

Cover letter

[LREV-ParlaMint-II-cover-letter.pdf](#)

Authors

[Edit](#)

Information	Institution	Funding
Tomaž Erjavec (Primary corresponding author) <tomaz.erjavec@ijs.si>	Jožef Stefan Institute (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, Research infrastructure CLARIN.SI (Jožef Stefan Institute), ARIS (Slovenian Research and Innovation Agency): Knowledge Technologies
Matyáš Kopp (Corresponding author) <kopp@ufal.mff.cuni.cz>	Charles University (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, Ministry of Education, Youth and Sports of the Czech Republic: LINDAT/CLARIAH-CZ, Research infrastructure CLARIN.SI (Jožef Stefan Institute), ARIS (Slovenian Research and Innovation Agency): "Basic national research program 'Digital Humanities' (2022 - 2027)"
Nikola Ljubešić <nikola.ljubasic@ijs.si>	Jožef Stefan Institute (Primary) Institute of Contemporary History	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, Research infrastructure CLARIN.SI (Jožef Stefan

Information	Institution	Funding
Taja Kuzman <taja.kuzman@ijs.si>	Jožef Stefan Institute (Primary)	Institute), ARIS (Slovenian Research and Innovation Agency): "Language Resources and Technologies for Slovene", ParlaSent: ARIS (Slovenian Research and Innovation Agency), "Flemish–Slovenian bilateral basic research project ‘Linguistic landscape of hate speech online’ (2019 – 2023)
Taja Kuzman <taja.kuzman@ijs.si>	Jožef Stefan Institute (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, Research infrastructure CLARIN.SI (Jožef Stefan Institute), ARIS (Slovenian Research and Innovation Agency): "Language Resources and Technologies for Slovene", ParlaSent: ARIS (Slovenian Research and Innovation Agency), "Flemish–Slovenian bilateral basic research project ‘Linguistic landscape of hate speech online’ (2019 – 2023)
Paul Rayson <p.rayson@lancaster.ac.uk>	Lancaster University (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, Oracle Cloud Infrastructure
Petya Osenova (Corresponding author) <petya@bultreebank.org>	Institute of Information and Communication Technologies (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora,

Information	Institution	Funding
Maciej Ogrodniczuk <maciej.ogrodniczuk@ipipan.waw.pl>	Institute of Computer Science (Primary)	Bulgarian Ministry of Education and Science: CLaDA-BG
Çağrı Çöltekin <ccoltekin@sfs.uni-tuebingen.de>	University of Tübingen (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, European Commission programme: European Regional Development Fund as a part of the 2014 - 2020 Smart Growth Operational Programme, Institute of Computer Science, Polish Academy of Sciences - "statutory research", Polish Ministry of Education and Science: "National contribution to CLARIN ERIC - European Research Infrastructure Consortium: Common Language Resources and Technology Infrastructure 2022 - 2023 (CLARIN Q)
Danijel Koržinek <danijel@pja.edu.pl>	Polish-Japanese Academy of Information Technology (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, Polish Ministry of Education and Science: "National contribution to CLARIN ERIC - European Research Infrastructure Consortium: Common Language Resources and Technology Infrastructure 2022 - 2023 (CLARIN Q)

Information	Institution	Funding
Katja Meden <katja.meden@ijs.si>	Jožef Stefan Institute (Primary) Jožef Stefan International Postgraduate School	Resources and Technology Infrastructure 2022 - 2023 (CLARIN Q) CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, Research infrastructure CLARIN.SI (Jožef Stefan Institute), ARIS (Slovenian Research and Innovation Agency): Knowledge Technologies
Jure Skubic <jure.skubic@inz.si>	Institute of Contemporary History (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, Research infrastructure CLARIN.SI (Jožef Stefan Institute)
Peter Rupnik <peter.rupnik@ijs.si>	Jožef Stefan Institute (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, Research infrastructure CLARIN.SI (Jožef Stefan Institute)
Tommaso Agnoloni <agnoloni@igsg.cnr.it>	Institute of Legal Informatics and Judicial Systems (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora
José Aires <jagc@edu.ulisboa.pt>	University of Lisbon (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, Fundação para a Ciência e a

Information	Institution	Funding
		Tecnologia
Starkaður Barkarson <starkadur.barkarson@arnastofnun.is>	Árni Magnússon Institute for Icelandic Studies (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, The Árni Magnússon Institute for Icelandic Studies
Roberto Bartolini <roberto.bartolini@ilc.cnr.it>	Institute for Computational Linguistics "A. Zampolli" (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora
Núria Bel <nuria.bel@upf.edu>	Pompeu Fabra University (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora
María Calzada Pérez <calzada@uji.es>	Jaume I University (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, Spanish Ministry of Science and Innovation: Original, translated and interpreted representations of the refugee crisis: methodological triangulation within corpus-based discourse studies
Roberts Dargis <roberts.dargis@lumii.lv>	University of Latvia (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, The EU Recovery and Resilience Facility project "Language technology Initiative"
Sascha Diwersy <sascha.diwersy@univ-	Université Paul-Valéry Montpellier (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable

Information	Institution	Funding
montp3.fr>		Parliamentary Corpora
Maria Gavriilidou <maria@athenarc.gr>	Institute for Language and Speech Processing (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, Institute for Language and Speech Processing / ATHENA Research Centre
Ruben van Heusden <r.j.vanheusden@uva.nl>	University of Amsterdam (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, Nederlandse Organisatie voor Wetenschappelijk Onderwijs: "Access to City Councils using Exploratory Search Systems"
Mikel Iruskieta <mikel.iruskieta@ehu.eus>	University of the Basque Country (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, HiTZ - Ixa Group (UPV/EHU)
Neeme Kahusk <neeme.kahusk@gmail.com >	University of Tartu (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora
Anna Kryvenko <ganna.kryvenko@inz.si>	Institute of Contemporary History (Primary) The National Institute for Strategic Studies	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, Research infrastructure CLARIN.SI (Jožef Stefan Institute), ARIS (Slovenian Research and Innovation Agency), "The MSCA Seal of Excellence postdoctoral proj 'The Changing Discursive

Information	Institution	Funding
		Semantics of EU Representations' (2022 - 2024) "
Noémi Ligeti-Nagy <ligeti-nagy.noemi@nytud.hunren.hu>	Hungarian Research Centre for Linguistics (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, Hungarian Research Centre for Linguistics
Carmen Magariños <mariadelcarmen.magarinos@usc.gal>	University of Santiago de Compostela (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, Galician Language Institute, University of Santiago de Compostela, Xunta de Galicia - University of Santiago de Compostela: Nós: Galician in the society and economy of artificial intelligence
Martin Mölder <martin.molder@ut.ee>	University of Tartu (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora
Costanza Navarretta <costanza@hum.ku.dk>	University of Copenhagen (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, Department of Nordic Studies and Linguistics (NorS), University of Copenhagen CLARIN-DK
Kiril Simov <kivs@bultreebank.org>	Institute of Information and Communication Technologies (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora,

Information	Institution	Funding
		Bulgarian Ministry of Education and Science: CLaDA-BG
Lars Magne Tunland <lars.tunland@nb.no>	National Library of Norway (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, National Library of Norway
Jouni Tuominen <jouni.tuominen@helsinki.fi >	University of Helsinki (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, The Postimees Foundation grant "Developments and contradictions on the Estonian political landscape 2011 – 2023 through the analysis of big textual data", Research Council of Finland: DIGIHUM research programme project "Semantic Parliament: Linked Open Data Service for Studying Political Culture (SEMPARL)" (2020 – 2022); FIN-CLARIAH
John Vidler <j.vidler@lancaster.ac.uk>	Lancaster University (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, Oracle Cloud Infrastructure
Adina Ioana Vladu <adina.vladu@usc.gal>	University of Santiago de Compostela (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, Galician Language Institute, University of Santiago de Compostela, Xunta de Galicia – University Santiago de Compostela: Nós:

Information	Institution	Funding
		Galician in the society and economy of artificial intelligence
Tanja Wissik <tanja.wissik@oeaw.ac.at>	Austrian Centre for Digital Humanities and Cultural Heritage (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, Austrian Academy of Sciences
Väinö Yrjänäinen <vaino.yrjanainen@statistik.uu.se>	Uppsala University (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora
Darja Fišer <darja.fiser@inz.si>	Institute of Contemporary History (Primary)	CLARIN ERIC, ParlaMint: Towards Comparable Parliamentary Corpora, ARIS (Slovenian Research and Innovation Agency): "Basic national research program 'Digital Humanities' (2022 - 2027)"

Author Contributions Statement

– T. Erjavec wrote Sections 1, 2 (except for 2.3), and most of 5. In the project, he was the co-leader of the work package WP1: Documentation, interoperability, metadata and lead of WP2: Corpus expansion. He performed most of the work in the tasks T1.1: Harmonisation of encoding and T2.3: Data distribution and oversaw the work in T1.3: Adding metadata to existing corpora, T2.1: Adding new corpora and T2.2: Extending existing corpora. – M. Kopp prepared the Figures and Tables in the paper. He was the co-leader of the work package WP1: Documentation, interoperability, metadata, where he was involved in tasks T1.1: Harmonisation of encoding. He performed most of the work in task T1.2: Git management and oversaw tasks T2.1: Adding new corpora and T2.2: Extending existing corpora. He produced the ParlaMint-CZ and ParlaMint-UA corpora and contributed to ParlaMint-ES. – N. Ljubešić wrote Section 4.4. He was the lead of WP3: Corpus enrichment and led task T.3.2: Multimodality. He was involved in the production of the

ParlaMint-BA, ParlaMint-BG, ParlaMint-HR, ParlaMint-SI, and ParlaMint-RS corpora. – T. Kuzman wrote the machine translation part of Section 4.3. She performed the complete MT, as part of T3.1 Machine translation and semantic tagging. – P. Rayson wrote the semantic annotation part of Section 4.3. He led the semantic annotation task in T3.1 Machine translation and semantic tagging. He led the production of the ParlaMint-GB corpus. – P. Osenova wrote the dissemination part of Section 5. She was the co-leader of WP5: Coordination and was central to the tasks T5.1: Management and T5.2: Dissemination. She led the production of the ParlaMint-BG corpus. – M. Ogrodniczuk was the co-leader of WP5: Coordination and was central to the tasks T5.1: Management. He led the production of the ParlaMint-PL corpus. – Ç. Çöltekin was the co-leader of WP4: Engagement activities and had the sole responsibility for T4.3: Shared task. He led the production of the ParlaMint-TR corpus. – D. Koržinek performed a significant part of the work on task T3.2: Multimodality. – K. Meden helped write the Sections 2.2 and 4. In the project, she helped with task T1.1: Harmonisation of encoding and contributed to task T1.3: Adding metadata to existing corpora. She led the production of the ParlaMint-SI corpus. – J. Skubic performed most of the work in T1.3: Adding metadata to existing corpora. – P. Rupnik compiled the ParlaMint-BA, ParlaMint-HR, and ParlaMint-RS corpora and participated in the work on task T3.2: Multimodality. – J. Vidler performed the semantic annotation task in T3.1 Machine translation and semantic tagging. – The other authors wrote the part of the paper that pertains to their corpus and compiled the individual corpora. – and D. Fišer was the co-leader of WP4: Engagement activities, and centrally contributed to T4.1: Tutorial and T4.2: Hackathon. She was also the driving force behind the ParlaMint projects.

Declarations

[Edit](#)

Hybrid

I confirm that I understand journal Language Resources and Evaluation is a hybrid journal. When research is accepted for publication, there is a choice to publish using either immediate gold open access or the traditional publishing route.

Competing Interests

No, I declare that the authors have no competing interests as defined by Springer, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

Dual Publication

The results/data/figures in this manuscript have not been published elsewhere, nor are they under consideration (from you or one of your Contributing Authors) by another publisher.

Authorship

I have read the Nature Portfolio journal policies on author responsibilities and submit this manuscript in accordance with those policies.

Third Party Material

All of the material is owned by the authors and/or no permissions are required.

Data Availability

Yes, I have research data to declare.


“The 4.0 release comprises: – the “plain text” corpora (Erjavec, Kopp, Ogrodniczuk, Osenova, Agirrezabal, et al., 2023): <http://hdl.handle.net/11356/1859>, – the linguistically annotated corpora (Erjavec, Kopp, Ogrodniczuk, Osenova, Agerri, et al.: <http://hdl.handle.net/11356/1860>, 2023), – the corpora machine-translated to English (Kuzman et al., 2023): <http://hdl.handle.net/11356/1864> The ParlaMint encoded 4.1 is ready but the full corpus build and service integration takes several weeks, with possibly more than one iteration. Version 4.1 will be released in time for the revised paper, if accepted.”


Acknowledgements

“The authors are indebted to the following collaborators, who contributed to the compilation of the ParlaMint corpora: Martin Kirnbauer, Hannes Pirker, Daniel Schopper (AT); Michal Mochtak (BA, HR, RS); Griet Depoorter, Henk van der Pol (BE); Jesse de Does (BE, NL); Ilko Grigorov, Vladislava Grigorova (BG); Dorte Haltrup Hansen, Bart Jongejan (DK); Kadri Vider (EE); Monica Albini, Maria del Mar Bonet Ramos, Ruben de Libano (ES); Iván Antiba-Cartazo, Marilina Pisani, Rodolfo Zevallos (ES-CT); Mario Barcala, Daniel Bardanca, Elisa Fernández Rei, Marcos Garcia, Noelia García Díaz, Pedro García Louzao, María Pérez Lago, Xosé Luís Regueira, Marta Vázquez Abuín, Adrián Vidal Miguéns, Ainhoa Vivel Couso (ES-GA); Rodrigo Agerri, Manex Agirrezabal, Jon Alkorta, Ekain Arrieta, Kike Fernandez (ES-PV); Laura Sinikallio, Minna Tamper (FI); Giancarlo Luxardo (FR); Matthew Coole, Amanda Nwadukwe (GB); Dimitris Gkoumas, Vassilis Papavassiliou, Stelios Piperidis, Prokopis Prokopidis (GR); Réka Dodé, Kinga Jelencsik-Mátyus, Zsófia Varga (HU); Roberto Battistoni, Francesca Frontini, Carlo Marchetti, Simonetta Montemagni, Valeri...




Quochi, Manuela Ruisi, Giulia Venturi (IT); Katrien Depuydt, Maarten Marx (NL); Bartłomiej Nitoń, Michał Rudolf (PL); Aida Cardoso, Amália Mendes, Rui Pereira (PT); Johan Jarlbrink, Måns Magnusson, Fredrik Mohammadi Norén (SE); Mihael Ojsteršek, Andrej Pančur (SI); Andriana Rii (UA).”

Share your work as a preprint (optional)

Springer Nature and Research Square are partnering to provide a free preprinting service to all authors. [Learn more about In Review.](#)  (opens in a new window)

- Preprints are assigned a DOI and posted openly online on [Research Square.](#) 
- This process establishes the priority of your findings over competing findings that have not yet been publicly shared.
- Preprints are posted right away and remain permanently available, enabling collaboration, citation, and community feedback.
- Your preprint will show where your manuscript is under consideration and allow others to track its status.
- If your submission is subsequently published, the preprint will link to the final journal article.

By opting in you agree:

- your manuscript and associated personal data will be shared with our partner Research Square for posting as a preprint. For more information on how we use your personal data, please see Springer Nature's [privacy policy.](#) 
- to grant Research Square the right to post your manuscript publicly and permanently as a preprint under a [CC-BY License Agreement 4.0](#) 
- that all authors are aware and have agreed to posting the submission as a preprint.
- to Research Square's [terms of service.](#) 

In Review is an optional service and is not required for your manuscript to be considered by the journal for publication.

- Yes, all authors agree to post this work as a preprint on Research Square
- No, we don't want to post our work as a preprint at this time

Submit manuscript

[← Previous](#)

[Your privacy choices/Manage cookies](#) [Help and support](#) [Privacy policy](#)

[Terms and conditions](#) [Accessibility statement](#)

© 2024 Springer Nature