

# Nikola Ljubešić – Curriculum Vitae

## Personal information

Name: Nikola Ljubešić

Birth date: March 26 1979

Nationality: Croatian

## Research and education

**2023** Senior research associate at Jožef Stefan Institute, Ljubljana

**8/2022** Visiting researcher at University of Antwerp

**2019** Research associate at Faculty of Computer and Information Science, University of Ljubljana

**2017** Research associate at Jožef Stefan Institute, Ljubljana

**5/2017 10/2017** Visiting scholar at University of Groningen

**11/2016** Visiting scholar at Lexical Computing

**6/2016** Visiting scholar at University of Zürich

**6/2015** Visiting scholar at University of Zürich

**2014-2017** Postdoctoral researcher at Jožef Stefan Institute, Ljubljana

**3/2014-6/2014** Visiting scholar at Jožef Stefan Institute, Ljubljana

**2004-2009** University of Zagreb, Faculty of Humanities and Social Sciences,  
PhD program in Information science

PhD thesis: Event detection from parallel information sources

PhD advisor: Damir Boras, full professor

**1/2007-6/2007** Visiting scholar at University of Pennsylvania, USA, Department of linguistics and Department of computer science

**11/2004-12/2004** Visiting student at City University, London, Department of information science

**1998-2003** University of Zagreb, Faculty of Philosophy,  
MA in Information science

MA in German language and literature

## Employment

**2022-** Additional employment at Institute for Contemporary History

**2019-** Shared-time researcher between Jožef Stefan Institute and University of Ljubljana

**2018-2019** Full-time researcher at Jožef Stefan Institute, Ljubljana

**2016-2017** Postdoctoral researcher at Jožef Stefan Insitute, Ljubljana

**2014-2015** Part-time postdoctoral researcher at Jožef Stefan Institute, Ljubljana

**2013-2016** Four-month secondment at Prompsit Language Engineering, Elche, Spain

**2011-** Assistant professor at Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb

**2009-2011** Postdoctoral researcher at Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb

**2007-2010** Lecturer at VERN' polytechnic

**2004-2006** Lecturer at na Technical polytechnic in Zagreb

**2004-2009** Research assisitant at Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb

## Active projects

### Scientific

**MEZZANINE: Basic Research for the Development of Spoken Language Resources and Speech Technologies for the Slovenian Language.** Slovene national project (2022-2025)

**ParlaMint: Towards Comparable Parliamentary Corpora.** CLARIN ERIC projects (2021-2023)

**MaCoCu: Massive Collection and Curation of Monolingual and Bilingual Data.** EU CEF project (2021-2023)

**The linguistic landscape of hate speech on social media.** Flemish-Slovene bilateral project (2019-2022)

## **Industrial**

**Lifesome** start-up – consultancy in the area of machine / deep learning for trait prediction from genomic data.

**Realnetworks** - consultancy in the area of deep learning for cross-lingual text classification.

## **Previous projects**

### **International**

**IMSyPP: Innovative Monitoring Systems and Prevention Policies of Online Hate Speech** EU REC AG project (2020-2022)

**ELEXIS: European Lexicographic Infrastructure** Horizon2020 (2018-2022)

**Automatic Building of Machine Translation (Abu-MaTran)** FP7-IAPP project (2013-2016), principal investigator.

**Regional linguistic data initiative (ReLDI)** Project funded by the Swiss National Science Foundation (2015-2017), principal investigator.

**Croatian-Slovene bilateral project** on semantic role labeling for Croatian and Slovene (2016-2017)

**CESAR** – Central and South-East European Resources (2011-2013)

**LetsMT!** – Platform for Online Sharing of Training Data and Building User Tailored Machine Translation (LetsMT!) (2010-2012)

**ACCURAT** – Analysis and evaluation of Comparable Corpora for UnderResourced Areas of machine Translation (2010-2012)

### **National**

**FRENK: Resources, methods, and tools for the understanding, identification, and classification of various forms of socially unacceptable discourse.** Slovene national project (2017-2020)

**MEGAHR: Modelling the mental grammar of Croatian.** Croatian national project (2017-2020)

**KOLOS: Collocations in Slovene.** Slovene national project (2018-2021)

**KAS: Slovene scientific texts: resources and description.** Slovene national project. (2016-2019)

**Computer assistant for facilitating text entry of people with language disabilities. (RAPUT)** Project funded through EU structural funds. (2014-2016), work package leader.

**Resources, Tools and Methods for the Research of Nonstandard Internet Slovene (JANES).** Slovene national project. (2014-2017)

## **Industrial**

**PropName** start-up – CTO.

**VizPren** start-up – consultancy in natural language processing.

**Ciklopea d.o.o.** - Implementation of a system for statistical machine translation as part of a computer-assisted translation environment.

**Neos d.o.o.** - Development of a system for opinion mining.

**KONČAR d.d.** – Concept project for knowledge management (2012)

**Neos d.o.o.** – Content extraction from the web, language identification and web data deduplication (2012)

**Imenik d.o.o.** – Development of a stemmer for information retrieval in Croatian (2010)

**Zavod za poslovna istraživanja d.o.o.** – Development of an algorithm for event detection (2008-2009)

**Zavod za poslovna istraživanja d.o.o.** – Development of a morphological lexicon of business subject names (2007)

## **Program, organization and editorial boards, reviews**

Reviewer for COLING, ACL, EACL, NAACL, EMNLP, IJCAI, IJCNLP, IJLRE, JNLE, Computational linguistics

2023 Organization board member of the Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial) at EACL, Dubrovnik, Croatia

2022 Organization board member of the Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial) at COLING, Gyeongju, Republic of Korea.

2021 Organization board member of the Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial) and the DSL7 shared task at EACL, Kiev, Ukraine

2020 Organization board member of the WMT shared task on similar languages translation, at WMT, Barcelona

2020 Organization board member of the SemEval shared task on Graded Word Similarity in Context, at SemEval, Barcelona

2020 Organization board member of the Workshop on NLP for Similar Languages, Varieties and Dialects(VarDial) and the DSL6 shared task at COLING, Barcelona

2019 Organization board member of the Workshop on NLP for Similar Languages, Varieties and Dialects(VarDial) and the DSL6 shared task at NAACL, Minneapolis

2018 Organization board member of the Workshop on NLP for Similar Languages, Varieties and Dialects(VarDial) and the DSL5 shared task at COLING, Santa Fe

2018 Area chair at COLING, Santa Fe

2017 Organization board member of the Workshop on NLP for Similar Languages, Varieties and Dialects(VarDial) and the DSL4 shared task at EACL, Valencia

2016 Organization board member of the Workshop on NLP for Similar Languages, Varieties and Dialects(VarDial) and the DSL3 shared task at COLING, Osaka

2015 Organization board member of the Joint Workshop on Language Technology for Closely Related Languages, Varieties and Dialects (LT4VarDial) and the DSL2 shared task at RANLP, Hissar

2014 Organization board member of the VARDIAL workshop and the DSL shared task at COLING, Dublin

## **Research interests**

- 1 linguistic processing of South Slavic languages
- 2 social media analytics
- 3 spatial and temporal variation of language
- 4 computational sociolinguistics

## **Memberships**

Association for Computational Linguistics

European Association for Computational Linguistics

Slovene society for language technologies

Croatian society for language technologies

## **Publications**

<https://scholar.google.hr/citations?user=zto4fTQAAAAJ&hl=en>