PERSONAL INFORMATION

Family name, First name: **Dobrovoljc, Kaja**

Date of birth: **17th February 1986**

Nationality: **Slovene**

ORCID: **0000-0002-5909-7965**

Google Scholar: http://tiny.cc/kajad

EDUCATION

2018	University of Ljubljana, Faculty of Arts, Department of Slovene Studies Doctoral programme in Slovene Studies PhD topic: Discourse-structuring devices in spoken and online communication Supervisors: Prof. Darinka Verdonik, Prof. Marko Stabej
2011	University of Ljubljana, Faculty of Arts, Department of Translation Studies University graduate translator for English and French

CURRENT POSITION

2018 –	Researcher, Jozef Stefan Institute, Artificial Intelligence Laboratory
2019 –	Research Associate, University of Ljubljana, Faculty of Arts, Department of Translation Studies
2021 –	Research Associate, University of Ljubljana, Faculty of Computer and Information Science, Laboratory for Cognitive Modelling

PREVIOUS POSITIONS

2018–2019	Researcher, Faculty of Electrical Engineering and Computer Science, University of Maribor
2018	Researcher, Faculty of Computer and Information Science, University of Ljubljana
2013–2018	Junior researcher, Trojina Institute for Applied Slovene Studies
2011–2013	Freelance lexicographer and translator

RESEARCH VISITS

2015	Prof. Joakim Nivre, Department of Linguistics and Philology, Uppsala University; PARSEME COST Action Short-Term Scientific Mission Grant (September-November 2015)

ONGOING PROJECTS

2021 –	**Development of Slovene in a Digital Environment** (Slovenian Ministry of Culture and the European Regional Development Fund). Slovenian UD treebank extension (WP leader), NLP pipeline evaluation and development, spoken corpus concondarcer development, consulting.
2019 –	**Language Resources and Technologies for Slovene** (National Research Agency Research Programme). Development of corpus linguistic methods for analysing contemporary Slovene; computational linguistic methods for multilingual research; methods for improving literacy in the digital world.
2019 –	**ELEXIS: European Lexicographic Infrastructure** (EU Horizon 2020 Project). Morphosyntactic and semantic annotation of Slovenian and English language resources, cross-lingual collocation extraction from dependency-parsed corpora.

FINISHED PROJECTS

2017–2020	**New grammar of contemporary standard Slovene: sources and methods** (National Research Agency Research Project). Development of methodology for extraction and linguistic description of formulaic sequences in Slovenian (WP leader), corpus-driven collocation and multi-word unit extraction.
2019–2020	**Collocation as a basis for language description: semantic and temporal perspectives** (National Research Agency Research Project). Optimization and evaluation of statistical methods for measuring collocations in Slovene.
2019	**STARK: dependency tree extraction tool** (CLARIN.SI). Development of a software tool for statistical analysis of dependency-parsed corpora (project leader).
2018–2019	**Slovenian in the palm of your hand** (Slovenian Ministry of Education, Science and Sport and the European Regional Development Fund). Development of a semi-automatic pipeline for sentence retrieval from dependency-parsed corpora for an interactive e-learning environment (WP leader).
2014–2015	**Language Resources Portal** (Slovenian Ministry of Culture). Development of an interactive online platform with descriptions and video tutorials to promote the awareness and use of language resources and technologies.
2012–2013	**Language Technology Seminars for Teachers** (Slovenian Ministry of Culture). Organization and material preparation for a series of nationwide seminars aimed at promoting the use of online language resources and technologies amongst teachers of Slovene in primary and secondary schools.
2012	**xLike - Cross Lingual Knowledge Extraction** (EU FP7-ICT Project). Slovene module for NooJ, a linguistic development environment for complex rule-based corpus processing.
2012	**PISA Dependency Treebank** (Faculty of Education, University of Ljubljana). Building and exploring a dependency treebank of Slovene PISA reading tests.
2010–2013	**Communication in Slovene** (Slovenian Ministry of Education, Science and Sport and the European Social Fund). Manual annotation of the reference treebank (tokenization, lemmatization, morphology and dependency syntax). Development and evaluation of basic NLP applications and lexical resources (morphological dictionary, POS tagger, dependency parser). Corpus data extraction and analysis for a corpus-based manual of style.

ORGANIZATION

2021	EACL 2021 Language Diversity Games, 20 April 2021
2019	22nd International Conference on Text, Speech and Dialogue, 11-13 September 2019, Ljubljana, Slovenia (Organizing Committee Member)
2018	18th EURALEX International Congress, 12-21 July 2018, Ljubljana, Slovenia (Organizing Committee Member)
2016	Chair of Student Research Track at the Language Technologies and Digital Humanities Conference, 29 September - 1 October 2016, Ljubljana, Slovenia
2016	Language Technologies and Digital Humanities Conference, 29 September - 1 October

2016, Ljubljana, Slovenia (Organizing Committee Member)

REVIEWING

2018 –	Journal Reviewer: Language Resources and Evaluation, Journal of Pragmatics, Slovenščina 2.0
2014 –	Conference Reviewer: Treebanks and Linguistic Theories, UD Workshop, SyntaxFest, KDD Project Showcase Track
2018	Best Student Paper Committee, Language Resources and Digital Humanities Conference
2018	Student Scholarship Committee, Slovenian Language Technologies Society

MEMBER

2015–2018	TextLink: Structuring Discourse in Multilingual Europe (COST Action)
2014–2017	Parseme: Parsing and multi-word expressions (COST Action)
2014 –	Slovenian Language Technologies Society (Secretary)
2014 –	ACL SIGLEX-MWE Section
2013 –	Young Academy

CO-EDITOR

2020 –	Sledilnik Language Monitor
2019 –	Sloleks 2.0: Slovene Morphological Lexicon
2019 –	Gigafida 2.0: Corpus of Written Standard Slovene
2018 –	Collocations Dictionary of Modern Slovene
2017 –	Thesaurus of Modern Slovene

TEACHING

2015 –	Lecturer at workshops on SketchEngine, WebAnno, NooJ, Q-CAT and other corpus-linguistic tools

AWARDS AND SCOLARSHIPS

2020	E-Pub Special Mention Award at the Slovenian Book Fair (Gigafida 2.0 Reference Corpus)
2018	University of Ljubljana Best Research Achievement (Thesaurus of Modern Slovene)
2013–2015	Škrabec Association Scholarship for promising students of linguistics
2001–2010	Zois National Merit Scholarship