

OKSANA DEREZA

✉ oksana.dereza@gmail.com ☎ +353899404168 📍 Galway, Ireland 🌐 github.com/ancatmara
ID ORCID: 0000-0003-2896-6037 📄 ResearcherID: Q-6140-2017 in linkedin.com/in/oksana-dereza/



EXPERIENCE

Graduate Teaching Assistant

University of Galway

📅 Sep 2022 – Nov 2022 📍 Galway

- Demonstrated labs for a masters' course 'Introduction to NLP'
- Created lab worksheets and assignments
- Graded assignments of 70+ students

Research Assistant

University College Dublin

📅 Jan 2021 – Apr 2021 📍 Online

- Researched DH methods and software for Medieval law studies
- Prototyped a linked data model for Medieval Irish legal tracts
- Facilitated writing a grant proposal for **FLEXI project** (received IRC and ERC awards)

Mentor

Skillfactory

📅 Sep 2019 – Jan 2020 📍 Online

- Provided help & feedback for students of 'Real Data Science' course
- Created promo Python and NLP tutorials

Lecturer

National Research University 'Higher School of Economics'

📅 Sep 2017 – Aug 2019 📍 Moscow

- Authored and taught 3 undergraduate courses for students in humanities, administered exams and assessed students' performance
 - *Python for Natural Language Processing*
 - *An Introduction to Python*
 - *Digital Literacy*
- Improved course materials and lectured 2 graduate courses for students with a major in NLP, graded assignments
 - *Machine Learning for Natural Language Processing*
 - *Functional and Cognitive Models in Linguistics*
- Developed syllabi for the 'Data Culture' project
- Co-authored and taught the NLP block of a further education course 'Practical Data Science'; supervised 3 final research projects of students
- Prepared and directed the 'Digital Spring' school in Digital Humanities
- Implemented and conducted the Network Analysis track at the 'Digital Spring' school
- Managed the III Moscow-Tartu School in Digital Humanities
- Published 3 research papers
- Delivered 5 conference talks and 2 guest lectures

TECHNICAL SKILLS

Python LaTeX R Unix NLPProc
Machine Learning Data Science

LANGUAGES

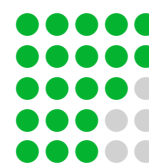
Russian — native speaker

English — C2

Irish

French

Danish



EDUCATION

PhD candidate in *Computer Science*

Insight Centre of Data Analytics, University of Galway

📅 Sep 2019 – now

Part of the **CARDAMOM project**

('Comparative Deep Models for Minority and Historical Languages')

Master in *Computational Linguistics and Language Theory*, with distinction

National Research University 'Higher School of Economics'

📅 Sep 2015 – Jun 2017

Thesis title: 'Morphological analysis of Old Irish data with neural networks'

Degree of *Specialist in Philology*.

Teacher of the Irish Language and Foreign Literature, with distinction

Lomonosov Moscow State University

📅 Sep 2010 – Jun 2015

Specialist = bachelor + master

Thesis title: 'Semantics of physical qualities in Goidelic: 'heavy' and 'light''

Exchange student

Trinity College Dublin, School of Irish

📅 Sep 2014 – Dec 2014

Exchange student

University of Copenhagen, Department for Scandinavian Studies and Linguistics

📅 Feb 2012 – Jun 2012

EXPERIENCE

Lecturer

Lomonosov Moscow State University

Jan 2018 – Jun 2019

Moscow

- Designed and lectured 3 courses for undergraduate students with a major in Irish, graded assignments and administered exams
 - Beginner Modern Irish
 - Readings in Modern Irish Literature
 - Early Modern Irish
- Acted as an official opponent of 2 senior theses

Data analyst & Computational linguist

Kaspersky Lab

Feb 2015 – Jul 2018

Moscow

- Evaluated and corrected automatic webpage classification results
- Performed 5+ NLP tasks: sentiment analysis, feedback classification, web query classification, content filtering, keyword extraction

AWARDS

GOI-IES Award

Higher Education Authority

2019

Ireland

High-potential Employee Award

National Research University 'Higher School of Economics'

2018, 2019

Russia

Oxford Russia Fund Postgraduate Award

Oxford Russia Fund

2018

Russia

Best Student Research Paper

National Research University 'Higher School of Economics'

2015

Russia

Certificate of Proficiency in English

Cambridge English

2014

Russia

REFEREES

Dr. John P. McCrae

@ john.mccrae@insight-centre.org

Unit for Linguistic Data

Insight Centre for Data Analytics, IDA
Business Park, Dangan, Galway

Dr. Fangzhe Qiu

@ fangzhe.qiu@ucd.ie

School of Irish Celtic Studies and Folklore

Newman Building, University College
Dublin, Belfield, Dublin 4

Dr. Victor Bayda

@ bajda-bajda@list.ru

Comhchoiste Ghaeltacht Uíbh Ráthaigh

Kineigh, Waterville Road, Cahersiveen,
Co. Kerry, V23 R894

SUMMER SCHOOLS

European Summer School in Logic,
Language, and Information (ESSLLI)

University of Galway

Aug 2022

Galway

Advanced Language Processing
Winter School (ALPS)

Université Grenoble Alpes & Naver Labs
Europe

Jan 2021

Online

II Summer School of Computational
Humanities: Mixed Methods in the
Making

Universität Heidelberg

Jul 2019

Heidelberg

II Moscow-Tartu School in Digital
Humanities

National Research University 'Higher
School of Economics' & University of Tartu

Oct 2017

Moscow

Summer school in Celtic studies

Dublin Institute for Advanced Studies

Jul 2011

Dublin

CONFERENCES

- Machine-Assisted Dating of Medieval Irish Texts with Cluster Analysis // *The XVIth International Congress Of Celtic Studies*. Bangor, 2019
- Automatic Dating of Medieval Irish Texts with Cluster Analysis // *International Medieval Congress*. Leeds, 2019
- A searchable database of Early Irish texts // *32nd Irish Conference of Medievalists*. Cork, 2018
- A Searchable Database of Early Irish Texts. // *The 9th International Colloquium of Societas Celto-Slavica*. Quimper, 2018
- Lemmatisation for under-resourced languages with sequence-to-sequence learning: A case of Early Irish // *The 3d Kolmogorov Workshop "Computational linguistics and language science"*. Moscow, 2018
- Lemmatizing Old Irish: Lexicons, Rules and Neural Networks. // *Spring Symposium on Computational and Corpus-based Linguistics (organized by Chronologicon Hibernicum)*. Maynooth, 2017
- A Lexicon-Based Lemmatizer for Old Irish (poster). // *AINL FRUCT: Artificial Intelligence and Natural Language Conference*. Saint-Petersburg, 2016
- Intratextual relations in the Ulster cycle. // *The 8th International Colloquium of Societas Celto-Slavica*. Heidelberg, 2016
- Verb-noun collocation and government model extraction from large corpora (poster, with Vladislav Tushkanov). // *The 3^d International Workshop on Concept Discovery in Unstructured Data*. Moscow, 2016
- Building a dictionary-based lemmatizer for Old Irish. // *The 2nd Celtic Language Technology Workshop, JEP-TALN-RECITAL*. Paris, 2016
- A homonymy-oriented metric for morphological analyzers (poster, with Alyona Fenogenova & Dmitri Kayutenko). // *Computational Linguistics and Language Science Workshop*. Moscow, 2016
- Automatic morphological analysis for Russian: a comparative study (poster, with Alyona Fenogenova & Dmitri Kayutenko). // *XXII International conference on Computational Linguistics and Intellectual Technologies 'DIALOGUE'*. Moscow, 2016
- A complex approach to spellchecking and autocorrection for Russian (with Alyona Fenogenova, Anna Marakasova & Dmitri Kayutenko). // *XXII International conference on Computational Linguistics and Intellectual Technologies 'DIALOGUE'*. Moscow, 2016
- Polysemous adjectives in Goidelic languages: 'heavy' and 'light'. // *The 7th International Colloquium of Societas Celto-Slavica*. Bangor, 2014
- The semantics of **k^w*enno- reflexes: history and typology (poster). // *Indian Summer School in Typology, HSE*. Moscow, 2014
- The polysemy of TRWM 'heavy' in Middle Welsh (with Dr. Elena Parina). // *Language Diversity in Wales*. Aberystwyth, 2014
- Semantic differences, meaning and distribution of the word *ceann* in Irish and Scottish Gaelic. // *Rannsachadh na Gàidhlig*. Edinburgh, 2014
- The semantics of Proto-Celtic **k^w*enno: history and typology. // *First European Symposium in Celtic Studies*. Trier, 2013

PUBLICATIONS

- A Multilingual Evaluation Dataset for Monolingual Word Sense Alignment // (with S. Ahmadi, J. McCrae, S. Nimb, F. Khan, M. Monachini, B. Pedersen, T. Declerck, T. Wissik, A. Bellandi, I. Pisani, T. Troelsgard, S. Olsen, S. Krek, V. Lipp, T. Váradi, L. Simon, A. Gyorffy, C. Tiberius, T. Schoonheim, Y. Ben Moshe, M. Rudich, R. Abu Ahmad, D. Lonke, K. Kovalenko, M. Langemets, J. Kallas, T. Fransen, D. Cillessen, D. Lindemann, M. Alonso, A. de Castro Salgado, J. L. Sancho, R. Ureña-Ruiz, J. Porta Zamorano, K. Simov, P. Osenova, Z. Kancheva, I. Radev, R. Stankovic, A. Perdih, D. Gabrovsek) // *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC)*, 2020
- Lemmatisation for Under-Resourced Languages with Sequence-to-Sequence Learning: A Case of Early Irish // *EPiC Series in Language and Linguistics*, 2019
- Lemmatization for Ancient Languages: Rules or Neural Networks? // *Series "Communications in Computer and Information Science"*, Springer, 2018
- Physical Qualities in Goidelic: A Corpus Study of Polysemy and Collocability. // *Proceedings of the Seventh International Colloquium of Societas Celto-Slavica*, 2018
- Building a Dictionary-Based Lemmatizer for Old Irish // *Actes de la conférence conjointe JEP-TALN-RECITAL 2016, volume 6: CLTW*. Paris, 2016
- Paired adjectives in the Middle Welsh *Ystorya Bown o Hamtwn*. // *Dafydd Johnston, Elena Parina and Maxim Fomin (eds.) 'Ynllaweniawn, ynllawniaith': Proceedings of the 6th International Colloquium of Societas Celto-Slavica. Studia Celto-Slavica 7*. Aberystwyth: University of Wales Centre for Advanced Welsh and Celtic Studies, 2015