

Dr. Ilia Markov

Curriculum Vitae

(short version, extended version available upon request)



Assistant Professor

Computational Linguistics & Text Mining Lab (CLTL),
Vrije Universiteit Amsterdam

[Google Scholar](#) • [ResearchGate](#) • [LinkedIn](#)

Personal information

- Phone: (+31) 628 54 99 08
- E-mail: i.markov@vu.nl

Work experience

- **Assistant professor**, January 2022 – present
Computational Linguistics & Text Mining Lab (CLTL), Vrije Universiteit Amsterdam,
Amsterdam, The Netherlands
- **Postdoctoral researcher**, June 2019 – December 2021
Centre for Computational Linguistics and Psycholinguistics (CLiPS), University of Antwerp,
Antwerp, Belgium
Project: [“The Linguistic Landscape of Hate Speech in Social Media”](#)
Advisor: [Dr. Walter Daelemans](#)
- **Postdoctoral researcher**, June 2018 – May 2019
Automatic Language Modelling and Analysis & Computational Humanities (ALMAAnaCH)
Team, French Institute for Research in Computer Science and Automation (INRIA), Paris, France
Project: “Automatic Analysis of Citizens’ Debates”
Advisor: [Dr. Eric Villemonte De la Clergerie](#)

Education

- **PhD (with honors), Computer Science**, 2014 – 2018, **GPA 10/10**
Natural Language Processing Laboratory, Center for Computing Research (CIC),
Instituto Politécnico Nacional (IPN), Mexico City, Mexico
PhD thesis: [“Automatic Native Language Identification”](#)
Advisors: [Dr. Grigori Sidorov](#), Dr. Obdulia Pichardo Lagunas
- **MSc, Language Sciences**, 2012 – 2014, **GPA 17/20**
University of Algarve, Faro, Portugal
MSc thesis: [“Automatic Identification of Whole-Part Relations in Portuguese”](#)
Advisors: [Dr. Jorge Baptista](#), [Dr. Nuno Mamede](#)
- **BSc (with honors), Computer Engineering**, 2001 – 2006, **GPA 5/5**
Kaliningrad State Technical University, Russia
BSc thesis: “Automatic Estimation of Stocking Norms”
Advisor: Dr. Olga Toporkova

Research experience (Natural Language Processing, AI)

- Hate speech detection, sentiment analysis, opinion mining
- Information retrieval, information extraction, text mining, argument mining, text similarity
- Native language identification, native language interference, discriminating between similar languages, author profiling, authorship attribution
- Corpus linguistics, lexical resources
- Computational semantics
- Human-computer interaction

Teaching experience

- 2024–2025: Text Mining for AI, XB_0085, Vrije Universiteit Amsterdam (6 ECTS) (upcoming)
- 2024–2025: Text Mining for AI, XB_0141, Vrije Universiteit Amsterdam (6 ECTS) (upcoming)
- 2024–2025: Forensic Linguistics 3: Methods of Authorship Analysis, L_AAMPLIN016, Vrije Universiteit Amsterdam (6 ECTS)
- 2024–2025: Subjectivity Mining, L_AAMPLIN018, Vrije Universiteit Amsterdam (6 ECTS)
- 2023–2024: Text Mining for AI, XB_0085, Vrije Universiteit Amsterdam (6 ECTS)
- 2023–2024: Forensic Linguistics 3: Methods of Authorship Analysis, L_AAMPLIN016, Vrije Universiteit Amsterdam (6 ECTS)
- 2023–2024: Subjectivity Mining, L_AAMPLIN018, Vrije Universiteit Amsterdam (6 ECTS)
- 2022–2023: Text Mining for AI, XB_0085, Vrije Universiteit Amsterdam (6 ECTS)
- 2022–2023: Applied Text Mining 1: Methods, L_PAMATLW004, Vrije Universiteit Amsterdam (6 ECTS)
- 2022–2023: Subjectivity Mining, L_AAMPLIN018, Vrije Universiteit Amsterdam (6 ECTS)

- 2021–2022: Natural Language Processing Technology, L_AAMAALG005, Vrije Universiteit Amsterdam (6 ECTS)
- 2021–2022: Text Mining, L_PABAALG002, Vrije Universiteit Amsterdam (6 ECTS)
- 2021–2022: Applied Text Mining 1: Methods, L_PAMATLW004, Vrije Universiteit Amsterdam (6 ECTS)
- 2021–2022: Natural Language Processing, UA_2007FLWDTA, University of Antwerp (6 ECTS)
- 2020–2021: Natural Language Processing, UA_2007FLWDTA, University of Antwerp (6 ECTS)
- 2019–2020: Computational Linguistics, UA_2010FLWTAA, University of Antwerp (6 ECTS)
- 2017–2018: Natural language processing, Instituto Politécnico Nacional
- 2016–2017: Information retrieval, Instituto Politécnico Nacional

Supervision

- 2023–present: Stefan F. Schouten, PhD thesis “Combining Intuitive and Reflective Inference in Natural Language Understanding Systems”, CLTL, Vrije Universiteit Amsterdam (co-supervisor)
- 2023–present: Wondimagegnhue Tufa, PhD thesis “Language Models for Low-Resource Languages, Dialectics, and Domains”, CLTL, Vrije Universiteit Amsterdam (co-supervisor)
- 2024: master’s thesis “A Comparative Study of Open-Source and Closed-Source Large Language Models for Native Language Identification”, CLTL, Vrije Universiteit Amsterdam
- 2024: master’s thesis “Multi-Label Topic Classification of Client Feedback in the Governance Domain”, CLTL, Vrije Universiteit Amsterdam
- 2024: master’s thesis “Exploring Models’ Performance for Abusive Language Detection: The Case of Obfuscated Abusive Terms”, CLTL, Vrije Universiteit Amsterdam
- 2023: master’s thesis “Exploring Cross-Lingual Language Models for Detecting Toxic Spans: The Case of Dutch”, CLTL, Vrije Universiteit Amsterdam
- 2023: master’s thesis “Automatic Generation of Personalized Counter Narratives Based on User Profile”, CLTL, Vrije Universiteit Amsterdam
- 2023: master’s thesis “Multimodal Hate Speech Detection with CLIP: Identifying Hateful Content in Memes”, CLTL, Vrije Universiteit Amsterdam
- 2023: master’s thesis “From Words to Images: Exploring the Crucial Role of Textual Modality in Multimodal Hate Speech Detection”, CLTL, Vrije Universiteit Amsterdam
- 2023: master’s thesis “Multi-Task Fine-Tuning for Hate Speech Detection”, CLTL, Vrije Universiteit Amsterdam
- 2022: master’s thesis “Evaluating a Transformer-Based Language Model for In-Domain and Cross-Domain Offensive Language Detection on Perturbed Data”, CLTL, Vrije Universiteit Amsterdam

- 2022: master's thesis "Automatic Retrieval of Topics Using Topic Modelling Techniques from Customer Conversations in the Airline Domain", CLTL, Vrije Universiteit Amsterdam
- 2022: master's thesis "Automatic Topic Classification of Customer Feedback in the Banking Domain", CLTL, Vrije Universiteit Amsterdam
- 2022: master's thesis "An Evaluation of Linguistic Features for Automatic Detection of Socially Unacceptable Discourse on Social Media", CLiPS, University of Antwerp
- 2022: master's thesis "The Effects of Posterior Cortical Atrophy on Written Language", CLiPS, University of Antwerp
- 2021: master's thesis "Linguistic Changes in Terry Pratchett's Writing as an Indication of Posterior Cortical Atrophy", CLiPS, University of Antwerp
- 2021: master's thesis "Context in Online Hate Speech: Error Analysis of the Performance of BERTje", CLiPS, University of Antwerp
- 2021: master's thesis "Detecting Vaccination Hesitancy in Online Comments", CLiPS, University of Antwerp
- 2021: master's thesis "A Close-up on BERT: Performing a Qualitative Error Analysis for Sarcasm Detection in Dutch Headlines", CLiPS, University of Antwerp
- 2020: master's thesis "Cross-Lingual Age Detection on English and German Tweets: Evaluating the Impact of Individual Parts of Speech", CLiPS, University of Antwerp
- 2018: BSc thesis "Syllables as Features for the Authorship Attribution Task", ESIME, Instituto Politécnico Nacional (IPN)

Keynotes, invited talks, PhD committee

- External PhD committee member. Jens Van Nooten, PhD thesis "Multi-label Text Classification", CLiPS, University of Antwerp, Belgium (upcoming).
- External PhD committee member. Jens Lemmens, PhD thesis "Sweeping the Social Minefield: Detecting Online Opinions and Emotions with NLP", CLiPS, University of Antwerp, Belgium, May 31, 2024.
- Keynote: "Combating Multilingual Hate Speech Online: Challenges & Research Directions". *Third International conference on Speech & Language Technologies for Low-resource Languages (SPELLL-2024)*, Chennai, India (virtual), 04–06 December, 2024 (upcoming).
- Invited talk: "An Interdisciplinary Perspective on Countering Harmful Multimodal Content". *Workshop on The Interdisciplinary Approaches to the Study of Safety and Danger in the Digital World*, The Hague, The Netherlands, 27–28 June, 2024.

- Invited talk: “Profiling Hate Speech Spreaders & Personalizing Counter Narratives against Hate Speech”. *Conference on Harmful Online Communication (CHOC 2023)*, Cologne, Germany, November 16–17, 2023.
- Invited talk: “Hate Speech Detection: Challenges & Research Directions”. *Huawei’s Cloud Service Helsinki NLP/NLU Workshop 2022*, Helsinki, Finland, December 16, 2022.
- Invited talk: “A Multilingual Perspective on Hate Speech Detection in Social Media”. *IPHSE-2: Interdisciplinary Perspectives on Hate Speech 2021*, Hildesheim, Germany, November 11–12, 2021.

Academic service

- Editorial board member: Computational Linguistics in the Netherlands Journal
- Senior area chair: The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024) – Offensive and Harmful Language Detection and Analysis Track

Publications

Journals

1. L. Hilte, I. Markov, N. Ljubešić, D. Fišer, W. Daelemans. Who are the Haters? A Corpus-Based Demographic Analysis of Authors of Hate Speech. *Frontiers in Artificial Intelligence*, vol. 6, 2023. ([DOI](#))
JCR impact factor: 0.264
2. I. Gevers, I. Markov, W. Daelemans. Linguistic Analysis of Toxic Language on Social Media. *Computational Linguistics in the Netherlands Journal*, vol. 12, pp. 33–48, 2022. ([PDF](#))
JCR impact factor: 0.47
3. J. Lemmens, T. Dejaeghere, T. Kreutz, J. Van Nooten, I. Markov, W. Daelemans. Vaccinpraat: Monitoring Vaccine Skepticism in Dutch Twitter and Facebook Comments. *Computational Linguistics in the Netherlands Journal*, vol. 11, pp. 173–188, 2022. ([PDF](#))
JCR impact factor: 0.47
4. J. Van Nooten, I. Markov, W. Daelemans. Evaluating the Impact of Word Classes on Cross-Domain Age Detection Models' Performance. *Computational Linguistics in the Netherlands Journal*, vol. 11, pp. 71–84, 2022. ([PDF](#))
JCR impact factor: 0.47

5. I. Markov, V. Nastase, C. Strapparava. Exploiting Native Language Interference for Native Language Identification. *Natural Language Engineering*, pp. 1–31, 2020. ([DOI](#))
JCR impact factor: 1.465
6. H. Gómez, R. Fuentes, I. Markov, G. Sidorov, A. Gelbukh. A Convolutional Neural Network Approach for Gender and Language Variety Identification. *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 5, pp. 4845–4855, 2019. ([DOI](#))
JCR impact factor: 1.426
7. G. Sidorov, I. Markov, O. Kolesnikova, L. Chanona. Human Interaction with Shopping Assistant Robot in Natural Language. *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 5, pp. 4889–4899, 2019. ([DOI](#))
JCR impact factor: 1.426
8. I. Markov, J. Baptista, O. Pichardo. Authorship Attribution in Portuguese Using Character N-grams. *Acta Polytechnica Hungarica*, vol. 14, no. 3, pp. 59–78, 2017. ([DOI](#), [PDF](#))
JCR impact factor: 0.745
9. G. Sidorov, M. Ibarra, I. Markov, R. Guzman, L. Chanona, F. Velásquez. Measuring Similarity Between Karel Programs Using Character and Word N-grams. *Programming and Computer Software*, vol. 43, no. 1, pp. 47–50, 2017. ([DOI](#), [PDF](#))
JCR impact factor: 0.230
10. H. Gómez, I. Markov, G. Sidorov, J.-P. Posadas, M. Sanchez, L. Chanona. Improving Feature Representation Based on a Neural Network for Author Profiling in Social Media Texts. *Computational Intelligence and Neuroscience*, vol. 2016, 13 pages, 2016. ([DOI](#))
JCR impact factor: 1.215
11. G. Sidorov, M. Ibarra, I. Markov, R. Guzman, L. Chanona, F. Velásquez. Automatic Detection of Similarity of Programs in Karel Programming Language based on Natural Language Processing Techniques. *Computación y Sistemas*, vol. 20, no. 2, pp. 279–288, 2016. ([DOI](#), [PDF](#))
Scopus Q2
12. H. Gómez, I. Markov, G. Sidorov, J.-P. Posadas, C. Fócil. Compiling a Lexicon of Social Media for the Author Profiling Task. *Research in Computing Science*, vol. 115, pp. 19–27, 2016. ([PDF](#))
DBLP
13. I. Markov, N. Mamede, J. Baptista. A Rule-Based Meronymy Extraction Module for Portuguese. *Computación y Sistemas*, vol. 19, no. 4, pp. 661–683, 2015. ([DOI](#), [PDF](#))
Scopus Q2

Conferences, workshops, and chapters in books

14. S. B. Santamaria, H. Gómez, I. Markov. Contextualized Graph Representations for Generating Counter-Narrative against Hate Speech. In: *Findings of the Association for Computational Linguistics: EMNLP 2024*, Miami, Florida, USA. ACL, pp. 7664–7674, November 12–16, 2024. ([PDF](#))
15. Y. Wang, I. Markov. CLTL at DIMEMEX Shared Task: Fine-Grained Detection of Hate Speech in Memes. In: *Iberian Languages Evaluation Forum (IberLEF 2024)*, colocated with the *Conference of the Spanish Society for Natural Language Processing (SEPLN 2024)*, Valladolid, Spain. CEUR-WS.org, vol. 3756, September 24, 2024. ([PDF](#))
Ranked 1st in the DIMEMEX 2024 Shared Task on Fine-Grained Detection of Hate Speech in Memes.
16. Y. Wang, I. Markov. CLTL at AraIEval Shared Task: Multimodal Propagandistic Memes Classification Using Transformer Models. In: *The Second Arabic Natural Language Processing Conference (ArabicNLP-WS 2024)*, Bangkok, Thailand, ACL, pp. 501–506, August 16, 2024. ([PDF](#))
Ranked 1st, 2nd, and 3rd per subtask in the AraIEval 2024 Shared Task on Multimodal Propagandistic Memes Classification.
17. W. T. Tufa, I. Markov, P. Vossen. Grounding Toxicity in Real-World Events Across Languages. In: *29th International Conference on Natural Language & Information Systems (NLDB 2024)*, Turin, Italy, LNCS, Springer, Cham, vol. 14762, pp. 197–210, June 25–27, 2024. ([DOI](#))
18. W. T. Tufa, I. Markov, P. Vossen. Unknown Script: Impact of Script on Cross-Lingual Transfer. In: *2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2024, Volume 4: Student Research Workshop)*, Mexico City, Mexico, ACL, pp. 124–129, June 16–21, 2024. ([DOI](#))
19. Y. Wang, I. Markov. CLTL@HarmPot-ID: Leveraging Transformer Models for Detecting Offline Harm Potential and Its Targets in Low-Resource Languages. In: *Fourth Workshop on Threat, Aggression & Cyberbullying @ LREC-COLING-2024 (TRAC 2024)*, Turin, Italy, ELRA and ICCL, pp. 21–26, May 20, 2024. ([PDF](#))
Ranked 1st in the TRAC 2024 Shared Task on Offline Harm Potential Identification.
20. W. T. Tufa, I. Markov, P. Vossen. The Constant in HATE: Toxicity in Reddit across Topics and Languages. In: *Fourth Workshop on Threat, Aggression & Cyberbullying @ LREC-COLING-2024 (TRAC 2024)*, Turin, Italy, ELRA and ICCL, pp. 1–11, May 20, 2024. ([PDF](#))

21. Y. Wang, I. Markov. CLTL@Multimodal Hate Speech Event Detection 2024: The Winning Approach to Detecting Multimodal Hate Speech and Its Targets. In: *7th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2024)*, St. Julians, Malta, ACL, pp. 73–78, March 21–22, 2024. ([PDF](#))
Ranked 1st in the Multimodal Hate Speech Event Detection shared task 2024.
22. S. F. Schouten, P. Bloem, I. Markov, P. Vossen. Reasoning about Ambiguous Definite Descriptions. In: *Findings of the Association for Computational Linguistics: EMNLP 2023*, Singapore, ACL, pp. 4479–4484, December 6–10, 2023. ([DOI](#))
23. M. Doğanç, I. Markov. From Generic to Personalized: Investigating Strategies for Generating Targeted Counter Narratives against Hate Speech. In: *1st Workshop on CounterSpeech for Online Abuse (CS4OA 2023)*, Prague, Czech Republic, ACL, pp. 1–12, September 11, 2023. ([PDF](#))
24. N. Yuzbashyan, N. Banar, I. Markov, W. Daelemans. An Exploration of Zero-Shot Natural Language Inference-Based Hate Speech Detection. In: *Third Workshop on Language Technology for Equality, Diversity and Inclusion (LT-EDI 2023)*, Varna, Bulgaria, ACL, pp. 1–9, September 7, 2023. ([PDF](#))
25. S. F. Schouten, B. Barbarestani, W. Tufa, P. Vossen, I. Markov. Cross-Domain Toxic Spans Detection. In: *28th International Conference on Natural Language & Information Systems (NLDB 2023)*, Derby, United Kingdom, LNCS, Springer, vol. 13913, pp. 533–545, June 21–23, 2023. ([DOI](#))
26. J. Lemmens, I. Markov, W. Daelemans. The LiLaH Emotion Lexicon of Greek, Kurdish, Turkish, Spanish, Farsi and Chinese. *CLiPS Technical Report Series*, CTRS-009, 2023. ([PDF](#))
27. I. Markov, W. Daelemans. The Role of Context in Detecting the Target of Hate Speech. In: *Third Workshop on Threat, Aggression and Cyberbullying (TRAC 2022)*, Gyeongju, Republic of Korea, ACL, pp. 37–42, October 17, 2022. ([PDF](#))
28. I. Markov, I. Gevers, W. Daelemans. An Ensemble Approach for Dutch Cross-Domain Hate Speech Detection. In: *27th International Conference on Natural Language & Information Systems (NLDB 2022)*, Valencia, Spain, LNCS, Springer, vol. 13286, pp. 3–15, June 15–17, 2022. ([DOI](#))
29. M. Kestemont, E. Manjavacas, I. Markov, J. Bevendorff, M. Wiegmann, E. Stamatatos, B. Stein, M. Potthast. Overview of the Cross-Domain Authorship Verification Task at PAN 2021. In: *Working Notes of CLEF 2021 – Conference and Labs of the Evaluation Forum*, Bucharest, Romania. CEUR, vol. 2936, pp. 1743–1759, September 21–24, 2021. ([PDF](#))

30. J. Lemmens, I. Markov, W. Daelemans. Improving Hate Speech Type and Target Detection with Hateful Metaphor Features. In: *Fourth Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda (NLP4IF 2021)*, Online, ACL, pp. 7–16, June 6, 2021. ([PDF](#))
31. I. Markov, W. Daelemans. Improving Cross-Domain Hate Speech Detection by Reducing the False Positive Rate. In: *Fourth Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda (NLP4IF 2021)*, Online, ACL, pp. 17–22, June 6, 2021. ([PDF](#))
32. I. Markov, N. Ljubešić, D. Fišer, W. Daelemans. Exploring Stylometric and Emotion-Based Features for Multilingual Cross-Domain Hate Speech Detection. In: *Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2021)*, Online, ACL, pp. 149–159, April 19, 2021. ([PDF](#))
33. J. Bevendorff, B. Chulvi, G. L. De La Peña Sarracén, M. Kestemont, E. Manjavacas, I. Markov, M. Mayerl, M. Potthast, F. Rangel, P. Rosso, E. Stamatatos, B. Stein, M. Wiegmann, M. Wolska, E. Zangerle. Overview of PAN 2021: Authorship Verification, Profiling Hate Speech Spreaders on Twitter, and Style Change Detection. In: *43rd European Conference on Information Retrieval (ECIR 2021)*, Online, LNCS, Springer, vol. 12657, pp. 567–573, March 28 – April 1, 2021. ([DOI](#))
34. N. Ljubešić, I. Markov, D. Fišer, W. Daelemans. The LiLaH Emotion Lexicon of Croatian, Dutch and Slovene. In: *Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media (PEOPLES 2020)*, Barcelona, Spain (Online), ACL, pp. 153–157, December 13, 2020. ([PDF](#))
35. E. Lotfi, I. Markov, W. Daelemans. A Deep Generative Approach to Native Language Identification. In: *28th International Conference on Computational Linguistics (COLING 2020)*, Barcelona, Spain (Online), International Committee on Computational Linguistics, pp. 1778–1783, December 8–13, 2020. ([DOI](#))
36. J. Bevendorff, B. Ghanem, A. Giachanou, M. Kestemont, E. Manjavacas, I. Markov, M. Mayerl, M. Potthast, F. Rangel, P. Rosso, G. Specht, E. Stamatatos, B. Stein, M. Wiegmann, E. Zangerle. Overview of PAN 2020: Authorship Verification, Celebrity Profiling, Profiling Fake News Spreaders on Twitter, and Style Change Detection. In: *Experimental IR Meets Multilinguality, Multimodality, and Interaction (CLEF 2020)*, Thessaloniki, Greece. LNCS, Springer, vol. 12260, pp. 372–383, September 22–25, 2020. ([DOI](#))
37. M. Kestemont, E. Manjavacas, I. Markov, J. Bevendorff, M. Wiegmann, E. Stamatatos, M. Potthast, B. Stein. Overview of the Cross-Domain Authorship Verification Task at PAN 2020. In: *Working Notes of CLEF 2020 – Conference and Labs of the Evaluation Forum*, Thessaloniki, Greece. CEUR, vol. 2696, September 22–25, 2020. ([PDF](#))

38. J. Lemmens, B. Burtenshaw, E. Lotfi, I. Markov, W. Daelemans. Sarcasm Detection Using an Ensemble Approach. In: *Second Workshop on Figurative Language Processing (FigLang 2020)*, Online, ACL, pp. 264–269, July 9, 2020. ([DOI](#))
39. I. Markov, V. Nastase, C. Strapparava. Anglicized Words and Misspelled Cognates in Native Language Identification. In: *14th Workshop on Innovative Use of NLP for Building Educational Applications (BEA14 2019)*, Florence, Italy. ACL, pp. 275–284, August 2, 2019. ([PDF](#))
40. I. Markov, E. De la Clergerie. INRIA at SemEval-2019 Task 9: Suggestion Mining Using SVM with Handcrafted Features. In: *13th International Workshop on Semantic Evaluation (SemEval-2019)*, Minneapolis, Minnesota, USA. ACL, pp. 1204–1207, June 6–7, 2019. ([PDF](#))
41. I. Markov, G. Sidorov. CIC-IPN@INLI2018: Indian Native Language Identification. In: *Working Notes of FIRE 2018 – 10th International Forum for Information Retrieval Evaluation*, Gandhinagar, India. CEUR-WS.org, vol. 2266, pp. 82–88, December 06–09, 2018. ([PDF](#))
42. I. Markov, V. Nastase, C. Strapparava, G. Sidorov. The Role of Emotions in Native Language Identification. In: *9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2018)*, Brussels, Belgium. ACL, pp. 123–129, October 31, 2018. ([DOI](#), [PDF](#))
43. I. Markov, H. Gómez, M. Jasso-Rosales, G. Sidorov. CIC-GIL Approach to Author Profiling in Spanish Tweets: Location and Occupation. In: *Third Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2018)*, Seville, Spain. CEUR-WS.org, vol. 2150, pp. 97–101, September 18, 2018. ([PDF](#))
44. I. Markov, V. Nastase, C. Strapparava. Punctuation as Native Language Interference. In: *27th International Conference on Computational Linguistics (COLING 2018)*, Santa Fe, New Mexico, USA. The COLING 2018 Organizing Committee, pp. 3456–3466, August 20–26, 2018. ([PDF](#))
45. I. Markov, H. Gómez, G. Sidorov, A. Gelbukh. The Winning Approach to Cross-Genre Gender Identification in Russian at RUSProfiling 2017. In: *Working Notes of FIRE 2017 – 9th International Forum for Information Retrieval Evaluation*, Bangalore, India. CEUR-WS.org, vol. 2036, pp. 20–24, December 08–10, 2017. ([PDF](#))
- Ranked 1st in the PAN RUSProfiling shared task 2017.**

46. I. Markov, L. Chen, C. Strapparava, G. Sidorov. CIC-FBK Approach to Native Language Identification. In: *12th Workshop on Innovative Use of NLP for Building Educational Applications (BEA12 2017)*, Copenhagen, Denmark. ACL, pp. 374–381, September 8, 2017. ([PDF](#))

Ranked 1st in the NLI shared task 2017

47. M. Sanchez, I. Markov, H. Gómez, G. Sidorov. Comparison of Character n-grams and Lexical Features on Author, Gender, and Language Variety Identification on the Same Spanish News Corpus. In: *Experimental IR Meets Multilinguality, Multimodality, and Interaction (CLEF 2017)*, Dublin, Ireland. LNCS, Springer, vol. 10456, pp. 145–151, September 11–14, 2017. ([DOI](#))

48. I. Markov, H. Gómez, G. Sidorov. Language- and Subtask-Dependent Feature Selection and Classifier Parameter Tuning for Author Profiling. In: *Working Notes of CLEF 2017 – Conference and Labs of the Evaluation Forum*, Dublin, Ireland. CEUR, vol. 1866, September 11–14, 2017. ([PDF](#))

49. I. Markov, E. Stamatatos, G. Sidorov. Improving Cross-Topic Authorship Attribution: The Role of Pre-Processing. In: *18th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2017)*, Budapest, Hungary. Springer, vol. 10762, pp. 289–302, April 17–23, 2017. ([DOI](#), [PDF](#))

Best poster award, third place

50. H. Gómez, I. Markov, J. Baptista, G. Sidorov, D. Pinto. Discriminating between Similar Languages Using a Combination of Typed and Untyped Character N-grams and Words. In: *4th Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2017)*, Valencia, Spain. ACL, pp. 137–145, April 3, 2017. ([PDF](#))

51. I. Markov, H. Gómez, J.-P. Posadas, G. Sidorov, A. Gelbukh. Author Profiling with Doc2vec Neural Network-Based Document Embeddings. In: *15th Mexican International Conference on Artificial Intelligence (MICAI 2016)*, Cancún, Mexico. Part II, LNAI, Springer, vol. 10062, pp. 117–131, October 23–29, 2016. ([DOI](#), [PDF](#))

Best paper award, first place

52. I. Markov, H. Gómez, G. Sidorov, A. Gelbukh. Adapting Cross-Genre Author Profiling to Language and Corpus. In: *Working Notes of CLEF 2016 – Conference and Labs of the Evaluation Forum*, Évora, Portugal. CEUR, vol. 1609, pp. 947–955, September 5–8, 2016. ([PDF](#))

53. J. Baptista, I. Markov. Morphosyntactic processes involving body-part nouns in Portuguese. In: *Perspectives Harrissiennes*. CRL - Cellule de Recherche en Linguistique, pp. 255–267, 2016. ([PDF](#))

54. G. Sidorov, H. Gómez, I. Markov, D. Pinto, N. Loya. Computing Text Similarity using Tree Edit Distance. In: *Annual Conference of the North American Fuzzy Information Processing Society (NAFIPS)*, joint with 2015 5th World Conference on Soft Computing (WConSC), Redmond, WA, USA. IEEE, pp. 1–4, August 17–19, 2015. ([DOI](#), [PDF](#))
55. H. Gómez, G. Sidorov, D. Pinto, I. Markov. A Graph Based Authorship Identification Approach. In: *Working Notes of CLEF 2015 – Conference and Labs of the Evaluation Forum*, Toulouse, France. CEUR, vol. 1391, September 8–11, 2015. ([PDF](#))
56. J.-P. Posadas, I. Markov, H. Gómez, G. Sidorov, I. Batyrshin, A. Gelbukh, O. Pichardo. Syntactic N-grams as Features for the Author Profiling Task. In: *Working Notes of CLEF 2015 – Conference and Labs of the Evaluation Forum*, Toulouse, France. CEUR, vol. 1391, September 8–11, 2015. ([PDF](#))
57. I. Markov, N. Mamede, J. Baptista. Whole-Part Relations Rule-Based Automatic Identification: Issues from Fine-Grained Error Analysis. In: *13th Mexican International Conference on Artificial Intelligence (MICA I 2014)*, Tuxtla Gutiérrez, Mexico. Springer, vol. 8856, pp. 37–50, November 16–22, 2014. ([DOI](#), [PDF](#))
58. I. Markov, N. Mamede, J. Baptista. Automatic Identification of Whole-Part Relations in Portuguese. In: *3rd Symposium on Languages, Applications and Technologies (SLATE 2014)*, Bragança, Portugal. Dagstuhl Publishing, vol. 38, pp. 225–232, June 19–20, 2014. ([DOI](#), [PDF](#))
59. J. Baptista, N. Mamede, I. Markov. Integrating Verbal Idioms into an NLP System. In: *11th International Conference on the Computational Processing of the Portuguese Language (PROPOR 2014)*, São Carlos, SP, Brazil. Springer, vol. 8775, pp. 250–255, October 6–9, 2014. ([DOI](#), [PDF](#))
60. I. Markov, N. Mamede, J. Baptista. Body-Part Nouns and Whole-Part Relations. In: *11th International Conference on the Computational Processing of the Portuguese Language (PROPOR 2014)*, São Carlos, SP, Brazil. Springer, vol. 8775, pp. 125–136, October 6–9, 2014. ([DOI](#), [PDF](#))
61. J. Baptista, N. Mamede, I. Markov. Integrating a Lexicon-Grammar of Verbal Idioms in a Portuguese NLP System. In: *2nd PARSEME General Meeting*, Athens, Greece, March 10–11, 2014.