahad Khan, Irene Murtagh, Sussi Olsen, Mike Rosner, Ineke Schuurman, Andon Tchechmedjiev and Andy Way *An Open-Source Gloss-Based Baseline for Spoken to Signed Language Translation* Amit Moryossef, Mathias Müller, Anne Göhring, Zifan Jiang, Yoav Goldberg and Sarah Ebling
12:30 – 13:30	Lunch break
13:30 – 15:00	**Round table** *The gap between MT for spoken and MT for signed languages: data and technology challenges* Moderator: Mathias Müller (UZH)
15:00 – 15:30	Coffee break
15:30 – 17:00	**Presentation Session 2** *A New English-Dutch-NGT Corpus for the Hospitality Domain* Mirella De Sisto, Vincent Vandeghinste and Dimitar Shterionov *BSL-Hansard: A parallel, multimodal corpus of English and interpreted British Sign Language data from parliamentary proceedings* Euan McGill and Horacio Saggion *Towards Accommodating Gerunds within the Sign Language Lexicon* Zaid Mohammed and Irene Murtagh
17:00 – 17:15	**Closing remarks**
17:15	**End of the workshop**