

Tilburg University

Preface by the Workshop Organizers

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

Published in:

AT4SSL 2023 - Proceedings of the 2nd International Workshop on Automatic Translation for Signed and Spoken Languages

Publication date:

2023

Document Version

Publisher's PDF, also known as Version of record

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):

Shterionov, D., De Sisto, M., Müller, M., Landuyt, D. V., Omardeen, R., Oboyle, S., Braffort, A., Roelofsen, F., Blain, F., Vanroy, B., & Avramidis, E. (2023). Preface by the Workshop Organizers. *AT4SSL 2023 - Proceedings of the 2nd International Workshop on Automatic Translation for Signed and Spoken Languages*, I-II. <https://aclanthology.org/2023.at4ssl-1.0/>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Contents

Preface by the Workshop Organizers	i
AT4SSL 2023 Committees	iii
Workshop Program	iv
Long Papers	1
Natalie Hollain, Martha Larson and Floris Roelofsen. <i>Analyzing the Potential of Linguistic Features for Sign Spotting: A Look at Approximative Features</i>	3
Thierry Declerck, Sam Bigeard, Fahad Khan, Irene Murtagh, Sussi Olsen, Mike Rosner, Ineke Schuurman, Andon Tchechmedjiev and Andy Way. <i>A Linked Data Approach for linking and aligning Sign Language and Spoken Language Data</i>	13
Amit Moryossef, Mathias Müller, Anne Göhring, Zifan Jiang, Yoav Goldberg and Sarah Ebling. <i>An Open-Source Gloss-Based Baseline for Spoken to Signed Language Translation</i>	25
Short Papers	37
Mirella De Sisto, Vincent Vandeghinste and Dimitar Shterionov. <i>A New English-Dutch-NGT Corpus for the Hospitality Domain</i>	39
Euan McGill and Horacio Saggion. <i>BSL-Hansard: A parallel, multimodal corpus of English and interpreted British Sign Language data from parliamentary proceedings</i>	43
Zaid Mohammed and Irene Murtagh. <i>Towards Accommodating Gerunds within the Sign Language Lexicon</i>	49

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 workshop³ 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/>

³Dimitar 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*.

SignON and EASIER This workshop is organised jointly by members of the SignON (www.signon-project.eu) and EASIER (www.project-easier.eu) projects. Both SignON and EASIER are Horizon 2020 projects, funded under the Horizon 2020 program ICT-57-2020 - “An empowering, inclusive, Next Generation Internet” with Grant Agreement number 101017255 and 101016982 respectively.

We sincerely thank everyone that contributed to this edition of the AT4SSL workshop: the authors of the submitted papers for their interest in the topic; the Programme Committee members for their valuable feedback and insightful comments; the EAMT organizers for their support.

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 Science Bar-Ilan University, Israel

Lyke Esselink, Faculty of Science University of Amsterdam, The Netherlands

Jampierre Rocha, Lenovo, Brazil

Mathieu De Coster, IDLab-AIRO Ghent University, Belgium

Myriam Vermeerbergen, Faculty of Arts, Katholieke Universiteit Leuven, Belgium

Amanda Duarte, Image Processing Group, Signal Theory and Communications Department Universitat Politècnica de Catalunya, Spain

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