

Dr. Ing. Vasile Florian PĂIȘ - Lista de lucrări

Lista este disponibilă online la adresele:

- <https://www.racai.ro/publications/>
- <https://scholar.google.com/citations?hl=ro&user=kyBZW7UAAAAJ>

Teză de doctorat

Păiș, Vasile (2019). **Contributions to semantic processing of texts; Identification of entities and relations between textual units; Case study on Romanian language.** Teză de doctorat, Academia Română, București – 2019, limba engleză. Rezumat: https://www.racai.ro/media/Vasile_Pais_TEZA_rezumat_ro-1.docx

Cărți

Păiș, Vasile (editor) (2022). **Speech Recognition Technology and Applications.** Nova Science Publishers, ISBN: 978-1-68507-929-1, <https://doi.org/10.52305/BKWM8996> .

Capitole de carte

Avram, Andrei-Marius and **Păiș, Vasile** and Tufiș, Dan (2022). **Self-supervised pre-training in speech recognition systems.** In Speech Recognition Technology and Applications, Nova Science Publishers, ISBN: 978-1-68507-929-1.

Păiș, Vasile and Avram, Andrei-Marius and Ion, Radu and Tufiș, Dan (2022). **The impact of speech recognition performance on human-computer interaction.** In Speech Recognition Technology and Applications, Nova Science Publishers, ISBN: 978-1-68507-929-1.

Păiș, Vasile (2022). **Punctuation recovery for Romanian transcribed documents.** In Speech Recognition Technology and Applications, Nova Science Publishers, ISBN: 978-1-68507-929-1.

Păiș, Vasile and Barbu-Mititelu, Verginica (2022). **Linguistic linked open data for speech processing.** In Speech Recognition Technology and Applications, Nova Science Publishers, ISBN: 978-1-68507-929-1.

Tuși, Dan and **Păiș, Vasile** and Barbu Mititelu, Verginica and Ion, Radu and Irimia, Elena and Avram, Andrei and Curea, Eric (2022). **MARCELL – A project to remember: hard work of a friendly consortium under wise coordination**. Nyelvtudományi Kutatóközpont, Budapest, ISBN: 978-963-9074-90-3, http://real.mtak.hu/126576/1/VT70_kotet_honlapra.pdf#page=130 .

Păiș, Vasile and Tuși, Dan (2022). **Language Report Romanian**. Chapter in Georg Rehm and Andy Way, editors. **European Language Equality: A Strategic Agenda for Digital Language Equality**. Cognitive Technologies. Springer. ISBN: 978-3-031-28818-0.

Păiș, Vasile and Ion, Radu and Avram, Andrei and Mitrofan, Maria and Tuși, Dan (2022). **RELATE: a modern processing platform for Romanian language**. Chapter in Tools, Methods, and Solutions for the Exploration of Romanian Corpora. Institutul European. ISBN: 978-606-24-0393-9

Mîrzea Vasile, Carmen, Elena Irimia, **Vasile Păiș** (ed.), 2024, Longitudinal learner corpus of Romanian (Long_LECOR), București: Editura Universității din București/Bucharest University Press. (în curs de publicare)

Editor proceedings

Gîfu, Daniela and Trandabăț, Diana and **Păiș, Vasile** and Cristea, Dan and Tuși, Dan (editors) (2018). **The Proceedings of the 13th Edition of the International Conference on Linguistic Resources and Tools for Processing Romanian Language**. Editura Universității "Alexandru Ioan Cuza" Iași, ISSN: 1843-911X.

Jurnale

Păiș, Vasile and Barbu Mititelu, Verginica and Irimia, Elena and Ion, Radu and Tuși, Dan (2024). **Underrepresented Speech Dataset from Open Data: Case Study on the Romanian Language**. Appl. Sci. 2024, 14(19), 9043; <https://doi.org/10.3390/app14199043>

Păiș, Vasile and Ion, Radu and Irimia, Elena and Barbu Mititelu, Verginica and Badea, Valentin and Tuși, Dan (2024). **System for the anonymization of Romanian jurisprudence**. Artificial Intelligence and Law. <https://link.springer.com/article/10.1007/s10506-024-09420-y>

Păiș, Vasile and Mitrofan, Maria and Gasan, Carol Luca and Ianov, Alexandru and Ghiță, Corvin and Coneschi, Vlad Silviu and Onuț, Andrei (2024). **LegalNERo: A linked corpus for named entity recognition in the Romanian legal domain**. SemanticWeb Journal. <https://content.iospress.com/articles/semantic-web/sw233351>

Avram, Andrei-Marius and Mititelu, Verginica Barbu and **Păiș, Vasile** and Cercel, Dumitru-Clementin and Trăușan-Matu, Ștefan (2023). **Multilingual Multiword**

Expression Identification Using Lateral Inhibition and Domain Adaptation. Mathematics. <https://www.mdpi.com/2227-7390/11/11/2548>

Păiș, Vasile and Ion, Radu and Avram, Andrei-Marius and Mitrofan, Maria and Tufiș, Dan (2021). **In-depth evaluation of Romanian natural language processing pipelines.** In Romanian Journal of Information Science and Technology (ROMJIST), vol. 24, no. 4, pp. 384–401, <https://www.romjist.ro/full-texts/paper700.pdf>.

Păiș, Vasile and Tufiș, Dan (2022). **Capitalization and punctuation restoration: a survey.** In: Artificial Intelligence Review 55.3, pp. 1681–1722. URL: <https://doi.org/10.1007/s10462-021-10051-x>.

Avram, Andrei-Marius and **Păiș, Vasile** and Tufiș, Dan (2020). **Towards a Romanian end-to-end automatic speech recognition based on DeepSpeech2.** In: Proceedings of the Romanian Academy Series A, 21, pp. 395–402. URL: https://acad.ro/sectii2002/proceedings/doc2020-4/11-Avram_Tufis.pdf.

Tufiș, Dan and Barbu Mititelu, Verginica and Irimia, Elena and **Păiș, Vasile** and Ion, Radu and Diewald, Nils and Mitrofan, Maria and Mihaela, Onofrei (2019). **Little Strokes Fell Great Oaks. Creating Corola, The Reference Corpus of Contemporary Romanian.** In: Revue Roumaine de Linguistique 64.3 (2019), pp. 227–240.

Păiș, Vasile and Tufiș, Dan (2018). **Computing distributed representations of words using the CoRoLa corpus.** In: Proceedings of the Romanian Academy Series A, 19.2 (2018), pp. 185–191.

Conferințe

Andrei-Marius Avram, Andreea Iuga, George-Vlad Manolache, Vlad-Cristian Matei, Răzvan-Gabriel Miclîuș, Vlad-Andrei Muntean, Manuel-Petru Sorlescu, Drago-Andrei Șerban, Adrian-Dinu Urse, **Vasile Păiș**, Dumitru-Clementin Cercel (2024). **HistNERo: Historical Named Entity Recognition for the Romanian Language.** In International Conference on Document Analysis and Recognition (pp. 126-144). Cham: Springer Nature Switzerland.

Păiș, Vasile, and Niță, Sara, and Jerpelea, Alexandru, and Pană, Luca, and Curea, Eric (2024). **RoMemes: A multimodal meme corpus for the Romanian language.** In Proceedings of the 19th International Conference on Linguistic Resources and Tools for Natural Language Processing (CONSILR 2024).

Radu Ion, Verginica Barbu Mititelu, **Vasile Păiș**, Elena Irimia, and Valentin Badea. (2024). **A Cross-model Study on Learning Romanian Parts of Speech with Transformer Models.** In Proceedings of the Sixth International Conference on Computational Linguistics in Bulgaria (CLIB 2024), pages 6–13, Sofia, Bulgaria. Department of Computational Linguistics, Institute for Bulgarian Language, Bulgarian Academy of Sciences.

Carol-Luca Gasan and Vasile Păis. (2024). **Multi-Model System for Effective Subtitling Compression**. In Proceedings of the 21st International Conference on Spoken Language Translation (IWSLT 2024), pages 57–64, Bangkok, Thailand (in-person and online). Association for Computational Linguistics.

Sara Niță and Vasile Păis. (2024). **RACAI at SemEval-2024 Task 10: Combining algorithms for code-mixed Emotion Recognition in Conversation**. In Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024), pages 1032–1037, Mexico City, Mexico. Association for Computational Linguistics.

Vasile Păis, Dan Tufis, Elena Irimia, and Verginica Barbu Mititelu. (2024). **Building a corpus for the anonymization of Romanian jurisprudence**. In Proceedings of The 18th Linguistic Annotation Workshop (LAW-XVIII), pages 71–76, St. Julians, Malta. Association for Computational Linguistics.

Vasile Păis (2024). **RACAI at ClimateActivism 2024: Improving Detection of Hate Speech by Extending LLM Predictions with Handcrafted Features**. In Proceedings of the 7th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2024), pages 67–72, St. Julians, Malta. Association for Computational Linguistics.

Vasile Păis, V. B. Mititelu, R. Ion and E. Irimia (2023). **Evaluating a Fine-Tuned Whisper Model on Underrepresented Romanian Speech**. 2023 International Conference on Speech Technology and Human-Computer Dialogue (SpeD), Bucharest, Romania, 2023, pp. 141-145, doi: 10.1109/SpeD59241.2023.10314923.

Ana Maria Barbu, Elena Irimia, Carmen Mîrzea Vasile, and Vasile Păis. (2023). **Designing the LECOR Learner Corpus for Romanian**. In Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing, pages 143–152, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Gasan, C. L., Păis, Vasile. (2023). **Investigation of Romanian speech recognition improvement by incorporating Italian speech data**. In Proceedings of the 18th International Conference on Linguistic Resources and Tools for Natural Language Processing (CONSILR 2023), 235.

Barbu Mititelu, Verginica and Irimia, Elena and Păis, Vasile and Avram, Andrei-Marius and Mitrofan, Maria (2022). **Use case: Romanian language resources in the LOD paradigm**. In: Proceedings of the Linked Data in Linguistics Workshop @ LREC2022. European Language Resources Association (ELRA), pp. 35–44. URL: <http://www.lrec-conf.org/proceedings/lrec2022/workshops/LDL/2022.ldl2022-1.0.pdf#page=43> .

Păis, Vasile and Mitrofan, Maria and Barbu Mititelu, Verginica and Irimia, Elena and Micu, Roxana and Gasan, Carol Luca (2022). **Challenges in Creating a Representative Corpus of Romanian Micro-Blogging Text**. In Proceedings of the Workshop on Challenges in the Management of Large Corpora (CMLC-10). European Language Resources Association, Marseille, France, pp. 1--7.

Váradi, Tamás and Nyéki, Bence and Koeva, Svetla and Tadić, Marko and Štefanec, Vanja and Ogrodniczuk, Maciej and Nitoń, Bartłomiej and Pęzik, Piotr and Barbu Mititelu, Verginica and Irimia, Elena and Mitrofan, Maria and **Păiș, Vasile** and Tufiș, Dan and Garabík, Radovan and Krek, Simon and Repar, Andraž (2022). **Introducing the CURLICAT Corpora: Seven-language Domain Specific Annotated Corpora from Curated Sources**. In Proceedings of the Language Resources and Evaluation Conference. European Language Resources Association, Marseille, France, pp. 100--108.

Avram, Andrei-Marius and Catrina, Darius and Cercel, Dumitru-Clementin and Dascalu, Mihai and Rebedea, Traian and **Păiș, Vasile** and Tufiș, Dan (2022). **Distilling the Knowledge of Romanian BERTs Using Multiple Teachers**. In Proceedings of the Language Resources and Evaluation Conference. European Language Resources Association, Marseille, France, pp. 374--384.

Mitrofan, Maria and **Păiș, Vasile** (2022). **Improving Romanian BioNER Using a Biologically Inspired System**. In Proceedings of the 21st Workshop on Biomedical Language Processing. Association for Computational Linguistics, Dublin, Ireland, pp. 316--322.

Costea, Daria and Moldovan, Andrei and **Păiș, Vasile** (2022). **Fine-Grained Legal References in the LegalNERo Corpus**. In Proceedings of the 17th Edition of the International Conference on Linguistic Resources and Tools for Natural Language Processing (CONSILR 2022). Chișinău, Moldova.

Păiș, Vasile (2022). **RACAI at SemEval-2022 Task 11: Complex named entity recognition using a lateral inhibition mechanism**. In Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022). Association for Computational Linguistics, Seattle, United States, pp. 1562--1569.

Avram, Andrei-Marius and Mitrofan, Maria and **Păiș, Vasile** (2022). **Species Entity Recognition Using a Neural Inhibitory Mechanism**. In Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022). A Coruña, Spain.

Micu, Roxana and Gasan, Carol Luca and **Păiș, Vasile** (2022). **Splitting Hashtags in Romanian Micro-Blogging Texts**. In Proceedings of the 17th Edition of the International Conference on Linguistic Resources and Tools for Natural Language Processing (CONSILR 2022). Chișinău, Moldova.

Avram, Andrei-Marius and **Păiș, Vasile** and Tufiș, Dan (2022). **Terminology Annotation Using a String-Matching Algorithm**. In Proceedings of the 17th Edition of the International Conference on Linguistic Resources and Tools for Natural Language Processing (CONSILR 2022). Chișinău, Moldova.

Avram, Andrei-Marius and **Păiș, Vasile** and Mitrofan, Maria (2022). **RACAI@SMM4H'22: Tweets Disease Mention Detection Using a Neural Lateral Inhibitory Mechanism**. In Proceedings of The Seventh Workshop on Social Media Mining for Health Applications, Workshop {&} Shared Task. Association for Computational Linguistics, Gyeongju, Republic of Korea, pp. 1--3 .

Păis, Vasile and Mitrofan, Maria and Avram, Andrei-Marius (2022). **Romanian Language Translation in the RELATE Platform**. In Proceedings of the Fifth Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2022). Association for Computational Linguistics, Gyeongju, Republic of Korea, pp. 64--74.

Păis, Vasile and Barbu Mititelu, Verginica and Irimia, Elena and Mitrofan, Maria and Gasan, Carol Luca and Micu, Roxana (2022). **Romanian micro-blogging named entity recognition including health-related entities**. In Proceedings of The Seventh Workshop on Social Media Mining for Health Applications, Workshop & Shared Task. Association for Computational Linguistics, Gyeongju, Republic of Korea, pp. 190--196.

Păis, Vasile and Mitrofan, Maria (2021). **Assessing multiple word embeddings for named entity recognition of professions and occupations in health-related social media**. In Proceedings of the Sixth Social Media Mining for Health (#SMM4H) Workshop and Shared Task. Association for Computational Linguistics, Mexico City, Mexico, pp. 128--130.

Avram, Andrei-Marius and **Păis, Vasile** and Tufiş, Dan (2021). **A Modular Approach for Romanian-English Speech Translation**. In Natural Language Processing and Information Systems (Métais, Elisabeth and Meziane, Farid and Horacek, Helmut and Kapetanios, Epaminondas). Springer International Publishing, Cham, pp. 57--63.

Păis, Vasile and Mitrofan, Maria (2021). **Towards a named entity recognition system in the Romanian legal domain using a linked open data corpus**. In Workshop on Deep Learning and Neural Approaches for Linguistic Data. Skopje, North Macedonia, pp. 16--17.

Păis, Vasile and Ion, Radu and Avram, Andrei-Marius and Irimia, Elena and Mititelu, Verginica Barbu and Mitrofan, Maria (2021). **Human-Machine Interaction Speech Corpus from the ROBIN project**. In 2021 International Conference on Speech Technology and Human-Computer Dialogue (SpeD). vol. , no. , pp. 91-96.

Pais, Vasile and Mitrofan, Maria and Gasan, Carol Luca and Coneschi, Vlad and Ianov, Alexandru (2021). **Named Entity Recognition in the Romanian Legal Domain**. In Proceedings of the Natural Legal Language Processing Workshop 2021. Association for Computational Linguistics, Punta Cana, Dominican Republic, pp. 9--18.

Avram, Andrei-Marius and **Păis, Vasile** and Tufiş, Dan (2021). **PyEuroVoc: A Tool for Multilingual Legal Document Classification with EuroVoc Descriptors**. In Proceedings of Recent Advances in Natural Language Processing (RANLP2021). pp. 92-101.

Barbu Mititelu, Verginica and Irimia, Elena and **Păis, Vasile** and Mitrofan, Maria and Avram, Andrei-Marius and Curea, Eric (2021). **Linked open data resources for Romanian**. In The 16th International Conference on Linguistic Resources and Tools for Natural Language Processing. pp. 7--19.

Păis, Vasile and Irimia, Elena and Ion, Radu and Tufiş, Dan and Mitrofan, Maria and Barbu Mititelu, Verginica and Avram, Andrei-Marius and Curea, Eric (2021). **Romanian text anonymization experiments from the CURLICAT project**. In The 16th International Conference on Linguistic Resources and Tools for Natural Language Processing. pp. 165--178.

Barbu Mititelu, Verginica and Irimia, Elena and **Păis, Vasile** and Avram, Andrei-Marius and Mitrofan, Maria and Curea, Eric (2020). **Romanian Resources in LLOD format**. In Proceedings of the 15th International Conference Linguistic Resources and Tools for Natural Language Processing (Barbu Mititelu, Verginica and Irimia, Elena and Tufiş, Dan and Dan, Cristea). online, pp. 29-40.

Tufiş, Dan and Mitrofan, Maria and **Păis, Vasile** and Ion, Radu and Coman, Andrei (2020). **Collection and Annotation of the Romanian Legal Corpus**. In Proceedings of The 12th Language Resources and Evaluation Conference. European Language Resources Association, Marseille, France, pp. 2766-2770.

Păis, Vasile and Tufiş, Dan and Ion, Radu (2020). **A Processing Platform Relating Data and Tools for Romanian Language**. In Proceedings of The 12th Language Resources and Evaluation Conference. European Language Resources Association, Marseille, France, pp. 81-88.

Păis, Vasile and Tufiş, Dan and Ion, Radu (2020). **MWSA Task at GlobaLex 2020: RACAI's Word Sense Alignment System using a Similarity Measurement of Dictionary Definitions**. In Proceedings of The 12th Language Resources and Evaluation Conference. European Language Resources Association, Marseille, France, pp. 69-75.

Tamás Váradi, Svetla Koeva, Martin Yamalov, Marko Tadić, Bálint Sass, Bartłomiej Nitoń, Maciej Ogrodniczuk, Piotr Pezik and Barbu Mititelu, Verginica and Radu Ion and Irimia, Elena and Mitrofan, Maria and **Păis, Vasile** and Tufiş, Dan and Radovan Garabík, Simon Krek, Andraz Repar, Matjaž Rihtar and Janez Brank (2020). **The MARCELL Legislative Corpus**. In Proceedings of The 12th Language Resources and Evaluation Conference. European Language Resources Association, Marseille, France, pp. 3754-3761.

Păis, Vasile and Ion, Radu (2020). **TermEval 2020: RACAI's automatic term extraction system**. In Proceedings of the 6th International Workshop on Computational Terminology. European Language Resources Association, Marseille, France, pp. 101-105.

Păis, Vasile (2020). **Multiple annotation pipelines inside the RELATE platform**. In The 15th International Conference on Linguistic Resources and Tools for Natural Language Processing. pp. 65--75.

Avram, Andrei-Marius and **Păis, Vasile** and Tufiş, Dan (2020). **Romanian speech recognition experiments from the ROBIN project**. In The 15th International Conference on Linguistic Resources and Tools for Natural Language Processing. pp. 103--114.

Barbu Mititelu, Verginica and Irimia, Elena and Păis, Vasile and Avram, Andrei-Marius and Mitrofan, Maria and Curea, Eric (2020). **Romanian Resources in Linguistic Linked Open Data format**. In The 15th International Conference on Linguistic Resources and Tools for Natural Language Processing. pp. 29–40.

Ion, Radu and Păis, Vasile and Mitrofan, Maria (2019). **RACAI's System at PharmaCoNER 2019**. In In Proceedings of the 5th Workshop on BioNLP Open Shared Tasks, Association for Computational Linguistics, pp. 90-99.

Păis, Vasile and Tufiş, Dan and Ion, Radu (2019). **Integration of Romanian NLP tools into the RELATE platform**. In International Conference on Linguistic Resources and Tools for Natural Language Processing.

Păis, Vasile and Tufiş, Dan (2018). **More Romanian word embeddings from the RETEROM project**. In Proceedings of the International Conference on Linguistic Resources and Tools for Processing Romanian Language - CONSILR. pp. 91-100.

Boroş, Tiberiu and Dumitrescu, Ştefan Daniel and Păis, Vasile (2018). **Tools and resources for Romanian text-to-speech and speech-to-text applications**. In Proceedings of the International Conference on Human-Computer Interaction (RoCHI), pp. 46–53.

Data
08.01.2025

Semnătura

A handwritten signature in black ink, consisting of a large, stylized initial 'P' followed by a horizontal line extending to the right.