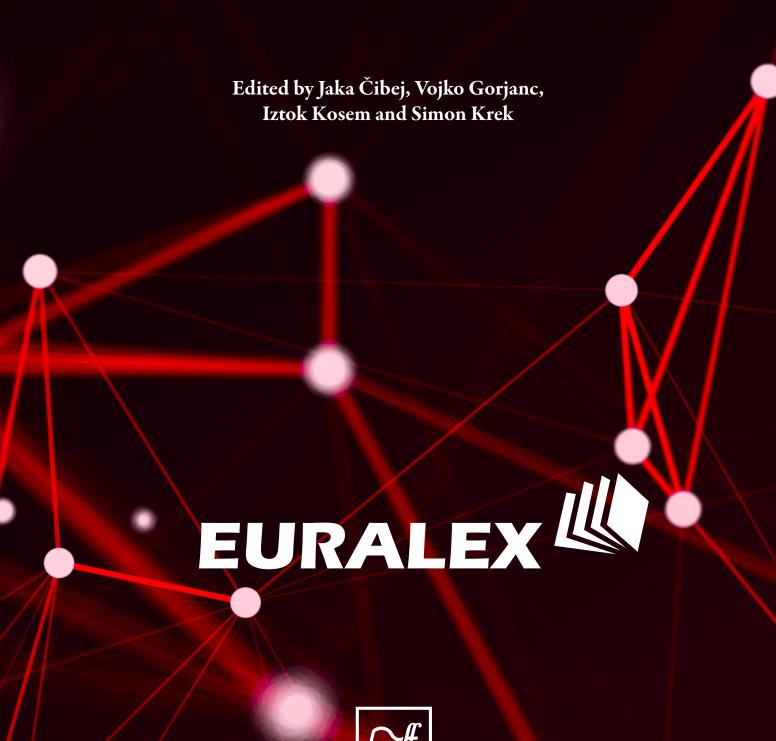
# Proceedings of the XVIII EURALEX International Congress Lexicography in Global Contexts

17-21 July 2018, Ljubljana





#### Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts

Edited by: Jaka Čibej, Vojko Gorjanc, Iztok Kosem and Simon Krek

Reviewers: Andrea Abel, Zoe Gavriilidou, Robert Lew and Tinatin Margalitadze

English language proofreading: Paul Steed

Technical editor: Aleš Cimprič



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Published by: Znanstvena založba Filozofske fakultete Univerze v Ljubljani / Ljubljana University Press, Faculty of Arts

Issued by: University of Ljubljana, Centre for language resources and technologies For the publisher: Roman Kuhar, Dean of the Faculty of Arts, University of Ljubljana

Ljubljana, 2018
First edition, e-edition
Publication is free of charge.

The editors acknowledge the financial support from the Slovenian Research Agency (research core funding No. P6-0215).

This text was written using the ZRCola input system (ZRCola.zrc-sazu.si), developed at the Research Centre of the Slovenian Academy of Sciences and Arts in Ljubljana (www.zrc.sazu.si) by Peter Weiss.

DOI: 10.4312/9789610600961

E-book is available at: https://e-knjige.ff.uni-lj.si/

Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani

COBISS.SI-ID=24619494

ISBN 978-961-06-0097-8 (epub) ISBN 978-961-06-0096-1 (pdf)

### **Acknowledgements**

We would like to thank all those who have made the XVIII EURALEX International Congress possible, by contributing to the reviewing, to the logistics and by financially supporting the event. In particular, we would like to thank our sponsoring partners and patrons:

A.S. Hornby Educational Trust

Ingenierie Diffusion Multimedia Inc.

Oxford University Press

ELEXIS – European Lexicographic Infrastructure

CLARIN.SI - Common Language Resources and Technology Infrastructure, Slovenia

Alpineon d.o.o.

DELIGHT d.o.o.

TshwaneDJe

Faculty of Arts, University of Ljubljana

#### **Programme Committee**

Andrea Abel (European Academy of Bozen/Bolzano, EURAC)

Polona Gantar (University of Ljubljana, Faculty of Arts)

Zoe Gavriilidou (Democritus University of Thrace, Xanthi)

Vojko Gorjanc (University of Ljubljana, Faculty of Arts)

Iztok Kosem (University of Ljubljana, Faculty of Arts / Trojina)

Simon Krek (Chair) (University of Ljubljana, Center for Language Resources and Technologies / Jožef Stefan Institute, Artificial Intelligence Laboratory)

Robert Lew (Adam Mickiewicz University in Poznań, Faculty of English)

Tinatin Margalitadze (Ivane Javakhishvili Tbilisi State University)

#### **Reviewers**

Adam Rambousek, Agáta Karčová, Agnes Tutin, Ales Horak, Alexander Geyken, Amália Mendes. Amanda Laugesen, Andrea Abel, Anne Dykstra, Annette Klosa, Antton Gurrutxaga, Arvi Tavast, Carla Marello, Carole Tiberius, Carolin Müller-Spitzer, Chris Mulhall, Christine Moehrs, Corina Forascu, Danie Prinsloo, Edward Finegan, Egon Stemle, Elena Volodina, Francesca Frontini, Francis Bond, Frieda Steurs, Geoffrey Williams, Henrik Lorentzen, Ilan Kernerman, Iztok Kosem, Janet Decesaris, Jelena Kallas, Jette Hedegaard Kristoffersen, John McCrae, Jorge Gracia, Julia Bosque-Gil, Julia Miller, Klaas Ruppel, Kristina Strkalj Despot, Kseniya Egorova, Lars Trap-Jensen, Lionel Nicolas, Lothar Lemnitzer, Lut Colman, Magali Paquot, Margit Langemets, Maria Khokhlova, Marie-Claude L'Homme, Michal Měchura, Michal Kren, Miloš Jakubiček, Monica Monachini, Nataša Logar Berginc, Nicoletta Calzolari, Nikola Ljubešić, Nora Aranberri, Oddrun Grønvik,

Orin Hargraves, Orion Montoya, Patrick Hanks, Patrick Drouin, Paul Cook, Paz Battaner, Philipp Cimiano, Pilar León Araúz, Piotr Zmigrodzki, Pius ten Hacken, Polona Gantar, Radovan Garabík, Robert Lew, Roberto Navigli, Ruben Urizar, Rufus Gouws, Sass Bálint, Sara Može, Simon Krek, Stella Markantonatou, Svetla Koeva, Sylviane Granger, Špela Arhar Holdt, Tamás Váradi, Tanneke Schoonheim, Tatjana Gornostaja, Thierry Fontenelle, Tinatin Margalitadze, Tomaž Erjavec, Ulrich Heid, Valentina Apresjan, Vincent Ooi, Vojko Gorjanc, Xabier Artola Zubillaga, Xabier Saralegi, Yongwei Gao, Yukio Tono, Zoe Gavriilidou





















# **Contents**

Foreword	
PLENARY LECTURES	15
Has Lexicography Reaped the Full Benefit of the (Learner) Corpus Revolution?  Sylviane Granger	17
Lexicography between NLP and Linguistics: Aspects of Theory and Practice  Lars Trap-Jensen	25
PAPERS	39
RESEARCH INTO DICTIONARY USE	41
Investigating the Dictionary Use Strategies of Greek-speaking Pupils  Elina Chadjipapa	43
Everything You Always Wanted to Know about Dictionaries (But Were Afraid to Ask): A Massive Open Online Course Sharon Creese, Barbara McGillivray, Hilary Nesi, Michael Rundell, Katalin Sule	59
Researching Dictionary Needs of Language Users Through Social Media: A Semi-Automatic Approach Jaka Čibej, Špela Arhar Holdt	67
The DHmine Dictionary Work-flow: Creating a Knowledge-based Author's Dictionary Tamás Mészáros, Margit Kiss	77
Analyzing User Behavior with Matomo in the Online Information System Grammis Saskia Ripp, Stefan Falke	87
Combining Quantitative and Qualitative Methods in a Study on Dictionary Use Sascha Wolfer, Martina Nied Curcio, Idalete Maria Silva Dias, Carolin Müller-Spitzer, María José Domínguez Vázquez <sup>4</sup>	101
DICTIONARY-MAKING PROCESS	113
Nathanaël Duez lexicographe : l'art de (re)travailler les sources Antonella Amatuzzi	115
A Workflow for Supplementing a Latvian-English Dictionary with Data from Parallel Corpora and a Reversed English-Latvian Dictionary Daiga Deksne, Andrejs Veisbergs	127
Towards a Representation of Citations in Linked Data Lexical Resources  Anas Fahad Khan, Federico Boschetti	137
The Sounds of a Dictionary: Description of Onomatopoeic Words in the Academic Dictionary of Contemporary Czech Magdalena Kroupová, Barbora Štěpánková, Veronika Vodrážková	149
Comparing Orthographies in Space and Time through Lexicographic Resources Christian-Emil Smith Ore, Oddrun Grønvik	159
A Universal Classification of Lexical Categories and Grammatical Distinctions for Lexicographic and Processing Purposes Roser Saurí, Ashleigh Alderslade, Richard Shapiro	173

	Commonly Confused Words in Contrastive and Dynamic Dictionary Entries  Petra Storjohann	187
	Slovenian Lexicographers at Work Alenka Vrbinc, Donna M. T. Cr. Farina, Marjeta Vrbinc	199
	Methodological issues of the compilation of the Polish Academy of Sciences Great Dictionary of Polish Piotr Żmigrodzki	209
L	EXICOGRAPHICAL PROJECTS AND PHRASEOLOGY	221
	Shareable Subentries in Lexonomy as a Solution to the Problem of Multiword Item Placement Michal Boleslav Měchura	223
	A Good Match: a Dutch Collocation, Idiom and Pattern Dictionary Combined Lut Colman, Carole Tiberius	233
	ColloCaid: A Real-time Tool to Help Academic Writers with English Collocations Robert Lew, Ana Frankenberg-Garcia, Geraint Paul Rees, Jonathan C. Roberts, Nirwan Sharma	247
	Looking for a Needle in a Haystack: Semi-automatic Creation of a Latvian Multi-word Dictionary from Small Monolingual Corpora Inguna Skadiņa	255
T.	ERMINOLOGY, TERMINOGRAPHY AND SPECIALISED LEXICOGRAPHY	267
	Semantic-based Retrieval of Complex Nominals in Terminographic Resources Melania Cabezas-García, Juan Carlos Gil-Berrozpe	269
	Towards a Glossary of Rum Making and Rum Tasting Cristiano Furiassi	283
	Russian Borrowings in Greek and Their Presence in Two Greek Dictionaries  Zoe Gavriilidou	297
	Frame-based Lexicography: Presenting Multiword Terms in a Technical E-dictionary Laura Giacomini	309
	Dictionaries of Linguistics and Communication Science / Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK) Stefan J. Schierholz	319
	When Learners Produce Specialized L2 Texts: Specialized Lexicography between Communication and Knowledge Patrick Leroyer, Henrik Køhler Simonsen	329
	New Platform for Georgian Online Terminological Dictionaries and Multilingual Dictionary Management System  Tinatin Margalitadze	339
	Building a Portuguese Oenological Dictionary: from Corpus to Terminology via Co-occurrence Networks  William Martinez, Silvia Barbosa	351
	Using Diachronic Corpora of Scientific Journal Articles for Complementing English Corpus-based Dictionaries and Lexicographical Resources for Specialized Languages <i>Katrin Menzel</i>	363

ELeFyS: A Greek Illustrated Science Dictionary for School Maria Mitsiaki, Ioannis Lefkos	373
Terms Embraced by the General Public: How to Cope with Determinologization in the Dictionary?  Jana Nová	387
REPORTS ON LEXICOGRAPHICAL PROJECTS	
Thesaurus of Modern Slovene: By the Community for the Community Špela Arhar Holdt, Jaka Čibej, Kaja Dobrovoljc, Polona Gantar, Vojko Gorjanc, Bojan Klemenc, Iztok Kosem, Simon Krek, Cyprian Laskowski, Marko Robnik-Šikonja	401
Dictionary of Verbal Contexts for the Romanian Language  Ana-Maria Barbu	411
A Sample French-Serbian Dictionary Entry based on the ParCoLab Parallel Corpus Saša Marjanović, Dejan Stosic, Aleksandra Miletic	423
HISTORICAL LEXICOGRAPHY, ETYMOLOGY	437
Lexicography in the Eighteenth-century Gran Chaco: the Old Zamuco Dictionary by Ignace Chomé  Luca Ciucci	439
Historical Corpus and Historical Dictionary: Merging Two Ongoing Projects of Old French by Integrating their Editing Systems  Sabine Tittel	453
Heritage Dictionaries, Historical Corpora and other Sources: Essential And Negligible Information Alina Villalva	467
SIGN LANGUAGE LEXICOGRAPHY	481
Authentic Examples in a Corpus-Based Sign Language Dictionary – Why and How Gabriele Langer, Anke Müller, Sabrina Wähl, Julian Bleicken	483
Multimodal Corpus Lexicography: Compiling a Corpus-based Bilingual Modern Greek Greek Sign Language Dictionary Anna Vacalopoulou, Eleni Efthimiou, Kiki Vasilaki	K – 499
PHRASEOLOGY AND COLLOCATION	509
Bilingual Corpus Lexicography: New English-Russian Dictionary of Idioms  Guzel Gizatova	511
Computer-aided Analysis of Idiom Modifications in German Elena Krotova	523
NEOLOGISMS	533
On the Detection of Neologism Candidates as a Basis for Language Observation and Lexicographic Endeavors: the STyrLogism Project Andrea Abel, Egon W. Stemle	535
Neologisms in Online British-English versus American-English Dictionaries Sharon Creese	545

	New German Words: Detection and Description  Annette Klosa, Harald Lüngen	559
	"Brexit means Brexit": A Corpus Analysis of Irish-language BREXIT Neologisms in The Corpus of Contemporary Irish Katie Ni Loingsigh	571
L	EXICOGRAPHY OF LESSER USED LANGUAGES	583
	Synonymy in Modern Tatar reflected by the Tatar-Russian Socio-Political Thesaurus Alfiia Galieva	585
	Revision and Extension of the OIM Database – The Italianisms in German Anne-Kathrin Gärtig	595
	The Treatment of Politeness Elements in French-Korean Bilingual Dictionaries Hae-Yun Jung, Jun Choi	607
	Lexicography in the French Caribbean: An Assessment of Future Opportunities Jason F. Siegel	619
V	ARIOUS TOPICS	629
	The Dictionary of the Learned Level of Modern Greek Anna Anastassiadis-Symeonidis, Asimakis Fliatouras, Georgia Nikolaou	631
	In Praise of Simplicity: Lexicographic Lightweight Markup Language Vladimír Benko	641
	Corpus-based Cognitive Lexicography: Insights into the Meaning and Use of the Verb Stagger  Thomai Dalpanagioti	649
	Polysemy and Sense Extension in Bilingual Lexicography  Janet DeCesaris	663
	Associative Experiments as a Tool to Construct Dictionary Entries  Ksenia S. Kardanova-Biryukova	675
	Lexicographic Potential of the Syntactic Properties of Verbs: The Case of Reciprocity in Czech Václava Kettnerová, Markéta Lopatková	685
	LexBib: A Corpus and Bibliography of Metalexicographical Publications  David Lindemann, Fritz Kliche, Ulrich Heid	699
	Process Nouns in Dictionaries: A Comparison of Slovak and Dutch Renáta Panocová, Pius ten Hacken	713
	Definitions of Words in Everyday Communication: Associative Meaning from the Pragmatic Point of View Svitlana Pereplotchykova	723
	Verifying the General Academic Status of Academic Verbs: An Analysis of co-occurrence and Recurrence in Business, Linguistics and Medical Research Articles  Natassia Schutz	735
	Unified Data Modelling for Presenting Lexical Data: The Case of EKILEX  Arvi Tavast, Margit Langemets, Jelena Kallas, Kristina Koppel	749

	On the Interpretation of Etymologies in Dictionaries  Pius ten Hacken	763
	The Virtual Research Environment of VerbaAlpina and its Lexicographic Function Christina Mutter, Aleksander Wiatr	775
7	OSTER PRESENTATIONS	787
	Lexicographie et terminologie au XIX <sup>e</sup> siècle :  Vocabularu romano-francesu [Vocabulaire roumain-français], de Ion Costinescu (1870) 789  Maria Aldea	
	Developing a Russian Database of Regular Semantic Relations Based on Word Embeddings Ekaterina Enikeeva, Andrey Popov	<b>79</b> 9
	Semantic Classification of Tatar Verbs: Selecting Relevant Parameters  Alfiia Galieva, Ayrat Gatiatullin, Zhanna Vavilova	811
	Word2Dict – Lemma Selection and Dictionary Editing Assisted by Word Embeddings Nicolai Hartvig Sørensen, Sanni Nimb	819
	Building a Lexico-Semantic Resource Collaboratively Mercedes Huertas-Migueláñez, Natascia Leonardi, Fausto Giunchiglia	827
	The CPLP Corpus: A Pluricentric Corpus for the Common Portuguese Spelling Dictionary (VOC)  Maarten Janssen, Tanara Zingano Kuhn, José Pedro Ferreira, Margarita Correia	835
	Málið.is: A Web Portal for Information on the Icelandic Language Halldóra Jónsdóttir, Ari Páll Kristinsson, Steinþór Steingrímsson	841
	Multilingual Generation of Noun Valency Patterns for Extracting Syntactic-Semantical Knowledge from Corpora (MultiGenera)  María José Domínguez Vázquez, Carlos Valcárcel Riveiro, David Lindemann	847
	A Lexicon of Albanian for Natural Language Processing Besim Kabashi	855
	Building a Gold Standard for a Russian Collocations Database  Maria Khokhlova	863
	Rethinking the role of digital author's dictionaries in humanities research  Margit Kiss, Tamás Mészáros	871
	European Lexicographic Infrastructure (ELEXIS) Simon Krek, Iztok Kosem, John P. McCrae, Roberto Navigli, Bolette S. Pedersen, Carole Tiberius, Tanja Wissik	881
	The EcoLexicon English Corpus as an Open Corpus in Sketch Engine Pilar León-Araúz, Antonio San Martín, Arianne Reimerink	893
	A Call for a Corpus-Based Sign Language Dictionary: An Overview of Croatian Sign Language Lexicography in the Early 21st Century Klara Majetić, Petra Bago	903
	Exploring the Frequency and the Type of Users' Digital Skills Using S.I.E.D.U.  Stavroula Mayrommatidou	909

	From Standalone Thesaurus to Integrated Related Words in The Danish Dictionary Sanni Nimb, Nicolai H. Sørensen, Thomas Troelsgård	915
	<b>Exploratory and Text Searching Support in the Dictionary of the Spanish Language</b> <i>Jordi Porta-Zamorano</i>	925
	Interactive Visualization of Dialectal Lexis Perspective of Research Using the Example of Georgian Electronic Dialect Atlas  Marine Beridze, Zakharia Pourtskhvanidze, Lia Bakuradze, David Nadaraia	931
	The Dictionary of the Serbian Academy: from the Text to the Lexical Database Ranka Stanković, Rada Stijović, Duško Vitas, Cvetana Krstev, Olga Sabo	941
S	OFTWARE DEMONSTRATIONS	951
	An Overview of FieldWorks and Related Programs for Collaborative Lexicography and Publishing Online or as a Mobile App David Baines	953
	Wortschatz und Kollokationen in "Allgemeine Reisebedingungen". Eine intralinguale und interlinguale Studie zum fachsprachlich-lexikographischen Projekt "Tourlex". Carolina Flinz, Rainer Perkuhn	959
	Advances in Synchronized XML-MediaWiki Dictionary Development in the Context of Endangered Uralic Languages  Mika Hämäläinen, Jack Rueter	967
	Linking Corpus Data to an Excerpt-based Historical Dictionary Tarrin Wills, Ellert Þór Jóhannsson, Simonetta Battista	979
	Collocations Dictionary of Modern Slovene Iztok Kosem, Simon Krek, Polona Gantar, Špela Arhar Holdt, Jaka Čibej, Cyprian Laskowski	989
	Computerized Dynamic Assessment of Dictionary Use Ability Osamu Matsumoto	999
	Creating a List of Headwords for a Lexical Resource of Spoken German Meike Meliss, Christine Möhrs, Dolores Batinić, Rainer Perkuhn	1009
	fLexiCoGraph: Creating and Managing Curated Graph-Based Lexicographical Data Peter Meyer, Mirjam Eppinger	1017
	Wordnet Consistency Checking via Crowdsourcing Aleš Horák, Adam Rambousek	1023

## **Foreword**

EURALEX, European Association for Lexicography was founded in 1983 and the year 2018 marks its thirty-fifth anniversary. From its second congress in 1986, the association organises a biannual congress series. Its 18<sup>th</sup> edition, EURALEX 2018 International Congress, was held between 17<sup>th</sup>-21<sup>st</sup> July 2018 in Ljubljana, Slovenia. It was organised jointly by the Centre for Language Resources and Technologies (CLRT) at the University of Ljubljana, and Trojina Institute for Applied Slovene Studies. Both institutions are dedicated to scientific research, and the development and maintenance of digital language resources and language technology applications for contemporary Slovene. Trojina Institute was founded in 2004 with the primary objective of promoting contemporary, goal-oriented research of the Slovene language, and the University of Ljubljana founded the Centre in 2015 to ensure a systematic long-term development of technologies, resources and tools for Slovene.

The motto of EURALEX 2018 was "Lexicography in global contexts", emphasising changes in the field of lexicography related to digital transformation, and the associated need to bring together lexicographic efforts on a global level. This has been done in recent years through the Globalex initiative, a constellation of lexicographic associations that includes representatives from all continental associations of lexicography: Afrilex, Asialex, Australex, Dictionary Society of North America, and Euralex. Similar development can be witnessed in the decision of European Commission in 2017 to fund a four-year project dedicated to the establishment of the European Lexicographic Infrastructure (ELEXIS), which was also presented at the congress.

This volume of proceedings includes congress papers submitted in three categories: papers, posters, and software demonstrations. During the review process each submitted contribution was evaluated by two independent blind referees. In case of doubt, a third independent opinion was involved. Similar to previous congresses, contributions were submitted on various topics of lexicography, including, but not limited to, the following fields:

- The Dictionary-Making Process
- Research on Dictionary Use
- Lexicography and Language Technologies
- Lexicography and Corpus Linguistics
- Bi- and Multilingual Lexicography
- Lexicography for Specialised Languages, Terminology and Terminography
- Lexicography of Lesser Used languages
- Phraseology and Collocation
- Historical Lexicography and Etymology
- Lexicological Issues of Lexicographical Relevance
- Reports on Lexicographical and Lexicological Projects

Four plenary lectures were given at the congress, with two plenary papers also included in this volume. In the Hornby lecture and paper, Sylviane Granger from the Centre for English Corpus Linguistics, Université catholique de Louvain, discusses the value of adding learner corpus data to the lexicographer's monolingual and bilingual corpus base. Plenary lecture and paper by Lars Trap-Jensen from Danish Society of Language and Literature, also former president of Euralex, discusses three major revolutions that lexicography has witnessed in the last hundred years. The remaining two plenary lectures were presented by Judy Pearsall, Dictionaries Director at Oxford University Press, titled "One model, many languages? An approach to developing global language content" and Edward Finegan, professor emeritus of linguistics and law at the University of Southern California, on "Legal Interpretation via Corpora: Are Judges Failing Lexicography 101?"

The organising committee would like to thank all plenary speakers for setting the tone of the congress, and to other contributors for submitting very interesting work. We would also like to thank all the colleagues who reviewed the papers and the colleagues who participated in the work of the EURALEX 2018 programme committee. As in past EURALEX editions, the Hornby Trust generously sponsored one of the plenary lectures in honour of A.S. Hornby, a pioneering figure in learner's dictionaries for non-native speakers. All patrons and sponsors who supported us for this edition are listed on a dedicated page within these proceedings.

As the chair of the congress, I would like to acknowledge precious work of the members of the organising committee who joined efforts with me to make EURALEX 2018 a successful event: Špela Arhar Holdt, Jaka Čibej, Kaja Dobrovoljc, Polona Gantar, Vojko Gorjanc, Iztok Kosem and Nataša Logar.

Simon Krek Chair, XVIII EURALEX International Congress July 5, 2018