Annette Klosa-Kückelhaus, Stefan Engelberg,
Christine Möhrs, Petra Storjohann (eds.)

Dictionaries and Society

Book of Abstracts of the
XX EURALEX International Congress,
12-16 July 2022,
Mannheim, Germany

IDS-Verlag

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.dnb.de abrufbar.

IDS IDS-Verlag

IDS LEIBNIZ-INSTITUT FÜR DEUTSCHE SPRACHE

Leibniz Gemeinschaft

IDS-Verlag · Leibniz-Institut für Deutsche Sprache · R 5, 6–13 · 68161 Mannheim

Redaktion: Melanie Kraus
Satz: Annett Patzschewitz, Joachim Hohwieler
Titelbild: Norbert Cußler-Volz

Die Online-Version dieser Publikation ist auf den Webseiten des Leibniz-Instituts für Deutsche Sprache (www.ids-mannheim.de) dauerhaft frei verfügbar (Open Access). doi: https://doi.org/10.14618/t86d-e088

Die gesetzliche Verpflichtung über die Ablieferung digitaler Publikationen als Pflichtexemplare wird durch die manuelle Ablieferung der Netzpublikation an die Badische Landesbibliothek (BLB) erfüllt.

ISBN: 978-3-937241-86-9 (PDF)

FOREWORD

This book contains the abstracts of keynotes, talks, posters, and software demonstrations presented at the XX EURALEX International Congress, Mannheim, Germany, July 12–6, 2022. The abstracts were either submitted and accepted as extended abstracts or were taken from full papers accepted for publication in the Proceedings of the XX EURALEX International Congress.

In “Part I: Overview,” the abstracts of keynotes, talks, posters, and software demonstrations are listed in each case in alphabetical order by the authors’ last names. Each entry in these lists is linked to the corresponding abstract.

Part II contains the abstracts of the keynote presentations. Part III lists the abstracts of the talks, posters, and software demonstrations in topical order. Here you will find all abstracts belonging to one of the following topics (in alphabetical order by authors’ last names for each topic):

- Dictionaries and Society
- Lexicography: Status, Theory and Methods
- Corpora in Lexicography
- Data Models and Databases in Lexicography
- Dictionary Writing Systems and Lexicographic Tools
- Design and Publication of Dictionaries
- (Promoting) Dictionary Use
- Dictionary Projects
- Bilingual Dictionaries
- Specialised Dictionaries
- Historical Lexicography: German
- Historical Lexicography: Romance and Other Languages
- (Historical) Lexicology
- Neologisms and Lexicography
- Phraseology & Collocations

Finally, an alphabetical index contains all authors’ names.

TABLE OF CONTENTS

Acknowledgements

Main Sponsors ... 15
Sponsors ... 15
Programme Committee ... 16
Scientific Committee ... 16

Part I: Overview on Keynotes, Talks, Poster Presentations, and Software Demonstrations

Keynotes ... 20
Talks ... 20
Poster Presentations ... 24
Software Demonstrations ... 26

Part II: Abstracts of Keynotes

Thomas Gloning
Ways of living, communication and the dynamics of word usage
How did German dictionaries cope with socio-cultural aspects and evolution of word usage and how could future systems do even better? ... 28

Rufus H. Gouws
Dictionaries: bridges, dykes, sluice gates ... 29

Nicola McLelland
Women in the history of lexicography
An overview, and the case of German ... 30

Martina Nied Curcio
Dictionaries, foreign language learners and teachers
New challenges in the digital era ... 32

Ben Zimmer
The evolving definition of „racism“ and its trail of text-artifacts ... 33

Part III: Abstracts of Talks, Poster Presentations, and Software Demonstrations

Dictionaries and Society

Mariona Arnau Garcia/Mercè Lorente Casafont
Are emerging economies a reality reflected in our dictionaries? ... 36

Mark Dang-Anh/Stefan Scholl
Basic socio-political concepts in German parliamentary debates from the 20th century ... 38

Alexandre Ecker
Equality between women and men, a societal and lexicographical issue
Bringing the content of the *Lëtzebuerger Online Dictionnaire* in line with the realities of the world and society ... 41

Stefan Engelberg
Lexicography's entanglement with colonialism: The history of Tok Pisin lexicography as colonial history ... 46

Christine Ganslmayer
Using dialect dictionaries as a database
Stereotypes of people in the 'Franconian Dictionary' ... 47

Laura Giacomini/Paolo DiMuccio-Failla/Patrizio De Martin Pinter
The representation of culture-specific lexical items in monolingual learner's lexicography
The case of the electronic Phrase-Based Active Dictionaries ... 50

Annette Klosa-Kückelhaus
Lexicography for society and with society – COVID-19 and dictionaries ... 51

Gloria Mambelli
Manorial society in multilingual medieval England
An onomasiological approach ... 52

Carolin Müller-Spitzer/Jan Oliver Rüdiger
The influence of the corpus on the representation of gender stereotypes in the dictionary. A case study of corpus-based dictionaries of German ... 54

Enakshi Nandi
Secrecy and the ethical question
Some reflections on documenting Ulti, a secret transgender language ... 55

Laura Pinnavaia
Identifying ideological strategies in the making of monolingual English language learner's dictionaries ... 57

Petra Storjohann
The public as linguistic authority: Why users turn to internet forums to differentiate between words ... 58

Marcin Zabawa
What do we learn about the society from the examples of usage in dictionaries? On (non-) stereotypical roles of men and women in English and Polish monolingual general dictionaries: a contrastive study ... 59

Lexicography: Status, Theory and Methods

Laura Balbiani/Anne-Kathrin Gärtig-Bressan/Martina Nied Curcio/ Stefan Schierholz
Dictionaries for the future – the future of dictionaries
The 15 Villa Vigoni Theses on Lexicography ... 63

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Konan Kouassi
Mensch-Maschine-Interaktion im lexikographischen Prozess zu lexikalischen Informationssystemen 66

Ivana Filipović Petrović
A corpus-driven approach to lexicographic definitions: The representation of meaning in the electronic *Dictionary of Croatian Idioms* 67

Ana Salgado/Rute Costa/Toma Tasovac
Applying terminological methods to lexicographic work: terms and their domains 69

Gilles-Maurice de Schryver
Metalexicography: an existential crisis 70

Michal Škrabal/Michaela Lišková/Martin Šemelík
On defining vocabulary in a monolingual online dictionary
Some remarks from the lexicographical practice on the *Academic Dictionary of Contemporary Czech* 71

Corpora in Lexicography

Nils Diewald/Marc Kupietz/Harald Lüngen
Tokenizing on scale
Preprocessing large text corpora on the lexical and sentence level 76

Ana-Maria Gînsac/Mihai-Alex Moruz/Mădălina Ungureanu
The first Romanian dictionaries (17th century). Digital aligned corpus 77

Iztok Kosem
Trendi – a monitor corpus of Slovene 78

Simon Krek/Polona Gantar/Iztok Kosem
Extraction of collocations from the Gigafida 2.1 corpus of Slovene 79

Meike Meliss/Vanessa González Ribao
Vergleichbare Korpora für multilinguale kontrastive Studien
Herausforderungen und Desiderata 80

Irene Renau/Rogelio Nazar
Towards a multilingual dictionary of discourse markers
Automatic extraction of units from parallel corpus 81

Jan Oliver Rüdiger/Sascha Wolfer/Alexander Koplenig/Frank Michaelis/Carolin Müller-Spitzer/Samira Ochs/Louis Cotgrove
OWIDplusLIVE
Day-to-day collection, exploration, analysis, and visualization of N-Gram frequencies in German (online press) language 82

Data Models and Databases in Lexicography

Thierry Declerck
Integration of sign language lexical data in the OntoLex-Lemon framework 86

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Birgit Füreder
Überlegungen zur Modellierung eines multilingualen ‚Periphrastikons‘:
Ein französisch-italienisch-spanisch-englisch-deutscher Versuch 87

Yevhen Kupriianov/Iryna Ostapova/Volodymyr Shyrokov/Mykyta Yablochkov
Virtual lexicographic laboratory as a linguist assistant in conducting dictionary-based researches: Case of VLL DLE 23 88

David Lindemann/Penny Labropoulou/Christiane Klaes
Introducing LexMeta: A metadata model for lexical resources 93

Manuel Márquez
Un modelo estructural de datos lexicográficos para la codificación en XML de un diccionario de aprendizaje de Latín: La DTD de los lemas verbales........ 94

Michal Měchura
Document or database? The search for the perfect storage paradigm for lexical data 97

Christian-Emil Smith Ore/Oddrun Grønvik/Trond Minde
Word banks, dictionaries and research results by the roadside........ 99

Ana Ostroški Anić/Ivana Brač
AirFrame
Mapping the field of aviation through semantic frames 100

Kristel Proost/Arne Zeschel/Frank Michaelis/Jan Oliver Rüdiger
MAP (Musterbank Argumentmarkierender Präpositionen)
A patternbank of argument-marking prepositions in German 101

Vasyl Starko
USL: a cognitively inspired lexicon for semantic tagging........ 102

Philipp Stöckle/Sabine Wahl
Lexicography and corpus linguistics
The case of the Dictionary of Bavarian Dialects in Austria (WBÖ) and its database 105

Anna Vacalopoulou/Eleni Efthimiou/Stavroula-Evita Fotinea/Theodoros Goulas/Athanasia-Lida Dimou/Kiki Vasilaki
Organizing a bilingual lexicographic database with the use of WordNet........ 108

Dictionary Writing Systems and Lexicographic Tools

Nico Dorn
An automated cluster constructor for a narrated dictionary
The cross-reference clusters of *Wortgeschichte digital* 110

Mireille Ducassé/Archil Elizbarashvili
Finding lemmas in agglutinative and inflectional language dictionaries with logical information systems
The case of Georgian verbs........ 111

Velibor Ilić/Lenka Bajčetić/Snežana Petrović/Ana Španović
SCyDia – OCR for Serbian Cyrillic with diacritics........ 112

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Julian Jarosch
Digitale Lexikografie mit Hausmitteln
Ein Fallbeispiel zum digitalen Arbeiten ohne Sachmittel 113

Dorielle Lonke/Ilan Kernerman/Vova Dzhuranyuk
Lexical data API 116

Takahiro Makino/Rei Miyata/Seo Sungwon/Satoshi Sato
Designing and building a Japanese controlled language for the automotive domain
Toward the development of a writing assistant tool 117

Adriane Orenha Ottaiano/Maria Eugênia Olímpio de Oliveira Silva/ Carlos Roberto Valêncio/João Pedro Quarado
Developing a Collocation Dictionary Writing System (COLDWS) for an Online Multilingual Collocations Dictionary Platform (PLATCOL) 118

Annabella Schmitz
RDF-Lifting von OntoLex-Lemon aus dem Digitalen Familiennamenwörterbuch Deutschlands mit XTriples 122

Alberto Simões/Ana Salgado
Smart dictionary editing with LeXmart 125

Chris A. Smith
Are phonesthemes evidence of a sublexical organising layer in the structure of the lexicon?
Testing the OED analysis of two phonesthemes with a corpus study of collocational behaviour of *sw-* and *fl-* words in the OEC 126

Tanara Zingano Kuhn/Špela Arhar Holdt/Rina Zviel Girshin/Ana R. Luís/Carole Tiberius/Kristina Koppel/Branislava Šandrih Todorović/ Iztok Kosem
Introducing CrowLL – the Crowdsourcing for Language Learning game 128

Design and Publication of Dictionaries

Bridgitte Le Du
Towards a user-centered design model to enhance usability in electronic lexicography
Some guiding principles applied in the adaptation of *A Dictionary of South African English on Historical Principles* 132

Zita Hollós
Cross-Media-Publishing in der korpusgestützten Lernerlexikographie
Entstehung eines Lernerwörterbuchportals DaF 134

(Promoting) Dictionary Use

Andrea Abel
Wörterbücher der Zukunft in Bildungskontexten der Gegenwart
Eine Fallstudie aus dem Südtiroler Schulwesen 136

Carolina Flinz/Sabrina Ballestracci
Das LBC-Wörterbuch: Eine erste Benutzerstudie 137

Valeria Caruso/Alessandra Chervino/Giulia Daniele
Disseminating dictionary skills with e-lex tools 138

Ida Dringó-Horváth/Katalin P. Márkus
Dictionary skills in teaching English and German as a foreign language in Hungary
A questionnaire study 140

Zoe Gavriilidou/Evi Konstandinidou
The effect of an explicit and integrated dictionary awareness intervention program on dictionary use strategies 143

Theresa Kruse/Ulrich Heid
Learning from students
On the design and usability of an e-dictionary of mathematical graph theory........ 145

Sascha Wolfer/Robert Lew
Predicting English Wiktionary consultations........ 146

Manuel Raaf
Evaluation des User-Centered Designs eines Sprachinformationssystems
Planung, Durchführung und Ergebnisse einer Benutzerumfrage zu Usability und User Experience 149

Geraint Paul Rees
Online dictionaries and accessibility for people with visual impairments 152

Silga Sviķe
Survey analysis of dictionary-using skills and habits among translation students........ 155

Carole Tiberius/Jelena Kallas/Svetla Koeva/Margit Langemets/Iztok Kosem
An insight into lexicographic practices in Europe
Results of the extended ELEXIS survey on user needs........ 156

Agnes Wigestrand Hoftun
Consultation behavior in L1 error correction
An exploratory study on the use of online resources in the Norwegian context........ 157

Dictionary Projects

Hauke Bartels
The long road to a historical dictionary of Lower Sorbian
Towards a lexical information system 159

Hanno Biber
“Bloody word ripping” – practical and theoretical prospects of a corpus-based lexicographic exploration of the Texts by Thomas Bernhard........ 160

Isidora Despotidou/Zoe Gavriilidou
An online school dictionary in Greek Sign Language for senior elementary pupils 162

Carolina Flinz/Laura Giacomini/Weronika Szemińska
TermiKnowledge: Ein Einblick in die Datenbeschaffung und Datenaufbereitung eines Online-Fachwörterbuchs zum Thema COVID-19........ 164

Polona Gantar/Simon Krek
Creating the lexicon of multi-word expressions for Slovene
Methodology and structure 168

Zoe Gavriilidou/Apostolos Garoufos
The lexicographic protocol of Mikaela_Lex: A free online school dictionary of Greek accessible for visually-impaired senior elementary children 169

Vanessa González Ribao
Fachlexikografie in digitalem Zeitalter
Ein metalexikografisches Forschungsprojekt 170

Peter Meyer
Lehnwortportal Deutsch: a new architecture for resources on lexical borrowings 171

Iryna Ostapova/Volodymyr Shyrokov/Yevhen Kupriianov/Mykyta Yablochkov
Etymological dictionary in digital environment 172

Anna Pavlova
Mehrsprachige Datenbank der Phrasem-Konstruktionen 173

Ralf Plate
Word Families in Diachrony
An epoch-spanning structure for the word families of Older German 174

Kyriaki Salveridou/Zoe Gavriilidou
Compilation of an ancient Greek – modern Greek online thesaurus for teaching purposes: microstructure and macrostructure 175

Bilingual Dictionaries

Voula Giouli/Anna Vacalopoulou/Nikos Sidiropoulos/Christina Flouda/ Athanasios Doupas/Gregory Stainhaouer
From mythos to logos: A bilingual thesaurus tailored to meet users' needs within the ecosystem of cultural tourism 177

Iztok Kosem
The Comprehensive Slovenian-Hungarian Dictionary:Bilingual lexicography meets monolingual lexicography 178

Anke Müller/Gabriele Langer/Felicitas Otte/Sabrina Wähl
Creating a dictionary of a signed minority language
A bilingualized monolingual dictionary of German Sign Language 180

Specialised Dictionaries

Maria Aldea
Bien écrire, bien parler au XIX^{e} siècle. le rôle du dictionnaire dans l'apprentissage de la langue maternelle: Le cas du roumain 182

Harald Bichlmeier
Almanca tuhfe/Deutsches Geschenk (1916)
oder: Wie schreibt man deutsch mit arabischen Buchstaben? 183

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/ Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Walter Amaru Flores Flores/Daniel Kroiß
Das Digitale Familiennamenwörterbuch Deutschlands (DFD) 184

Dominika Kováříková/Michal Škrabal
The dictionary of Czech core academic vocabulary 186

Lorna Morris
The treatment of human reproductive organs in school dictionaries, with recommendations for South African primary school dictionaries 189

María Pozzi
Design of a dictionary to help school children to understand basic mathematical concepts 191

Stefan J. Schierholz/Monika Bielinska/Maria José Domínguez Vázquez/ Rufus H. Gouws/Martina Nied Curcio
The EMLex Dictionary of Lexicography (EMLexDictoL) 192

Historical Lexicography: German

Andreas Deutsch
Fremdwörter im Deutschen Rechtswörterbuch (DRW) 194

Volker Harm
Wortgeschichte digital: A historical dictionary of New High German 197

Andrea Moshövel
Skatologischer Wortschatz im Frühneuhochdeutschen als kulturgeschichtliche und lexikographische Herausforderung 198

Historical Lexicography: Romance and Other Languages

Maria Arapopoulou/Georgios Kalafikis/Dimitra Karamitsou/Efstratios Sarischoulis/Sotiris Tselikas
“Vocabula Grammatica”: threading a digital Ariadne’s String in the labyrinth of Ancient Greek scholarship 200

Anaïs Chambat
La lignée « Capuron-Nysten-Littré » Entre ruptures et continuités doctrinales 201

Sarah Mantegna/Carla Marello
The multilingual appendix of Le ricchezze della lingua volgare (1543) by Francesco Alunno
A lexicographer’s “service list” and an intercomprehension tool 202

Mihai-Alex Moruz/Mădălina Ungureanu
17th-century Romanian lexical resources and their influence on Romanian written tradition 206

Clarissa Stincone
Usage labels in Basnage’s *Dictionnaire universel* (1701) 207

Marija Žarković
The legal lexicon in the first dictionary of the Spanish Royal Academy (1726–1739)
The concept of the judge 208

(Historical) Lexicology

Rita Calabrese/Katherine E. Russo
Metaphorical constructions in Indian English and Australian Aboriginal English
From compositionality to grammaticalization 210

Zoe Gavriilidou/Asimakis Fliatouras/Elina Chadjipapa
Arabic loanwords in Greek 213

Ellert Thor Johannsson
Old words and obsolete meanings in Modern Icelandic 215

Claudia Lauer/Birgit Herbers
Hapax legomena in der deutschsprachigen Literatur des Mittelalters. Bedingungen, Verfahren und Bedeutungen
Ein Projektbericht 216

Geda Paulsen/Ene Vainik/Maria Tuulik/Ahti Lohk
The morphosyntactic profile of prototypical adjectives in Estonian 218

Pius ten Hacken/Renáta Panocová
The etymology of internationalisms
Evidence from German and Slovak 221

Neologisms and Lexicography

Ieda Maria Alves/Bruno Maroneze
From society to neology and lexicography
Relationships between morphology and dictionaries 223

Mikyung Baek/Jinsan An/Yelin Go
A distributional approach to Korean semantic neologisms
Identifying their first occurrences and investigating their spread 224

Jun Choi/Hae-Yun Jung
On loans in Korean new word formation and in lexicography 228

Emmanuel Cartier
Diachronic semantic evolution automatic tracking: a pilot study in modern and contemporary French combining dependency analysis and contextual embeddings 229

Lars Trap-Jensen/Henrik Lorentzen
Recent neologisms provoked by COVID-19 in the Danish language and in The Danish Dictionary 232

Gilles-Maurice de Schryver/Minah Nabirye
Towards a monitor corpus for a Bantu language
A case study of neology detection in Lusoga 233

Urška Vranjek Ošlak/Helena Dobrovoljc
Neologisms in the light of the new Slovenian normative guide ... 234

Phraseology & Collocations

Amparo Alcina
Representing collocations using ontologies ... 238

Valeria Caruso/Angela Caiazza
Lexicography and phraseology of Romance languages
The case of procomplement verbs ... 242

Maria Ermakova/Alexander Geyken/Lothar Lemnitzer/Bernhard Roll
Integration of multi-word expressions into the digital dictionary of German Language (DWDS)
Towards a lexicographic representation of phraseological variation ... 245

Larysa Kovbasyuk
Corona bekennt Farbe: Phraseologische Neologismen im Deutschen und Ukrainischen aus kulturlinguistischer Sicht ... 246

Semantics

Paolo DiMuccio-Failla
Disambiguating word senses through semantic conditions
A project in learner's lexicography ... 250

Daniele Franceschi
Lexicographic representations of Anglo-Saxon and Latinate near-synonyms in English monolingual and English-Italian bilingual learners' dictionaries ... 253

Robert Krovetz
An investigation of sense ordering across dictionaries with respect to lexical semantic relationships ... 256

Index of Authors

Index of authors ... 258

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Acknowledgements

We would like to thank all those who have supported the XX EURALEX International Conference financially:

Main Sponsors

Funded by
DFG Deutsche Forschungsgemeinschaft
German Research Foundation

Hornby
A. S. Hornby Educational Trust

elexis
european lexicographic infrastructure

IDS LEIBNIZ-INSTITUT FÜR DEUTSCHE SPRACHE
Freunde des Leibniz-Instituts für Deutsche Sprache e. V.

Sponsors

BLOOMSBURY ACADEMIC

DE G DE GRUYTER

DUDEN

ESV ERICH SCHMIDT VERLAG

F Frank & Timme

OED
Oxford English Dictionary

BUSKE

lexicala
BY K DICTIONARIES

Universitätsverlag
WINTER
Heidelberg

We would like to thank all those who have contributed to reviewing the submissions and papers:

Programme Committee

Gilles-Maurice **de Schryver** (Ghent University, Belgium & University of Pretoria, South Africa)

Stefan **Engelberg** (The Leibniz Institute for the German Language, Germany)

Annette **Klosa-Kückelhaus** (The Leibniz Institute for the German Language, Germany)

Iztok **Kosem** (Jožef Stefan Institute / University of Ljubljana, Slovenia, Germany)

Robert **Lew** (Adam Mickiewicz University, Poland)

Christine **Möhrs** (The Leibniz Institute for the German Language, Germany)

Petra **Storjohann** (The Leibniz Institute for the German Language, Germany)

Kristina **Štrkalj Despot** (Institute of Croatian Language and Linguistics, Croatia)

Scientific Committee

Andrea **Abel** (EURAC, Italy)

Arleta **Adamska-Sałaciak** (Adam Mickiewicz University, Poland)

Hauke **Bartels** (Sorbian Institute, Germany)

Hans **Bickel** (Schweizerisches Idiotikon, Switzerland)

Anna **Braasch** (University of Copenhagen, Denmark)

Dominik **Brückner** (The Leibniz Institute for the German Language, Germany)

Thomas **Burch** (Trier Center for Digital Humanities, Germany)

Lut **Colman** (Dutch Language Institute, Netherlands)

Paul **Cook** (University of New Brunswick, Canada)

Gilles-Maurice **de Schryver** (Ghent University, Belgium & University of Pretoria, South Africa)

Janet **DeCesaris** (Pompeu Fabra University, Spain)

Idalete Maria Silva **Dias** (University of Minho, Portugal)

María José **Domínguez Vázquez** (University of Santiago de Compostela, Spain)

Philip **Durkin** (Oxford University Press, Great Britain)

Anne **Dykstra** (Fryske Academy, Netherlands)

Anna **Dziemianko** (Adam Mickiewicz University, Poland)

Ilse **Feinauer** (Stellenbosch University, South Africa)

Edward **Finegan** (University of Southern California, USA)

Carolina **Flinz** (University of Milan, Italy)

Thierry **Fontenelle** (European Investment Bank, Belgium)

Polona **Gantar** (University of Ljubljana, Slovenia)

Zoe **Gavriilidou** (Democritus University of Thrace, Greece)

Alexander **Geyken** (Berlin-Brandenburg Academy of Sciences, Germany)

Sylviane **Granger** (Catholic University of Louvain, Belgium)

Oddrun **Grønvik** (University of Oslo, Norway)

Volker **Harm** (The Göttingen Academy of Sciences and Humanities, Germany)

Ulrich **Heid** (Hildesheim University, Germany)

Zita **Hollós** (Károli Gáspár University, Hungary)

Miloš **Jakubíček** (Lexical Computing CZ s.r.o., Czech Republic)

Maarten **Janssen** (University of Vienna, Austria)

Besim **Kabashi** (Friedrich-Alexander University Erlangen, Germany)

Jelena **Kallas** (Institute of the Estonian Language, Estonia)

Heidrun **Kämper** (The Leibniz Institute for the German Language, Germany)

Ilan **Kernerman** (K Dictionaries, Israel)

Alexander **Koplenig** (The Leibniz Institute for the German Language, Germany)

Iztok **Kosem** (Jožef Stefan Institute / University of Ljubljana, Slovenia)

Simon **Krek** (Jožef Stefan Institute / University of Ljubljana, Slovenia)

Tanara Zingano **Kuhn** (University of Coimbra, Portugal)

Kathrin **Kunkel-Razum** (Duden-Verlag, Germany)

Margit **Langemets** (Institute of the Estonian Language, Estonia)

Lothar **Lemnitzer** (Berlin-Brandenburg Academy of Sciences, Germany)

Robert **Lew** (Adam Mickiewicz University, Poland)

Marie-Claude **L'Homme** (University of Montreal, Canada)

Anja **Lobenstein-Reichmann** (The Göttingen Academy of Sciences and Humanities, Germany)

Henrik **Lorentzen** (The Danish Language and Literature Society, Denmark)

Carla **Marello** (University of Turin, Italy)

Tinatin **Margalitadze** (Ilia State University, Georgia)

John P. **McCrae** (National University of Ireland, Ireland)

Peter **Meyer** (The Leibniz Institute for the German Language, Germany)

Frank **Michaelis** (The Leibniz Institute for the German Language, Germany)

Julia **Miller** (University of Adelaide, Australia)

Fabio **Mollica** (University of Milan, Italy)

Orion **Montoya** (Brandeis University, USA)

Rosamund **Moon** (University of Birmingham, Great Britain)

Carolin **Müller-Spitzer** (The Leibniz Institute for the German Language, Germany)

Kilim **Nam** (Kyungpook National University, South Korea)

Hilary **Nesi** (Coventry University, Great Britain)

Vincent **Ooi** (National University of Singapore, Singapore)

Maike **Park** (The Leibniz Institute for the German Language, Germany)

Ralf **Plate** (The Academy of Sciences and Literature Mainz / University of Trier, Germany)

Kristel **Proost** (The Leibniz Institute for the German Language, Germany)

Natascia **Ralli** (EURAC, Italy)

Stefan J. **Schierholz** (Friedrich-Alexander University Erlangen, Germany)

Thomas **Schmidt** (The Leibniz Institute for the German Language, Germany)

Hindrik **Sijens** (Fryske Academy, Netherlands)

Egon W. **Stemle** (EURAC, Italy)

Frieda **Steurs** (Dutch Language Institute, Netherlands)

Kathrin **Steyer** (The Leibniz Institute for the German Language, Germany)

Philipp **Stöckle** (Austrian Academy of Sciences, Austria)

Kristina **Štrkalj Despot** (Institute of Croatian Language and Linguistics, Croatia)

Janusz **Taborek** (Adam Mickiewicz University, Poland)

Elsabé **Taljard** (University of Pretoria, South Africa)

Pius **ten Hacken** (University of Innsbruck, Austria)

Carole **Tiberius** (Dutch Language Institute, Netherlands)

Yukio **Tono** (Tokyo University of Foreign Studies, Japan)

Lars **Trap-Jensen** (The Danish Language and Literature Society, Denmark)

Anna **Vacalopoulou** (Institute for Language and Speech Processing, Greece)

Carlos **Valcárcel Riveiro** (University of Vigo, Spain)

Ruth **Vatvedt Fjeld** (University of Oslo, Norway)

Craig **Volker** (James Cook University Cairns, Australia)

Sabine **Wahl** (Austrian Academy of Sciences, Austria)

Geoffrey **Williams** (Université Bretagne Sud, France)

Sascha **Wolfer** (The Leibniz Institute for the German Language, Germany)

Part I: Overview on Keynotes, Talks, Poster Presentations, and Software Demonstrations

Keynotes

Thomas Gloning: Ways of living, communication and the dynamics of word usage. How did German dictionaries cope with socio-cultural aspects and evolution of word usage and how could future systems do even better?

Rufus Gouws: Dictionaries: bridges, dykes and sluice gates

Nicola McLelland: Women in the history of lexicography – an overview, and the case of German

Martina Nied Curcio: Dictionaries, foreign language learners and teachers. New challenges in the digital era

Ben Zimmer: The evolving definition of “racism” and its trail of text-artifacts

Talks

Andrea Abel: Wörterbücher der Zukunft in Bildungskontexten der Gegenwart – Eine Fallstudie aus dem Südtiroler Schulwesen

Amparo Alcina: Representing collocations using ontologies

Maria Aldea: Bien écrire, bien parler au XIX^{e} siècle. Le rôle du dictionnaire dans l’apprentissage de la langue maternelle: Le cas du roumain

Ieda Maria Alves/Bruno Maroneze: From society to neology and lexicography: relationships between morphology and dictionaries

Maria Arapopoulou/Georgios Kalafikis/Dimitra Karamitsou/Efstratios Sarischoulis/Sotiris Tselikas: “Vocabula Grammatica”: threading a digital Ariadne’s string in the labyrinth of Ancient Greek scholarship

Hauke Bartels: The long road to a historical dictionary of Lower Sorbian – towards a lexical information system

Harald Bichlmeier: *Almanca tuhfe / Deutsches Geschenk* (1916) – oder: Wie schreibt man deutsch mit arabischen Buchstaben?

Rita Calabrese/Katherine E. Russo: Metaphorical constructions in Indian English and Australian Aboriginal English: from compositionality to grammaticalization

Emmanuel Cartier: Diachronic semantic evolution automatic tracking: a pilot study in modern and contemporary French combining dependency analysis and contextual embeddings

Valeria Caruso/Angela Caiazza: Lexicography and phraseology of Romance languages: the case of procomplement verbs

Valeria Caruso/Alessandra Chervino/Giulia Daniele: Disseminating dictionary skills with e-lex tools

Anaïs Chambat: La lignée « Capuron-Nysten-Littré »entre ruptures et continuités doctrinales

Mark Dang-Anh/Stefan Scholl: Basic socio-political concepts in German parliamentary debates from the 20th century

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Andreas Deutsch: Fremdwörter im Deutschen Rechtswörterbuch (DRW)

Paolo DiMuccio-Failla: Disambiguating word senses through semantic conditions: a project in learner's lexicography

Ida Dringó-Horváth/Katalin P. Márkus: Dictionary skills in teaching English and German as a foreign language in Hungary – a questionnaire study

Bridgitte Le Du: Towards a user-centered design model to enhance usability in electronic lexicography: some guiding principles applied in the adaptation of *A Dictionary of South African English on Historical Principles*

Stefan Engelberg: Lexicography's entanglement with colonialism: the history of Tok Pisin lexicography as colonial history

Maria Ermakova/Alexander Geyken/Lothar Lemnitzer/Bernhard Roll: Integration of multiword expressions into the digital dictionary of German Language (DWDS) – towards a lexicographic representation of phraseological variation

Ivana Filipović Petrović: A corpus-driven approach to lexicographic definitions: the representation of meaning in the electronic Dictionary of Croatian Idioms

Carolina Flinz/Sabrina Ballestracci: Das LBC-Wörterbuch: Eine erste Benutzerstudie

Walter Amaru Flores Flores/Daniel Kroiß: Das Digitale Familiennamenwörterbuch Deutschlands (DFD)

Daniele Franceschi: Lexicographic representations of Anglo-Saxon and Latinate near-synonyms in English monolingual and English-Italian bilingual learners' dictionaries

Christine Ganslmayer: Using dialect dictionaries as a data base: stereotypes of people in the ‚Franconian Dictionary'

Polona Gantar/Simon Krek: Creating the lexicon of multi-word expressions for Slovene: methodology and structure

Zoe Gavriilidou/Asimakis Fliatouras/Elina Chadjipapa: Arabic loanwords in Greek

Zoe Gavriilidou/Evi Konstandinidou: The effect of an explicit and integrated dictionary awareness intervention program on dictionary use strategies

Laura Giacomini/Paolo DiMuccio-Failla/Patrizio De Martin Pinter: The representation of culture-specific lexical items in monolingual learner's lexicography: the case of the electronic Phrase-Based Active Dictionaries

Voula Giouli/Anna Vacalopoulou/Nikos Sidiropoulos/Christina Flouda/Athanasios Doupas/Gregory Stainhaouer: From mythos to logos: a bilingual thesaurus tailored to meet users' needs within the ecosystem of cultural tourism

Pius ten Hacken/Renáta Panocová: The etymology of internationalisms: evidence from German and Slovak

Volker Harm: *Wortgeschichte digital*: a historical dictionary of New High German

Zita Hollós: Cross-Media-Publishing in der korpusgestützten Lernerlexikographie. Entstehung eines Lernerwörterbuchportals DaF

Julian Jarosch: Digitale Lexikografie mit Hausmitteln – ein Fallbeispiel zum digitalen Arbeiten ohne Sachmittel

Ellert Thor Johannsson: Old words and obsolete meanings in modern Icelandic

Jun Choi/Hae-Yun Jung: On loans in Korean new word formation and in lexicography

Annette Klosa-Kückelhaus: Lexicography for society and with society – COVID-19 and dictionaries

Iztok Kosem: The Comprehensive Slovenian-Hungarian Dictionary: bilingual lexicography meets monolingual lexicography

Iztok Kosem: Trendi – a monitor corpus of Slovene

Konan Kouassi: Mensch-Maschine-Interaktion im lexikographischen Prozess zu lexikalischen Informationssystemen

Dominika Kováříková/Michal Škrabal: The Dictionary of Czech Core Academic Vocabulary

Simon Krek/Polona Gantar/Iztok Kosem: Extraction of collocations from the Gigafida 2.1 corpus of Slovene

Robert Krovetz: An investigation of sense ordering across dictionaries with respect to lexical semantic relationships

Theresa Kruse/Ulrich Heid: Learning from students: on the design and usability of an e-dictionary of mathematical graph theory

Yevhen Kupriianov/Iryna Ostapova/Volodymyr Shyrokov/Mykyta Yablochkov: Virtual Lexicographic Laboratory as a linguist assistant in conducting dictionary-based researches: Case of VLL DLE 23

Claudia Lauer/Birgit Herbers: *Hapax legomena* in der deutschsprachigen Literatur des Mittelalters. Bedingungen, Verfahren und Bedeutungen – ein Projektbericht

David Lindemann/Penny Labropoulou/Christiane Klaes: Introducing LexMeta: a metadata model for lexical resources

Gloria Mambelli: Manorial society in multilingual medieval England: an onomasiological approach

Takahiro Makino/Rei Miyata/Seo Sungwon/Satoshi Sato: Designing and building a Japanese controlled language for the automotive domain: toward the development of a writing assistant tool

Manuel Márquez: Un modelo estructural de datos lexicográficos para la codificación en XML de un diccionario de aprendizaje de latín: la DTD de los lemas verbales

Michal Měchura: Document or database? The search for the perfect storage paradigm for lexical data

Lorna Morris: The treatment of human reproductive organs in school dictionaries, with recommendations for South African primary school dictionaries

Mihai-Alex Moruz/Mădălina Ungureanu: 17th century Romanian lexical resources and their influence on Romanian written tradition

Andrea Moshövel: Skatologischer Wortschatz im Frühneuhochdeutschen als kulturgeschichtliche und lexikographische Herausforderung

Anke Müller/Gabriele Langer/Felicitas Otte/Sabrina Wähl: Creating a dictionary of a signed minority language: a bilingualized monolingual dictionary of German Sign Language

Carolin Müller-Spitzer/Jan Oliver Rüdiger: The influence of the corpus base on the representation of gender stereotypes in the dictionary. A case study for corpus-based dictionaries of German

Enakshi Nandi: Secrecy and the ethical question: some reflections on documenting Ulti, a secret transgender language

Adriane Orenha Ottaiano/Maria Eugênia Olímpio de Oliveira Silva/Carlos Roberto Valêncio/João Pedro Quadrado: Developing a Collocation Dictionary Writing System (COLDWS) for an Online Multilingual Collocations Dictionary Platform (PLATCOL)

Iryna Ostapova/Volodymyr Shyrokov/Yevhen Kupriianov/Mykyta Yablochlov: Etymological dictionary in digital environment

Ana Ostroški Anić/Ivana Brač: AirFrame: mapping the field of aviation through semantic frames

Geda Paulsen/Ene Vainik/Maria Tuulik/Ahti Lohk: The morphosyntactic profile of prototypical adjectives in Estonian

Anna Pavlova: Mehrsprachige Datenbank der Phrasem-Konstruktionen

Laura Pinnavaia: Identifying ideological strategies in the making of English language learners' dictionaries

Ralf Plate: Word Families in Diachrony. An epoch-spanning structure for the word families of Older German

María Pozzi: Design of a dictionary to help school children to understand basic mathematical concepts

Kristel Proost/Arne Zeschel/Frank Michaelis/Jan Oliver Rüdiger: MAP (Musterbank Argumentmarkierender Präpositionen): A patternbank of argument-marking prepositions in German

Manuel Raaf: Evaluation des User-Centered Designs eines Sprachinformationssystems: Planung, Durchführung und Ergebnisse einer Benutzerumfrage zu Usability und User Experience

Geraint Paul Rees: Online dictionaries and accessibility for people with visual impairments

Irene Renau/Rogelio Nazar: Towards a multilingual dictionary of discourse markers: automatic extraction of units from parallel corpus

Ana Salgado/Rute Costa/Toma Tasovac: Applying terminological methods to lexicographic work: terms and their domains

Kyriaki Salveridou/Zoe Gavriilidou: Compilation of an Ancient Greek – Modern Greek online thesaurus for teaching purposes: microstructure and macrostructure

Stefan J. Schierholz/Monika Bielinska/Maria José Domínguez Vázquez/Rufus H. Gouws/Martina Nied Curcio: The EMLex Dictionary of Lexicography (EMLexDictoL)

Gilles-Maurice de Schryver: Metalexicography: an existential crisis

Gilles-Maurice de Schryver/Minah Nabirye: Towards a monitor corpus for a Bantu language: a case study of neology detection in Lusoga

Alberto Simões/Ana Salgado: Smart dictionary editing with LeXmart

Michal Škrabal/Michaela Lišková/Martin Šemelík: On defining vocabulary in a monolingual online dictionary. Some remarks from the lexicographical practice on the *Academic Dictionary of Contemporary Czech*

Christian-Emil Smith Ore/Oddrun Grønvik/Trond Minde: Word banks, dictionaries and research results by the roadside

Vasyl Starko: USL: a cognitively inspired lexicon for semantic tagging

Clarissa Stincone: Usage labels in Basnage de Beauval's Dictionnaire universel of 1701

Philipp Stöckle/Sabine Wahl: Lexicography and corpus linguistics – the case of the Dictionary of Bavarian Dialects in Austria (WBÖ) and its database

Petra Storjohann: The public as linguistic authority: why users turn to internet forums to differentiate between words

Carole Tiberius/Jelena Kallas/Svetla Koeva/Margit Langemets/Iztok Kosem: An insight into lexicographic practices in Europe: results of the extended ELEXIS Survey on User Needs

Lars Trap-Jensen/Henrik Lorentzen: Recent neologisms provoked by COVID-19 – in the Danish language and in The Danish Dictionary

Anna Vacalopoulou/Eleni Efthimiou/Stavroula-Evita Fotinea/Theodoros Goulas/Athanasia-Lida Dimou/Kiki Vasilaki: Organizing a bilingual lexicographic database with the use of WordNet

Urška Vranjek Ošlak/Helena Dobrovoljc: Neologisms in the light of the new Slovenian normative guide

Agnes Wigestrand Hoftun: Consultation behavior in L1 error correction: an exploratory study on the use of online resources in the Norwegian context

Marcin Zabawa: What do we learn about the society from the examples of usage in dictionaries? On (non-)stereotypical roles of men and women in English and Polish monolingual general dictionaries: a contrastive study

Marija Žarković: The legal lexicon in the first dictionary of the Royal Spanish Academy (1726–1739) – The concept of the judge

Poster Presentations

Mariona Arnau Garcia/Mercè Lorente Casafont: Are emerging economies a reality reflected in our dictionaries?

Mikyung Baek/Jinsan An/Yelin Go: A distributional approach to Korean semantic neologisms: identifying their first occurrences and investigating their spread

Laura Balbiani/Anne-Kathrin Gärtig-Bressan/Martina Nied Curcio/Stefan Schierholz: Dictionaries for the future – the future of dictionaries: the 15 Villa Vigoni Theses on Lexicography

Hanno Biber: "Bloody word ripping" – practical and theoretical prospects of a corpus-based lexicographic exploration of the texts by Thomas Bernhard

Thierry Declerck: Integration of sign language lexical data in the OntoLex-Lemon framework

Isidora Despotidou/Zoe Gavriilidou: An online school dictionary in Greek Sign Language for senior elementary pupils

Nils Diewald/Marc Kupietz/Harald Lüngen: Tokenizing on scale – preprocessing large text corpora on the lexical and sentence level

Alexandre Ecker: Equality between women and men, a societal and lexicographical issue – Bringing the content of the *Lëtzebuerger Online Dictionnaire* in line with the realities of the world and society

Carolina Flinz/Laura Giacomini/Weronika Szemińska: TermiKnowledge: Ein Einblick in die Datenbeschaffung und Datenaufbereitung eines Online-Fachwörterbuchs zum Thema COVID-19

Birgit Füreder: Überlegungen zur Modellierung eines multilingualen 'Periphrastikons' – ein französisch-italienisch-spanisch-englisch-deutscher Versuch

Zoe Gavriilidou/Apostolos Garoufos: The lexicographic protocol of Mikaela_Lex: a free online school dictionary of Greek accessible for visually-impaired senior elementary children

Ana-Maria Gînsac/Mihai-Alex Moruz/Mădălina Ungureanu: The first Romanian dictionaries (17th century). Digital aligned corpus

Vanessa González Ribao: Fachlexikografie in digitalem Zeitalter: Eine metalexikografisches Forschungsprojekt

Velibor Ilić/Lenka Bajčetić/Snežana Petrović/Ana Španović: SCyDia: OCR for Serbian Cyrillic with diacritics

Sarah Mantegna/Carla Marello: The multilingual appendix of *Le ricchezze della lingua volgare* (1543) by Francesco Alunno. A lexicographer's "service list" and an intercomprehension tool

Meike Meliss/Vanessa González Ribao: Vergleichbare Korpora für multilinguale kontrastive Studien: Herausforderungen und Desiderata

Chris Smith: Are phonesthemes evidence of a sublexical organising layer in the structure of the lexicon? Testing the OED analysis of two phonesthemes with a corpus study of collocational behaviour of *sw-* and *fl-* words in the OEC

Silga Sviķe: Survey analysis of dictionary-using skills and habits among translation students

Sascha Wolfer/Robert Lew: Predicting English Wiktionary consultations

Tanara Zingano Kuhn/Špela *Arhar Holdt/Rina Zviel Girshin/Ana R. Luis/Carole Tiberius/ Kristina Koppel/Branislava Šandrih Todorović/Iztok Kosem*: Introducing CrowLL – the Crowdsourcing for Language Learning game

Software Demonstrations

Nico Dorn: An automated cluster constructor for a narrated dictionary: the cross-reference clusters of „Wortgeschichte digital"

Mireille Ducassé/Archil Elizbarashvili: Finding lemmas in agglutinative and inflectional language dictionaries with logical information systems: the case of Georgian verbs

Dorielle Lonke/Ilan Kernerman/Vova Dzhuranyuk: Lexical data API

Larysa Kovbasyuk: Corona bekennt Farbe: phraseologische Neologismen im Deutschen und Ukrainischen aus kulturlinguistischer Sicht

Peter Meyer: Lehnwortportal Deutsch: a new architecture for resources on lexical borrowings

Jan Oliver Rüdiger/Sascha Wolfer/Alexander Koplenig/Frank Michaelis/Carolin Müller-Spitzer/Samira Ochs/Louis Cotgrove: OWIDplusLIVE – Day-to-day collection, exploration, analysis, and visualization of N-Gram frequencies in German (online press) language

Annabella Schmitz: RDF-Lifting von OntoLex-Lemon aus dem Digitalen Familiennamenwörterbuch Deutschlands mit XTriples

Part II:
Abstracts of Keynotes

Thomas Gloning

WAYS OF LIVING, COMMUNICATION AND THE DYNAMICS OF WORD USAGE

How did German dictionaries cope with socio-cultural aspects and evolution of word usage and how could future systems do even better?

Abstract Words and their usages are in many cases closely related to or embedded in social, cultural, technical and ideological contexts. This does not only apply to individual words and specific senses, but to many vocabulary zones as well. Moreover, the development of words is often related to aspects of socio-cultural evolution in a broad sense. In this paper I will have a look at traditional dictionaries and digital lexical systems focussing on the question how they deal with socio-cultural and discourse-related aspects of word usage. I will also propose a number of suggestions how future digital lexical systems might be enriched in this respect.

Keywords Digital lexical systems; faceted search; vocabulary organization in dictionaries, forms of representation in digital lexicography

Contact information

Thomas Gloning
Justus-Liebig-Universität Gießen
thomas.gloning@germanistik.uni-giessen.de

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Rufus H. Gouws

DICTIONARIES: BRIDGES, DYKES, SLUICE GATES

Abstract In a multilingual and multicultural society, dictionaries play an important role to enhance interlingual communication. A diversity of languages and different levels of dictionary culture demand innovative lexicographic approaches to establish a dictionary landscape that responds to the needs of the various speech communities. Focusing on the South African situation this paper discusses some aspects of a few dictionaries that contributed to an improvement of the local dictionary landscape. Using the metaphors of bridges, dykes and sluice gates it is shown how lexicographers need a balanced approach in their lemma selection and treatment. Whilst a too strong prescriptive approach can be to the detriment of the macrostructural selection, a lack of regulatory criteria could easily lead to a data overload. The lexicographer should strive to give a reflection of the actual language use and enable the users to retrieve the information that can satisfy their specific communication and cognitive needs. Such lexicographic products will enrich and improve the dictionary landscape.

Keywords Bilingualised dictionary; dictionary+; dictionary culture; dictionary portal; monolingualised dictionary; prescription

Contact information

Rufus H. Gouws
Department of Afrikaans and Dutch
Stellenbosch University
rhg@sun.ac.za

Nicola McLelland

WOMEN IN THE HISTORY OF LEXICOGRAPHY

An overview, and the case of German

Abstract This paper first attempts a state-of-the art overview of what is known about women in the history of lexicography up to the early twentieth century. It then focusses more closely on the German and German-English lexicographical traditions to 1900, examining them from three different perspectives (following Russell's 2018 study of women in English lexicography): women as users and dedicatees of dictionaries; women as contributors to and compilers of lexicographical works; and (in a very preliminary way) women and female sexuality as represented in German/English bilingual dictionaries of the eighteenth and early nineteenth centuries.

Russell (2018) was able to identify some 24 dictionaries invoking women as patrons, dedicatees or potential users before 1700, and some 150 works in English lexicography by women between 1500 and 1900, besides the contribution of hundreds of women as supporters and helpers, not least as unpaid readers and sub-editors for the *Oxford English Dictionary*. Equivalent research in other languages is lacking, but this paper presents some of the known examples of women as lexicographers. The evidence tends to support Russell's finding for English, that women were more likely to find a place in lexicography outside the mainstream: sometimes in a more private sphere (like Hester Piozzi); often in bilingual lexicography (such as Margrethe Thiele, working on a Danish-French dictionary), including missionary and or colonizing activity (such as Cinie Louw in Africa, Daisy Bates in Australia); and in dialect description (Coronedi Berti in Italy, Luisa Lacal and María Moliner in Spain).

Within the German-speaking context, women who participated in lexicographical work themselves are hard to identify before the late nineteenth century, though those few women who did have access to education were often engaged in language learning, including translation activity, and they were likely users of bilingual and multilingual dictionaries. Christian Ludwig's (1706) English-German dictionary – the first of its kind – was dedicated to the Electoral Princess Sophia of Hanover. Elizabeth Weir may have been the first named female compiler of a German dictionary, with her bilingual *New German Dictionary* (1888). Rather better known are the cases of Agathe Lasch and Luise Pusch, who, as pioneering women in the field of German linguistics, ultimately led major lexicographical projects documenting German regional varieties in the first half of the twentieth century (Middle Low German and Hamburgish in the case of Lasch; the Hessisch-Nassau dialect dictionary in the case of Berthold).

In the light of existing research on gender and sexuality in the history of English lexicography (e.g. Iamartino 2010, Turton 2019), I conclude with a preliminary exploration how woman and sexuality have been represented in dictionaries of German and English, taking the words *Hure* and *woman* in bilingual German-English dictionaries of the eighteenth and nineteenth centuries as my case studies.

Keywords Lexicography, German, women, Hester Piozzi, Margrethe Thiele, Cinie Louw, Theodor Arnold, Christian Ludwig, Elizabeth Weir

References

Russell, L. R. (2018): Women and dictionary making: gender, genre, and English language lexicography. Cambridge.

Iamartino, G. (2010): Words by women, words on women in Samuel Johnson's dictionary of the English language. In: Considine, J. (ed): Adventuring in dictionaries: new Studies in the history of lexicography. Cambridge, pp. 94–124.

Ludwig, C. (1706): A dictionary English, German, and French […]. Leipzig.

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Turton, S. (2019): Unlawful entries: buggery, sodomy, and the construction of sexual normativity in early English dictionaries. In: Dictionaries: Journal of the Dictionary Society of North America 40 (1), pp. 81–112.

Weir, E. (1888): Heath's new German dictionary: in two parts, German-English--English-German. Boston. [Also 1888 as Cassell's new German fictionary [...]. London.]

Contact information

Nicola McLelland
University of Nottingham
nicola.mclelland@nottingham.ac.uk

Martina Nied Curcio

DICTIONARIES, FOREIGN LANGUAGE LEARNERS AND TEACHERS

New challenges in the digital era

Abstract In foreign language teaching the use of dictionaries, especially bilingual, has always been related to the hypotheses concerning the relationship between the native language (L1) and second language acquisition method. If the bilingual dictionary was an obvious tool in the grammar-translation method, it was banned from the classroom in the direct, audiolingual and audiovisual methods. Also in the communicative method, foreign language learners are discouraged from using a dictionary. Its use should not obstruct the goals of communicatively oriented foreign language learning – a view still held by many foreign language teachers.

Nevertheless, the reality has been different: Foreign language learners have always used dictionaries, even if they no longer possess a print dictionary and mainly use online resources and applications. Dictionaries and online resources will continue to play an important role in the future. In the Council of Europe's language policy, with its emphasis on multilingualism and lifelong learning, the adequate use of reference tools as a strategic skill is highlighted. In several European countries, educational guidelines refer to the use of dictionaries in the context of media literacy, both in mother tongue and foreign language teaching. Not only is their adequate use important, but so too is the comparison, assessment and evaluation of the information presented, in order to develop Language Awareness and Language Learning Awareness. This is good news. However, does this mean that dictionaries are actually used in class? What role do dictionaries play in foreign language teaching in schools and universities? Are foreign language learners in the digital era really competent users? And how competent are their teachers? Are they familiar with the current (online) dictionary landscape? Can they support their students? After a more in-depth study of the status quo of dictionary use by foreign language learners and teachers and the gap between their needs and the reality, this contribution discusses the challenges facing lexicographers and meta-lexicographers and what educational policy measures are necessary to make their efforts worthwhile in turning foreign language learners – and their teachers – into competent users in a multilingual and digital world.

Keywords Dictionaries; dictionary use; dictionary teaching; dictionary didactics; online resources; foreign language learner; foreign language teacher; language awareness; foreign language teaching; lifelong learning, reference tools, media literacy

Contact information

Martina Nied Curcio
Università degli Studi Roma Tre
martina.nied@uniroma3.it

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Ben Zimmer

THE EVOLVING DEFINITION OF „RACISM“ AND ITS TRAIL OF TEXT-ARTIFACTS

Abstract In 2020, Merriam-Webster announced that it would be updating the entry for “racism” in its online dictionary, partially in response to the critique of a young Black activist, Kennedy Mitchum. The revised entry foregrounded “systemic oppression” and “white supremacy,” further elucidating the significance of institutionalized racism, beyond what Camara Phyllis Jones has called “personally mediated” or “internalized” racism. While news coverage of the revision tended to portray the story as “the dictionary gets woke,” the history of how “racism” has been defined reveals a much more complex narrative. That narrative dates back to 1938 when lexicographers at Merriam-Webster first considered adding the word to its dictionaries, at a time when “racism” was chiefly associated with the policies of Nazi Germany. The publisher’s archives contain documentation of the in-house discussions about “racism” among editors at the time, in the form of handwritten slips with notes back and forth about how to define the word. The slips reveal the decision-making process that began when assistant editor Rose Frances Egan noticed that “racism” was missing from the second edition of the unabridged New International Dictionary, published in 1934. Egan’s discovery, made while she was preparing Webster’s Dictionary of Synonyms, set in motion an editorial chain of events resulting in the addition of “racism” to the Addenda section of the 1939 printing of the New Unabridged, a first for any major English dictionary. By analyzing the materiality of extant “text-artifacts,” to borrow a term from Michael Silverstein, we can better understand how the seemingly monolithic authority of “the dictionary” in fact consists of a series of editorial judgments by lexicographers at work. The practice of defining “racism” can be seen as emerging from a kind of communicative interplay, with each generation bringing its own discursive tools to the effort of framing and contesting the word’s definition.

Keywords Digital lexical systems; faceted search; vocabulary organization in dictionaries, forms of representation in digital lexicography

Contact information

Ben Zimmer
Linguist, lexicographer and language columnist, New York
bgzimmer@gmail.com

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Part III: Abstracts of Talks, Poster Presentations, and Software Demonstrations

Dictionaries and Society

XX EURALEX INTERNATIONAL CONGRESS
DICTIONARIES AND SOCIETY
12-16 July 2022, Mannheim, Germany

Mariona Arnau Garcia/Mercè Lorente Casafont

ARE EMERGING ECONOMIES A REALITY REFLECTED IN OUR DICTIONARIES?

Keywords Dictionaries; economics; emerging economies; new economies; ideology

Our world is constantly changing. These changes bring about new realities and have a direct impact on people's daily lives. And these new realities need words that represent them. If these new words respond to the designative and communicative needs of the community of speakers and if they are stabilised in use, they need to be included into reference dictionaries. That also happens in specialised fields, such as economics, a field which has the particularity of being part of the daily lives of speakers in general.

In times of crisis, the tendency to rethink the economic model is inherent to it. This is why in recent years, when it seems that the economic crisis has turned into a systemic crisis, some models have appeared that can be included within the umbrella of the "new economies" (Hernández/Serrano 2021). Terms such as "emerging economies", "social economies" or "conscious capitalism" are now leaving a mark on our society. Social economic models such as green economies, feminist economies or social and solidarity-based economies are increasingly being introduced into our everyday life. As this phenomenon takes place, and emerging economies are now leaving their mark on our society, traditional economies based on capitalism are no longer the only economic model in our society. In that sense we need to take in mind the concept of "economic ideology", which refers to a current of economics that expresses the perspective on the way in which the economy should work, always with a specific purpose. Also, it is important to mention that the followers of an economic ideology, think it is the correct one. That coincides with the notion that in discourses, even in lexicographical ones, any structure or strategy can have ideological marks, which denote a person's beliefs, but also can be used with a persuasive function (Van Dijk 1999).

To develop this experimental analysis, we need to consider that, sometimes, these concepts that we think are new, are not really new or are connected with concerns related to the establishment of a delimited scientific space or a specific current of thought (Chaves/Monzón 2018). We also need to bear in mind that there is a terminological plurality around these concepts that is not linked to a consensual conceptualisation (Defourny/Nyssens 2017).

Our contribution for this conference is part of a bigger study in which we intend to analyse from an experimental point of view general and terminological dictionaries in five languages (English, Catalan, Spanish, French and Italian). But for this contribution to test our methodology we will only address the dictionaries and corpora in Catalan (CAT) and English (ENG). We have created a corpus for each language with articles from the last three years (2019/2020/2021) of two specialised academic journals related with emerging economies: *Review of Social Economy* [ENG] and *Nexe* [CAT]. From these corpora, we automatically extracted the specialised terminology using Terminus 2.0, a web application for corpus and terminology management. Our hypothesis is that we will find terms that traditionally have not been associated with economics that now have these connotations. We will consult the extracted terms in general dictionaries (*DIEC2* in Catalan; *Oxford Advanced Learner's Dictionary* in English) in order to see if these terms have any categorical label from economics.

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

And we will also consult them in terminological economics dictionaries (*Cercaterm* and *Fonaments d'Economia* in Catalan; *The New Palgrave Dictionary of Economics* and *A Dictionary of Economics* in English) to see if we find in them these terms that traditionally have not been associated with the economics field.

We want to examine whether these terms are being included in these dictionaries or whether, on the contrary, all the terminology related to economics that we find is based on the traditional or capitalist system. We will be able to make a comparison between general and terminological dictionaries, but also between different languages.

To set an example, we have seen that the term *vulnerability* with economical connotations only appears as part of an example in its entry in the *Oxford Advanced Learner's Dictionary*: "financial vulnerability". We also have seen that "decrease", only appears with that connotation in *Lèxic de la crisi econòmica*[1] (Cercaterm). Both terms are recurring in our economics corpora, but we saw that they are not fully established in our lexicography and terminology, nor in our society, as terms related to economics.

References

Cercaterm: https://www.termcat.cat/ca/cercaterm (last access: 23-03-2022).

Chaves, R./Monzón, J. L. (2018): La economía social ante los paradigmas económicos emergentes: innovación social, economía colaborativa, economía circular, responsabilidad social empresarial, economía del bien común, empresa social y economía solidaria. In: CIRIEC-España: Revista de Economía Pública, Social y Cooperativa 93, pp. 5–50. https://doi.org/10.7203/CIRIEC-E.93.12901 (last access: 23-03-2022).

Defourny, J./Nyssens, M. (2017): Fundamentals for an international typology of social enterprise models. In: VOLUNTAS – International Journal of Voluntary and Nonprofit Organizations 28 (6), pp. 2469–2497. https://doi.org/10.1007/s11266-017-9884-7 (last access: 23-03-2022).

Federació de Cooperatives de Treball de Catalunya (2019, 2020, 2021): Nexe. https://nexe.coop/revista-en-paper (last access: 23-03-2022).

Hernández, I./Serrano, E. (2021): Breu introducció a conceptes relacionats amb les "noves economies". In: Terminàlia 23, pp. 56–59.

Oxford advances learner's dictionary. https://www.oxfordlearnersdictionaries.com/definition/english/ (last access: 23-03-2022).

Van Dijk, T. A. (1999): Ideología: una aproximación multidisciplinaria. Barcelona.

Dutt, A. K. et al. (2019/2020/2021): Review of Social Economy. https://www.tandfonline.com/journals/rrse20 (last access: 23-03-2022).

Contact information

Mariona Arnau Garcia
IULATERM (IULA-CER), Universitat Pompeu Fabra
mariona.arnau@upf.edu

Mercè Lorente Casafont
IULATERM (IULA-CER), Universitat Pompeu Fabra
merce.lorente@upf.edu

[1] In English: *Lexicon of the economic crisis.*

Mark Dang-Anh/Stefan Scholl

BASIC SOCIO-POLITICAL CONCEPTS IN GERMAN PARLIAMENTARY DEBATES FROM THE 20TH CENTURY

Keywords Conceptual history; corpus pragmatics; discourse semantics

Germany's diverse history in the 20th century raises the question of how social upheavals were constituted in and through political discourse. By analysing basic concepts, the research network "The 20th century in basic concepts" (based at the Leibniz institutes IDS, ZfL, ZZF) aims to identify continuities and discontinuities in political and social discourse. In this way, historical sediments of the present are to be uncovered and those challenges identified that emerged in the course of the 20th century and continue to shape political discourse until the present.

One of the projects of the outlined research network will be presented in the talk: "Basic Socio-Political Concepts with Large Scope and Duration". In the tradition of historical semantics (Müller/Schmieder 2016), the project conceives of conceptual history ("Begriffsgeschichte") as an undertaking to be explored linguistically, which, complementary to the classical hermeneutic investigation of highbrow literature, must be oriented towards the empirical analysis of wider linguistic corpora (cf. Busse/Teubert 2014). The project presented focuses on the parliamentary protocols of the 20th century in a discourse-semantic and corpus-pragmatic way. With the intention of tracing modes of use and changes in meaning of central political-social concepts in the political communication space of parliament, the project addresses the important desideratum of a linguistically-empirically founded conceptual history for a clearly defined core area of political discourse.

On the basis of German-language parliamentary debates (Reichstag and Bundestag minutes), the project looks at basic political-social concepts of the 20th century and examines their modes of use in parliamentary communication by means of corpus-pragmatic and discourse-semantic methods. The focus is on parliament as a space of communication in which the constitution of meaning(s), competing and shifting patterns of interpretation and identificatory use of political-social basic concepts can be observed in a specifically condensed but homogenous communicative setting. The central aim of the project is to analyse how these semantic processes took place in the parliamentary communication space against the background of their pragmatic contexts and how they can be described in terms of conceptual history.

The talk will outline the project corpus, which consists of the Reichstag minutes of the German Empire, the Weimar Republic and National Socialism as well as the Bundestag minutes of the 20th century.

For the project, which combines quantitative-corpus linguistic and qualitative-hermeneutic analytical perspectives, concepts are selected that had considerable social and political reach throughout the 20th century and were also at the centre of (party-)political interpretive struggles (e. g. 'democracy', 'people'/'nation'). A three-stage discourse-semantic heuristic will be presented in the talk:

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

1) explorative identification of semantic clusters and diachronic condensations,
2) contextualisation of concept-constituting lexical items,
3) qualitative-hermeneutic analysis of parliamentary concept formation.

In the talk we present an excerpt from our ongoing work on the concept of *democracy*. The resulting article connects to the *democracy* article in "Geschichtliche Grundbegriffe" (Conze et al. 1972), which, however, only marginally refers to the 20th century. Using data from 20th century parliamentary minutes, we present exploratory steps towards a corpus-based understanding of democracy. Following the idea that the corpus-based identification of candidates of concept-constituting lexical items complements the forthcoming lexicon articles, we identify semantic clusters relevant to the concept of democracy by means of corpus-linguistic analyses.

We conduct the analysis of explorative identification of semantic clusters and diachronic condensations in four steps: Our exploration starts following the assumption that *-demokrati-* is highly productive with regard to possible concept-relevant (1a) word formations. We identify (1b) collocates and (1c) n-gram clusters that co-occur with those word formations. And finally, we conduct a diachronic, cohort-wise (1d) keyword analysis based on the different parliamentary session periods in the German Reichstag and Bundestag.

The second step consists of selecting concept-relevant candidates for (2) concordance lines on the basis of the corpus linguistic evaluations and annotating them with regard to their pragmatic uses. On the one hand, this makes the data hermeneutically accessible in a manually editable way; on the other hand, a qualified quantification is carried out.

The concepts can thus be traced back to the lexical usage patterns, which in a further step are analysed (3) qualitatively-hermeneutically on the basis of selected key phases and sessions of parliamentary discourse, also with regard to their interactional-communicative dynamics. The results of the micro-analysis will be checked for their patternedness or semantic equivalences by means of quantitative queries of longer periods. The results of the analysis form the basis for articles that will be the outcome of the research network in the form of a conceptual history lexicon.

References

Busse, D./Teubert, W. (2014): Using corpora for historical semantics. In: Angermuller, J./Maingueneau, D./Wodak, R. (eds.), The discourse studies reader: main currents in theory and analysis. Amsterdam, pp. 340–349.

Conze, W./Meier, C./Koselleck, R./Maier, H./Reimann, H. L. (1972): Demokratie. In: Brunner, O./Conze, W./Koselleck, R. (eds.), Geschichtliche Grundbegriffe: Historisches Lexikon zur politisch-sozialen Sprache in Deutschland. Stuttgart, pp. 821–899.

Müller, E./Schmieder, F. (2016): Begriffsgeschichte und historische Semantik. Ein kritisches Kompendium. Berlin.

Contact information

Mark Dang-Anh
Leibniz-Institut für Deutsche Sprache
dang@ids-mannheim.de

Stefan Scholl
Leibniz-Institut für Deutsche Sprache
scholl@ids-mannheim.de

Alexandre Ecker

EQUALITY BETWEEN WOMEN AND MEN, A SOCIETAL AND LEXICOGRAPHICAL ISSUE

Bringing the content of the *Lëtzebuerger Online Dictionnaire* in line with the realities of the world and society

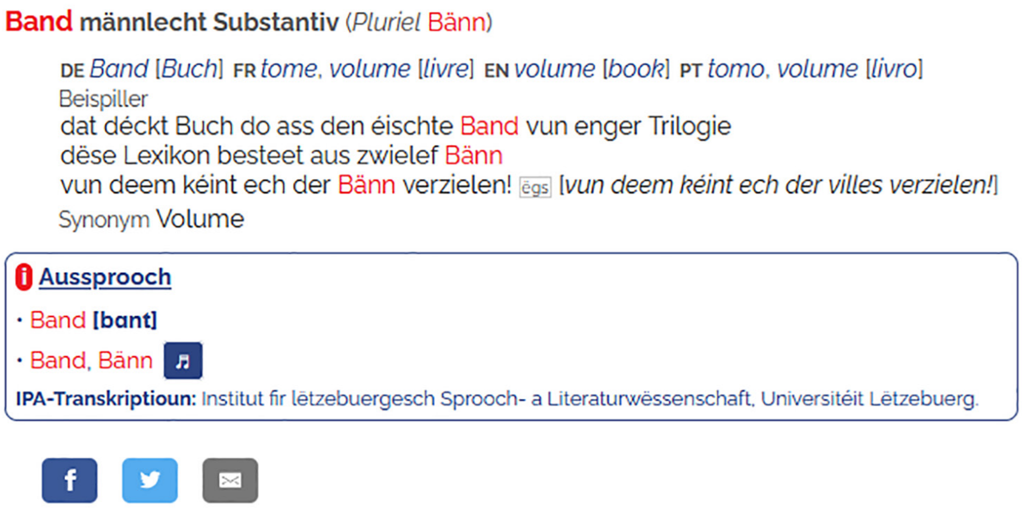
Keywords Luxembourgish; multilingualism; gender equality; lesser used language; XML/XSL technologies

The *Lëtzebuerger Online Dictionnaire* (*LOD*, https://lod.lu) is a quintilingual dictionary (Luxembourgish, German, French, English, Portuguese) describing the Luxembourgish language. It is edited by the *Zenter fir d'Lëtzebuerger Sprooch* (*ZLS*, https://portal.education.lu/zls), a department of the Luxembourgish Ministry of Education.

The *LOD* is a specific lexicographical production as it is set in a rather particular multilingual context:

Luxembourg, a small country in terms of surface area, is officially trilingual (Luxembourgish, German, French). Economic activity depends on more than 200,000 cross-border workers from France, Germany and Belgium, which corresponds to ~46% of employees working in Luxembourg. Some sectors also employ a large number of resident expatriates whose mother tongue is not one of the three official languages.

In the *LOD*, the use of translation equivalents, disambiguated by the addition of semantic clarifiers, enables the dictionary to function as a tool for decoding the Luxembourg language both for native speakers (many of whom are bilingual or more often trilingual) as well as for allophones mastering at least one of the four target languages.



Band männlecht Substantiv (*Pluriel* Bänn)

DE *Band* [*Buch*] FR *tome*, *volume* [*livre*] EN *volume* [*book*] PT *tomo*, *volume* [*livro*]

Beispiller

dat déckt Buch do ass den éischte Band vun enger Trilogie

dëse Lexikon besteet aus zwielef Bänn

vun deem kéint ech der Bänn verzielen! ëgs [*vun deem kéint ech der villes verzielen!*]

Synonym Volume

Aussprooch

· Band **[bɑnt]**

· Band, Bänn ♬

IPA-Transkriptioun: Institut fir lëtzebuergesch Sprooch- a Literaturwëssenschaft, Universitéit Lëtzebuerg.

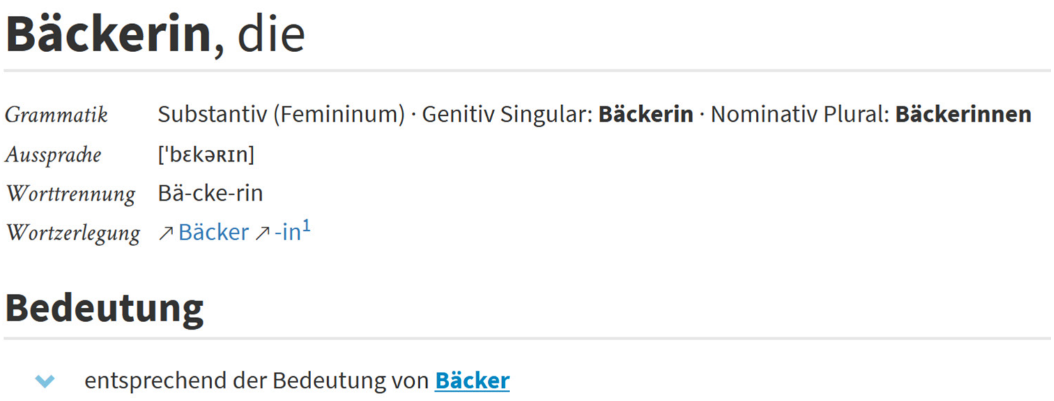
Fig. 1: *LOD* entry **Band**

Examples, glossed phrases, synonyms and phonetic information (IPA transcriptions and an audio feature) complete the dictionary entries and make it possible to envisage secondary lexicographical functions (encoding, translation, etc.) (Fig. 1):

The dictionary is becoming increasingly successful with almost one million entries consulted per month.

In order to strengthen the societal roots of this popular service, the team responsible for its production has undertaken a major project to improve the lexicographical treatment of terms describing female persons in their professions.

These terms were initially treated within cross-reference entries, a method used for instance for some terms in the *Digitales Wörterbuch der deutschen Sprache* (*DWDS*, https://dwds.de):



Bäckerin, die

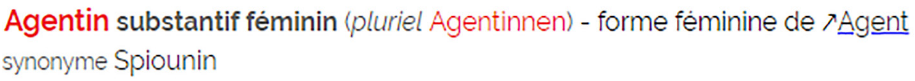
Grammatik	Substantiv (Femininum) · Genitiv Singular: **Bäckerin** · Nominativ Plural: **Bäckerinnen**
Aussprache	[ˈbɛkəʀɪn]
Worttrennung	Bä-cke-rin
Wortzerlegung	↗Bäcker ↗-in^{1}

Bedeutung

entsprechend der Bedeutung von **Bäcker**

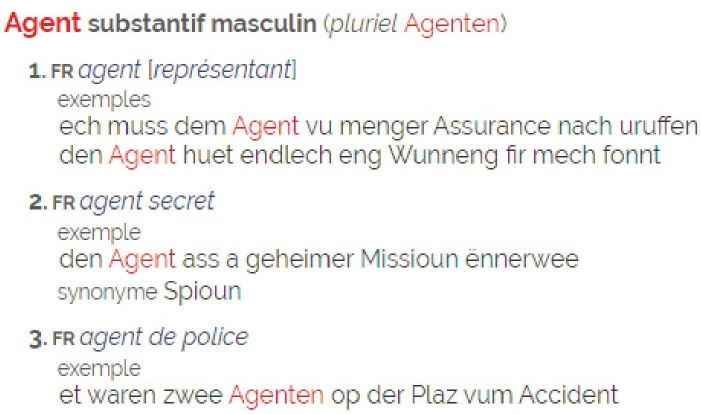
Fig. 2: *DWDS* entry **Bäckerin**

This led to a faster production and result. The female form of a profession would simply refer the user back to the male form of a given term:



Agentin **substantif féminin** (*pluriel* Agentinnen) - forme féminine de ↗Agent
synonyme Spiounin

Fig. 3: *LOD* entry **Agentin**



Agent **substantif masculin** (*pluriel* Agenten)

1. **FR** *agent [représentant]*
exemples
ech muss dem Agent vu menger Assurance nach uruffen
den Agent huet endlech eng Wunneng fir mech fonnt

2. **FR** *agent secret*
exemple
den Agent ass a geheimer Missioun ënnerwee
synonyme Spioun

3. **FR** *agent de police*
exemple
et waren zwee Agenten op der Plaz vum Accident

Fig. 4: *LOD* entry **Agent**

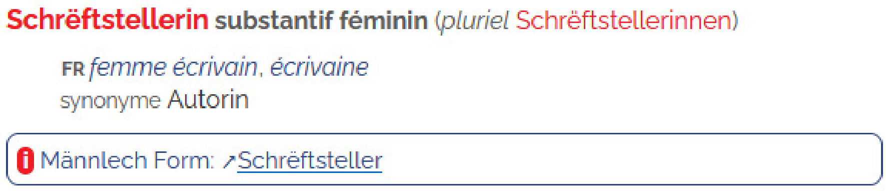
This treatment is not satisfactory for the following reasons:

In the example above (Fig. 3 and Fig. 4), the user can only hypothesize about the relevance of the semantic descriptions of the word *Agent* for *Agentin*.

Furthermore, the feminine forms of the French translational equivalents (in this case *agent*, *agent secret* and *agent de police*) are not found in the translation indexes, nor are the German, English and Portuguese translations. As a consequence, the user's access to various lexicographical features for the terms in question remains limited.

The improvement was planned in two stages:

1) Systematic addition of translation equivalents and cross-references in both directions (from female to male and vice versa), breaking with the previous cross-reference structure:



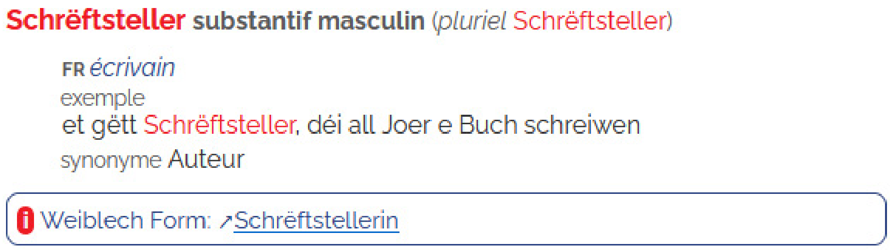
Schrëftstellerin substantif féminin (*pluriel* Schrëftstellerinnen)

FR *femme écrivain*, *écrivaine*

synonyme Autorin

ℹ Männlech Form: ↗Schrëftsteller

Fig. 5: *LOD* entry **Schrëftstellerin**



Schrëftsteller substantif masculin (*pluriel* Schrëftsteller)

FR *écrivain*

exemple

et gëtt Schrëftsteller, déi all Joer e Buch schreiwen

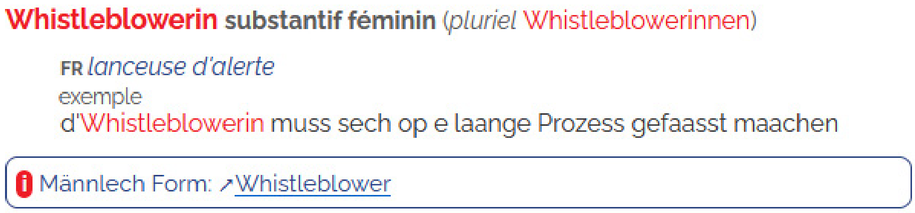
synonyme Auteur

ℹ Weiblech Form: ↗Schrëftstellerin

Fig. 6: *LOD* entry **Schrëftsteller**

This step revealed differences in the female form of job names across languages.

2) Systematic addition of examples in order to achieve a perfectly identical treatment:



Whistleblowerin substantif féminin (*pluriel* Whistleblowerinnen)

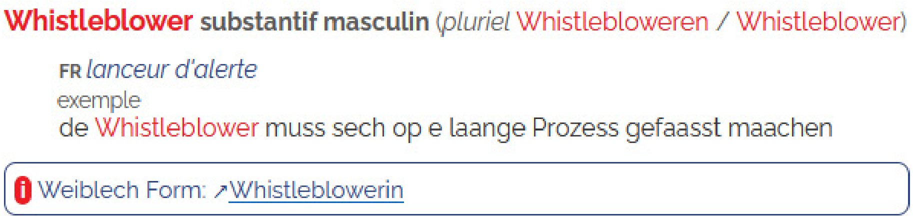
FR *lanceuse d'alerte*

exemple

d'Whistleblowerin muss sech op e laange Prozess gefaasst maachen

ℹ Männlech Form: ↗Whistleblower

Fig. 7: *LOD* entry **Whistleblowerin**



Whistleblower substantif masculin (*pluriel* Whistlebloweren / Whistleblower)

FR *lanceur d'alerte*

exemple

de Whistleblower muss sech op e laange Prozess gefaasst maachen

ℹ Weiblech Form: ↗Whistleblowerin

Fig. 8: *LOD* entry **Whistleblower**

This step led to an in-depth reflection on the use of the generic masculine form in Luxembourgish.

The undertaking has had effects on both the LOD workflow and output:

c) during the steps taken, from a (meta)lexicographical point of view

- identification of candidate terms,
- necessary differentiation between mono- and polysemous terms,
- impact on the numerous variants (orthographic and homosemic),

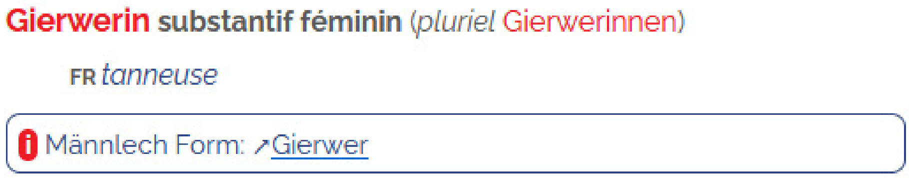
d) during the steps taken, from a technical point of view

- use of XML/XSL technologies,
- partial automation,
- control procedures,

e) due to the difficulties encountered

- in the target languages (feminine form of job names),
- examples (generic masculine).

It is interesting to note that in an effort to systematically propose a feminine equivalent for professions, the *ZLS* team has broken with a basic *LOD* principle by showing certain Luxembourgish terms without a trace in the corpora that feed the content of the dictionary:



Gierwerin substantif féminin (*pluriel* Gierwerinnen)

FR *tanneuse*

ℹ Männlech Form: ↗Gierwer

Fig. 9: *LOD* entry Gierwerin

The word *Gierwerin* is easily formed on the basis of other common words (*Astronaut*, *Astronautin* [*astronaut*]), but could not be attested.

More than a thousand entries are currently affected by these changes.

The results obtained may help other publishers to better plan similar work, aiming to achieve a more equal treatment of women and men in their lexicographic production.

References

Digitales Wörterbuch der deutschen Sprache (2022). https://dwds.de (last access: 11-04-2022).

Lëtzebuerger Online Dictionnaire (2022). https://lod.lu (last access: 23-03-2022).

Contact information

Alexandre Ecker
Zenter fir d'Lëtzebuerger Sprooch
alexandre.ecker@lod.lu

Stefan Engelberg

LEXICOGRAPHY'S ENTANGLEMENT WITH COLONIALISM: THE HISTORY OF TOK PISIN LEXICOGRAPHY AS COLONIAL HISTORY

Abstract Tok Pisin is a pidgin/creole language spoken since the late 19th century in most of the area that nowadays constitutes Papua New Guinea where it emerged under German colonial rule. Unusual for a pidgin/creole, Tok Pisin is characterized by a extensive lexicographic history. The Tok Pisin Dictionary Collection at the Leibniz Institute for the German Language, described in this article, includes about fifty dictionaries. The collection forms the basis for the sketch of the history of Tok Pisin lexicography as part of colonial history presented here. The basic thesis is that in the history of Tok Pisin, lexicographic strategies, dictionary structures, and publication patterns reflect the interest (and disinterest) of various groups of colonial actors. Among these colonial actors, European scientists, Catholic missionaries, and the Australian and US militaries played important roles.

Keywords Pidgin; Tok Pisin; colonialism; history of lexicography; lexicography and war; missionary linguistics; colonial linguistics

Contact information

Stefan Engelberg
Leibniz-Institut für Deutsche Sprache
engelberg@ids-mannheim.de

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Christine Ganslmayer

USING DIALECT DICTIONARIES AS A DATABASE

Stereotypes of people in the 'Franconian Dictionary'

Keywords Dialect dictionary; database; stereotypes

The 'Franconian Dictionary' (Fränkisches Wörterbuch = WBF) pursues the expected objective for a dialect dictionary, namely to research, preserve and pass on the cultural heritage of dialect (cf. https://wbf.badw.de/zielsetzung.html). However, the function of this dictionary does not have to be limited to preserving an "intangible cultural heritage". For usage-based research questions, this dictionary of the Franconian administrative districts is a valuable database: a substantial part of the archival material consists of questionnaires that the informants have filled out themselves.

Since the WBF is not published as a multi-volume book publication, but as an online database, the data of the indirect questionnaire surveys have been successively made accessible online since 2017. So far, only parts of the so-called "post-war questionnaires" have been recorded. These comprise 123 different questionnaires from the period 1960 to 2001 with a total of 6,678 questions, which were sent to respondents (mostly primary school teachers at rural schools) and filled out by them personally. The return rate of these indirect questionnaire surveys amounts to a total of approx. 48,000 questionnaires. Currently (03-2022), the database already contains the records of 1,358 questions.

Thus, the lemmatised word material offers direct access to authentic utterances and linguistic constructions of competent dialect speakers, in which individual experiences and ideas are manifested as well as collective, generalizing attributions.

In the presentation, it will be shown on the basis of the WBF how dictionary data can be methodically used as an empirical database. The limits and possibilities of the dictionary as a research corpus will also be problematised (cf. Kürschner/Habermann/Müller (eds.) 2019).

As an exemplary object of investigation, stereotypes of people are evaluated:

> Das Stereotyp umfasst, wenn es auf Menschen bezogen ist, einen Sachverhalt, in dem in ungerechtfertigter Weise Personengruppen verallgemeinernd, bewertend (meist pejorativ) und vereinfachend Eigenschaften zugeordnet werden (Pümpel-Mader 2010, p. 10).[1]

Research on stereotypes is now widespread across disciplines, so that a precise definition of the term depends on the particular perspective.[2] However, in stereotyping, qualities or actions that are assumed to be characteristic are associated with particular groups of people (Hinton 2000, p. 64), not least in order to contrastingly delimit one's own social group. Therefore, stereotypes can reflect positive or negative assumptions about social groups (Brandt/Reyna 2011, p. 49) and are often based on binary (opposing) thought patterns.

[1] "The stereotype, when related to people, is based on a process in which characteristics are unjustifiably assigned to groups of people in a generalizing, evaluative (usually pejorative) and simplifying manner."

[2] A first approach to different aspects of the concept of stereotypes can be found, for example, in Hahn (2007), and from a linguistic point of view in Quasthoff (1973) or Hentschel (1995).

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Stereotypes can be assigned to different social contexts such as nation, ethnicity, occupation, age, gender, etc. and refer to different aspects such as physical characteristics, traits, behaviours, attitudes or life outcomes of the stereotyped group (Merk-Carinci 2020, p. 53; Brandt/Reyna 2011, p. 49).

The aim of this study is to elicit such person-related schematizations, to analyse them formally and functionally, and to reconstruct culture-specific stereotypes that have been handed down in the dialect dictionary.

Stereotypes can be realised implicitly or explicitly in language use (cf. summarising and problematising Quasthoff/Hallsteinsdóttir 2016, pp. 350–351). Explicit stereotypes manifest themselves through certain linguistic forms of expression at all linguistic levels, from phonetics, word-formation patterns, the use of certain types of words and syntactic structures to sentence and verbal patterns (Pümpel-Mader 2010). With reference to the dictionary data of the WBF, which consists of lemmatised lexemes, this means that the respondents have noted down lexemes or sometimes phrases in response to questions that could be used at any time as a linguistic utterance in a concrete linguistic context. Or in other words, the question in the questionnaire replaces the concrete linguistic context, which supports the detection and interpretation of possible stereotypes.

When searching the WBF database, one can set various filters so that, among other things, the tokens for certain questions can be displayed. This provides initial access to relevant dictionary data for the study of stereotypes. Certain questions about persons or social groups contain typical linguistic markers for stereotypes, such as negatively evaluative adjectives (What do you *derogatorily* call a “resident of the Arab countries”?, cf. questionnaire 69, question 39; dialect *mocking* name for city dwellers?, cf. questionnaire 35, question 39) or pick up on negative characteristics: Dialectal for *careless*, female person who works *untidily*? (cf. questionnaire 97, question 5). Other questions are contrastively related to each other: Catholic and Protestant children like to call each other mocking names. Which ones are common among them? for Catholic children? for Protestant children? (cf. questionnaire 36, question 16), and enable a comparison between different social groups. Sometimes autostereotypes are asked in contrast to heterostereotypes or metaheterostereotypes (cf. Thiele 2015 for differentiation): What are the inhabitants of your home landscape called? Are other names used for immigrants and refugees? Which ones? What do migrants and refugees call the long-established inhabitants of the landscape? (cf. questionnaire 18, questions 55–57).

References

Brandt, M./Reyna, C. (2011): Stereotypes as attributions. In: Simon, E. L. (ed.): Psychology of stereotypes. New York, pp. 47–80.

Fränkisches Wörterbuch (WBF). https://wbf.badw.de/das-projekt.html (last access: 25-03-2022).

Hahn, H. H. (2007): 12 Thesen zur Stereotypenforschung. In: Hahn, H. H./Mannová, E. (eds.): Nationale Wahrnehmungen und ihre Stereotypisierung. Beiträge zur Historischen Stereotypenforschung. Frankfurt a. M. a. o., pp. 15–24.

Hentschel, G. (1995): Stereotyp und Prototyp: Überlegungen zur begrifflichen Abgrenzung vom linguistischen Standpunkt. In: Hahn, H. H. (ed.): Historische Stereotypenforschung. Methodische Überlegungen und empirische Befunde. Oldenburg, pp. 14–40.

Hinton, P. R. (2000): Stereotypes, cognition and culture. Hove.

Kürschner, S./Habermann, M./Müller, P. O. (eds.) (2019): Methoden moderner Dialektforschung. Erhebung, Aufbereitung und Auswertung von Daten am Beispiel des Oberdeutschen. Hildesheim a. o.

Merk-Carinci, D. (2020): Bilder der Anderen: Kritische Diskursanalyse der westdeutschen und britischen Presseberichterstattung zur Zeit der zweiten Berlin-Krise (1958–62). Würzburg.

Pümpel-Mader, M. (2010): Personenstereotype. Eine linguistische Untersuchung zu Form und Funktion von Stereotypen. Heidelberg.

Quasthoff, U. (1973): Soziales Vorurteil und Kommunikation. Eine sprachwissenschaftliche Analyse des Stereotyps. Frankfurt a. M.

Quasthoff, U./Hallsteinsdóttir, E. (2016): Stereotype in Webkorpora: Strategien zur Suche in sehr großen Datenmengen. In: Linguistik online 79 (5), pp. 347–379. http://dx.doi.org/10.13092/lo.79.3349 (last access: 25-03-2022).

Thiele, M. (2015): Medien und Stereotype. Konturen eines Forschungsfeldes. Bielefeld.

Contact information

Christine Ganslmayer
Friedrich-Alexander-Universität Erlangen-Nürnberg
christine.ganslmayer@fau.de

Laura Giacomini / Paolo DiMuccio-Failla /
Patrizio De Martin Pinter

THE REPRESENTATION OF CULTURE-SPECIFIC LEXICAL ITEMS IN MONOLINGUAL LEARNER'S LEXICOGRAPHY

The case of the electronic Phrase-Based Active Dictionaries

Abstract This paper focuses on the treatment of culture-bound lexical items in a novel type of online learner's dictionary model, the Phrase-Based Active Dictionary (PAD). A PAD has a strong phraseological orientation: each meaning of a word is exclusively defined in a typical phraseological context. After introducing the relevant theory of *realia* in translation studies, we develop a broader notion of culture-specific lexical items which is more apt to serve the purposes of learner's lexicography and thus to satisfy the needs of a larger and often undefined target group. We discuss the treatment of such words and expressions in common English learner's dictionaries and then present various excerpts from PAD entries in English, German, and Italian which display different strategies for coping with cultural contents in the lexicon. Our aim is to demonstrate that the phraseological approach at the core of the PAD model turns out to be extremely important to convey cultural knowledge in a suitable way for users to fully grasp cultural implications in language.

Keywords Learner's lexicography; phraseology; culture-specific items; realia; multimedia

Contact information

Laura Giacomini
University of Hildesheim
laura.giacomini@uni-hildesheim.de

Paolo DiMuccio-Failla
University of Hildesheim
muccio@uni-hildesheim.de

Patrizio De Martin Pinter
University of Heidelberg
patrizio.de_martin_pinter01@stud.uni-heidelberg.de

This page is part of the publication: Klosa-Kückelhaus, Annette / Engelberg, Stefan / Möhrs, Christine / Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Annette Klosa-Kückelhaus

LEXICOGRAPHY FOR SOCIETY AND WITH SOCIETY – COVID-19 AND DICTIONARIES

Abstract Not only professional lexicographers, but also people without a professional background in lexicography, have reacted to the increased need for information on new words or medical and epidemiological terms being used in the context of the COVID-19 pandemic. In this study, corona-related glossaries published on German news websites are presented, as well as different kinds of responses from professional lexicography. They are compared in terms of the amount of encyclopaedic information given and the methods of definition used. In this context, answers to corona-related words from a German question-answer platform are also presented and analyzed. Overall, these different reactions to a unique challenge shed light on the importance of lexicography for society and vice versa.

Keywords Lay-lexicography; professional lexicography; glossaries; general language dictionaries; neologism dictionaries

Contact information

Annette Klosa-Kückelhaus
Leibniz-Institut für Deutsche Sprache
klosa@ids-mannheim.de

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Gloria Mambelli

MANORIAL SOCIETY IN MULTILINGUAL MEDIEVAL ENGLAND

An onomasiological approach

Keywords Onomasiology; multilingualism; manorial society; historical lexicography; Middle English; Anglo-French; Medieval Latin

The English language underwent an unprecedented lexical enrichment during the period of linguistic contact following the Norman Conquest, when a significant proportion of borrowings of French origin entered the lexicon (Durkin 2014). This is still often referred to as a phenomenon mainly affecting domains related to the aristocratic milieu (cf Baugh/Cable 2002; Barber/Beal/Shaw 2009), even though the investigation of non-literary mixed-language texts, prompted by recent interest in late medieval multilingualism (cf Trotter (ed.) 2000; Wright (ed.) 2020), demonstrates that language contact was likely to occur in everyday life activities carried out by sub-aristocratic classes as well (cf Sylvester/Marcus 2017; Ingham/Marcus 2016). In particular, contact-induced variation is found in business documents such as accounts and inventories, a "text type where a mixing of two or more languages is the norm" (Wright 2000). Notwithstanding the numerous instances of code-switching from Latin into English and French contained especially in manorial accounts, suggesting that manorial officials worked in a multilingual environment (Ingham 2009), lexical fields pertaining to rural life still play a marginal role in contact linguistic research.

By focusing on the multilingual lexis denoting post-Conquest manorial society, this paper illustrates how an onomasiological approach can be adopted for combining linguistic and sociolinguistic investigation. Making use of structures such as taxonomical hierarchies to explore the lexical items with which a concept can be expressed (Grondelaers et al. 2010), onomasiology provides an appropriate framework for studying lexical variation within a specific domain while taking the sociolinguistic context into account. In view of the difficulty of accessing the multilingual vocabulary to be investigated, scattered in semasiological dictionaries and thesauri, an onomasiological categorisation was carried out by semantically and chronologically arranging Anglo-French, Medieval Latin, and Middle English terms used to refer to the people living and working on manorial estates. Lexical items and dates of attestation were extracted from the online editions of the *Anglo-Norman Dictionary*, the *Dictionary of Medieval Latin from British Sources*, and the *Middle English Dictionary*. Only nouns, the most commonly borrowed word-class (Matras 2009, p. 167), were included.

This paper reports on the progress of an ongoing project whose output is a trilingual thesaurus assembling terms connected to people and locations of manorial estates of post-Conquest England, designed to carry out lexicological and lexicographical analyses. The mapping of this lexical domain allows for the tracking of the outcomes of language contact in rural contexts of late medieval England and shows how the presence of officials, whose profession required competence in three languages, contributed to spreading multilingualism in such areas notwithstanding the majority of English monolingual speakers.

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

References

Anglo-Norman Dictionary (2022): https://anglo-norman.net/ (last access: 04-03-2022).

Barber, C./Beal, J. C./Shaw, P. A. (2009): The English language: a historical introduction. 2nd edition. Cambridge.

Baugh, A./Cable, T. (2002): A history of the English language. 5th edition. London.

Dictionary of Medieval Latin from British Sources (2015): http://clt.brepolis.net/dmlbs/ (last access: 04-03-2022).

Durkin, P. (2014): Borrowed words: a history of loanwords in English. Oxford.

Grondelaers, S./Speelman, D./Geeraerts, D. (2010): Lexical variation and change. In: Geeraerts, D./Cuyckens, H. (eds.): The Oxford handbook of cognitive linguistics. Oxford, pp. 988–1011.

Ingham, R. (2009): Mixing languages on the manor. In: Medium Ævum 78 (1), pp. 80–97.

Ingham, R./Marcus, I. (2016): Vernacular bilingualism in professional spaces, 1200 to 1400. In: Classen, A. (ed.): Multilingualism in the Middle Ages and Early Modern Age. Berlin/Boston, pp. 145–164.

Matras, Y. (2009): Language contact. Cambridge.

Middle English Dictionary (2019): https://quod.lib.umich.edu/m/middle-english-dictionary (last access: 04-03-2022).

Sylvester, L./Marcus, I. (2017): Studying French-origin Middle English lexis using the Bilingual Thesaurus of Medieval England: a comparison of the vocabulary of two occupational domains. In: Delesse, C./Louviot, E. (eds.): Studies in language variation and change 2: shifts and turns in the history of English. Newcastle upon Tyne, pp. 217–228.

Trotter, D. (ed.) (2000): Multilingualism in later medieval Britain. Cambridge.

Wright, L. (2000): Bills, Accounts, inventories: everyday trilingual activities in the business world of later medieval England. In: Trotter, D. (ed.): Multilingualism in later medieval Britain. Cambridge, pp. 149–156.

Wright, L. (ed.) (2020): The multilingual origins of Standard English. Berlin/Boston.

Contact information

Gloria Mambelli
Università degli Studi di Verona
gloria.mambelli@univr.it

Carolin Müller-Spitzer/Jan Oliver Rüdiger

THE INFLUENCE OF THE CORPUS ON THE REPRESENTATION OF GENDER STEREOTYPES IN THE DICTIONARY. A CASE STUDY OF CORPUS-BASED DICTIONARIES OF GERMAN

Abstract Dictionaries are often a reflection of their time; their respective (socio-)historical context influences how the meaning of certain lexical units is described. This also applies to descriptions of personal terms such as *man* or *woman*. Lexicographers have a special responsibility to comprehensively investigate current language use before describing it in the dictionary. Accordingly, contemporary academic dictionaries are usually corpus-based. However, it is important to acknowledge that language is always embedded in cultural contexts. Our case study investigates differences in the linguistic contexts of the use of man and woman, drawing from a range of language collections (in our case fiction books, popular magazines and newspapers). We explain how potential differences in corpus construction would therefore influence the "reality"[1] depicted in the dictionary. In doing so, we address the far-reaching consequences that the choice of corpus-linguistic basis for an empirical dictionary has on semantic descriptions in dictionary entries. Furthermore, we situate the case study within the context of gender-linguistic issues and discuss how lexicographic teams can engage with how dictionaries might perpetuate traditional role concepts when describing language use.

Keywords Gender linguistics; corpus-based lexicography; collocations; lexicography equality; gender equality

Contact Information

Carolin Müller-Spitzer
Leibniz-Institut für Deutsche Sprache Mannheim
mueller-spitzer@ids-mannheim.de

Jan Oliver Rüdiger
Leibniz-Institut für Deutsche Sprache Mannheim
ruediger@ids-mannheim.de

[1] What can be seen as "linguistic reality" is a very complex matter that goes beyond the scope of this paper. When we use in the following the term "linguistic reality", we are aware that texts or corpora are not a "description" or "representation" of this assumed reality, but serve to construct and interpret one possible part of this reality from language use (e.g. simply in reading or in specific work such as lexicography).

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Enakshi Nandi

SECRECY AND THE ETHICAL QUESTION

Some reflections on documenting Ulti, a secret transgender language

Keywords Ulti; secret language; semantic domains; hijra; koti; ethics

Ulti is a secret language spoken by the marginalized and oppressed transfeminine/femme hijra-koti community in West Bengal, India. It comprises of a set of lexical items that have emerged historically from the socially segregated hijra households, to give voice to the range of queer identities and experiences that have been silenced in mainstream languages and discursive spaces, which are predominantly conservative, cisnormative, and hetero-patriarchal. Thus, these lexical items are secret and used in restricted domains specific to the lives of hijras and kotis.

The domains of use in Ulti include sexual activities (*dhurano* – to have sex, *khumur kora* – to have oral sex), private body parts (*likom* – penis, *cipti* – vagina), sexual and gender identities (*bhobrashi* – intersex person, *chibri* – hijra), and other aspects exclusive to the hijra subculture (including kinship terms, professions, rituals and customs) and the hijra-koti lifestyle (including clothes, wigs, accessories, currency, biological kinship terms). The Ulti lexicon is limited (around 300 lexical items) but is rapidly expanding due to the presence of a symbolic gender marker *mashi* (lit. "aunt"), that combines with nouns in Bangla, Hindi, and English to create new compound nouns in Ulti.

In this paper, I reflect on the methodological, practical, and ethical questions that I was confronted with during my fieldwork as I embarked upon the project of documenting Ulti. The lexical data was collected using two data elicitation tools – the Rapid Word Collection method (Moe 2003) which uses the concept of semantic domains, and the word list and sentence list developed by Abbi (2001) for South Asian languages. Significant modifications were made to both to accommodate the secret nature of the language and the socio-political context of its speakers.

The central question I had to contend with was of the ethics of publishing the lexicon of a language that was not only secret, but which – as a secret lect – served to protect its speakers from scrutiny, censure, and harassment in public spaces. Ulti is also a professional lect, used in the course of traditional hijra professions such as *badhai*, *cholla/mangti*, and *khajra*. To publish that information would compromise the autonomy of hijras going about their daily lives. However, many speakers of Ulti are excited to share their language with the academic and mainstream community, claiming that their language has already been "revealed" in recent works of fiction and is no longer as secret as it was.

The motivation to document Ulti is an academic quest to learn more about a language that has existed for centuries but escaped detection and documentation until recently. The lack of information on what these languages may have looked like initially, how they may have evolved, and how they may have been used constitutes a major gap in the literature, which has consequently hampered our understanding of the community and contributed to their continued marginalization and segregation from mainstream spaces.

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

The ideal solution from the vantage point of the hijra-koti community would be to take a theoretical lexicographic approach and publish only as much of the language as is necessary to gain an understanding of its grammar and its role and significance within the community, steering clear of constructions that would put the community in any immediate danger. That precludes the development of a dictionary and/or a primer in the language under study.

This research work and paper does not present a conclusive and ethically infallible solution to the problem of documenting secret lects, It reflects on the working solution found by one researcher to one socio-cultural context, fully aware of its potential limitations in other research contexts. More significantly, this paper presents a problem that demands close attention and deep introspection by the academic community about the merits and ethics of working with a marginalized community and publishing findings of a sensitive nature in future research projects, especially in the form of an exhaustive dictionary.

References

Abbi, A. (2001): The manual of linguistic fieldwork and structures of Indian languages. New Delhi.

Atkins, B. T. (1992): Theoretical lexicography and its relation to dictionary-making. In: Dictionaries: Journal of the Dictionary Society of North America 14, pp. 4–43.

Awan, M. S./Muhammad, S. (2011): Hijra Farsi wordlist. Islamabad.

Bergenholtz, H./Gouws, R. (2012): What is llexicography? In: Lexikos 22, pp. 31–42.

Campbell, L. (1994): Linguistic reconstruction and unwritten languages. In: Asher, R. E./Simpson, J. Y.: Encyclopedia of language and linguistics. London, pp. 3475–3480.

Connell, B. (1998): Lexicography, linguistics, and minority languages. In: Journal of the Anthropological Society of Oxford (JASO) 29/3, pp. 231–242.

Keymeulen, J. v. (2010): Compiling a dictionary of an unwritten language: a non-corpus-based approach. In: Lexikos 13, pp. 183–205.

Lahiry, S. (2008): Koti language. Dhaka.

Moe, R. (2003): Compiling dictionaries using semantic domains. In: Lexikos 13, pp. 215–223.

Mosel, U. (2004): Dictionary making in endangered speech communities. In: Austin, P. K. (ed.): Language documentation and description, Vol. 2. London, pp. 39–54.

Planning a rapid word collection workshop (2017): Retrieved from Rapid Words SIL: http://rapidwords.net/resources/planning-rapid-word-collection-workshop.

Sorensen, J. (2004): Vulgar tongues: canting dictionaries and the language of the people in eighteenth-century Britain. In: Eighteenth-Century Studies 37 (3) (Critical Networks), pp. 435–454.

Storch, A. (2011): Secret manipulations: language and context in Africa. Oxford.

van den Berg, R./Shore, S. (2006): A new mass elicitation technique: the Dictionary Development Program. In: Tenth International Conference on Austronesian Linguistics (10ICAL). Puerto Princesa City, Palawan, Phillipines.

Contact information

Enakshi Nandi
Jawaharlal Nehru University
enakshi.nandi@gmail.com

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Laura Pinnavaia

IDENTIFYING IDEOLOGICAL STRATEGIES IN THE MAKING OF MONOLINGUAL ENGLISH LANGUAGE LEARNER'S DICTIONARIES

Abstract The aim of this paper is to show how lexicographical choices reflect ideological thinking, singled out by Eagleton (2007) into the strategies of rationalizing, legitimating, action-orienting, unifying, naturalizing and universalizing. It will be carried out by examining two twenty-first century editions of each of the five English monolingual learner's dictionaries published by Cambridge, Collins, Longman, Macmillan, and Oxford. The synchronic and diachronic analyses of the dictionaries and their different editions at the macro-structural level (the wordlists) and at the micro-structural level (the definitional styles) will show how the reduction and change of data, derived from heterogeneous social and cultural contexts of language use, to abstract essential forms, involves decisions about the central and peripheral aspects of the lexicon and the meaning of words.

Keywords English monolingual learner's dictionaries; ideology; British twenty-first-century lexicography

References

Eagleton, T. (2007): Ideology. London/New York.

Contact information

Laura Pinnavaia
University of Milano (Italy)
Laura.pinnavaia@unimi.it

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Petra Storjohann

THE PUBLIC AS LINGUISTIC AUTHORITY: WHY USERS TURN TO INTERNET FORUMS TO DIFFERENTIATE BETWEEN WORDS

Abstract This paper addresses the question of why we face unsatisfactory German dictionary entries when looking up and comparing two similar lexical terms that are loan words, new words, (near)-synonyms, or confusables. It explains how users are aware of existing reference works but still search or post on language forums, often after consulting a dictionary and experiencing a range of dictionary-based problems. Firstly, these dictionary-based difficulties will be scrutinised in more detail with respect to content, function, presentation, and the language of definitions. Entries documenting loan words and commonly confused pairs from different lexical reference resources serve as examples to show the shortcomings. Secondly, I will explain why learning about your target group involves studying discussion forums. Forums are a valuable source for detailed user studies, enabling the examination of different communicative needs, concrete linguistic questions, speakers' intuitions, and people's reactions to posts and comments. Thirdly, with the help of two examples I will describe how the study of chats and forums had a major impact on the development of a recently compiled German dictionary of confusables. Finally, that same problem-solving approach is applied to the idea of a future dictionary of neologisms and their synonyms.

Keywords Internet forums; synonyms; confusables; sense discrimination; problem-solving approach

Contact information

Petra Storjohann
Leibniz-Institut für Deutsche Sprache
storjohann@ids-mannheim.de

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Marcin Zabawa

WHAT DO WE LEARN ABOUT THE SOCIETY FROM THE EXAMPLES OF USAGE IN DICTIONARIES? ON (NON-) STEREOTYPICAL ROLES OF MEN AND WOMEN IN ENGLISH AND POLISH MONOLINGUAL GENERAL DICTIONARIES: A CONTRASTIVE STUDY

Keywords Gender roles in society; English lexicography; Polish lexicography; examples of usage in dictionaries

In the past, television commercials, movies, school textbooks, dictionaries, etc., often presented traditional and oversimplified gender roles. For instance, women were far more frequently portrayed as parents, homemakers, etc., than men (cf. e. g. McArthur/Resko 1974; Fullerton/Kendrick 2000; Arabski 2010; Karwatowska/Szpyra-Kozłowska 2010).

Nowadays the situation pictured above appears to have been slowly, yet consistently, changing. Dictionaries are one of the areas in which the portrayal of men and women is of great importance since dictionaries are often perceived as a model for a good language use. Thus, the aim of the present paper is to investigate gender differences in dictionaries. More specifically, the paper examines to what extent English and Polish monolingual dictionaries stick to the stereotypical (and thus oversimplified) view of the roles of men and women in society. This will be done through the analysis of the use of pronouns *he*/*she*, *him*/*her*, etc., and other gender-specific nouns, in the examples provided by the dictionaries of English and Polish.

The procedure is as follows: certain words have been chosen, connected with the stereotypical (and oversimplified at the same time) roles of men and women in the society. Three main semantic areas have been taken into consideration: activities done at home, such as cleaning, activities denoting addiction or not accepted in the society, such as smoking, and activities connected with the career and earning money.

To be more precise, the study includes verbs connected with household activities, traditionally and stereotypically connected with women (*cook*, *bake*, *fry*, *wash*, *clean*, *iron*, *hoover*, *sweep*, plus, naturally, their Polish equivalents) as well as verbs describing the engagement in certain behaviours, stereotypically connected with men (*smoke*, *drink*, *fight*). In addition, certain nouns have also been taken into consideration; these denote the concepts connected with earning money and are thus stereotypically connected with men (*career*, *job*, *money*, *salary*, *courage*). The study focuses not on the definitions themselves, but on the examples of usage provided by the dictionaries; specifically, the contexts have been analyzed in detail, and it has been counted how many examples for each word given above refer to men, how many to women, and how many are gender-neutral. The procedure has been done separately for English and Polish dictionaries.

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

The study aims at providing answers to the following questions:

Do the examples provided by the English and Polish dictionaries stick to the traditional and oversimplified gender roles in the society or not? To what extent are the examples provided (non-)stereotypical? Are there differences in this respect between English and Polish dictionaries?

The study is based on the following English and Polish monolingual dictionaries: *Oxford English Dictionary* (online edition), *Oxford Advanced Learner's Dictionary* (online edition), *Lexico Dictionary* (online edition), *Longman Dictionary of Contemporary English* (online edition), *Collins Dictionary* (online edition), *Wielki słownik języka polskiego* [A Great Polish Dictionary] (online edition), *Wielki słownik języka polskiego PWN* [A Great PWN Polish Dictionary] (print edition, 2018), *Inny słownik języka polskiego* [A Different Dictionary of Polish] (print edition, 2017). Some of the English dictionaries enumerated above are intended for language learners, but it does not seem to have any effect on the final conclusions.

The study is, in general, synchronic in nature (all the dictionaries upon which the study is based are from 2017 onwards), but there is also a diachronic component, as the results of the present study are going to be compared with the results of a similar study done by the author in the past (Zabawa 2012). The study in question was a small one and published only locally in Polish. It was connected with dictionaries published in the years 2000–2010. The results of the study indicated that the dictionaries of Polish were far more stereotypical with respect to traditional and oversimplified gender roles in the society: for instance, the examples provided by the Polish dictionaries for the verb *zarabiać* 'to earn (money)' were in the majority connected with men rather than women; in the case of the verb *earn* and the English dictionaries, however, there was a far more balanced proportion between men and women.

The present study (dictionaries 2017 onwards) indicates that the contrast between English and Polish dictionaries is not as evident as it used to be. In fact, many examples in today's dictionaries (both English and Polish) are gender-neutral, i. e. they do not specify a particular sex; instead, dictionaries tend to provide examples in the plural, use the passive voice, use subjectless constructions, etc. This, while a continuation of the practices seen in the case of English dictionaries, is a novelty in Polish.

References

Arabski, J. (2010): Język a płeć [Language and gender]. In: Arabski, J./Ziębka, J. (eds.): Płeć języka – język płci [The gender of the language – The language of gender]. Katowice, pp. 11–30.

Collins Dictionary (2022): https://www.collinsdictionary.com (last access: 10-03-2022).

Fullerton, J. A./Kendrick, A. (2000): Portrayal of men and women in U.S. Spanish-language television commercials. In: Journalism & Mass Communication Quarterly 77 (1), pp. 128–142.

Inny słownik języka polskiego [A Different Dictionary of Polish] (2017): Warsaw.

Karwatowska, M./Szpyra-Kozłowska, J. (2010): Lingwistyka płci. Ona i on w języku polskim [Linguistics and gender. She and he in the Polish language]. Lublin.

Lexico Dictionary (2022): http://www.lexico.com (last access: 10-03-2022).

Longman Dictionary of Contemporary English (2022): https://www.ldoceonline.com (last access: 10-03-2022).

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

McArthur, L. Z./Resko, B. G. (1975). The portrayal of men and women in American television commercials. In: The Journal of Social Psychology 97 (2), pp. 209–220.

Oxford Advanced Learner's Dictionary (2022): https://www.oxfordlearnersdictionaries.com (last access: 10-03-2022).

Oxford English Dictionary (2022): https://www.oed.com (last access: 10-03-2022).

Wielki słownik języka polskiego [A Great Polish Dictionary] (2022): https://wsjp.pl (last access: 10-03-2022).

Wielki słownik języka polskiego PWN [A Great PWN Polish Dictionary] (2018): Warsaw.

Zabawa, M. (2012): *W tradycyjnym domu* żona *sprząta i gotuje* vs. *He cooked lunch for me*: o (nie) stereotypowym postrzeganiu świata w polskich i angielskich słownikach [*The wife cooks and cleans in a traditional house* vs. *He cooked lunch me*: on (non-)stereotypical view of the world in Polish and English dictionaries]. In: Pstyga, A. (ed.): Słowo z perspektywy językoznawcy i tłumacza IV [The word from the perspective of a linguist and a translator 4]. Gdańsk, pp. 225–232.

Contact information

Marcin Zabawa
University of Silesia in Katowice
marcin.zabawa@us.edu.pl

Lexicography: Status, Theory and Methods

XX EURALEX INTERNATIONAL CONGRESS
DICTIONARIES AND SOCIETY
12-16 July 2022, Mannheim, Germany

Laura Balbiani/Anne-Kathrin Gärtig-Bressan/
Martina Nied Curcio/Stefan Schierholz

DICTIONARIES FOR THE FUTURE – THE FUTURE OF DICTIONARIES

The 15 Villa Vigoni Theses on Lexicography

Keywords Villa Vigoni Theses; dictionary use; dictionary research; dictionary quality; lexicography in society; dictionaries for the future

The dramatic transformation from print to online dictionaries during the last two decades and the users' preferences for costless online data, waiving proof of the data reliability, has changed the academic and commercial worlds of dictionaries, lexicography as well as the relevant research. In a rapidly changing global and digital society the crucial question that needs to be asked is: what role do dictionaries play in cultural education today? How are dictionaries actually perceived by today's society and how are they used in the process of cultural mediation? How should the reliability and traceability of lexicographic data be assessed? Which lexicographic approaches are at the center of scholarship and where are the publishing houses heading? What should dictionaries for the future look like?

In order to provide answers to these questions, a particular workshop was held in November 2018 at the Centro Italo-Tedesco per l'Eccellenza Europea/Deutsch-Italienisches Zentrum für Europäische Exzellenz, *Villa Vigoni* on Lake Como, Northern Italy. Eighteen experts from the fields of Dictionary Research, Practical Lexicography, German as a Foreign Language, Italian Studies, Translation Science and Empirical Linguistics from Germany, Austria and Italy congregated to debate the following topics in an innovative format of short kick-off presentations and intensive discussion groups:

- lexicography and its status in society;
- the production of dictionaries and the issue of their quality;
- dictionary types and forms of presentation (printed, online and app format);
- dictionary use and didactics.

At the conclusion of the workshop the participants formulated *The 15 Villa Vigoni Theses on Lexicography*:

1) **Dictionaries of the future** are lexical or linguistic information systems in which existing lexicographic data are conflated, multilingualism and linguistic variety are entrenched, and which provide people, when they are confronted with gaps in their knowledge, with an answer as well as support in the writing and formulation processes of texts.
2) Lexical information systems must become a significant topic of **public discourse**. **Awareness** of the fact that the respective online data available should provide the **requisite high quality** must be publicly nurtured.
3) Practical lexicography must constantly be aware of its **social responsibility** and must strive for a comprehensive, pluralistic **description** of **linguistic** and **factual realities**. In the process, the demarcation between the subject area and the selective lexicographic prioritization must be rendered perceptible.

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

4) As independent social institutions, universities and public research facilities must actively participate in **critical discussions and evaluations** regarding lexical information systems.
5) Lexicographic **amendments** in online information systems must be **chronicled** and **preserved** so that they remain permanently available as well-documented evidence of academic processes.
6) Lexicography requires **partners** and **allies**: the solutions and challenges for the lexicography of the future demand, with a view to European perspectives, an **interdisciplinary exchange** between research institutes, academies, publishing houses and other representatives of the private sector.
7) One significant **task for the lexicography** of the digital future is the orderly conflation of data which has been generated automatically by text corpora and specifically processed as well as a user-orientated presentation. The social relevance of such information systems will be consolidated once the underlying corpora mirror the entire linguistic diasystem and are freely available to researchers.
8) Dictionary research must be considered a **cultural science** which, through interdisciplinary projects, conflates practical lexicography, linguistics, computer science, book science and documentation science.
9) In a modern information society, we require academic studies to advance a **standardization process** for metalexicographic **core terminology**, as a solid theory induces multifaceted improvements in practical lexicography.
10) **Academic lexicography** should be increasingly visually creative and with regard to digital formats, it should venture into **experiments**, thereby availing itself of the interest of people in linguistic questions. **State funding** must concentrate on lexicographic innovations.
11) Lexicographic projects should be oriented towards the **specific needs** of the **users** (towards the first language and the foreign language, towards translating *et al.*) as well as towards the users' linguistic acts or communicative intentions, as language is the subject matter of lexicography and learning and understanding languages is a central competence in a globalised world.
12) Academic findings regarding the use of lexicographic information systems as well as **teaching practice** and **translation practice** should be increasingly incorporated into the **lexicographic process**.
13) Lexicography is called upon to develop concepts for a productive **user participation** in lexicographic information systems.
14) The digital supply of data in the information systems of the future must be regarded as a significant means for 'lifelong learning' so that the **critical use of resources** can be established as a **strategic key competence**. This must also be firmly entrenched in the **training and continued education of teachers**.
15) Lexicography requires **pedagogical concepts** in order to be able to accomplish the **didactic implementation** of lexicographic information systems. In the process, this should integrate the media competence of the users.

Although originated in a German-Italian context, the issues raised by the *Villa Vigoni Theses* seem to be relevant for the future direction of lexicography in a European context and beyond. The individual conference participants have disseminated and shared the theses within their networks and in publications and at conferences. The EMLex colloquium "Lexicography at a crossroads" in Tbilisi and the conference "New Challenges in Dictionary Teaching" in Rome in 2021 are just two examples. The theses have also found their way into popular scientific discourse, e. g. in April 2019 in the form of a radio contribution in the series *Dimensions* on the Austrian radio station Ö1.

The workshop participants included the following persons:

Andrea Abel, Laura Balbiani, Wiebke Blanck, Gualtiero Boaglio, Stefan Engelberg, Anne-Kathrin Gärtig-Bressan, Luisa Giacoma, Laura Giacomini, Christine Konecny, Kathrin Kunkel-Razum, Fabio Mollica, Carolin Müller-Spitzer, Martina Nied Curcio, Lorenza Rega, Elmar Schafroth, Rüdiger Scherpe, Stefan Schierholz, Francesco Urzí

Table 1: Workshop Participants

Contact information

Laura Balbiani
Università della Valle d'Aosta
l.balbiani@univda.it

Anne-Kathrin Gärtig-Bressan
Università degli Studi di Trieste
akgaertig@units.it

Martina Nied Curcio
Università degli Studi Roma Tre
martina.nied@uniroma3.it

Stefan Schierholz
Friedrich-Alexander Universität Erlangen-Nürnberg
Stefan.Schierholz@fau.de

Konan Kouassi

MENSCH-MASCHINE-INTERAKTION IM LEXIKOGRAPHISCHEN PROZESS ZU LEXIKALISCHEN INFORMATIONSSYSTEMEN

Abstract Dictionaries of today and tomorrow are rather digital products than print dictionaries. From the user's perspective, electronic dictionary applications and in particular „lexical information systems", also referred to as „digital word information systems" are coming to the fore alongside Google searches. Given the rapid developments in the area of the automated provision of lexicographic information, more precisely the automatic creation of online dictionaries, the new role of the lexicographer in the modern lexicographic process is questionable. This article addresses this issue.

Keywords Beispielangaben; computergenerierte Angabe; Funktion des Lexikographen; lexikalisches Informationssystem; redaktionell erstellte Angaben

Contact information

Konan Jean Mermoz Kouassi
Friedrich-Alexander-Universität Erlangen-Nürnberg
konan.kouassi@fau.de

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Ivana Filipović Petrović

A CORPUS-DRIVEN APPROACH TO LEXICOGRAPHIC DEFINITIONS: THE REPRESENTATION OF MEANING IN THE ELECTRONIC *DICTIONARY OF CROATIAN IDIOMS*

Keywords Corpus-driven approach; idioms; representation of meaning in the dictionary; defining strategies; true electronic features

One of the major concerns in lexicography, both paper and electronic, are defining strategies. Several studies investigated what dictionary users want most from their dictionaries and the results showed that finding meaning is among their primary needs (cf. Wingate 2002; Tarp 2009; Lew 2010). The notion of meaning is especially relevant in the case of idioms[1] – conventionalized multiword expressions with figurative meaning, or traditionally, meaning which is 'not the sum of its parts' (see Fernando 1996; Moon 1998). When it comes to the lexicographic treatment of idioms, there were two major milestones: computer corpora, given that corpus data provide many examples of real usage and context in which idioms tend to occur, and digital medium, which increased search options and eliminated space constraints.

In this text we are presenting functionalities that contribute to the representation of meaning in the electronic corpus-driven *Dictionary of Croatian Idioms*, which is currently available in a beta version on the Lexonomy platform. The macrostructure is organized as follows: the headword is the most frequent variant form of the idiom, i. e., the whole construction. The dictionary entry also contains other variant forms of the idiom, explanation in the form of a reduced sentence (which is best suited to Croatian) and examples of use. Boxes with additional information regarding usage or wordplay are placed at the bottom of entry. A search box gives results if the user inserts any component of an idiom, as well as any variant component which is noted in the entry. It is a true electronic feature (cf. Prinsloo/van Graan 2021), which significantly upgraded the issue of searching in regard to printed dictionaries.

In addition, given that Lexonomy enables cross referencing, idioms with a similar and/or opposite meaning are connected via hyperlinks, thus creating a conceptual network of idioms around a common concept, such as anger, mental condition, or happiness. Idioms are included in a conceptual network depending on their meaning, structure, and use, according to corpus data from the Croatian web corpus hrWaC. For example, idioms that are connected through the concept of anger are very frequent, so they are further grouped according to structural features which contribute to the meaning of the whole construction: one group contains idioms with adjectives and the meaning 'angry': *crven od bijesa* (lit. red with rage) and *ljut kao pas/ris* (lit. angry as a lynx/dog), and the other group contains idioms with perfective verbs and the meaning 'get angry': *pozelenjeti od bijesa* (lit. turn green with rage) and *pao je mrak na oči komu* (lit. darkness fell on someone's eyes).

Furthermore, according to the corpus data, significant number of highly flexible idioms occur with a common lexis, but different grammatical forms (*biti u škripcu* 'be in a corner',

[1] In this paper we use the term 'idiom' in its narrower uses, as a translation equivalent for the Croatian word 'frazem'.

doći u škripac 'get into a corner', *dovesti u škripa koga* 'back someone into the corner', *izvući se iz škripca* 'get out of a corner'). In the *Dictionary of Croatian Idioms*, they are treated as variations and are listed in a single entry, in line with the cognitive linguistic view that variations present the same event in different ways (Langlotz 2006; Parizoska/Omazić 2020). The defining technique implemented here includes a definition of the first, most frequent variant listed in the entry: 'to be in a difficult situation', and other variants are shown underneath with examples of use. Leaving the definition out, we emphasize the role of a variant form which shows lexico-grammatical pattern of use and the role of examples as well, in order to provide not only decoding information in the *Dictionary*, but also encoding. Although studies on the role of examples in language production (Frankenberg-Garcia 2014, 2015) point out that examples that help with encoding and decoding simultaneously are difficult to find, this is exactly what we do. By careful hand-picking, we are looking for examples that bring sufficient context for comprehension and characteristic collocation at the same time.

Overall, this text makes two contributions. Firstly, it shows the connection between corpus data and meaning representation in the electronic *Dictionary of Croatian Idioms*, which is directly reflected in defining strategies. Secondly, it shows that the representation of meaning is organized using available digital functionalities in order to create an up-to-date user-friendly dictionary of Croatian figurative language.

References

Fernando, C. (1996): Idioms and idiomaticity. Oxford.

Frankenberg-Garcia, A. (2014): The use of corpus examples for language comprehension and production. In: ReCALL 26, pp. 128–146.

Frankenberg-Garcia, A. (2015): Dictionaries and encoding examples to support language production. In: International Journal of Lexicography 28 (4), pp. 490–512.

Langlotz, A. (2006): Idiomatic creativity: a cognitive-linguistic model of idiom-representation and idiom-variation in English. Amsterdam.

Lew, R. (2010): Multimodal lexicography: the representation of meaning in electronic dictionaries. In: Lexikos 20, pp. 290–306.

Moon, R. (1998): Fixed expressions and idioms in English. A corpus-based approach. Oxford.

Parizoska, J./Omazić, M. (2020): Sheme dinamike sile i promjenjivost glagolskih frazema. In: Jezikoslovlje 20 (2), pp. 179–205.

Prinsloo, D./van Graan, N. D. (2021): Principles and practice of cross-referencing in paper and electronic dictionaries with specific reference to African languages. In: Lexicography: Journal of ASIALEX 8 (1), pp. 32–58.

Tarp, S. (2009): Reflections on lexicographical user research. In: Lexikos 19, pp. 275–296.

Wingate, U. (2002): The effectiveness of different learners dictionaries. Tübingen.

Contact information

Ivana Filipović Petrović
Linguistic Research Institute
Croatian Academy of Sciences and Arts
ifilipovic@hazu.hr

Ana Salgado / Rute Costa / Toma Tasovac

APPLYING TERMINOLOGICAL METHODS TO LEXICOGRAPHIC WORK: TERMS AND THEIR DOMAINS

Abstract Applying terminological methods to lexicography helps lexicographers deal with the terms occurring in general language dictionaries, especially when it comes to writing the definitions of concepts belonging to special fields. In the context of the lexicographic work of the *Dicionário da Língua Portuguesa*, an updated digital version of the last Academia das Ciências de Lisboa' dictionary published in 2001, we have assumed that terminology – in its dual dimension, both linguistic and conceptual – and lexicography are complementary in their methodological approaches. Both disciplines deal with lexical items, which can be lexical units or terms. In this paper, we apply terminological methods to improve the treatment of terms in general language dictionaries and to write definitions as a form of achieving more precision and accuracy, and also to specify the domains to which they belong. Additionally, we highlight the consistent modelling of lexicographic components, namely the hierarchy of domain labels, as they are term identification markers instead of a flat list of domains. The need to create and make available structured, organised and interoperable lexicographic resources has led us to follow a path in which the application of standards and best practices of treating and representing specialised lexicographic content are fundamental requirements.

Keywords Definition; domain label; general language dictionary; lexicography; term; terminology

Contact information

Ana Salgado
NOVA CLUNL, Centro de Linguística da Universidade NOVA de Lisboa, Portugal/Academia das Ciências de Lisboa, Portugal
ana.salgado@fcsh.unl.pt

Rute Costa
NOVA CLUNL, Centro de Linguística da Universidade NOVA de Lisboa, Portugal
rute.costa@fcsh.unl.pt

Toma Tasovac
BCDH – Belgrade Center for Digital Humanities, Serbia
ttasovac@humanistika.org

This page is part of the publication: Klosa-Kückelhaus, Annette / Engelberg, Stefan / Möhrs, Christine / Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Gilles-Maurice de Schryver

METALEXICOGRAPHY: AN EXISTENTIAL CRISIS

Abstract While there was arguably a need for multi-authored, multi-volume, metalexicographic handbooks three decades ago – when the field of metalexicography was still 'young' – it is a bit puzzling to make sense of the current output-flurry in this field. Is it simply a matter of 'every publisher trying to fill its shelves'? or is there really a need in the scientific community for more and (continuously) updated reference works? And once available, are such works also consulted? Which parts? By whom? How often? For what purposes? In this paper we look at an ongoing, real-world metalexicographic handbook project to answer these questions.

Keywords Metalexicography; major reference work; publishing model; download vs. citation patterns

Contact information

Gilles-Maurice de Schryver
BantUGent – UGent Centre for Bantu Studies, Ghent University
&
Department of African Languages, University of Pretoria
gillesmaurice.deschryver@UGent.be

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Michal Škrabal/Michaela Lišková/Martin Šemelík

ON DEFINING VOCABULARY IN A MONOLINGUAL ONLINE DICTIONARY

Some remarks from the lexicographical practice on the *Academic Dictionary of Contemporary Czech*

Keywords Defining vocabulary; definition; metalanguage; hapax legomenon; Academic Dictionary of Contemporary Czech

1. Theoretical background

Our study focuses on defining vocabulary in the context of monolingual lexicography with the *Academic Dictionary of Contemporary Czech* (ADCC) as the main subject of interest. It is anchored in corpus linguistic research based on the Czech National Corpus (Charles University) and lexicographical practice at the Czech Language Institute (Czech Academy of Sciences). We aim to demonstrate how even a simple statistic can improve dictionary definitions from the user perspective and to offer some recommendations for the authors' future work.

The ADCC is an alphabetical, monolingual general-purpose dictionary. In every monolingual general-purpose dictionary, the meaning description of lexical units plays a crucial role. Consequently, "a systematically selected range of words to be used for describing the content of a larger number of words" (Svensén 2009, p. 246), the defining vocabulary, poses a relevant research topic, especially in connection with the user aspect.

Although there is not a strictly predefined metalanguage for the meaning description for the ADCC, it can be said, however, that (a) the defining vocabulary consists of lexemes, which are included in the ADCC main register and that (b) within certain lexical-semantic classes, standardised basic "pillar" words for the expression of the *genus proximum* are determined (cf. Kochová/Opavská 2016, p. 88). Lexicographers understand and comply with the basic rule that metalanguage should not be too complicated. Apart from distinctive meaning features, some other elements are considered a part of lexical meaning in the ADCC. Among those, we find non-distinctive, facultative features that reflect a complex of knowledge that language users have at the level of common knowledge about denominated non-linguistic facts.

On the contrary, in the Anglo-Saxon tradition, the defining vocabulary is a frequent feature of many dictionaries, especially those that can be termed as learners' dictionaries (Kamiński 2021 for English; Töpel 2021 for German). These have defined their defining vocabularies "to ensure that the definitions are clear and easy to understand and that words used in explanations are easier than the words being defined" (LDoCe p. B17; cf. Xu 2012). Within the Czech lexicography, we have noted only one exception so far (significantly under an English influence): Sinclair et al.'s (1998) *English-Czech Explanatory Dictionary*, which is a bilingualised dictionary based on the original work *Collins COBUILD Student's Dictionary* (1990), provided with a COBUILD Word List that consists of words which appear in mean-

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

ing descriptions at least ten times. In this list, there are 1860 lemmas, respectively 2591 words. (Sinclair et al. 1998, pp. IX, 1162).

Within our experiment, we compare entries with four initial letters, i. e. A–Č, in the ADCC. After publication, A-entries were criticised for being too encyclopaedic, but this could also be due to the prevalence of words of foreign origin, often terms. Conversely, Č-entries are mostly of domestic origin. Besides, the ADCC's conception has changed in the meanwhile. These changes included, among others, the following:

a) On the basis of stricter inclusion rules, the terms are given less prominence as compared to previous practice.

b) With regard to the user aspect, we avoid cognitively overloaded definitions. Definitions undergo a gradual process of "de-encyclopedisation".

2. Quantitative analysis of the ADCC's metalanguage

We used the following procedure for our analysis:

1) In the dictionary editorial system, we exported the definitions of all currently published entries from the Definitions field (omitting synonyms, for which there is another column).

We lemmatised the individual text files and performed simple frequency statistics, which resulted in Table 1.

Initial letter	Tokens	Types	Type-token ratio	H-point	Hapaxes	Hapax-token ratio	Number of entries
A	39,822	7,202	0.181	63	3,490	0.088	2,897
B	56,832	9,546	0.168	72	4,458	0.078	3,806
C	17,702	4,434	0.25	40	2,342	0.132	1,308
Č	20,004	4,304	0.215	46	2,156	0.108	1,236
A–Č	134,360	16,063	0.12	122	6,967	0.052	9,247

Table 1: Frequency statistics of the ADCC's metalanguage (entries A–Č)

Reviewers' comments and instructions given in the reviews should be taken into account.

2) We compared individual groups with a focus on prominent content words that form the pillar of every definition) and, at the same time, on *hapax legomena* that are most prone to be eliminated from ADCC's defining vocabulary.

3) On the basis of qualitative analysis, as well as a comparison with the defining vocabularies from English dictionaries, we aim to make recommendations for authors' future work. As far as hapaxes are concerned, these can be as follows:

   (a) A word can be deleted with no substitute:

   borka I vrchní odumřelá *zkorkovatělá* vrstva kůry kmene dřevin ('the upper dead, *corky* layer of bark of a tree trunk')

   The word *zkorkovatělá* is not only rare and terminological but also redundant in the sense that even without this particular word the meaning delimitation is functional.

(b) A word may be substituted for another, more common word**:**

blaťácké zlato měkký sýr s pružnou konzistencí, *zlatooranžově* zbarveným povrchem a hořkomandlovou, nakyslou chutí (‘a soft cheese with an elastic consistency, a *golden-orange* surface and a bitter, sourish taste’)

The word *zlatooranžově* is rare (*zlatooranžov.** 88 hits (ipm 0.01) in SYN v10). Its more frequent synonym *žlutooranžově* (‘yellow-orange’, *žlutooranžov.** 1,374 hits (ipm 0.23) in SYN v10) is, as additional analyses have revealed, factually more appropriate.

(c) A definition needs to be re-formulated:

brukev 2. rostlina (odrůda brukve zelné) s listy na dlouhých řapících a *ztlustlým* stonkem, pěstovaná jako zelenina; syn. kedlubna 1 (‘a plant (a variety of kohlrabi) with leaves on long leafstalks and a *thickened* stem, grown as a vegetable; syn. turnip cabbage 1’)

The word *ztlustlý* is not common (*ztlustl.** 337 hits (ipm 0.06) in SYN v10). Additional evaluation of the definition resulted in the conclusion that the word *bulva/hlíza* (‘tuber’) should be included in the definition leading to post-editional measures taken in this respect.

3. Conclusion

Our study shows how a simple frequency statistic of defining vocabulary can improve lexicographic definitions and serve the user-friendliness of a dictionary. One can also ask more general questions, e. g.: Where do the words in the defining vocabulary actually come from? Do they necessarily overlap with the “core general vocabulary” (Brezina/Gablasová 2015) of a given language? In fact, such a vocabulary already exists for Czech (Čermák/Křen 2011), and besides its pedagogical use, its application to the field of lexicography is logically suggested. The result of this case study will be considered in the further modification of the ADCC’s conception.

References

ADCC (2017–2020): Akademický slovník současné češtiny. A–Č. Prague. http://slovnikcestiny.cz/uvod.php (last access: 24-03-2022).

Brezina, V./Gablasova, D. (2015): Is there a core general vocabulary? Introducing the New General Service List. Applied Linguistics 36 (1), pp. 1–22.

Collins COBUILD student’s dictionary (1990). London.

Čermák, F./Křen, M. (2011): Frekvenční slovník češtiny. Prague.

Kamiński, M. P. (2021): Defining with simple vocabulary in English dictionaries. Amsterdam.

Kochová, P./Opavská, Z. (2016): Kapitoly z koncepce Akademického slovníku současné češtiny. [released by Jazairiová, P./Opavská, Z.]. Prague.

LDoCE (1995): Longman dictionary of contemporary English. 3rd edition. Munich.

Sinclair, J. et al. (1998): Anglicko-český výkladový slovník. Prague.

Svensén, B. (2009): A handbook of lexicography. The theory and practice of dictionary-making. Cambridge.

SYN v10 (2022): Křen, M. et al.: Korpus SYN, verze 10 z 22. 2. 2022. Prague. http://www.korpus.cz (last access: 24-03-2022).

Töpel, A. (2021): Der Definitionswortschatz im einsprachigen Lernerwörterbuch des Deutschen. Anspruch und Wirklichkeit. Tübingen.

Xu, H. (2012): A critique of the controlled defining vocabulary in Longman dictionary of contemporary English. In: Lexikos 22, pp. 367–381.

Contact information

Michal Škrabal
Institute of the Czech National Corpus, Charles University, Prague
michal.skrabal@ff.cuni.cz

Michaela Lišková
Czech Language Institute, Czech Academy of Sciences, Prague
liskova@ujc.cas.cz

Martin Šemelík
Czech Language Institute, Czech Academy of Sciences, Prague
semelik@ujc.cas.cz

Acknowledgements

The preparation of this article was financed within the statutory activity of the Czech Language Institute of the Czech Academy of Sciences (RVO No. 68378092) and the European Regional Development Fund-Project “Creativity and Adaptability as Conditions of the Success of Europe in an Interrelated World” (No. CZ.02.1.01/0.0/0.0/16_019/0000734).

Corpora in Lexicography

XX EURALEX INTERNATIONAL CONGRESS
DICTIONARIES AND SOCIETY
12-16 July 2022, Mannheim, Germany

Nils Diewald/Marc Kupietz/Harald Lüngen

TOKENIZING ON SCALE

Preprocessing large text corpora on the lexical and sentence level

Abstract When comparing different tools in the field of natural language processing (NLP), the quality of their results usually has first priority. This is also true for tokenization. In the context of large and diverse corpora for linguistic research purposes, however, other criteria also play a role – not least sufficient speed to process the data in an acceptable amount of time. In this paper we evaluate several state-of-the-art tokenization tools for German – including our own – with regard to theses criteria. We conclude that while not all tools are applicable in this setting, no compromises regarding quality need to be made.

Keywords Corpora; tokenization; software

Contact information

Nils Diewald
Leibniz-Institut für Deutsche Sprache
diewald@ids-mannheim.de

Marc Kupietz
Leibniz-Institut für Deutsche Sprache
kupietz@ids-mannheim.de

Harald Lüngen
Leibniz-Institut für Deutsche Sprache
luengen@ids-mannheim.de

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Ana-Maria Gînsac/Mihai-Alex Moruz/Mădălina Ungureanu

THE FIRST ROMANIAN DICTIONARIES (17TH CENTURY). DIGITAL ALIGNED CORPUS

Abstract This paper presents the project "The first Romanian bilingual dictionaries (17th century). Digitally annotated and aligned corpus" (eRomLex) which deals with the editing of the first bilingual Romanian dictionaries. The aim of the project is to compile an electronic corpus comprising six Slavonic-Romanian lexicons dating from the 17th century, based on their relatedness and the fact that they follow a common model in order to highlight the characteristics of this lexicographical network (the affiliations between the lexicons, the way they relate to the source, the innovations towards it, their potential uses) and to facilitate the access to their content. A digital edition allows exhaustive data extraction and comparison and link with other digitized resources for old Romanian or Church Slavonic, including dictionaries. After presenting the corpus, we point to the necessary stages in achieving this project, the techniques used to access the material and the challenges and obstacles we encountered along the way. We describe how the corpus was created, stored, indexed and can be searched over; we will also present and discuss some statistical analyses highlighting relations between the Romanian lexicons and their Slavonic-Ruthenian source.

Keywords Romanian lexicography; 17th century; Church Slavonic; bilingual dictionaries in electronic format

Contact information

Ana-Maria Gînsac
Institute of Interdisciplinary Research, Department of Social Sciences and Humanities, "Alexandru Ioan Cuza" University, Iași
anamaria_gansac@yahoo.com

Mihai-Alex Moruz
Faculty of Computer Science, "Alexandru Ioan Cuza" University, Iași
mmoruz@info.uaic.ro

Mădălina Ungureanu
Institute of Interdisciplinary Research, Department of Social Sciences and Humanities, "Alexandru Ioan Cuza" University, Iași
madandronic@gmail.com

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Iztok Kosem

TRENDI – A MONITOR CORPUS OF SLOVENE

Abstract In this paper we present Trendi, a monitor corpus of written Slovene, which has been compiled recently as part of the SLED (Monitor corpus and related resources) project. The methodology and the contents of the corpus are presented, as well as the findings of the survey that aimed to identify the needs of potential users related to topical language use. The Trendi corpus currently contains news articles and other web content from 110 different sources, with the texts being collected and linguistically annotated on a daily basis. The corpus complements Gigafida 2.0, a 1.13-billion-word reference corpus of standard written Slovene. Also discussed are the ways in which the corpus will be integrated into various lexicographic projects, helping not only in the identification of neologisms but also in monitoring changes in already identified language phenomena.

Keywords Monitor corpus; language use; trends; Slovene; neologisms; lexicography; newsfeed

Contact information

Iztok Kosem
Jožef Stefan Institute & Faculty of Arts, University of Ljubljana
iztok.kosem@ijs.si

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Simon Krek/Polona Gantar/Iztok Kosem

EXTRACTION OF COLLOCATIONS FROM THE GIGAFIDA 2.1 CORPUS OF SLOVENE

Abstract This paper describes a method for extracting collocation data from text corpora based on a formal definition of syntactic structures, which takes into account not only the POS-tagging level of annotation but also syntactic parsing (syntactic treebank model) and introduces the possibility of controlling the canonical form of extracted collocations based on statistical data on forms with different properties in the corpus. Specifically, we describe the results of extraction from the syntactically tagged Gigafida 2.1 corpus. Using the new method, 4,002,918 collocation candidates in 81 syntactic structures were extracted. We evaluate the extracted data sample in more detail, mainly in relation to properties that affect the extraction of canonical forms: definiteness in adjectival collocations, grammatical number in noun collocations, comparison in adjectival and adverbial collocations, and letter case (uppercase and lowercase) in canonical forms. The conclusion highlights the potential of the methodology used for the grammatical description of collocation and phrasal syntax and the possibilities for improving the model in the process of compilation of a digital dictionary database for Slovene.

Keywords Collocations; discovering collocations in corpora; digital collocation database

Contact Information

Polona Gantar
University of Ljubljana
apolonija.gantar@guest.arnes.si

Iztok Kosem
Jožef Stefan Institute & Faculty of Arts, University of Ljubljana
iztok.kosem@ijs.si

Simon Krek
Jožef Stefan Institute
simon.krek@guest.arnes.si

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Meike Meliss/Vanessa González Ribao

VERGLEICHBARE KORPORA FÜR MULTILINGUALE KONTRASTIVE STUDIEN

Herausforderungen und Desiderata

Abstract This contribution aims to show the necessity of working in the development of multilingual corpora and appropriate tools for multilingual contrastive studieS. We take the corpus of the lexicographical project COMBIDIGILEX as example to show, how difficult it is to build a suitable data basis to study and compare linguistic phenomena in German, Spanish and Portuguese. Despite the availability of big reference corpora for the three languages (at least for written language), it is not able to obtain a comparable data basis from, because the mentioned corpora are created according to different requirements and they are also powered by disparate information systems and analyse toolS. To break the status quo, we plead for increasing research infrastructures by means of compatible language technology and sharing data.

Keywords Corpus linguistics; comparative corpora; contrastive multilingual linguistics; language technologies

Contact information

Meike Meliss
Universidad de Santiago de Compostela
meike.meliss@usc.es

Vanessa González Ribao
Postdoc-Stipendiatin der Fritz-Thyssen-Stiftung
vanessina_gr@hotmail.com

Irene Renau/Rogelio Nazar

TOWARDS A MULTILINGUAL DICTIONARY OF DISCOURSE MARKERS

Automatic extraction of units from parallel corpus

Abstract This paper presents a multilingual dictionary project of discourse markers. During its first stage, consisting of collecting the list of headwords, we used a parallel corpus to automatically extract units from texts written in Spanish, Catalan, English, French and German. We also applied a method to create a taxonomy structure for automatically organising the markers in clusters. As a result, we obtain an extensive, corpus-driven list of headwords. We present a prototype of the microstructure of the dictionary in the form of a standard XML database and describe the procedure to automatically fill in most of its fields (e. g., the type of DM, the equivalents in other languages, etc.), before human intervention.

Keywords Computational lexicography; corpus-driven lexicography; discourse markers; multilingual lexicography

Contact Information

Irene Renau
Pontificia Universidad Católica de Valparaíso, Chile
irene.renau@gmail.com

Rogelio Nazar
Pontificia Universidad Católica de Valparaíso, Chile
rogelio.nazar@pucv.cl

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Jan Oliver Rüdiger/Sascha Wolfer/Alexander Koplenig/
Frank Michaelis/Carolin Müller-Spitzer/Samira Ochs/
Louis Cotgrove

OWIDplusLIVE

Day-to-day collection, exploration, analysis, and visualization of N-Gram frequencies in German (online press) language

Keywords LIVE-Data; N-Gram; German; data exploration; visualization

With OWIDplusLIVE, we would like to introduce the EURALEX community to two resources that provide analytical access to daily updated data (data: frequency data and N-grams – reference point: previous day).

The project started following the first confirmed COVID-19 cases in Germany. It was already clear at the time that the social impact of the pandemic would be immense. And yet, in retrospect, it is very surprising how broad and wide-reaching the influence of the pandemic has been, especially at the level of German-language vocabulary (see Wolfer et al. 2022). It remains to be seen how persistent and lasting these influences are, i. e., how many of the words that have found their way into everyday usage will continue to be consistently used in the future.

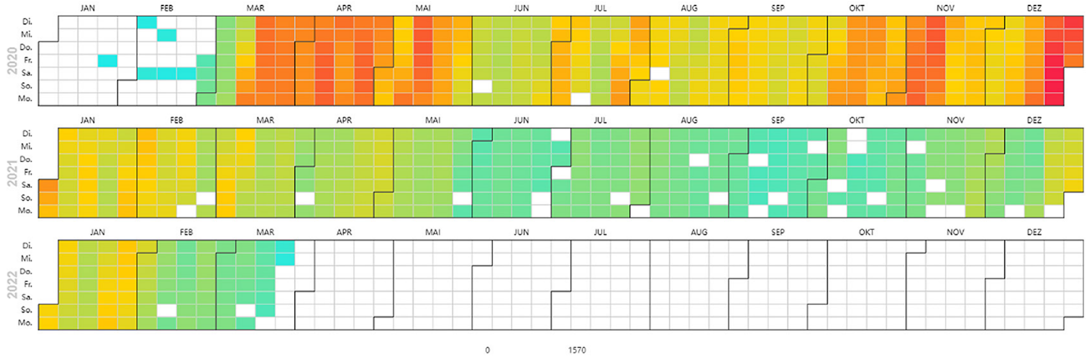


Fig. 1: OWIDplusLIVE – calendar view – query “corona VVFIN”

Against this background, it is necessary to develop instruments for monitoring language as close to real time as possible. This will allow the analysis of new events, conflicts and socially relevant topics.

Most “traditional” corpora don’t allow to analyze corpus material in real-time (first notable projects in this field are e. g.: Davies 2013; Vogel et al. 2021; Barbaresi 2022). Therefore, we present two resources to track language use in a limited subset of the German language. The first resource is an RSS corpus containing titles and so-called “descriptions” (leads or teasers) to online newspaper articles from 13 German-language online sources (one each from Switzerland and Austria, 11 from Germany). Currently (2022-Mar-23), the corpus contains about 67.9 million tokens in about 1.5 million feed items since January 1, 2020. The second resource is called OWIDplusLIVE (https://www.owid.de/plus/live-2021/), a web application that allows to investigate the frequency of uni-, bi- and trigrams in the mentioned RSS corpus. Users can perform searches at three different layers: word-form, lemma and part-of-speech (within our NLP pipeline we use (Schmid 1995) for automatic annotation). Wildcard searches are also possible. In order to make the application accessible to a wide range of user groups, we have deliberately avoided implementing a complex search syntax. We also provide a (German) video tutorial for a quick start.

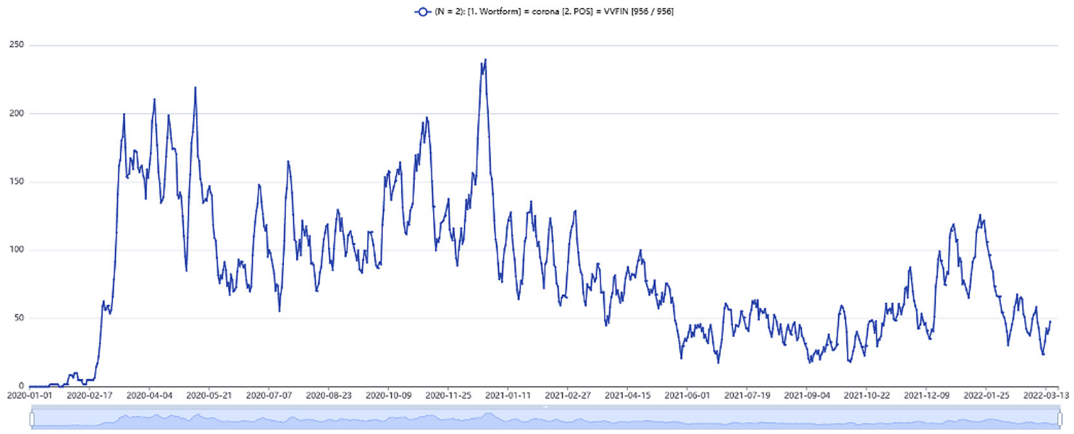


Fig. 2: OWIDplusLIVE – line chart – query “corona VVFIN”

Currently, three visualizations are available: 1) a calendar view (see Fig. 1) where each day is colored according to the (relative or absolute) frequency of the current day; 2) a line chart (see Fig. 2) showing the (relative or absolute) frequencies. Here, frequencies can also be compared between multiple queries; 3) a Sankey diagram (see Fig. 3) that reveals language usage patterns, especially for more complex queries. All visualizations have interactive elements such as mouseover information and/or zoom and selection options. All queries can be further explored in a list of all results. This feature is a key feature of this tool (e. g., compared to other similar N-Gram viewers) enabling the transparent exploration of results. In other words, you not only get an aggregated total frequency for the query “corona VVFIN” (word form “corona” followed by a finite verb) but also separate time series (these can be manually de/selected). This enables results to be fine-tuned according to the needs of the researcher. Furthermore, we want to avoid to have a ‘black box’-software. Researchers can understand which time series are included in the visualization and what impact a time series has on the overall result.

Results, queries, visualizations, and result data can be shared and exported via a URL, JSON, TSV (tab-delimited text) format.

The source code and a documentation are freely available. In the context of lexicography, OWIDplusLIVE can be used for all issues related to vocabulary change and lexical innovation.

References

Barbaresi, A. (2022): Webmonitor. https://www.dwds.de/d/korpora/webmonitor (last access: 23-03-2022).

Davies, M. (2013): Corpus of News on the Web (NOW): 3+ billion words from 20 countries, updated every day. https://corpus.byu.edu/now/ (last access: 23-03-2022).

Schmid, H. (1995): TreeTagger. https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/ (last access: 23-03-2022).

Vogel, F./Bäumer, B./Deus, F./Knobloch, C./Rüdiger, J. O./Schmallenbach, J./Schölzel, H./Tripps, F./Weber, S./Wilton, A. (2021): www.Diskursmonitor.de – gemeinschaftlich erarbeitete Online-Plattform zur Aufklärung und Dokumentation strategischer Kommunikation. http://doi.org/10.5281/ZENODO.5780230 (last access: 23-03-2022).

Wolfer, S./Koplenig, A./Michaelis, F./Müller-Spitzer, C./Rüdiger, J. O. (2022): Wie können wir den Einfluss der Corona-Pandemie auf die Verteilungen im deutschen Online-Pressewortschatz messen und explorieren? In: Kämper, H./Plewnia, A. (eds.): Sprache in Politik und Gesellschaft. Perspektiven und Zugänge. (= Jahrbuch des Instituts für Deutsche Sprache 2021). Berlin/Boston, pp. 331–338. http://doi.org/10.1515/9783110774306-022 (last access: 23-03-2022).

Contact information

Jan Oliver Rüdiger
Leibniz-Institut für Deutsche Sprache
ruediger@ids-mannheim.de

Sascha Wolfer
Leibniz-Institut für Deutsche Sprache
wolfer@ids-mannheim.de

Alexander Koplenig
Leibniz-Institut für Deutsche Sprache
koplenig@ids-mannheim.de

Frank Michaelis
Leibniz-Institut für Deutsche Sprache
michaelis@ids-mannheim.de

Carolin Müller-Spitzer
Leibniz-Institut für Deutsche Sprache
mueller-spitzer@ids-mannheim.de

Samira Ochs
Leibniz-Institut für Deutsche Sprache
ochs@ids-mannheim.de

Louis Cotgrove
Leibniz-Institut für Deutsche Sprache
cotgrove@ids-mannheim.de

Data Models and Databases in Lexicography

XX EURALEX INTERNATIONAL CONGRESS
DICTIONARIES AND SOCIETY
12-16 July 2022, Mannheim, Germany

Thierry Declerck

INTEGRATION OF SIGN LANGUAGE LEXICAL DATA IN THE ONTOLEX-LEMON FRAMEWORK

Abstract We describe the status of work intending at including sign language lexical data within the OntoLex-Lemon framework. Our general goal is to provide for a multimodal extension to this framework, which was originally conceived for covering only the written and phonetic representation of lexical data. Our aim is to achieve in the longer term the same type of semantic interoperability between sign language lexical data as this is achieved for their spoken or written counterparts. We want also to achieve this goal across modalities: between sign language lexical data and spoken/written lexical data.

Keywords Sign Languages; OntoLex-Lemon; lexical data

Contact information

Thierry Declerck
DFKI GmbH, Multilinguality and Language Technology, Saarland Informatics Campus D3 2,
Saarland, Germany
declerck@dfki.de

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Birgit Füreder

ÜBERLEGUNGEN ZUR MODELLIERUNG EINES MULTILINGUALEN ‚PERIPHRASTIKONS‘:

Ein französisch-italienisch-spanisch-englisch-deutscher Versuch

Abstract In the course of the last years, digital lexicography has opened up a variety of avenues fostering the conceptualisation, application and use of constructicons, a type of lexicographical reference work which has revealed itself highly promising in terms of connectivity and flexibility, at the same time, however, also challenging as to its technical implementation. The present paper takes up the ambitious aim to propose some reflections as well as a first draft for a possible model of a multilingual 'periphrasticon' as a subtype of a bigger constructicon focusing on a specific typology-related structural feature, i.e. periphrasticity. Taking periphrastic verbal constructions in French, Italian and Spanish as a starting point, it tries to sketch out a unified constructional network including not only equivalent (or corresponding) constructions within Romance, but also establishing (formal and functional) cross-linguistic connections to German and English. Comprising the major languages available to most language learners in (at least) German-speaking environments, the model is also supposed to pave the way for multilingual constructicography which, on the one hand, is able to account for intra- and cross-linguistic relations and, on the other hand, can also prove a valuable tool for language learning and use.

Keywords Multilingual lexicography; periphrastic constructions; French; Italian; Spanish; English; German

Contact information

Birgit Füreder
Universität Salzburg
birgitursula.fuereder2@plus.ac.at

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Yevhen Kupriianov/Iryna Ostapova/Volodymyr Shyrokov/Mykyta Yablochkov

VIRTUAL LEXICOGRAPHIC LABORATORY AS A LINGUIST ASSISTANT IN CONDUCTING DICTIONARY-BASED RESEARCHES: CASE OF VLL DLE 23

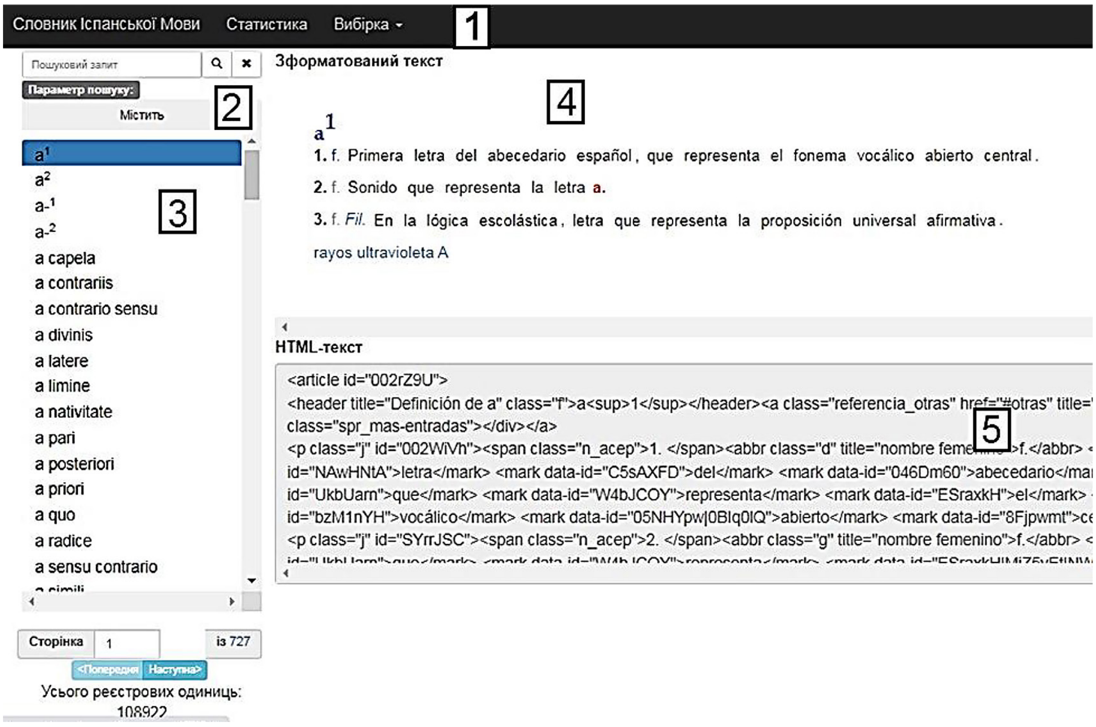
Keywords Virtual lexicographic laboratory; explanatory dictionary; lexicographic text; lexicographic system; dictionary-based researches

Introduction. The virtual lexicographic laboratories (VLL) developed in the Ukrainian Language Information Fund offer effective tools to conduct dictionary-based researches. In this regard, lexicographic text is considered not only as a reference product but a tool assisted basis for creating and updating a dictionary as well as a means of professional communication and transfer of linguistic knowledge. This mainly concerns comprehensive explanatory dictionaries that are characterized by detailed and multi-aspect description of language. Hence, the problem in question is providing these dictionaries with appropriate tools that make it possible to derive from the dictionary text any linguistic information while conducting linguistic researches. The present paper shares the authors' experience in elaborating such tools while carrying out the project of VLL DLE 23, a virtual lexicographic laboratory to perform linguistic researches on the basis of the Spanish dictionary "Diccionario de la lengua Española. 23ª edición" (DLE 23).

To implement the project of VLL DLE 23 the following tasks were to be resolved: 1) building up a formal model of the dictionary (hereinafter referred to as lexicographic system) on the basis of which the VLL DLE 23 database and interface elements are to be developed; 2) implementing parsing technology to convert the structured dictionary text into database format; 3) selecting the format for the database containing all the information elements of the dictionary entry (the choice was made in favor of a document database); and 4) elaboration of the user interface of VLL DLE 23 which provides the access and work with the information elements of the dictionary entries in the database.

Current state of VLL DLE 23 project. The pilot prototype of the VLL DLE 23 allows users to analyze the vocabulary contained in the word list and to examine the entry structure, as well as to provide access to some information elements of the dictionary entries. The main interface (Fig. 1) consists of the menu bar (1), the search panel (2), the wordlist panel (3), the entry area (4) and the area to view dictionary entry text in HTML format (5) to be used for full-text search in the dictionary.

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.



a1

1. f. Primera letra del abecedario español, que representa el fonema vocálico abierto central.

2. f. Sonido que representa la letra **a**.

3. f. *Fil.* En la lógica escolástica, letra que representa la proposición universal afirmativa.

rayos ultravioleta A

```
<article id="002rZ9U">
<header title="Definición de a" class="f">a<sup>1</sup></header><a class="referencia_otras" href="#otras" title="
class="spr_mas-entradas"></div></a>
<p class="j" id="002WiVh"><span class="n_acep">1. </span><abbr class="d" title="nombre feme  >f.</abbr> <
id="NAwHNtA">letra</mark> <mark data-id="C5sAXFD">del</mark> <mark data-id="046Dm60">abecedario</mar
id="UkbUarn">que</mark> <mark data-id="W4bJCOY">representa</mark> <mark data-id="ESraxkH">el</mark> <
id="bzM1nYH">vocálico</mark> <mark data-id="05NHYpw|0BIq0IQ">abierto</mark> <mark data-id="8Fjpwmt">ce
<p class="j" id="SYrrJSC"><span class="n_acep">2. </span><abbr class="g" title="nombre femenino">f.</abbr> <
```

Fig. 1: General view of the main window of VLL DLE 23

As the VLL development is still underway, the language of the interface remains Ukrainian. However, in final version the users will be able to switch between English, Spanish and Ukrainian. With the current interface the users can: 1) generate statistics on the whole dictionary or specific sample of the dictionary entries; 2) perform linguistic researches aimed at studying lexical meanings, etymology, grammar and usage peculiarities of the Spanish language units; and 3) build sub-dictionaries on the basis of DLE 23, for example: sub-dictionary of morphemes, homonyms, collocations, etc.

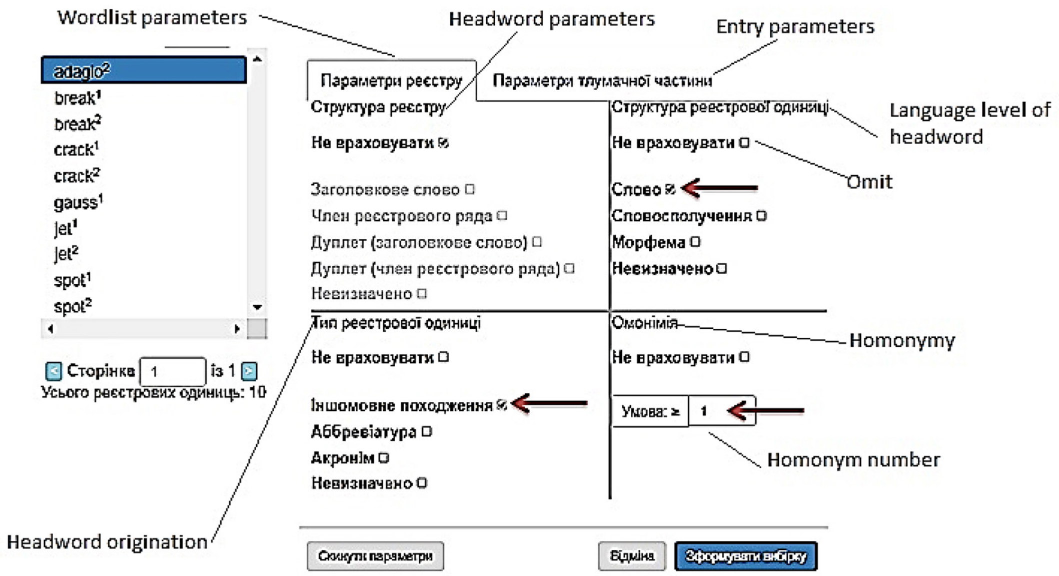
The menu bar contains two main tools: “Статистика” (Statistics) and “Вибірка” (Sample). The first automatically gets indicators for a separate sample or the entire dictionary. The second one offers parameters to form the samples of DLE 23 entries:

- **Headword parameters:** lemma, masculine, feminine, regional variant, not defined;
- **Language level of the headword:** word, collocation, morpheme, not defined;
- **Headword origination:** borrowing, contraction, acronym, not defined;
- **Homonym:** numerical value.

For example, we can get a sample of homonymous foreign words in the modern Spanish language. To do this, the following parameters are to be selected:

- **Language level of the headword:** word;
- **Headword origination:** borrowing;
- **Homonym:** ≥ 1.

The results are shown in Figure 2 (on the left).



adagio^2
break^1
break^2
crack^1
crack^2
gauss^1
jet^1
jet^2
spot^1
spot^2

Сторінка 1 із 1
Усього реєстрових одиниць: 10

Параметри реєстру | Параметри тлумачної частини

Структура реєстру
Не враховувати ☑
Заголовкове слово □
Член реєстрового ряда □
Дуплет (заголовкове слово) □
Дуплет (член реєстрового ряда) □
Невизначено □

Структура реєстрової одиниці
Не враховувати □
Слово ☑
Словосполучення □
Морфема □
Невизначено □

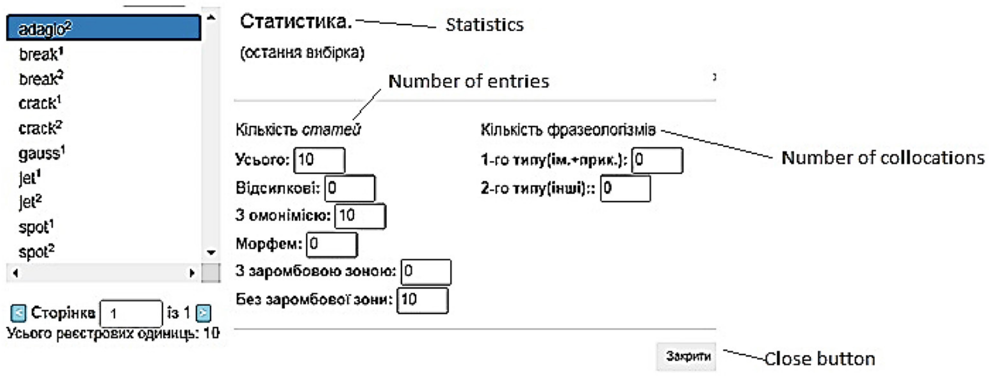
Тип реєстрової одиниці
Не враховувати □
Іншомовне походження ☑
Аббревіатура □
Акронім □
Невизначвно □

Омонімія
Не враховувати □
Умова: ≥ 1

Скинути параметри | Відміна | Зформувати вибірку

Fig. 2: List of the entries with homonymic words of foreign origin (left) and parameters selected (right)

With "Statistics" (Fig. 3) we get the following figures for the above sample:



adagio^2
break^1
break^2
crack^1
crack^2
gauss^1
jet^1
jet^2
spot^1
spot^2

Сторінка 1 із 1
Усього реєстрових одиниць: 10

Статистика.
(остання вибірка)

Кількість *статей*
Усього: 10
Відсилкові: 0
З омонімією: 10
Морфем: 0
З заромбовою зоною: 0
Без заромбової зони: 10

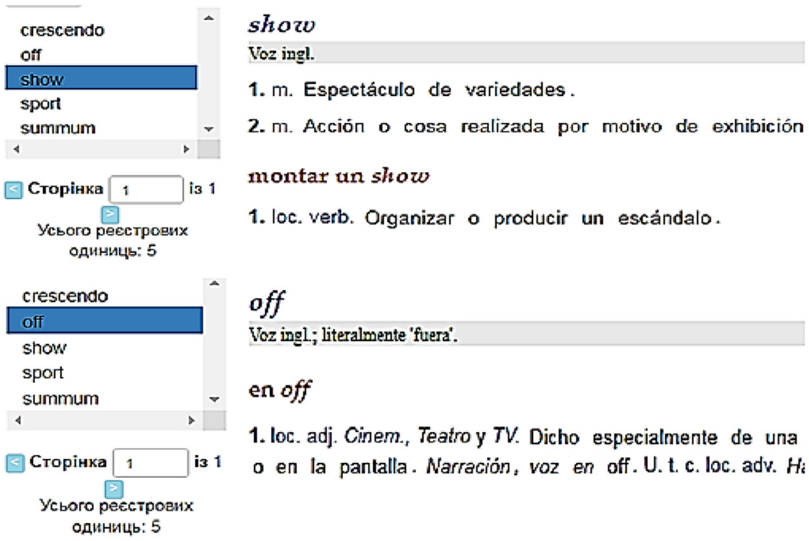
Кількість фразеологізмів
1-го типу(ім.+прик.): 0
2-го типу(інші):: 0

Закрити

Fig. 3: Statistics for the sample of homonymous words of foreign origin

- Total number of the entries: 10;
- Referential entries: 0;
- Homonyms: 10;
- Entries with collocations: 0;
- Entries without collocations: 0.

Another example is getting a sample of Spanish collocations composed of foreign words (Fig. 4.).



show

Voz ingl.

1. m. Espectáculo de variedades.

2. m. Acción o cosa realizada por motivo de exhibición

montar un *show*

1. loc. verb. Organizar o producir un escándalo.

off

Voz ingl.; literalmente 'fuera'.

en *off*

1. loc. adj. *Cinem.*, *Teatro* y *TV.* Dicho especialmente de una o en la pantalla. *Narración, voz en* off. U. t. c. loc. adv. *Hi*

Fig. 4: Sample of dictionary entries containing Spanish collocations composed of foreign words

Future development of VLL DLE 23 project. The next version will ensure the access to all the information elements of the entry provided by the dictionary lexicographic system. The users will be able to specify the content of any information element. Further extension of the functionality will be determined by the request for solving linguistic problems.

References

Diccionario de la lengua española (2022): https://dle.rae.es/ (last access: 23-03-2022).

Kupriianov, Y./Ostapova, I./Yablochkov, M. (2020): Design of the user's interface of virtual lexicographic laboratory for explanatory dictionary of the Spanish language. In: Lytvyn, S. (ed.): Proceedings of the 4th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2020), Lviv, 23–24 April 2020. Volume I: Main Conference. Lviv. http://ceur-ws.org/Vol-2604/paper10.pdf (last access: 23-03-2022).

Kupriianov, Y./Shyrokov, V./Ostapova, I./Yablochkov, M. (2021): Digital toolkit to develop research potential of explanatory dictionary (case of Spanish language dictionary). In: Lytvyn, S. (ed.): Proceedings of the 5th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2021), Kharkiv, 22–23 April 2021. Volume I: Main Conference. Kharkiv. http://ceur-ws.org/Vol-2870/paper28.pdf. (last access: 23-03-2022).

Stanković, R./Stijović, R./Duško, V./Krstev, C./Sabo, O. (2018): The dictionary of the Serbian academy: from the text to the lexical database. In: Krek, S./Jaka C./Gorjanc, V./Kosem, I. (eds.): Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts, Ljubljana, 17–21 July 2018. Ljubljana, p. 941. https://euralex.org/publications/the-dictionary-of-the-serbian-academy-from-the-text-to-the-lexical-database/ (last access: 23-03-2022).

Contact information

Yevhen Kupriianov
National Technical University "Kharkiv Polytechnic Institute"
eugeniokuprianov@gmail.com

Iryna Ostapova
Ukrainian Lingua-Information Fund, NAS of Ukraine
eugeniokuprianov@gmail.com

Volodymyr Shyrokov
Ukrainian Lingua-Information Fund, NAS of Ukraine
vshirokov48@gmail.com

Mykyta Yablochkov
Ukrainian Lingua-Information Fund, NAS of Ukraine
gezartos@gmail.com

David Lindemann/Penny Labropoulou/Christiane Klaes

INTRODUCING LexMeta: A METADATA MODEL FOR LEXICAL RESOURCES

Abstract In this paper, we present **LexMeta**, a metadata model for the description of human-readable and computational lexical resources in catalogues. Our initial motivation is the extension of the **LexBib** knowledge graph with the addition of metadata for dictionaries, making it a catalogue *of* and *about* lexicographical works. The scope of the proposed model, however, is broader, aiming at the exchange of metadata with catalogues of Language Resources and Technologies and addressing a wider community of researchers besides lexicographers. For the definition of the LexMeta core classes and properties, we deploy widely used RDF vocabularies, mainly **Meta-Share**, a metadata model for Language Resources and Technologies, and **FRBR**, a model for bibliographic records.

Keywords Lexical resources metadata; linked data; Wikibase; semantic web

Contact information

David Lindemann
UPV/EHU University of the Basque Country (Spain)
david.lindemann@ehu.eus

Penny Labropoulou
Institute for Language and Speech Processing (ILSP)/Athena R. C. (Greece)
penny@athenarc.gr

Christiane Klaes
TU Braunschweig (Germany)
University Library
c.klaes@tu-braunschweig.de

Manuel Márquez

UN MODELO ESTRUCTURAL DE DATOS LEXICOGRÁFICOS PARA LA CODIFICACIÓN EN XML DE UN DICCIONARIO DE APRENDIZAJE DE LATÍN: LA DTD DE LOS LEMAS VERBALES

Keywords Computational lexicography; DTD; XML; learner’s dictionaries

La evolución tecnológica ha contribuido notablemente a la mejora de la praxis lexicográfica, especialmente – pero no solo – durante las dos últimas décadas. La versatilidad que proporciona, por ejemplo, la codificación de los datos lexicográficos en un formato XML permite tanto su libre transferencia como su análisis y gestión por diferentes aplicaciones sin necesidad de modificar dichos datos. Asimismo, el desarrollo de los Dictionary Writing Systems (DWS) facilita el proceso de confección de un diccionario, toda vez que proporcionan un entorno de trabajo adecuado y productivo con vistas al almacenamiento, la gestión y la explotación de los datos lexicográficos (Rubio López et al. 2021; Woldrich et al. 2020; Simões et al. 2019; Abel 2012). No obstante, la tecnología no es el único pilar que sustenta la confección de una obra lexicográfica. Es necesario que unos principios teóricos sólidos fundamenten la estructura de los datos con los que se trabaja.

Teniendo en cuenta lo expuesto en las líneas precedentes y la experiencia adquirida durante cinco proyectos de innovación educativa dedicados a la confección y al testado de dos diccionarios de aprendizaje de lenguas declinables, nos planteamos la posibilidad de mejorar la explotación de este tipo de diccionarios nivelando la información que proporcionan.

Afrontamos el reto de diseñar una DTD que define una estructura de datos lexicográficos codificados en un formato XML que permite una codificación de los contenidos fundamentales de un diccionario de aprendizaje de lenguas (Heuberger, 2018) al tiempo que una explotación variada de esos datos en atención al nivel cognitivo del usuario. El prototipo diseñado reutiliza una selección de los datos básicos de un diccionario electrónico de latín-español de iniciación al estudio de dicha lengua ya validado, reorganizando y añadiendo nueva información codificada en XML. La DTD diseñada permite validar la estructura y la sintaxis del documento para que tanto las búsquedas como la información devuelta san correctas.

Como logros del diseño y de su posterior experimentación se citan los siguientes:

1) Crear un modelo estándar de codificación de datos lexicográficos que permita una explotación nivelada, complementando la DTD de diccionarios de TEI.

2) Explotar los datos XML mediante el interfaz de un único diccionario electrónico que resulte efectivo para responder a las diferentes necesidades cognitivas del alumnado de latín (atención a la diversidad) en un mismo contexto educativo.

3) Facilitar diferentes ritmos de aprendizaje mediante el uso de los datos lexicográficos.

4) Diseñar un sistema de estructura de datos que, una vez testado, pueda ser aplicado a otras lenguas, declinables y no declinables, como el griego (clásico y moderno), el alemán y el inglés.

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Presentamos en este trabajo un fragmento de una DTD, concretamente aquel que corresponde al tratamiento de los verbos, unidad léxica principal de la estructura lexicográfica. Dicha DTD viene a complementar y mejorar la propuesta de TEI (https://tei-c.org/release/doc/tei-p5-doc/es/html/DI.html), por cuanto potencia la medioestructura del diccionario resultante (estableciendo vínculos entre verbos que pertenecen a una misma clase) y nivela la información relativa a cada verbo: clase de verbo (se sigue, a grandes rasgos, la clasificación proporcionada por ADESSE: http://adesse.uvigo.es/data/clases.php), significados, complementación verbal (descripción de su valencia cuantitativa y cualitativa, esta última en atención a la descripción morfológica de los argumentos verbales y a su ontología), colocaciones, etimología y ejemplos. La cantidad de información y su grado de complejidad se incrementa según el nivel de conocimientos lingüísticos del estudiante: n1 (básico), n2 (medio), n3 (avanzado)

El diseño de la DTD se fundamenta en los siguientes principios lingüísticos: la Gramática Dependencial de Tesnière (1959) en lo que se refiere a la descripción de la valencia verbal; la Lingüística Funcional de Dik (1997) en cuanto a los roles semánticos que desempeñan los argumentos verbales (agente, procesado, fuerza, meta…); por último, una ontología adaptada de las entidades de primer y segundo orden de Lyons (1980) utilizada para la descripción de la caracterización léxica de los argumentos verbales (+/− animado, +/− humano, +/− concreto, lugar). En cuanto a los principios lexicográficos, entendemos que, en el ámbito del aprendizaje de lenguas, la Teoría Funcional de la Lexicografía proporciona un marco teórico idóneo por cuanto establece unas bases teórico-prácticas adecuadas para determinar las fases de confección del diccionario y la estructura de la información codificada en atención a los diferentes tipos de necesidades de los potenciales usuarios de los datos lexicográficos (Tarp 2013). Por último, confirmamos que está previsto volcar la estructura XML en *Lexonomy*, un DWS que permite la confección y publicación de diccionarios electrónicos reduciendo a mínimos las habilidades técnicas requeridas en sus usuarios, con el objetivo de facilitar en un futuro la ampliación de los datos por parte del equipo de lexicógrafos y la apertura del sistema en forma de diccionario electrónico para su libre consulta.

Bibliografía

Abel, A. (2012): Dictionary Writing Systems and beyond. In: Granger, S./Paquot, M. (eds.): Electronic lexicography. Oxford, p. 83–106.

Dik, S. C. (1997): The theory of functional grammar. Part 1: The structure of the clause. Berlin/New York.

Heuberger, R. (2018): Dictionaries to assist teaching and learning. In: Fuertes Olivera, P. A. (ed.): The Routledge handbook of lexicography. New York, p. 300–316.

Lyons, J. (1980): Semántica. Barcelona.

Měchura, M. (2017): Introducing lexonomy: an open-source dictionary writing and publishing system. https://www.lexonomy.eu/docs/elex2017.pdf (last access: 22-03-2022).

Rubio López, R. J./Bonilla Huérfano, J. E./Bernal Chávez, J. A. (2021): Dictionary Writing Systems y otras herramientas informáticas para la elaboración, administración y publicación de diccionarios. In: Lingüística y Literatura 80, p. 340–360.

Simões, A./Salgado, A./Costa, R./Almeida, J. J. (2019): LeXmart: a smart tool for lexicographers. In: Kosem, I. et al. (eds.): Electronic Lexicography in the 21st Century: Proceedings of the eLex 2019 Conference: Smart Lexicography. Sintra, p. 453–466.

Tarp, S. (2013): Necesidad de una teoría independiente de la lexicografía: el complejo camino de la lingüística teórica a la lexicografía práctica. In: Círculo de Lingüística Aplicada a la Comunicación 56, p. 110–154.

Tésnière, L. (1959): Éléments de syntaxe structurale. Paris.

Woldrich, A./Goli, T./Kosem, I./Matuška, O./Wissik, T. (2020): ELEXIS: Technical and social infrastructure for lexicography. In: K Lexical News 28, p. 45–52.

Información del contacto

Manuel Márquez
Universidad Complutense de Madrid
manmarqu@ucm.es

Agradecimientos

Agradecemos al grupo de investigación ILSA de la Universidad Complutense de Madrid (http://ilsa.fdi.ucm.es/) su inestimable labor en el campo de la investigación sobre gramáticas formales como herramienta de desarrollo de software, así como sobre los repositorios y colecciones de objetos digitales.

Michal Měchura

DOCUMENT OR DATABASE? THE SEARCH FOR THE PERFECT STORAGE PARADIGM FOR LEXICAL DATA

Keywords E-lexicography; data structures in lexicography; lexical databases

When building a dictionary writing system or indeed any software system for lexical data, one decision the software engineers need to make early in the project is, which storage format are we going to persist our data in? Are we going to store each entry as a separate XML document (or, more modernly, as a JSON object)? Or would it be better to store everything in a relational database with tables and relations between them? Or perhaps a graph database? An RDF triple store? Your answer to these questions will affect how easy or difficult certain things will be in the rest of the software, for example how easy or difficult it will be to search the dictionary, to perform bulk edits on multiple entries at the same time, to make changes to the entry schema, or to enforce referential integrity on cross-references. This paper will review the storage design patterns often found in lexicography and discuss their advantages and disadvantages. Broadly speaking, there are two major patterns: the document pattern and the database pattern.

- In the document pattern, each entry is stored as a document in XML, JSON or some other markup/serialization language. The entries may and typically do have a highly formal internal structure which is made explicit in an entry schema such as a DTD, but this structure is unknown and invisible to the storage environment, apart from some indexes. Examples of the document pattern are file-based storage and XML databases, as well as relational databases where most tables are used only for storing indexes about the entries.
- In the database pattern, each entry is analyzed into its components (such as senses, example sentences, translations) and these are stored in the database as individual entities with their own identities. Examples of the database pattern are relational databases, graph databases and RDF triple stores.

When a lexicographer is editing, he or she is not editing the entry as a whole, he or she is editing the individual entities. And each time an entry is to be shown to a human user, it is composed dynamically from the entities as an impermanent "view".

The database pattern is strong where the document pattern is weak, and vice versa. For example, enforcing referential integrity on cross-references is easy in the database pattern (most database software has built-in support for that) and difficult in the document pattern (needs to be programmed manually). On the other hand, changing the entry schema and customizing the software for a different dictionary is easy in the document paradigm (we just need to provide an entry schema plus some rules for updating the indexes) and difficult in the database paradigm (the entry schema is hard-coded in the database structure).

The document pattern is very common in lexicography; practically all well-known dictionary writing systems are based on it, including Lexonomy, the IDM Dictionary Production System, and tLex. The database pattern is rare in lexicography and can mostly be found in

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

single-purpose systems developed for a specific dictionary project where the schema is not likely to change. Hybrid approaches are occasionally seen too, where some (but not all) entry components are analyzed into entities and stored à la the database pattern while the rest of the entry is left unanalyzed and stored à la the document pattern (this is how subentries are handled in Lexonomy, for example).

The paper will conclude by speculating that, while the document pattern is dominant currently, it is likely that lexicography will start embracing the database pattern more enthusiastically in the future, given current trends in the standardization of entry schemas and given a general tendency away from understanding dictionaries as static documents towards understanding them as dynamic repositories of structured content.

Contact information

Michal Měchura
Natural Language Processing Centre, Masaryk University
valselob@gmail.com

Christian-Emil Smith Ore/Oddrun Grønvik/Trond Minde

WORD BANKS, DICTIONARIES AND RESEARCH RESULTS BY THE ROADSIDE

Abstract Many European languages have undergone considerable changes in orthography over the last 150 years. This hampers the application of modern computer-based analysers to older text, and hence computer-based annotation and studies of text collections spanning a long period. As a step towards a functional analyser for Norwegian texts (Nynorsk standard) from the 19th century, funding was granted in 2020 for creating a full form generator for all inflected forms of headwords found in Ivar Aasen's dictionary published in 1873 (Aasen 1873) and his grammar from 1864 (Aasen 1864).

Creating this word bank led to new insight in Aasen (1873), its structure, internal organisation, and ambition level as well as its link to Aasen (1864). As a test, the full form list generated from this new word bank was used to analyse the word inventory of texts by Aa. O. Vinje, written in the period 1850–1870. The Vinje texts were also analysed using a full form list of modern standard Norwegian, to study the differences in applicability and see how Vinje's language relates to the written standard of modern Norwegian.

Keywords Dictionary and text analysis; full form systems; close reading of dictionaries

References

Aasen, I. (1864): Norsk Grammatik. Kristiania.

Aasen, I. (1873): Norsk Ordbog. Med dansk Forklaring. Kristiania.

Contact information

Christian-Emil Smith Ore
Department of Linguistics and Scandinavia Studies, University of Oslo
c.e.s.ore@iln.uio.no

Oddrun Grønvik
Department of Linguistic, Literary and Aesthetic Studies, University of Bergen
Oddrun.Gronvik@uib.no

Trond Minde
Department of Linguistic, Literary and Aesthetic Studies, University of Bergen
Trond.Minde@uib.no

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Ana Ostroški Anić/Ivana Brač

AirFrame

Mapping the field of aviation through semantic frames

Abstract The paper presents the process of developing the AirFrame database, a specialized lexical resource in which aviation terminology is defined in the form of semantic frames, following the methodology of the Berkeley FrameNet (FN). First, the structure of the database is presented, and then the methodology applied in developing and populating the database is described. The link between specialized aviation frames and general language semantic frames, of which frames defining entities, processes, attributes and events are particularly relevant, is discussed on the example of the semantic frame of Flight and its related frames. The paper ends with discussing possibilities of using AirFrame as a model for further developing resources in which general and specialized knowledge are linked.

Keywords Terminology; aviation terminology; semantic frames; specialized knowledge; specialized lexicography

Contact information

Ana Ostroški Anić
Institute of Croatian Language and Linguistics
aostrosk@ihjj.hr

Ivana Brač
Institute of Croatian Language and Linguistics
ibrac@ihjj.hr

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Kristel Proost/Arne Zeschel/Frank Michaelis/
Jan Oliver Rüdiger

MAP (Musterbank Argumentmarkierender Präpositionen)

A patternbank of argument-marking prepositions in German

Abstract Recent years have seen a growing interest in linguistic phenomena that challenge the received division of labour between lexicon and grammar, and hence often fall through the cracks of traditional dictionaries and grammars. Such phenomena call for novel, pattern-based types of linguistic reference works (see various papers in Herbst 2019). The present paper introduces one such resource: MAP ("Musterbank argumentmarkierender Präpositionen"), a web-based corpus-linguistic patternbank of prepositional argument structure constructions in German. The paper gives an overview of the design and functionality of the MAP-prototype currently developed at the Leibniz-Institute for the German Language in Mannheim. We give a brief account of the data and our analytic workflow, illustrate the descriptions that make up the resource and sketch available options for querying it for specific lexical, semantic and structural properties of the data.

Keywords Argument structure; valency; prepositions; constructicography; construction grammar

Reference

Herbst, T. (ed.) (2019): From lexicography to constructicography/Von der Lexikographie zur Konstruktikographie/De la lexicographie à la constructographie. Thematic Part of Lexicographica: International Annual for Lexicography/Revue Internationale de Lexicographie/Internationales Jahrbuch für Lexikographie. Vol. 35.

Contact information

Kristel Proost
Leibniz-Institut für Deutsche Sprache (IDS)
proost@ids-mannheim.de

Arne Zeschel
Leibniz-Institut für Deutsche Sprache (IDS)
zeschel@ids-mannheim.de

Frank Michaelis
Leibniz-Institut für Deutsche Sprache (IDS)
michaelis@ids-mannheim.de

Jan Oliver Rüdiger
Leibniz-Institut für Deutsche Sprache (IDS)
ruediger@ids-mannheim.de

Vasyl Starko

USL: A COGNITIVELY INSPIRED LEXICON FOR SEMANTIC TAGGING

Keywords Semantic tagging; semantic lexicon; Ukrainian; corpus annotation; cognitive semantics

Semantics is a language level that is extremely important and yet notoriously difficult for natural language processing. Approaches to semantic annotation have been dominated by lexicon-based solutions such as FrameNet and WordNet (Piao et al. 2015). A significant number of corpora are semantically annotated using a version of the WordNet for the language in question. A potential weakness of the WordNet is the low level of granularity, which may complicate downstream tasks. At least for some tasks, a more coarse-grained classification scheme based on the so-called supersenses has been utilized to reduce the average number of senses per word (Ciaramita/Altun 2006).

A different type of a semantic lexical resource for semantic tagging is exemplified by the USAS semantic lexicon (Rayson et al. 2004), which assigns semantic tags based on a universal annotation scheme. However, its dichotomous classification scheme can at times be too rigid, while the top layers of its hierarchy may be excessively abstract for a non-specialist user.

Keeping this in mind, a relatively coarse-grained semantic lexicon has been created based on a classification scheme that reflects the types of categories used by ordinary speakers in speech acts. This lexicographic resource, titled the Ukrainian Semantic Lexicon, has been developed for the Ukrainian language with a possibility of extension to other languages. The USL is a machine-readable dictionary in which each lemma (more precisely, each sense of a given lemma) is supplied with a string of semantic tags. For example,

> *naukovets* ‘scholar’ 1:conc:hum:prof,

where **1** is the number of the sense, the **conc** tag means ‘concrete noun’, **hum** ‘human being’, and **prof** ‘profession’. Another example illustrates semantic tags for an adjective:

> *velykyi* ‘large, great’ 1:size:2:degree:3:age,

where the second size refers to such contexts as *velyka radist* ‘great joy’.

Three more classes of words—verbs, adverbs, and numerals—are represented in the USL:

> *stoiaty* ‘to stand’ 1:loc:body:noncaus:2:loc:noncaus,

where **loc:body** refers to a body position, **loc** expressed the idea of location in general, and **noncaus** points to the noncausative nature of the verb’s senses.

> *povnistiu* ‘completely’ 1:physqual:2:degree:max,

where the Ukrainian adverbs mirrors two senses of its English equivalent.

Numerals in the USL include both quantifiers and absolute quantities:

> *bahato* ‘many, a lot’ 1:quantif

> *simdesiat* ‘seventy’ 1:abst:quantity:absol

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

A cognitive linguistic approach has been adopted to develop a system of semantic classification for the USL. In the center of attention are the so-called basic-level categories, which enjoy a privileged status in natural language categorization and are characterized by a convergence of perceptual, behavioral, and abstract features (Taylor 1995). From this level, relations may extend up and down but only to a limited degree, allowing for shallow hierarchies. In general, the semantic classification in USL is based on a faceted, rather than hierarchical, approach and multiclass membership is possible.

The Ukrainian Semantic Lexicon is now in its second iteration (USL 2.0) and contains 80,000 entries. It is suitable for NLP applications and, indeed, has been used in conjunction with the TagText tagger that performs both morphological and semantic annotation of Ukrainian texts. Both resources are available to lexicographers, computational linguists, and NLP researchers from their respective github repositories (Rysin 2022), (Ukrainian Semantic Lexicon). Both tools have been utilized to semantically tag a large reference corpus of Ukrainian – the General Regionally Annotated Corpus of Ukrainian (GRAC). The full classification scheme is presented on GRAC’s website (Shvedova et al. 2017–2022), while a discussion of the theoretical foundations is provided in Starko (2020) and (2021).

In the future, the USL will be further expanded, fine-tuned, and applied in projects involving semantic tagging.

References

Ciaramita, M./Altun, Y. (2006): Broad-coverage sense disambiguation and information extraction with a supersense sequence tagger. In: Jurafsky, D./Gaussiere, E. (eds.): Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP). Sydney, pp. 594–602.

Piao, S./Bianchi, F./Dayrell, C./D’Egidio, A./Rayson, P. (2015): Development of the multilingual semantic annotation system. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT 2015). Denver, pp. 1268–1274.

Rayson, P./Archer, D./Piao, S./McEnery, T. (2004): The UCREL semantic analysis system. In: Proceedings of LREC-04 Workshop: Beyond Named Entity Recognition Semantic Labeling for NLP Tasks. Lisbon, pp. 7–12.

Rysin, A. (2022): LanguageTool API NLP UK Project. https://github.com/brown-uk/nlp_uk (last access: 25-03-2022).

Shvedova, M./von Waldenfels, R./Yarygin, S./Rysin, A./Starko, V./Nikolajenko, T. (2017–2022): GRAC: General Regionally Annotated Corpus of Ukrainian. Kyiv/Lviv/Jena. http://uacorpus.org/ (last access: 25-03-2022).

Starko, V. (2021): Implementing semantic annotation in a Ukrainian corpus. In: Proceedings of the 5th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2021), April 22–23, Kharkiv, Ukraine. Volume I: Main conference. Kharkiv, pp. 435–447.

Starko, V. (2020): Semantic annotation for Ukrainian: categorization scheme, principles, and tools. In: Procededings of the 4th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2020), April 23–24, Lviv, Ukraine. Volume I: Main conference. Lviv, pp. 239–248.

Taylor, J. R. (1995): Linguistic categorizaton. 2nd edition. Oxford.

Ukrainian Semantic Lexicon. https://github.com/brown-uk/dict_uk/tree/master/data/sem (last access: 25-03-2022).

Contact information

Vasyl Starko
Ukrainian Catholic University
v.starko@ucu.edu.ua

Acknowledgements

This project has been supported by a grant from the Believe in Yourself Foundation at the Ukrainian Catholic University in Lviv.

Philipp Stöckle/Sabine Wahl

LEXICOGRAPHY AND CORPUS LINGUISTICS

The case of the Dictionary of Bavarian Dialects in Austria (WBÖ) and its database

Keywords Lexicography; corpus linguistics; Bavarian dialects; diminutives; digital humanities

The "Dictionary of Bavarian Dialects in Austria" (WBÖ) is a long-term project at the Austrian Academy of Sciences (ÖAW). Its main goal is the lexicographical documentation of the richly structured (historical) Bavarian dialects in Austria and South Tyrol.

The data for the WBÖ was collected in the first half of the 20th century, mainly with the help of dialect speakers who volunteered to support the dictionary by filling in numerous questionnaires. In addition, dialect literature as well as dialectological studies were extracted. This led to a data collection of about 3.6 million hand-written paper slips, the so-called "Hauptkatalog" (main catalogue), with often very detailed information on a huge number of dialectal words, including their meanings, their pronunciation as well as example sentences with which the collocations and the usage of the words can be studied. From the 1960s until 2015, the letters *A*, *B*/*P*, *C*, *D*/*T* and *E* were published in five volumes. From the letter *F*/*V* onwards, the Dictionary of Bavarian Dialects in Austria is published online via the platform "Lexikalisches Informationssystem Österreich (LIÖ; Lexical Information System Austria)".

Parallel to writing dictionary articles and with the intention of facilitating and speeding up the lexicographical work, the hand-written paper slips were manually transferred into the electronic database system TUSTEP (starting in the 1990s with the letter *D*). From 2014 to 2019, this data format was changed into a standard XML/TEI format (cf. Bowers/Stöckle 2018). This new database contains about 2.4 million entries and is also available online and open access via LIÖ. By exploiting the full potential of digital publications, LIÖ offers direct linking between the dictionary articles and the data as well as a mapping tool. Moreover, the database (although not being annotated systematically) can be used for linguistic research beyond lexicography (e. g., Stöckle/Hemetsberger/Stütz 2021).

In our contribution, the functionalities of LIÖ are demonstrated as well as its potential to answer research questions using the example of nominal diminutive suffixes, which appear in many different variants in Bavarian dialects (e. g., diminutives of *Fleisch* 'meat': *Fleischl*, *Fleischerl*, *Fleischele* or *Fleischi*). While variation in nominal diminutives has been studied for contemporary (Standard) German in Austria (cf. Korecky-Kröll in print; Schwaiger et al. 2019), the WBÖ material represents a huge corpus of historical base dialects.

In the lexicographic process, diminutives can be dealt with in various ways. In the case of the WBÖ, the diminutives are included in the dictionary article of the base noun. All phonetic variants of the diminutives in the WBÖ database are subsumed under the suffix types EL (for diminutives ending in -*el* or syncopated -*l*), ERL, ELEIN (for -*ele*/-*le*), and I.

From a variationist point of view, corpus-based analyses were conducted using the WBÖ database to detect geographic patterns of diminutives in the Bavarian dialect area in Austria and South Tyrol.

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

For this analysis, 20 lemmas from recently published articles on nouns starting with *F* were selected. In the underlying database entries (n = 9,055), 1,717 diminutive forms were detected which could be clearly assigned to one of the diminutive types mentioned above.[1]

The following map shows the distribution of the nominal diminutive variants:

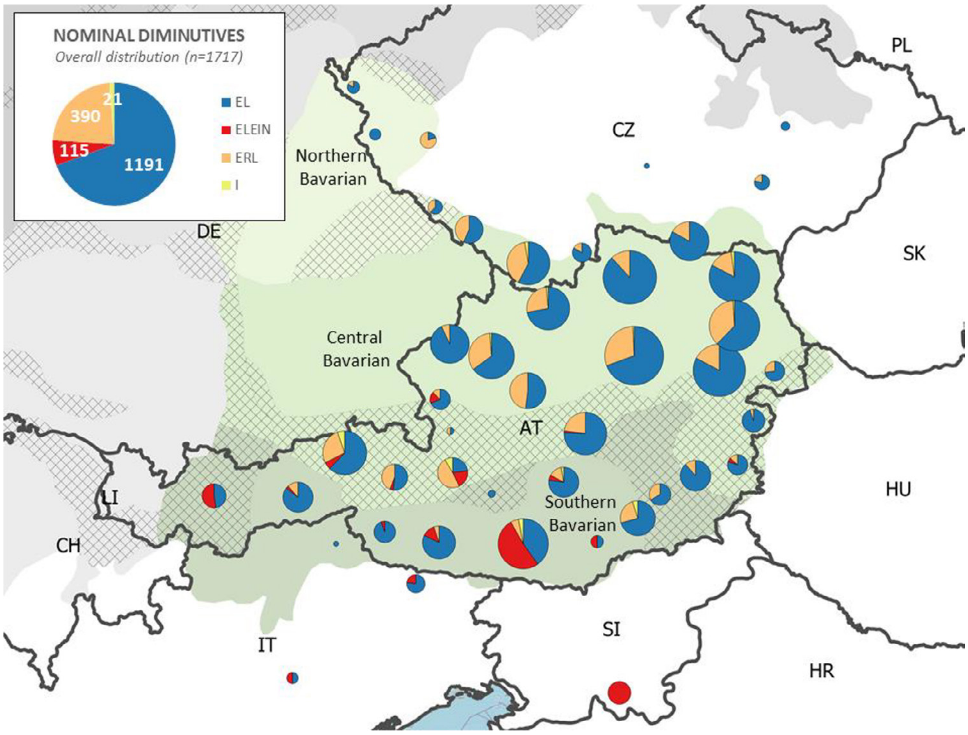


Fig. 1: Geographic distribution of diminutives

As shown in Figure 1, type EL (*Fleischl*) is the most common variant, and it is found in all but a few regions. Almost a quarter of the diminutives belong to type ERL (*Fleischerl*), but they appear to form a geographic pattern with type ELEIN (*Fleischele*), which almost only appears in Southern Bavarian. The diminutives of type I (*Fleischi*) are generally rare, and their geographic distribution does not seem to follow a clear pattern.

In future research, our results on the morphology of diminutives in the WBÖ and – in a further step – on semantic and pragmatic aspects will be compared to findings on contemporary nominal diminutives in a variety of electronic corpora (of Standard German) in Austria (e. g., Schwaiger et al. 2019).

References

Bowers, J./Stöckle, P. (2018): TEI and Bavarian dialect resources in Austria: updates from the DBÖ and WBÖ. In: Frank, A. U. et al. (eds.): Proceedings of the Second Workshop on Corpus-Based Research in the Humanities (CHR-2), 25–26 January 2018 Vienna, Austria. Vienna, pp. 45–54.

Korecky-Kröll, K. (in print): "Ma tuat net so vüü vernledlichen!" – oder doch? Verweigerung und Hinzufügung von Diminutiven als Schnittstellenprobleme von mündlichen "Wenker"-Übersetzungsaufgaben. In: Zeitschrift für Dialektologie und Linguistik.

[1] We would like to thank Theresa Eiweck for her help with data preparation.

Lexikalisches Informationssystem Österreich (LIÖ). https://lioe.dioe.at (last access: 25-03-2022).

Schwaiger, S./Barbaresi, A./Korecky-Kröll, K./Ransmayr, J./Dressler, W. U. (2019): Diminutivvariation in österreichischen elektronischen Korpora. In: Bülow, L./Fischer, A. K./Herbert, K. (eds.): Dimensions of linguistic space: Variation – multilingualism – conceptualisations/Dimensionen des sprachlichen Raums: Variation – Mehrsprachigkeit – Konzeptualisierung. Berlin, pp. 147–162.

Stöckle, P./Hemetsberger, C./Stütz, M. (2021): Die WBÖ-Belegdatenbank als Quelle für syntaktische Analysen – Möglichkeiten, Grenzen, Perspektiven. In: Wiener Linguistische Gazette 89, pp. 579–626. https://wlg.univie.ac.at/fileadmin/user_upload/p_wlg/892021/FSLenz_Stoeckle_etal.pdf (last access: 25-03-2022).

Contact information

Philipp Stöckle
Österreichische Akademie der Wissenschaften
philipp.stoeckle@oeaw.ac.at

Sabine Wahl
Österreichische Akademie der Wissenschaften
sabine.wahl@oeaw.ac.at

Anna Vacalopoulou/Eleni Efthimiou/
Stavroula-Evita Fotinea/Theodoros Goulas/
Athanasia-Lida Dimou/Kiki Vasilaki

ORGANIZING A BILINGUAL LEXICOGRAPHIC DATABASE WITH THE USE OF WORDNET

Abstract This paper reports on the restructuring of a bilingual (Greek Sign Language, GSL – Modern Greek) lexicographic database with the use of the WordNet semantic and lexical database. The relevant research was carried out by the Institute for Language and Speech Processing (ILSP) / Athena R.C. team within the framework of the European project Easier. The project will produce a framework for intelligent machine translation to bring down language barriers among several spoken/written and sign languages. This paper describes the experience of the ILSP team to contribute to a multilingual repository of signs and their corresponding translations and to organize and enhance a bilingual dictionary (GSL – Modern Greek) as a result of this mapping; this will be the main focus of this paper. The methodology followed relies on the use of WordNet and, more specifically, the Open Multilingual WordNet (OMW) tool to map content in GSL to WordNet synsets.

Keywords WordNet; semantic network; lexicographic database; multimodal database; sign language resources; Greek Sign Language

Contact information

Anna Vacalopoulou
Institute for Language and Speech Processing, ATHENA RC
avacalop@athenarc.gr

Eleni Efthimiou
Institute for Language and Speech Processing, ATHENA RC
eleni_e@athenarc.gr

Stavroula-Evita Fotinea
Institute for Language and Speech Processing, ATHENA RC
evita@athenarc.gr

Theodoros Goulas
Institute for Language and Speech Processing, ATHENA RC
tgoulas@athenarc.gr

Athanasia-Lida Dimou
Institute for Language and Speech Processing, ATHENA RC
ndimou@athenarc.gr

Kiki Vasilaki
Institute for Language and Speech Processing, ATHENA RC
kvasilaki@athenarc.gr

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Dictionary Writing Systems and Lexicographic Tools

XX EURALEX INTERNATIONAL CONGRESS
DICTIONARIES AND SOCIETY
12-16 July 2022, Mannheim, Germany

Nico Dorn

AN AUTOMATED CLUSTER CONSTRUCTOR FOR A NARRATED DICTIONARY

The cross-reference clusters of *Wortgeschichte digital*

Abstract *Wortgeschichte digital* (Digital Word History) is an emerging historical dictionary of the German language that focuses on describing semantic shifts from about 1600 through today. This article provides deeper insight into the dictionary's "cross-reference clusters," one of its software tools that performs visualization of its reference network. Hence, the clusters are a part of the project's macrostructure. They serve as both a means for users to find entries of interest and a tool to elucidate relations among dictionary entries. Rather than delve into technical aspects, this article focuses on the applied logics of the software and discusses the approach in light of the dictionary's microstructure. The article concludes with some considerations about the clusters' advantages and limitations.

Keywords Historical lexicography; dictionary; word history; visualization; digital humanities; graph theory

Contact information

Nico Dorn
Akademie der Wissenschaften zu Göttingen
ndorn@gwdg.de

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Mireille Ducassé/Archil Elizbarashvili

FINDING LEMMAS IN AGGLUTINATIVE AND INFLECTIONAL LANGUAGE DICTIONARIES WITH LOGICAL INFORMATION SYSTEMS

The case of Georgian verbs

Abstract Looking up for an unknown word is the most frequent use of a dictionary. For languages both agglutinative and inflectional, such as Georgian, this can be quite challenging because an inflected form can be very far from the lemmas used by the target dictionary. In addition, there is no consensus among Georgian lexicographers on which lemmas represent a verb in dictionaries. It further complicates dictionaries access. Kartu-Verbs is a base of inflected forms of Georgian verbs accessible by a logical information system. It currently contains more than 5 million inflected forms related to more than 16 000 verbs for 11 tenses; each form can have 11 properties; there are more than 80 million links in the base. This demonstration shows how, from any inflected form, we can find the relevant lemma to access any dictionary. Kartu-Verbs can thus be used as a front-end to any Georgian dictionary.

Keywords E-dictionary; lemma; Georgian language; under-resourced language; inflected forms; logical information systems; semantic web

Contact information

Mireille Ducassé
IRISA-INSA Rennes
mireille.ducasse@irisa.fr

Archil Elizbarashvili
Ivane Javarishvili Tbilisi State university
archil.elizbarashvili@tsu.ge

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Velibor Ilić/Lenka Bajčetić/Snežana Petrović/Ana Španović

SCyDia – OCR FOR SERBIAN CYRILLIC WITH DIACRITICS

Abstract In the currently ongoing process of retro-digitization of Serbian dialectal dictionaries, the biggest obstacle is the lack of machine-readable versions of paper editions. Therefore, one essential step is needed before venturing into the dictionary-making process in the digital environment – OCRing the pages with the highest possible accuracy. Successful retro-digitization of Serbian dialectal dictionaries, currently in progress, has shown a dire need for one basic yet necessary step, lacking until now – OCRing the pages with the highest possible accuracy. OCR processing is not a new technology, as many open-source and commercial software solutions can reliably convert scanned images of paper documents into digital documents. Available software solutions are usually efficient enough to process scanned contracts, invoices, financial statements, newspapers, and books. In cases where it is necessary to process documents that contain accented text and precisely extract each character with diacritics, such software solutions are not efficient enough. This paper presents the OCR software called "SCyDia", developed to overcome this issue. We demonstrate the organizational structure of the OCR software "SCyDia" and the first results. The "SCyDia" is a web-based software solution that relies on the open-source software "Tesseract" in the background. "SCyDia" also contains a module for semi-automatic text correction. We have already processed over 15,000 pages, 13 dialectal dictionaries, and five dialectal monographs. At this point in our project, we have analyzed the accuracy of the "SCyDia" by processing 13 dialectal dictionaries. The results were analyzed manually by an expert who examined a number of randomly selected pages from each dictionary. The preliminary results show great promise, spanning from 97.19% to 99.87%.

Keywords OCR; Cyrillic; Serbian language; retro-digitization; convolutional neural networks

Contact information

Velibor Ilić
The Institute for Artificial Intelligence Research and Development of Serbia, 21000 Novi Sad, Serbia
velibor.ilic@ivi.ac.rs

Lenka Bajčetić
Institute for the Serbian Language of SASA
lenka.bajcetic@gmail.com

Snežana Petrović
Institute for the Serbian Language of SASA
snezzanaa@gmail.com

Ana Španović
Institute for the Serbian Language of SASA
tesicana@gmail.com

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Julian Jarosch

DIGITALE LEXIKOGRAFIE MIT HAUSMITTELN

Ein Fallbeispiel zum digitalen Arbeiten ohne Sachmittel

Keywords Tuareg; Citizen Science; Retrodigitalisierung; Digitale Geisteswissenschaften; Onomastik; Linked Open Data

Die lexikografischen Arbeiten des französischen Forschungsreisenden und Mönchs Charles de Foucauld (1858–1916) sind bis heute unübertroffen in der Forschungsgeschichte der Berberlinguistik (Ritter 2009, S. XVI) zur Varietät des Tamahaq (exonymisch: Ahaggar-Tuareg; afroasiatische Familie; gesprochen in der zentralen Sahara, heute im Süden von Algerien), da sie als Produkt von über zehnjähriger immersiver Feldforschung (ca. 1904–1916) entstanden sind. Im Kontrast zu ihrer wissenschaftlichen Bedeutung ist ihre Zugänglichkeit begrenzt: Sie sind überwiegend seit Jahrzehnten nicht mehr im Druck, und in einem Fall bisher nur als Faksimile der handschriftlichen druckfertigen Endfassung des Autors erschienen.

Die zentralen lexikografischen Werke sind ein Eigennamenwörterbuch (Foucauld 1940) und ein umfangreiches allgemeinsprachliches Wörterbuch (Foucauld 1951–1952). Das Eigennamenwörterbuch enthält u.a. Toponyme, Personennamen und Tiernamen. Das allgemeinsprachliche Wörterbuch ist eine detaillierte lexikografische Ressource mit tiefgehenden semantischen Informationen, die in Ansätzen sogar ein kulturelles Lexikon bilden, das eine fast noch vorkoloniale Epoche widerspiegelt. Dieses 2000-seitige zentrale Hauptwerk ist nur als Handschrift-Faksimile publiziert.

Seit 2009 verfolge ich die Erschließung der lexikografischen Werke Foucaulds. Es handelt sich nicht um ein institutionalisiertes oder gefördertes Projekt – der Einblick in die laufenden und abgeschlossenen Arbeiten fungiert deshalb als Beispiel für Forschung mit öffentlich und kostenfrei verfügbaren Werkzeugen und Infrastrukturen. Die Auswahl der Teilbereiche deckt dabei einschlägige Eckpunkte des digitalen Forschungsprozesses ab: Text-Retrodigitalisierung, digitale Editorik, semantische Anreicherung und wissenschaftliche Auswertung.

Für die Digitalisierung des handschriftlichen Textes verwende ich automatische Handschrifterkennung in der Software Transkribus (https://readcoop.eu/transkribus/). Dieses Werkzeug erlaubt es, neuronale Netze auf die Erkennung beliebiger Handschriften zu trainieren. Voraussetzung ist, dass transkribierte und korrigierte Textseiten als Trainingsmaterial vorliegen, was hier durch die manuelle Transkription von mehreren hundert Seiten umgesetzt wurde. Diese Methode kann im vorliegenden Anwendungsfall auch die Mischung von Sprachen, Schriftsystemen und Schreibrichtungen adäquat berücksichtigen.

Die Edition und langfristige Publikation der Wörterbuchtexte erfolgt in Wikisource, einem Schwesterprojekt der Wikipedia. Wikisource steht als kostenfreie öffentliche Infrastruktur zur Textdigitalisierung zur Verfügung und bietet als Arbeitsplattform wesentliche und anpassbare Werkzeuge wie Sonderzeichen-Tastaturen und Templates für wiederkehrende Text- oder Layoutbausteine. Darüber hinaus handelt es sich um eine offene, Crowdsourcing-basierte Plattform. Dies bedeutet, dass die digitalisierten Texte nicht in einer End-

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

fassung fossilisieren, sondern dass zu jedem Zeitpunkt die Genauigkeit der Transkription gegen die Originale geprüft und, wenn nötig, weiter verbessert werden kann – ohne Zugangsschwellen.

Ein Forschungszweig im hier vorgestellten Projektkontext ist die semantische Anreicherung des Toponyme-Gazetteers im Eigennamenwörterbuch durch Georeferenzierung – also die Zuordnung der im Namenbuch (mit groben geografischen Informationen) genannten Orte zu präzisen geoinformatischen Objekten. Zur eindeutigen und dauerhaften Identifikation der Orte wird die offene geografische Normdatenbank GeoNames (http://www.geonames.org/) verwendet, die Geografika unveränderliche Identifikatoren zuweist und langfristig verfügbar vorhält. Die punktförmig groben Verortungen von GeoNames ermöglichen im nächsten Schritt die Zuordnung der Wörterbucheinträge zu geografischen Objekten in der Geo-Datenbank und Online-Karte OpenStreetMap (https://www.openstreetmap.org/), was eine deutlich genauere und adäquatere Repräsentation von linienförmigen und flächigen Orten ermöglicht (beispielsweise Trockenflüsse und Gebirge). OpenStreetMap ist ebenfalls ein durch Crowdsourcing und Nutzerbeiträge getragenes Projekt und kann deshalb in der Untersuchungsregion um die benötigten Daten ergänzt werden. Die Zusammenführung der drei genannten Datensätze – der Gazetteer von Foucauld, die eindeutigen Identifizierungen in GeoNames, und die geografischen Daten in OpenStreetMap – können in eine kartografische Darstellung der Ortsnamenbuch-Informationen resultieren.

Auf der linguistischen Ebene können die Toponyme auch als beispielhafte Auswertung der retrodigitalisierten Informationen zur Beantwortung einer wissenschaftlichen Fragestellung dienen. Während für Eigennamen eine referentielle Einzieligkeit als gesicherte wissenschaftliche Erkenntnis etabliert ist (Nübling/Fahlbusch/Heuser 2012, S. 17–20), scheint das Ortsnamenbuch von Foucauld diesem Monoreferenzprinzip zu widersprechen, indem es knapp über ein Viertel der Ortsnamen mehr als einen Referenten zuordnet. Eine Frequenz- und areale Verteilungsanalyse der Toponyme, ermöglicht durch die Digitalisierung der Daten, erlaubt die Interpretation des Phänomens als Funktion der nomadischen Lebensweise in der ariden Landschaft: Ausgeprägte Ortskenntnisse sind dort überlebensnotwendig, aber ein unbegrenzt umfangreiches Onomastikon ist mit der menschlichen Gedächtniskapazität unvereinbar; die Ortsnamen für mehrere geografische Objekte können also als mnemonisches Mittel zur Gedächtnisökonomie eingeschätzt werden (Jarosch/Späth 2021).

Das Projekt berührt also an mehreren Punkten gesellschaftliche Aspekte: Als überwiegend crowdsourcing-offenes Projekt ist es partizipatorisch angelegt für citizen scientists; die Wörterbücher enthalten eine Vielfalt an wissenschaftlich relevanten, historisch-kulturellen Einblicken in eine forschungsperspektivisch marginalisierte Sprechergemeinschaft; und als erste maßgebliche digitale lexikografische Ressource zum Tamahaq hat es das Potenzial, sprachpolitisch rezipiert zu werden und eine Rolle im Sprach- und Kulturerhalt zu spielen.

Literatur

Foucauld, C. de (1940): Dictionnaire abrégé touareg-français de noms propres. Dialecte de l'Ăhaggar. Paris: Larose. https://fr.wikisource.org/wiki/Dictionnaire_abrégé_touareg-français_de_noms_propres (Stand: 04-04-2022).

Foucauld, C. de (1951–1952): Dictionnaire touareg – français. Dialecte de l'Ăhaggar. 4 Bände. Paris: Imprimerie Nationale de France. https://fr.wikisource.org/wiki/Dictionnaire_touareg_–_français (Stand: 04-04-2022).

Jarosch, J./Späth, L. (2021): Toponyme einer nomadischen Gesellschaft – Orientierung in einer ariden Landschaft. In: Dräger, K./Heuser, R./Prinz, M. (Hg.): Toponyme. Berlin/New York, S. 87–108. https://doi.org/10.1515/9783110721140-005.

Nübling, D./Fahlbusch, F./Heuser, R. (2012): Namen. Eine Einführung in die Onomastik. Tübingen.

Ritter, H. (2009): Wörterbuch zur Sprache und Kultur der Twareg. Band I. Wiesbaden.

Kontaktinformationen

Julian Jarosch
Akademie der Wissenschaften und der Literatur Mainz / Johannes Gutenberg-Universität Mainz
julian.jarosch@adwmainz.de

This page is part of the publication: Klosa-Kückelhaus, Annette / Engelberg, Stefan / Möhrs, Christine / Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Dorielle Lonke/Ilan Kernerman/Vova Dzhuranyuk

LEXICAL DATA API

Abstract This API provides data from various dictionary resources of K Dictionaries across 50 languages. It is used by language service providers, app developers, and researchers, and returns data as JSON documents. A basic search result consists of an object containing partial lexical information on entries that match the search criteria, but further in-depth information is also available. Basic search parameters include the source resource, source language, and text (lemma), and the entries are returned as objects within the *results* array. It is possible to look for words with specific syntactic criteria, specifying the part of speech, grammatical number, gender and subcategorization, monosemous or polysemous entries. When searching by parameters, each entry result contains a unique entry ID, and each sense has its own unique sense ID. Using these IDs, it is possible to obtain more data – such as syntactic and semantic information, multiword expressions, examples of usage, translations, etc. – of a single entry or sense. The software demonstration includes a brief overview of the API with practical examples of its operation.

Keywords API; lexical data; search; dictionary

Contact information

Dorielle Lonke
K Dictionaries
dorielle@kdictionaries.com

Ilan Kernerman
K Dictionaries
ilan@kdictionaries.com

Vova Dzhuranyuk
K Dictionaries
vova@kdictionaries.com

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/
Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of
Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Takahiro Makino/Rei Miyata/Seo Sungwon/Satoshi Sato

DESIGNING AND BUILDING A JAPANESE CONTROLLED LANGUAGE FOR THE AUTOMOTIVE DOMAIN

Toward the development of a writing assistant tool

Abstract In this paper, we propose a controlled language for authoring technical documents and report the status of its development, while maintaining a specific focus on the Japanese automotive domain. To reduce writing variations, our controlled language not only defines approved and unapproved lexical elements but also prescribes their preferred location in a sentence. It consists of components of a) case frames, b) case elements, c) adverbial modifiers, d) sentence-ending functions, and e) connectives, which have been developed based on the thorough analyses of a large-scale text corpus of automobile repair manuals. We also present our prototype of a writing assistant tool that implements word substitution and reordering functions, incorporating the constructed controlled language.

Keywords Japanese controlled language; corpus-based lexicon building; variation management; writing support tool; automotive domain

Contact information

Rei Miyata
Nagoya University
miyata.rei.f2@f.mail.nagoya-u.ac.jp

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Adriane Orenha Ottaiano/Maria Eugênia Olímpio de Oliveira Silva/Carlos Roberto Valêncio/João Pedro Quarado

DEVELOPING A COLLOCATION DICTIONARY WRITING SYSTEM (COLDWS) FOR AN ONLINE MULTILINGUAL COLLOCATIONS DICTIONARY PLATFORM (PLATCOL)

Keywords Dictionary writing system; collocations dictionary; multilingual dictionary; collocations

This ongoing research project, funded by the São Paulo Research Foundation (FAPESP – 2020/01783-2), has the purpose of developing a phraseographical methodology and model for an Online Corpus-Based Multilingual Collocations Dictionary Platform (PLATCOL), in five different languages so far: English and Portuguese, French, Spanish, and Chinese. It is aimed to be customized for different target audiences according to their needs, such as language learners, pre- and in-service teachers, translators, material developers and researchers or lexicographers. To achieve this goal, we follow the theoretical assumptions of the function theory of lexicography (Bothma/Tarp 2012; Fuertes-Olivera/Tarp 2014; Tarp 2015). Hence, both the procedures chosen for the selection, organization and presentation of lexicographic data, as well as the determination of the content, form and access routes are adapted and subordinated to the users' preferences.

The methodology being developed for the PLATCOL relies on the combination of automatic methods to extract candidate collocations (Garcia et al. 2019a). The automatic approaches take advantage of NLP tools to annotate large corpora with lemmas, PoS-tags and dependency relations in the five languages. Using these data, we apply statistical measures (Evert et al. 2017; Garcia/García-Salido/Alonso-Ramos 2019b) and distributional semantics strategies to select the collocation candidates (Garcia/García-Salido/Alonso-Ramos 2019c) and to retrieve corpus-based examples (Kilgarriff et al. 2008). Having the lexicographers selected the suitable collocations, we follow Garcia/García-Salido/Alonso-Ramos (2019c) to carry out an automatic translation of the collocations. All automatically extracted data are being carefully post-edited by the lexicographers involved in this investigation (Orenha-Ottaiano et al. 2021).

In order to better enable or enhance the post-editing of all the automatically retrieved data, we developed an in-house Collocations Dictionary Writing System (COLDWS), a software aimed at specifically compiling and producing collocation dictionaries, so that all automatically extracted data can be automatically inserted into this COLDWS, post-edited by the lexicographers, as well as be afterwards exported to an end-user platform.

In this paper, we will focus on the theoretical as well as methodological aspects regarding the development of the COLDWS, duly created to fulfill the specific needs of our collocations dictionaries. We have knowledge of the dictionary writing system Lexonomy, a web-based dictionary writing system (Měchura 2017), and TshwaneLex (de Schryver 2007), two good quality DWSs. However, as previously said, we needed to rely on a DWS that would meet the specificities of a multilingual collocations dictionary and also deal with the specific data output generated for our project.

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

The COLDWS exclusively focuses on the management of entries, collocations and all the other dictionary data related to them, automatically retrieved and automatically inserted into this software. With this software, it is possible to register and edit all dictionary information, such as languages, corpora data, morphosyntactic structures, taxonomy of the collocations, translations, statistical measures etc. There is also a functionality with which reviewers can post-edit entries, collocations and translations. With respect to the validation process, the software will allow lexicographers to choose from three phases (traffic lights phases), indicating to users the status of the entries or collocations. In the first phase, data automatically inserted into the COLDWS (not revised yet) will be displayed with a red icon, even to users, when exported to the end-user platform. Phase 2 represents data revised by one member of the team (reviewer 1), but which may still need a second evaluation and/or some adjustments – an orange icon will be shown. In phase 3, data is checked by a second reviewer (reviewer 2) and now considered to be suitable – a green green icon will be displayed.

Besides that, the COLDWS will also be of valuable help when it comes to optimizing translation of entries or collocations. For example, once translation pairs between collocations are identified and registered in the system, making up a multilingual database, it becomes possible to identify and automatically suggest new translations among other languages. This process occurs through an inference-based algorithm, built from an inference hypothesis related to the composition of multiple translation dictionaries: if word or collocation A translates into word or collocation B which in turn translates into word or collocation C, what is the probability that C is a translation of A? Studies developed under this hypothesis (e. g. Mausam et al. 2010) presented significant results in relation to the analysis via inference of translation pairs between different languages. In this process, the algorithm performs the analysis of previously registered translations, identifies other translation pairs via inference, and shows lexicographers the possibilities of translations, who must analyze the reliability and quality of the translation found. To our knowledge, there is not any DWS functionality compared to this one and that may be a successful innovation with respect to DWS functionalities, enhancing the development of collocations dictionaries.

In what concerns computational development, the COLDWS was built using languages from current and widespread programming such as Java (with Model View Controller [MVC] architecture), HTML and Javascript (jQuery library). For the storage, consultation and deletion of entries, the relational data model with PostgreeSQL software was used in conjunction with SQL language. In addition, concepts of User Experience – UX were applied to provide a good experience for the lexicographers.

The DWS is still under evaluation and if the results are positive, we will also allow free access to and use of the software upon request.

References

Bothma, T. J. D./Tarp, S. (2012): Lexicography and the relevance criterion. In: Lexikos 22, pp. 86–108.

de Schryver, G.-M./De Pauw, G. (2007): Dictionary writing system (DWS) + corpus query package (CQP): the case of Tshwane Lex. In: Lexikos 17, pp. 226–246.

Evert, S./Uhrig, P./Bartsch, S./Proisl, T. (2017): E-VIEW-affilation – a large-scale evaluation study of association measures for collocation identification. In: Kosem, I./Tiberius, C./Jakubíček, M./Kallas, J./Krek, S./Baisa, V. (eds.): Electronic Lexicography in the 21st Century: Lexicography from Scratch. Proceedings of eLex 2017. Leiden, the Netherlands, 19–21 September 2017. Brno, pp. 531–549. https://elex.link/elex2017/proceedings/eLex_2017_Proceedings.pdf (last access: 12-03-2020).

Fuertes Olivera, P. A./Tarp, S. (2014): Theory and practice of specialised dictionaries. Lexicography versus terminography, Berlin/Boston.

Garcia, M./García-Salido, M./Alonso-Ramos, M. (2019a): Towards the automatic construction of a multilingual dictionary of collocations using distributional semantics. In?: Kosem, I./Kuhn,T. Z./Correia, M./Ferreira, J. P./Jansen, M./Pereira, I./Kallas, J./Jakubíček, M./Krek, S./Tiberius, C. (eds.): Electronic Lexicography in the 21st Century. Proceedings of eLex 2019. Sintra, Portugal, 1–3 October 2019. Brno, pp. 747–762. https://elex.link/elex2019/wp-content/uploads/2019/09/eLex_2019_42.pdf (last access: 10-03-2020).

Garcia, M./García-Salido, M./Alonso-Ramos, M. (2019b): A comparison of statistical association measures for identifying dependency-based collocations in various languages. In: Savary, A./Parra Escartín, C./Bond, F./Mitrović, J./Mititelu, V. B. (eds.): Proceedings of the Joint Workshop on Multiword Expressions and WordNet (MWE-WN 2019). Florence, Italy, August 2, 2019, Association for Computational Linguistics. Stroudsburg, pp. 49–59. https://www.aclweb.org/anthology/W19-5107.pdf (last access: 09-03-2020).

Garcia, M./García-Salido, M./Alonso-Ramos, M. (2019c): Weighted compositional vectors for translating collocations using monolingual corpora. In: Corpas Pastor, G./Mitkov, R. (eds.) Computational and corpus-based phraseology. Cham, pp. 113–128.

Kilgarriff, A./Husák, M./McAdam, K./Rundell, M./Rychly, P. (2008): GDEX: automatically finding good dictionary examples in a corpus. In: Bernal, E./DeCesaris, J. (eds.): Proceedings of the 13th EURALEX International Congress. Barcelona, 15–19 July 2008. Barcelona, pp. 425–432.

Mausam, S./Etzioni, S./Weld, O./Reiter, D. S./Skinner, K./Sammer, M./Vessier, S. (2010): Panlingual lexical translation via probabilistic inference. In: Artificial Intelligence 174 (9–10), pp. 619–637.

Měchura, M. B. (2017): Introducing lexonomy: an open-source dictionary writing and publishing system. In: Electronic Lexicography in the 21st Century: Lexicography from Scratch. Proceedings of eLex 2017. Leiden, the Netherlands, 19–21 September 2017. Brno, pp. 662–679.

Orenha-Ottaiano, A./Garcia-Gonzalez, M./Olímpio, M. E./L'Homme, M./Alonso Ramos, M./Valencio, C. R./Tenorio, W. (2021): Corpus-based methodology for an online multilingual collocations dictionary: first Steps. In: Kosem, I./Cukr, M./Jakubíček, M./Kallas, J./Krek, S./Tiberius, C. (eds.): Electronic Lexicography in the 21st Century. Proceedings of the eLex 2021 conference. 5–7 July 2021, virtual. Brno, pp. 1–28. https://elex.link/elex2021/wp-content/uploads/2021/08/eLex_2021_01_pp1-28.pdf (last access: 04-08-2021).

Tarp, S. (2015): La teoría funcional en pocas palabras. In: Estudios de Lexicografía 4, pp. 31–42.

Contact information

Adriane Orenha-Ottaiano
São Paulo State University (UNESP)
adriane.ottaiano@unesp.br

Maria Eugênia Olímpio de Oliveira Silva
University of Alcalá
eugenia.olimpio@uah.es

Carlos Roberto Valêncio
São Paulo State University (UNESP)
carlos.valencio@unesp.br

João Pedro Quadrado
São Paulo State University (UNESP)
jp.quadrado@unesp.br

Acknowledgments

We gratefully acknowledge the financial support provided by The São Paulo Research Foundation (FAPESP), Process ner 2020/01783-2.

Annabella Schmitz

RDF-LIFTING VON OntoLex-Lemon AUS DEM DIGITALEN FAMILIENNAMENWÖRTERBUCH DEUTSCHLANDS MIT XTriples

Keywords Linked Open Data; RDF; Digital Lexicography; XML; Digital Humanities

Das digitale Familiennamenwörterbuch Deutschlands (DFD) ist ein Forschungsprojekt zwischen der Akademie der Wissenschaften und der Literatur Mainz, der Technischen Universität Darmstadt und der Johannes Gutenberg-Universität Mainz. Anhand der Telefonanschlüsse der Telekom aus dem Jahr 2005 sollen in Deutschland vorkommende Familiennamen digital lexikographisch erfasst werden. Berücksichtigt werden hier Bedeutungen der Familiennamen, Zuordnung zu festgelegten Kategorien, Häufigkeit der Vorkommen eines Namens in Deutschland und, wenn vorhanden, im Ausland, Sprachvorkommen sowie die Kartierung der geographischen Verbreitung der Familiennamen in Deutschland.

Um die Bestandteile des Familiennamenwörterbuchs als Linked Open Data bereitzustellen und eine Vernetzung im Semantic Web zu etablieren, bietet sich eine Modellierung mit der OntoLex-Lemon Ontologie (Cimiano/McCrae/Buitelaar (Hg.) 2016) an. OntoLex-Lemon ist ein Vokabular zur Darstellung von lexikographischen Ressourcen als RDF (Resource Descriptional Framework). Es besteht aus mehreren Modulen, die sich je nach Intention des Lexikons wählen lassen. Für das Digitale Familiennamenwörterbuch eigneten sich das lexikographische Modul (Bosque-Gil/Gracia (Hg.) 2019) und das Kernmodul am besten. Das lexikographische Modul, da hiermit die Struktur der ursprünglichen Ressource beibehalten werden kann, und das Kernmodul zur Darstellung von weiteren linguistischen Eigenschaften und semantischen Verlinkungen mit anderen Webseiten, die nicht der genauen Struktur der Webseite des Digitalen Familiennamenwörterbuchs entsprechen. Zusätzlich wurden für das RDF-Modell mit OntoLex noch weitere Vokabulare herangezogen, um die Darstellung des Wörterbuchs vollständig zu machen. Letztendlich können mit OntoLex-Lemon semantische Aussagen über das Familiennamenwörterbuch und dessen lexikalische Einträge getroffen werden. So ist eine Modellierung mit OntoLex-Lemon sehr hilfreich, um die semantischen Aussagen des Digitalen Familiennamenwörterbuchs mit anderen lexikalischen Ressourcen vernetzbar und vergleichbar zu machen. Eine beispielhafte Modellierung eines Familiennamenartikels ist in der Abbildung 1 anhand des Familiennamens „Eis“ zu sehen.

Damit dann schließlich das OntoLex-Lemon-Modell für das Digitale Familiennamenwörterbuch in Turtle (Terse RDF Triple Language), einer Syntax für RDF, für alle publizierten Familiennamenartikel umgesetzt werden kann, kann der Webservice XTriples herangezogen werden.

Mit diesem Webservice lassen sich RDF-Tripel (semantische Aussagen, bestehend aus Subjekt, Prädikat und Objekt) aus XML-Daten extrahieren, die dann in verschiedenen Output-Formaten verfügbar gemacht werden. Neben Turtle, was für das DFD generiert werden soll, stehen noch RDF/XML, N-Triples, N-Quads, TriX, JSON und SVG Graph als Output zur Verfügung. Das RDF-Modell kann mit einer einfachen Konfiguration erstellt werden, indem die XML-Daten der Familiennamenartikel des Digitalen Familiennamenwörterbuchs

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

als Rest-Schnittstelle in der Konfiguration hinterlegt werden. Die Konfiguration für die XTriples besteht aus einer XML-ähnlichen Struktur und es wird für jede RDF-Aussage des gewünschten Ziel-Modells ein Subjekt, ein Prädikat und ein Objekt in der Konfiguration definiert. Mögliche semantische Aussagen eines Artikels des Digitalen Familiennamenwörterbuchs mit ihrer Umsetzung in Turtle wären:

1) Das DFD ist eine lexikographische Ressource.
   - :dfd a lexicog:LexicographicResource.

2) Das DFD hat den Eintrag „Eis".
   - :dfd lexicog:entry lexicog:LexicographicResource.

3) „Eis" ist ein lexikographischer Eintrag.
   - <http://www.namenforschung.net/id/name/9815/1> a lexicog:Entry.

So wird dann jeder einzelne hinterlegte Artikel auf dieser Schnittstelle gecrawlt und mittels XPath-Ausdrücken werden Informationen für die einzelnen RDF-Statements in der Konfiguration aus den XML-Daten gelesen und von einer XML-Struktur in einen RDF-Graphen bzw. ein anderes mögliches Ausgabeformat abgeleitet (RDF-Lifting).

Da alle Familiennamenartikel für das Digitale Familiennamenwörterbuch als TEI-XML-Artikel erstellt werden, sind alle dazu benötigten Daten schon in einer durchsuchbaren Struktur vorhanden und können anhand des Modells in ihrer Gesamtheit mithilfe einer beispielhaften Konfiguration extrahiert werden, sodass das beispielhafte Modell (Abb. 1) für jeden Familiennamenartikel in der gezeigten Struktur umgesetzt werden kann.

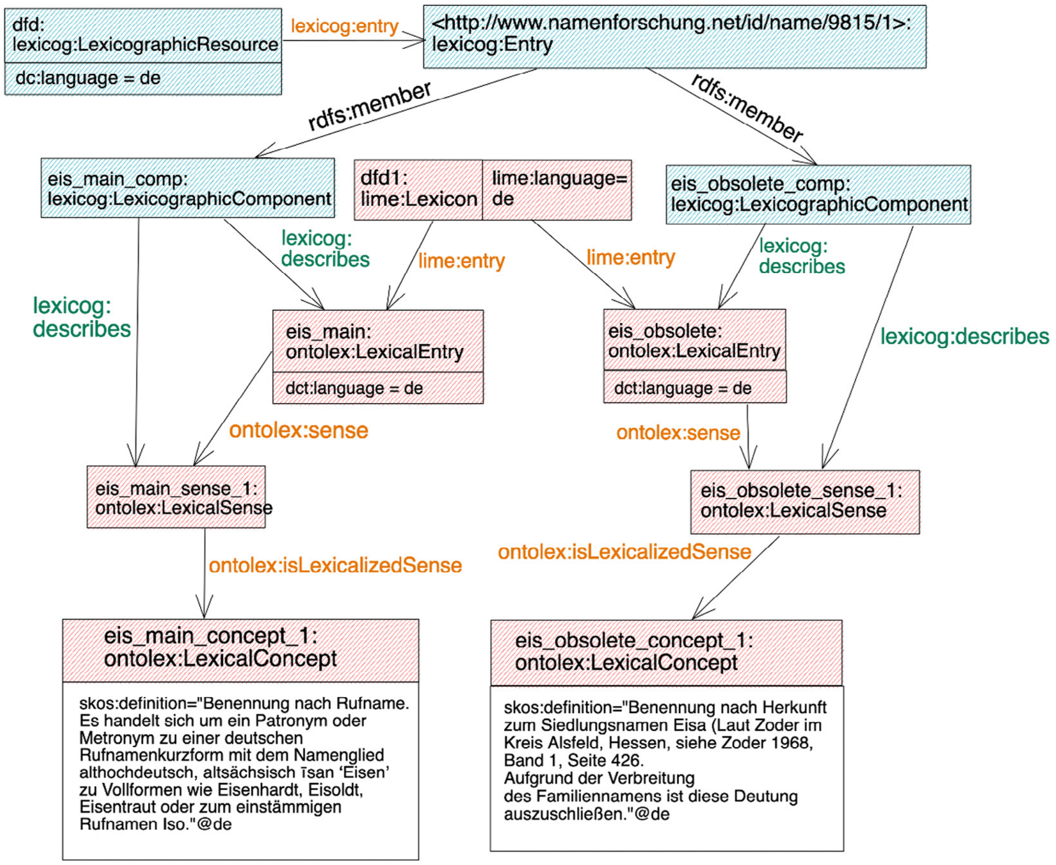

dfd:
lexicog:LexicographicResource
dc:language = de
lexicog:entry
<http://www.namenforschung.net/id/name/9815/1>:
lexicog:Entry
rdfs:member
rdfs:member
eis_main_comp:
lexicog:LexicographicComponent
dfd1:
lime:Lexicon
lime:language=
de
eis_obsolete_comp:
lexicog:LexicographicComponent
lexicog:
describes
lime:entry
lime:entry
lexicog:
describes
lexicog:
describes
eis_main:
ontolex:LexicalEntry
dct:language = de
eis_obsolete:
ontolex:LexicalEntry
dct:language = de
lexicog:describes
ontolex:sense
ontolex:sense
eis_main_sense_1:
ontolex:LexicalSense
eis_obsolete_sense_1:
ontolex:LexicalSense
ontolex:isLexicalizedSense
ontolex:isLexicalizedSense
eis_main_concept_1:
ontolex:LexicalConcept
skos:definition="Benennung nach Rufname.
Es handelt sich um ein Patronym oder
Metronym zu einer deutschen
Rufnamenkurzform mit dem Namenglied
althochdeutsch, altsächsisch Īsan 'Eisen'
zu Vollformen wie Eisenhardt, Eisoldt,
Eisentraut oder zum einstämmigen
Rufnamen Iso."@de
eis_obsolete_concept_1:
ontolex:LexicalConcept
skos:definition="Benennung nach Herkunft
zum Siedlungsnamen Eisa (Laut Zoder im
Kreis Alsfeld, Hessen, siehe Zoder 1968,
Band 1, Seite 426.
Aufgrund der Verbreitung
des Familiennamens ist diese Deutung
auszuschließen."@de

Abb. 1: OntoLex-Kernmodell des Familiennamens „Eis" (Griebel 2022)

Folglich kann mithilfe von XTriples und mit dem RDF-Vokabular OntoLex ein Wörterbuch wie das Digitale Familiennamenwörterbuch, welches die Artikel in XML-Daten vorliegen hat, schnell, einfach und effizient als RDF/Turtle und damit auch als Linked Open Data bereitgestellt werden.

Literatur

Bosque-Gil, J./Gracia, J. (Hg.) (2019): The OntoLex Lemon Lexicography Module: Final Community Group Report 17 September 2019. https://www.w3.org/2019/09/lexicog/ (Stand: 25.03.2022).

Cimiano, P./McCrae, J./Buitelaar, P. (Hg.) (2016): Lexicon model for ontologies: community report. https://www.w3.org/2016/05/ontolex/ (Stand: 24.03.2022).

Digitales Familiennamenwörterbuch Deutschlands:
https://www.namenforschung.net/dfd/woerterbuch/liste/ (Stand: 24.03.2022).

Griebel, J. (2022): Eis. Digitales Familiennamenwörterbuch Deutschlands.
http://www.namenforschung.net/id/name/9815/1 (Stand: 25.03.2022).

XTriples: https://xtriples.lod.academy/index.html (Stand: 25.03.2022).

Kontaktinformationen

Annabella Schmitz
Akademie der Wissenschaften und der Literatur Mainz
annabella.schmitz@adwmainz.de

Alberto Simões/Ana Salgado

SMART DICTIONARY EDITING WITH LeXmart

Abstract Given the relevance of interoperability, born-digital lexicographic resources as well as legacy retro-digitised dictionaries have been using structured formats to encode their data, following guidelines such as the Text Encoding Initiative or the newest TEI Lex-0. While this new standard is being defined in a stricter approach than the original TEI dictionary schema, its reuse of element names for several types of annotation as well as the highly detailed structure makes it difficult for lexicographers to efficiently edit resources and focus on the real content. In this paper, we present the approach designed within LeXmart to facilitate the editing of TEI Lex-0 encoded resources, guaranteeing consistency through all editing processes.

Keywords Dictionary encoding; Text Encoding Initiative; dictionary editing system

Contact information

Alberto Simões
2Ai – School of Technology, IPCA, Barcelos, Portugal
asimoes@ipca.pt

Ana Salgado
NOVA CLUNL, Centro de Linguística da Universidade NOVA de Lisboa, Portugal/Academia das Ciências de Lisboa, Portugal
ana.salgado@fcsh.unl.pt

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Chris A. Smith

ARE PHONESTHEMES EVIDENCE OF A SUBLEXICAL ORGANISING LAYER IN THE STRUCTURE OF THE LEXICON?

Testing the OED analysis of two phonesthemes with a corpus study of collocational behaviour of *sw-* and *fl-* words in the OEC

Abstract Phonesthemes (Firth 1930) are sublexical constructions that have an effect on the lexico-grammatical continuum: they are recurring form-meaning associations that occur more often than by chance but not systematically (Abramova et al. 2013). Phonesthemes have been shown (Bergen 2004) to affect psycholinguistic language processing; they organise the mental lexicon. Phonesthemes appear over time to emerge as driven by language use as indexical rather than purely iconic constructions in the lexicon (Smith 2016; Bergen 2004; Flaksman 2020). Phonesthemes are acknowledged in construction morphology (Audring/Booij/Jackendoff 2017) as motivational schemas. Some phonesthemes also tend to have lexicographic acknowledgment, as shown by etymologist Liberman (2010a, b), although this relevance and cohesion appears to be highly variable as we will show in this paper.

This paper seeks to compare two phonesthemes in a combined lexicographic and corpus study with a view to testing the results obtained. **Firstly**, following Smith (2016) which identified 11 semantic categories of *fl-* words in the OED, we analyse the OED entries for 245 *sw-* monomorphemes with a view to carrying out a key word analysis and a semantic trait analysis. The 245 monomorphemes have a total of 469 senses out of which 330 can be classified into 18 recurring semantic traits in Table 1.

semantic traits based on OED key words	**number of senses carrying the trait**
sway sweep swish	78
strike blow swipe	56
pressure swell swathe	57
sway swagger boast	11
compact cluster agitated	7
big fellow	4
flame burn waste	10
deceive sway swindle	11
faint swoon agitated	18
cool dark	7
drink	19
surface	9
hollow	10
exchange swap	6
labour toil sweat	12
deviate deflect	6

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

semantic traits based on OED key words	**number of senses carrying the trait**
sound	9
Total	330

Table 1: Lexicographic behaviour of sw- senses in the OED.

Then, in a **second step**, the comparison between the OED analysis of *fl-* and *sw-* monomorphemes shows that *sw-* words appear less likely to undergo any semantic change and therefore appear to be less indexical. In the light of these differing lexicographic behaviours, we aim, in a third step, to analyse the collocational behaviour of some common phonesthemic verbs carrying *fl-* and *sw-*. Collocational behaviour via a collexeme analysis will enable us to identify combinatorial patterns of use. For the study, we use the very large contemporary (2 billion words) OEC corpus (2000–2005) using Sketch Engine (Kilgarriff et al 2004). The results of the compared analysis allow us to discuss whether phonesthemes are actual (sub)lexical "chunks" deserving of a lexical status, or whether they belong to larger phraseological "chunks" or units. This question raises the issue of the architecture of the lexico-grammatical continuum, the "constructicon": does the constructicon accommodate or require a sublexical layer?

What are the repercussions for lexicography and phraseology?

Keywords Phonesthemes; analogy; collocational behaviour; OED; OEC; phraseological chunks

References

Abramova, E./Fernandez, A./Sangati, F. (2013): Automatic labeling of phonesthemic senses. In: UC Merced 35 (35). pp. 1696–1701.

Audring, J./Booij G./Jackendoff R. (2017): Menscheln, kibbelen, sparkle: Verbal diminutives between grammar and lexicon. In: Le Bruyn, B./Lestrade, S. (eds.): Linguistics in the Netherlands 2017. Amsterdam, pp. 1–15.

Bergen, B.-K. (2004): The psychological reality of phonaesthemes. In: Language 80 (2), pp. 290–311.

Firth, J. (1930): Speech. London.

Flaksman, M. (2020): Pathways of de-iconization: how borrowing, semantic evolution and sound change obscure iconicity. In: Perniss, P./Fischer, O./Ljungberg, C. (eds.): Operationalizing iconicity. Amsterdam, pp. 75–104.

Kilgarriff, A./Rychlý, P./Smrž, P./Tugwell, D. (2004): The Sketch Engine. In: Information Technology. Lorient.

Liberman, A. (2010a): Iconicity and etymology. In: Signergy, Iconicity in Language and Literature 9, pp. 243–258.

Liberman, A. (2010b): The state of English etymology (a few personal observations). In: Cloutier, R. A./Hamilton-Brehm, A-M./Kretzschmar, W. A, Jr. (eds.); Studies in the history of the English language V. Variation and change in English grammar and lexicon: contemporary approaches. Berlin/New York, pp. 161–182.

Smith, C. A. (2016): Tracking semantic change in fl- monomorphemes in the OED. In: Journal of Historical Linguistics 6 (2), pp. 165–200.

Contact information

Chris A. Smith
University of Caen Normandy, CRISCO EA4255
chris.smith@unicaen.fr

Tanara Zingano Kuhn/Špela Arhar Holdt/ Rina Zviel Girshin/Ana R. Luís/Carole Tiberius/Kristina Koppel/Branislava Šandrih Todorović/Iztok Kosem

INTRODUCING CrowLL – THE CROWDSOURCING FOR LANGUAGE LEARNING GAME

Keywords Crowdsourcing; dictionaries; game; pedagogical corpora

In this demo, we introduce CrowLL (Crowdsourcing for Language Learning), a game with a purpose for creating pedagogical corpora of Dutch, Estonian, Serbian, Slovene, and Portuguese. CrowLL is primarily meant for the publication of SkeLL (Sketch Engine for Language Learning) (Baisa/Suchomel 2014) for these languages, but also applicable for dictionary making and teaching materials development. With this game, we propose an alternative way of creating pedagogical corpora in which corpora are not cleaned of structure and content usually considered inappropriate for learners, but rather labelled. The design process of a pedagogical corpus is characterized by the 'pedagogic mediation of corpora' (Braun 2005), which can be the close monitoring of the content of the corpus to identify possible structural (grammar and spelling) problems as well as sensitive/offensive content. One possible approach to creating pedagogical corpora consists of excluding from the corpora sentences containing words from a blacklist of taboo and swear words, e. g. using GDEX (Kilgarriff et al. 2008). However, one of the greatest disadvantages of this method regards the fact that many words from the blacklist are polysemic. That means that the neutral sense of those words is not displayed in the corpus because all sentences containing those words have been excluded. Moreover, teachers might want to address sensitive/offensive content in their lessons depending on the characteristics of their learners and the teaching context (age, group level, unit topic, etc.). One way to create pedagogical corpora that still contain potentially problematic content and structure is to label, rather than remove these sentences. The end users of these corpora, such as dictionary makers and teachers, can then filter out the sentences according to their purposes. Considering that a) the process of labelling sentences in corpora is extremely time-consuming, if done manually; b) that automatic labelling can also be challenging given the polysemic nature of words; and that c) sensitivity and offensiveness are rather subjective concepts, this project is developing a game in which the crowd helps to achieve this task. With this game, players identify and label problematic sentences automatically extracted from existing corpora. CrowLL is a multimode game available as a webpage and a mobile app. At the time of writing, the single-player mode is being finished, with the dual-player mode expected by the time of the conference. Both modes have three levels, namely, level 1 (I'm curious!), level 2 (I'm eager to help!), and level 3 (I'm feeling enthusiastic!). In level 1, players identify problematic sentences according to their judgement; in level 2, they categorise those sentences, ranging from grammar/spelling problems to offensiveness and sensitivity; and in level 3, players mark in the sentence what they consider problematic. Players can choose to play the full game cycle (levels 1, 2 and 3), a combination of two levels (level 1 and level 2 or level 2 and level 3) or only one level (either level 1, or level 2 or level 3). In addition to demoing the game, we will present the method-

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/ Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

ology of the game project (Zingano Kuhn et al. 2021), which is organised in three stages, namely data preparation (stage 1), game design (stage 2), and machine learning preparation (stage 3). In the third stage, the plan is to use the sentences labelled by the players as a dataset to first train a binary machine learning model that will be able to automatically classify sentences as appropriate or inappropriate, and then to train a multi-class classifier that would be able to perform fine-grained labelling of the inappropriate sentences using the same categories as used in the game.

References

Baisa, V./Suchomel, V. (2014): SkELL: web interface for English language learning. In: Horák, A./ Rychlý, P. (eds.): Proceedings of the Eighth Workshop on Recent Advances in Slavonic Natural Language Processing, RASLAN 2014. Brno, pp. 63–70.

Braun, S. (2005): From pedagogically relevant corpora to authentic language learning contents. In: ReCALL 17, pp. 47–64.

Kilgarriff, A./Husák, M./McAdam, K./Rundell, M./Rychlý, P. (2008): GDEX: automatically finding good dictionary examples in a corpus. In: Bernal, E./DeCesaris, J. (eds.): Proceedings of the XIII EURALEX International Congress. Barcelona, pp. 425–432.

Zingano Kuhn, T./Todorović, B. Š./Arhar Holdt, Š./Zviel-Girshin, R./Koppel, K./Luís, A. R./Kosem, I. (2021): Crowdsourcing pedagogical corpora for lexicographical purposes. In: Gavriilidou, Z./Mitits, L./Kiosses, S. (eds.): Proceedings of the EURALEX XIX Congress. Volume 2. Alexandropoulos, pp. 771–779.

Contact information

Tanara Zingano Kuhn
Centre for the Studies of General and Applied Linguistics (CELGA-ILTEC),
University of Coimbra
tanarazingano@outlook.com

Špela Arhar Holdt
Centre for Language Resources and Technologies,
University of Ljubljana
spela.arharholdt@ff.uni-lj.si

Rina Zviel-Girshin
Ruppin Academic Center
rinazg@ruppin.ac.il

Ana R. Luís
Centre for the Studies of General and Applied Linguistics (CELGA-ILTEC),
University of Coimbra
aluis@fl.uc.pt

Carole Tiberius
Dutch Language Institute
carole.tiberius@ivdnt.org

Kristina Koppel
Institute of the Estonian Language
kristina.koppel@eki.ee

Introducing CrowLL – the Crowdsourcing for Language Learning game

Branislava Šandrih Todorović
University of Belgrade
branislava.sandrih@fil.bg.ac.rs

Iztok Kosem
Centre for Language Resources and Technologies, University of Ljubljana
iztok.kosem@ff.uni-lj.si

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Design and Publication of Dictionaries

XX EURALEX INTERNATIONAL CONGRESS
DICTIONARIES AND SOCIETY
12-16 July 2022, Mannheim, Germany

Bridgitte Le Du

TOWARDS A USER-CENTERED DESIGN MODEL TO ENHANCE USABILITY IN ELECTRONIC LEXICOGRAPHY

Some guiding principles applied in the adaptation of *A Dictionary of South African English on Historical Principles*

Keywords Online lexicography; user interface design; data aggregation; usability; user-centred design

In electronic dictionaries design trumps content: lexicographers find that no matter how authoritative its data, their product will see limited use unless it meets high standards of usability. In monolingual, historical dictionaries which have an elaborate structure and set of data types, user-oriented design becomes paramount.

Lexicographic theory is not always helpful in meeting this challenge; discussion has tended to remain focused on the technical innovations of dictionaries as digital artefacts, rather than on their usability. While the challenges involved in the shift from print to dynamic, functional online dictionaries should not be understated, theoretical and methodological lexicographical guidelines are not sufficiently directed at electronic dictionary designers' use cases. The result is a lack of project support in an area where it is later found it is most needed, namely design. This role, traditionally fulfilled by publishers in the print era, is often for pragmatic reasons assumed by programmers or lexicographers doubling as designers on electronic dictionaries, with mixed results. In the interim, design guidelines for online products such as search engines, news sites and social media have advanced rapidly, establishing usability conventions and forming user expectations, sometimes resulting in poorly designed dictionaries with high-quality data being rejected by users in favour of user-oriented resources with inferior content.

This presentation draws attention to the need for a guided approach to addressing fundamental design challenges brought about by the shift to a digital publishing medium, and suggests an emerging theoretical model for online dictionary design that addresses the parameter of usability as a key element in online dictionary User Experience (UX). The discussion draws on the development of a user-centred design approach as evolving from the print-to-electronic adaptation and enhancement of *A Dictionary of South African English on Historical Principles* (Silva et al. 1996).

Initial development phases in collaboration with the Universities of Hildesheim, Germany and Stellenbosch, South Africa focused on structuring use cases and enriching metadata, producing (1) dataset enhancements to support selective querying and data presentation changes (Van Niekerk/Stadler/Heid 2016); (2) prototypes of micro- and macro-visualisation devices using data aggregation (Van Niekerk/Stadler/Heid 2016; Van Niekerk/Le Du 2017); and (3) initial prototypes and wireframes following an early user survey and overall design review focused on simplified but more functional and accessible navigation features, as well as aesthetic and structural layout changes (Du Plessis/Van Niekerk 2016).

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Subsequently, assimilating the completed project's presentational and navigation components into a single, responsive interface for multiple platforms presented challenges dictated by factors that were both pragmatic (e.g. screen size) and user-centred (e.g. cognitive overload). The ensuing design process suggested fundamental parameters applicable to electronic lexicography, drawing on design patterns broadly employed in the online publishing industry. In the context of electronic lexicography, design considerations also reflect principles involving the compression of both the *navigation functions* of the dictionary interface (searching, sorting, accessing Help) and its *data* (individual entries).

Situating dictionaries within the context of generic reference-oriented web applications, and then adapting usability methodology and principles to lexicographic requirements, is an important step towards establishing effective dictionary design theory and practice. In the digital era users no longer accord dictionaries "the status of a kind of Bible" (Zaenen 2002); dictionaries have become one of many types of online resources consulted daily and unwieldy design is less likely to be accommodated.

References

Atkins, B. T. S./Rundell, M. (2008): The Oxford guide to practical lexicography. Oxford.

Dictionary of South African English (2020): http://dsae.co.za (last access: 24-03-2022).

Du Plessis, A./Van Niekerk, T. (2016): Adapting a historical dictionary for the modern online user: the case of the *Dictionary of South African English on Historical Principles*'s presentation and navigation features. In: Lexikos 26, pp. 82–102.

Silva, P./Dore, W./Mantzel, D./Muller, C./Wright, M. (eds.) (1996): A Dictionary of South African English on Historical Principles. Oxford.

Van Niekerk, T./Le Du, B. (2017): Data visualisation in the online Dictionary of South African English. Presentation at AFRILEX 22nd International Conference, Grahamstown, 26–29 June 2017.

Van Niekerk, T./Stadler, H./Heid, U. (2016): Enabling selective queries and adapting data display in the electronic version of a historical dictionary. In: Margalitadze, T./Meladze, G. (eds.): Lexicography and Linguistic Diversity. Proceedings of the XVII EURALEX International Congress, Tbilisi, 6–10 September 2016. Tbilisi, p. 635.

Zaenen, A. (2002): Musings about the impossible electronic dictionary. In: Corréard, M.-H. (ed.): Lexicography and natural language processing. A Festschrift in Honour of B. T. S. Atkins. Stuttgart, pp. 230–244.

Contact information

Bridgitte Le Du
Dictionary Unit for South African English, Rhodes University
b.ledu@ru.ac.za

Zita Hollós

CROSS-MEDIA-PUBLISHING IN DER KORPUSGESTÜTZTEN LERNERLEXIKOGRAPHIE

Entstehung eines Lernerwörterbuchportals DaF

Abstract Abstract This paper gives an insight into a cross-media publishing process on different stages: from a printed bilingual syntagmatic dictionary for GFL to an online learner's dictionary of German collocations to a German learner's dictionary portal. On the basis of an sql database specially developed for a corpus-guided dictionary of German collocations, the bilingual syntagmatic learner's dictionary KolleX was published in 2014. The first part of the article describes this lexicographic process, focusing the most relevant aspects of the dictionary concept, e. g. dictionary type, subject matter, corpus-guided data selection and microstructure. The second part introduces the first online version of KolleX from 2016 and the profound changes in the editing system – from a desktop version (2005) to a web-based editing system (2016) –, which resulted successively in a prototype of a German learner's dictionary portal, called E-KolleX DaF (2018–). Focusing on the aspects of dynamism and integration of different resources from a learner's perspective the paper shows the innovative features of this new online reference work. The contribution presents the solutions for the integration of new datatypes in the database of KolleX and the linking to different data in German monolingual dictionary platforms. The paper outlines the web design, functioning and technical improvements of E-KolleX DaF. The conclusions provide an outlook to the forthcoming challenges.

Keywords Wörterbuchportal; Kollokationen; Lernerwörterbuch; DaF; Datenbank; Webdesign

Contact information

Zita Hollós
Károli Gáspár University of the Reformed Church in Hungary (Budapest)
hollos.zita@kre.hu

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

(Promoting) Dictionary Use

XX EURALEX INTERNATIONAL CONGRESS
DICTIONARIES AND SOCIETY
12-16 July 2022, Mannheim, Germany

Andrea Abel

WÖRTERBÜCHER DER ZUKUNFT IN BILDUNGSKONTEXTEN DER GEGENWART

Eine Fallstudie aus dem Südtiroler Schulwesen

Abstract The focus of this paper will be on lexical information systems and the framework guidelines for the definition of the curricula within the educational system of the Autonomous Province of Bolzano/Bozen (Italy). In Italy, the competences to be achieved at different school levels are published in the form of general guidelines. On this basis each school has to specify the general competency goals and to spell them out in a concrete curriculum.

In this paper I will examine to what extent lexical information systems are represented in the framework guidelines within the German and the Italian educational system of the Autonomous Province, these being separate systems. In a second step, I will check the representations of the resources against the "Villa Vigoni Theses on Lexicography". Finally, I will discuss the results and give an outlook for further research.

Keywords Lexikalisches Informationssystem; Wörterbuchbenutzung; Wörterbücher in der Schule

Kontakt

Andrea Abel
Institut für Angewandte Sprachforschung – Eurac Research
andrea.abel@eurac.edu

Carolina Flinz/Sabrina Ballestracci

DAS LBC-WÖRTERBUCH: EINE ERSTE BENUTZERSTUDIE

Abstract This paper describes the results of an empirical investigation carried out within the project Lessico Multilingue dei Beni Culturali (LBC), whose aim is to create a multilingual online dictionary of the lexicon of the Italian artistic heritage. The dictionary, whose lexicographic process has already started, is intended for linguists and specialist translators as well as for professionals in the tourism sector and students of Foreign Languages and Literatures. The investigation conducted through a questionnaire submitted to undergraduate students at the University of Milan and at the University of Florence has a double aim: to research the habits in the use of lexicographic tools by possible users of the dictionary (Italian Learners of German Language), and to identify preferences regarding macro-, medio- and microstructural features of the future LBC-dictionary to realize a user-friendly tool. After a brief introduction on the state of the art of the survey in the field of Dictionary Users Studies, the article describes the questionnaire and the results obtained from the pilot study. A summary and a discussion on the future developments of the project conclude the work.

Keywords Dictionary use; LSP-Dictionary; Lexicon of Cultural Heritage

Contact information

Carolina Flinz
Università degli Studi di Milano
carolina.flinz@unimi.it

Sabrina Ballestracci
Università degli Studi di Firenze
sabrina.ballestracci@unifi.it

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Valeria Caruso/Alessandra Chervino/Giulia Daniele

DISSEMINATING DICTIONARY SKILLS WITH E-LEX TOOLS

Keywords Dictionary skills; collaborative lexicography; lexicography and public engagement; lexicography teaching; Covid-19; corpora; multilingual electronic dictionary

This paper reports on two intertwined activities carried out with Lexonomy, a web-based tool for publishing dictionaries originally developed by Michal Měchura (2017; Rambousek/Jakubíček/Kosem 2021) and then released as an open-source project. Despite the vast literature focused on the importance of teaching lexicographic skills (for a summary see for example Gavriilidou 2013), cursory attention has been paid to the activities and methods which can improve users' understanding of the dictionaries they use.

With this aim in mind, one hundred university students in Linguistic Mediation were guided through compiling a multilingual Covid-19 dictionary which, later on, was used in a lexicographic informational session during the European Researchers' Night events. These activities, held in virtual rooms of video conferencing web platforms during the pandemic lockdown, profited from online resources like Lexonomy, which ensure data crowdsourcing from authorised contributors. Besides, using Lexonomy students should easily understand the microstructural organisation of dictionary articles -from the type of linguistic items to be found in general language dictionaries to the way they are described or arranged in the entry- since the data is added through an XML visual editor, displaying the entry structure in the form of a clickable hierarchical tree with nested elements. The task of compiling entries should therefore be easy to learn and engaging to accomplish, therefore the activities of this experimental protocol were meant to ensure different learning goals: to practise metalinguistic and metalexicographical knowledge for enhancing linguistic and translation skills.

The first attempt to compile the *Dizionario Multilingue del Covid19 – Covid19 Multilingual Dictionary – Covid19 Mehrsprachiges Wörterbuch* was carried out during class hours of a course in Lexicology and Lexicography. Students collected articles from major newspapers for the languages involved in the project: Chinese, Dutch, English, German, Italian, Polish, Portuguese, Spanish and Russian. Texts were shared in online folders and were used to build comparable corpora using the *Sketch Engine* tools.

Later on, students were given the article microstructure schema, defined by the teacher, for compiling the multilingual dictionary. The article is made up by a meaning explanation and one example of use of an Italian lemma, followed by its collocates and licensed prepositions. The corresponding lemmas in the other languages covered by the dictionary are given on the same page as separate entries, arranged in the same structural organization as their Italian counterpart. Direct access to foreign words is possible by performing free searches in the search box since the multilingual macrostructure is not reversible.

As regards dictionary data, translation equivalents were identified by comparing the single and multi-word terms extracted from the comparable corpora collected. Yet additional searches on Google news or in the *Sketch Engine* "Covid-19 Corpus" were necessary when the language data proved to be insufficient.

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

With the aim of evaluating students' performance, specific assessment parameters were used:

1) the accuracy in compiling the dictionary entries,
2) the metalinguistic and metalexicographical skills acquired,
3) the effectiveness of peer learning.

In two of the three parameters analysed, the students achieved good results, whereas metalinguistic and metatextual skills, assessed by means of an open-ended questionnaire, were poor. Only a fraction of the students (around 30%) were able to explain fully some key concepts for the task, such as what a dictionary entry, microstructure or collocation is.

The presentation will provide further details of the experimental compilation carried out by the students and the ninety-three people who participated in the European Researchers' Night demonstration. The audience at this event rated the session as very interesting (4,68 on a five-point Likert scale). They also declared that it helped them understand that research in the humanities is carried out with scientific methods and has an impact on citizens' daily lives.

References

Dizionario Multilingue del Covid19 – Covid19 Multilingual Dictionary – Covid19 Mehrsprachiges Wörterbuch. https://www.lexonomy.eu/hpg58cpv (last access: 11-2021).

European Researchers' Night: https://ec.europa.eu/research/mariecurieactions/event/2021-european-researchers-night (last access: 11-2021).

Gavriilidou, Z. (2013): Development and validation of the Strategy Inventory for Dictionary Use (S.I.D.U.). In: International Journal of Lexicography 26 (2), pp. 135–153.

Lexonomy: https://www.lexonomy.eu (last access: 11-2021).

Měchura, M. B. (2017): Introducing Lexonomy: an open-source dictionary writing and publishing system. In: Electronic lexicography in the 21st century: lexicography from scratch. Proceedings of the eLex 2017 Conference, 19–21 September 2017. Leiden, pp. 662−689.

Rambousek, A./Jakubíček, M./Kosem, I. (2021): New developments in Lexonomy. In: Kosem, I./Cukr, M./Jakubíček, M./Kallas, J./Krek, S./Tiberius, C. (eds.): Electronic lexicography in the 21st century. Proceedings of the eLex 2021 conference. Brno, pp 455−462.

Contact information

Valeria Caruso
Naples-Università degli studi di Napoli “L'Orientale”
vcaruso@unior.it

Alessandra Chervino
Università di Bologna - Alma mater studiorum
alessandra.chervino@hotmail.it

Giulia Daniele
Università Telematica Pegaso
giulia.daniele@outlook.com

Ida Dringó-Horváth/Katalin P. Márkus

DICTIONARY SKILLS IN TEACHING ENGLISH AND GERMAN AS A FOREIGN LANGUAGE IN HUNGARY

A questionnaire study

Keywords Lexicography; quantitative research; methodology; dictionary use; electronic dictionary

Research into dictionary use has increased significantly in the recent past (e.g., Kosem et al. 2018; Lew/de Schryver 2014; Müller-Spitzer (ed.) 2014; Nesi 2012), yet Hungary seems to fall behind as there is a scarcity of research on dictionary use despite its necessity in the learning process (e.g., Dringó-Horváth 2017; Gaál 2017; P. Márkus 2020).

A dictionary comprises essential information offered to language learners, however, interpreting all the information obtained is not an easy task. Learners need special skills to succeed and if the challenges are high, but the skills are low, users may lose confidence in dictionaries. It looks as though the skills of dictionary users have been taken for granted. Nevertheless, nobody is born with the skills and knowledge needed to use a dictionary effectively. Students must learn how to communicate with their dictionaries to solve language-related problems as they support autonomous learning. The dictionary is widely regarded as a 'teacher who cannot talk' (Chi 1998). Teachers are not available continuously, so students need to develop the competences required to achieve independent and autonomous learning.

Results of previous studies point out that in Hungary 'dictionary awareness' is generally rather low (e.g., Dringó-Horváth 2017; P. Márkus 2020), and that more attention to the teaching of dictionary skills would be needed in the curricula for English language learning. The enormous challenges presented by the poor dictionary culture and the inability of teachers to integrate dictionary pedagogy into everyday teaching are clearly visible. With the aim of developing a method to improve the existing situation, the authors examined the dictionary-using behaviour of graduates in English and German Language as well as their attitudes towards teaching and learning dictionary skills in the classroom. The quantitative research aimed to investigate the participants' (N=197) preferences and attitudes regarding dictionary use, their dictionary consultation behaviour, and the role of dictionaries as an aid to language learning. The results of the research seem to confirm some trends revealed in previous studies (e.g., Dringó-Horváth 2017; Gaál 2017; Nied Curcio 2015; Kosem et al. 2018; Töpel 2015): a strong increase in the use of digital, especially online dictionaries; the preference of free online dictionaries over paid ones without being aware of the quality differences between the two; the frequent use of translation programs and search engines for dictionary purposes; and the low prestige of teaching dictionary use (dictionary skills are typically acquired in an autonomous, self-taught way). The results may also suggest that the participants do not or only superficially use the aids related to dictionaries, and that dictionaries are not used in sufficient quantity consciously. Based on previous research and observations of dictionary use "it is a truth universally acknowledged in lexicographic circles that user's guides are very seldom consulted" (Svensén 2009, p. 459), which hinders effective dictionary use (e.g., Atkins/Varantola 1997).

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

The authors were also curious how teaching of dictionary use is reflected in the educational practice of the participants with teaching experience. Unfortunately, it seems that dictionary didactics is not necessarily an essential part of foreign language lessons. However, different types of dictionaries and search techniques are to a certain extent part of teaching dictionary use. The answers indicate that teaching dictionary skills would be most facilitated by educational aids related to the topic and that the central educational policy should support a better presentation of the topic (cf. Lew/Galas 2008; P. Márkus 2020).

As for the electronic dictionaries, it may be concluded that participants typically do not take advantage of the new features (e.g., customization, upgradeability) and search techniques. Several other studies show a similar lack of knowledge of electronic dictionaries (Dringó-Horváth 2017; Nied Curio 2015), which indicates a great deal of uncertainty when choosing from the increasing range of dictionaries. Participants may not be aware of these new opportunities and their potential benefits, as teaching how to use electronic dictionaries lags behind the teaching of the use of traditional paper dictionaries. This seems problematic based on the results of classroom dictionary use, which shows that the use of online dictionaries is more prevalent. In addition to the lack of teaching dictionary skills, the responsibility of publishers and electronic dictionary providers also arises: Does the user interface provide sufficient description of the available functions? / Is there an adequate reference to the new functions?

Dictionary use and studying with electronic dictionaries are – despite many similarities – characterized by significant differences compared to their print counterparts. Therefore, special emphasis should be laid on the special features of electronic dictionaries: in the classroom, with the help of specific exercises instructors should demonstrate how specific features of electronic dictionaries can be used effectively (e.g., search strategies that can have a positive impact on the language learning process) (Dringó-Horváth 2021).

Based on the findings, we plan to design a core "dictionary skills module", which could be incorporated into different courses at the university. After the completion of the project, we hope to be able to give guidance on how to integrate (the teaching of) dictionary use into traditional classroom teaching, thus making the teaching and learning process more effective; improve the digital competences of students; as well as contribute to the reduction of inequalities among students from different socio-economic backgrounds.

References

Atkins, B. T. S./Varantola, K. (1997): Monitoring dictionary use. In: International Journal Lexicography 10 (1), pp. 1–45.

Dringó-Horváth, I. (2017): Digitális szótárak – szótárdidaktika és szótárhasználati szokások [Digital Dictionaries – Dictionary Didactics and Dictionary Use]. In: Alkalmazott Nyelvtudomány 17 (5), pp. 1–27. http://alkalmazottnyelvtudomany.hu/wordpress/wp-content/uploads/DringoHorvath.pdf (last access: 21-03-2022).

Dringó-Horváth, I. (2021): Digitale Wörterbücher – Auswahlkriterien und angepasste Wörterbuchdidaktik. In: Fremdsprache Deutsch 64 (Themenheft Wortschatz). https://fremdsprachedeutschdigital.de/download/fd/FD_64_online_Dringo-Horvath.pdf (last access: 03-21-2022).

Gaál, P. (2017): Onlineszótár-használat Magyarországon (OHM): Egy kérdőíves szótárhasználati felmérés eredményei [Online dictionary use in Hungary]. In: Alkalmazott Nyelvtudomány 17 (1), pp. 5–19.

Kosem, I./Lew, R./Müller-Spitzer, C./Ribeiro Silveira, M./Wolfer, S. (2018): The image of the monolingual dictionary across Europe. Results of the European survey of dictionary use and culture. In: International Journal of Lexicography 32 (1), pp. 92–114.

Lew, R./de Schryver, G.-M. (2014): Dictionary users in the digital revolution. In: International Journal of Lexicography 27 (4,), pp. 341–359.

Lew, R./Galas, K. (2008): Can dictionary skills be taught: the effectiveness of lexicographic training for primary-school-level Polish learners of English. In: Bernal, E./DeCesaris J. (eds.): Proceedings of the XIII EURALEX International Congress, 15–19 July 2008. Barcelona, pp. 1273–1285.

Müller-Spitzer, C. (ed.) (2014): Using online dictionaries. (= Lexicographica. Series Maior 145). Berlin/Boston.

Nesi, H. (2012): Alternative e-dictionaries: uncovering dark practices. In: Granger, S./Paquot, M. (eds.): Electronic lexicography. Oxford, pp. 363–378.

Nied Curcio, M. (2015): Wörterbuchbenutzung und Wortschatzerwerb. Werden im Zeitalter des Smartphones überhaupt noch Vokabeln gelernt? In: Info DaF 5, pp. 445–468.

P. Márkus, K. (2020): A szótárhasználat jelene és jövője a közoktatásban – a nyelvoktatást szabályozó dokumentumok és segédanyagok tükrében [Dictionary use in public education – present and future). In: Modern Nyelvoktatás 26 (1–2), pp. 59–79.

Svensén, B. (2009): A handbook of lexicography. The theory and practice of dictionary-making. Cambridge.

Töpel, A. (2015): Das Wörterbuch ist tot – es lebe das Wörterbuch?! In: Info DaF 5, pp. 515–534.

Contact information

Ida Dringó-Horváth
Károli Gáspár University of the Reformed Church in Hungary
dringo.horvath.ida@kre.hu

Katalin P. Márkus
Károli Gáspár University of the Reformed Church in Hungary
p.markus.kata@kre.hu

Zoe Gavriilidou/Evi Konstandinidou

THE EFFECT OF AN EXPLICIT AND INTEGRATED DICTIONARY AWARENESS INTERVENTION PROGRAM ON DICTIONARY USE STRATEGIES

Abstract There is a growing interest in pedagogical lexicography, and more specifically in the study of dictionary users' abilities and strategies (Prichard 2008; Gavriilidou 2010, 2011; Gavriilidou et al. 2020; Gavriilidou/Konstantinidou 2021; Chatjipapa et al. 2020). The purpose of this presentation is to investigate dictionary use strategy and the effect of an explicit and integrated dictionary awareness intervention program on upper elementary pupils' dictionary use strategies according to gender and type of school. A total of 150 students from mainstream and intercultural schools, aged 10–12 years old, participated in the study. Data were collected before and after the intervention through the Strategy Inventory for Dictionary Use (SIDU) (Gavriilidou 2013). The results showed a significant effect of the intervention program on Dictionary Use Strategies employed by the experimental group and support the claim that increased dictionary use can be the outcome of explicit strategy instruction. In addition, the effective application of the program suggests that a direct and clear presentation of DUS is likely to be more successful than an implicit presentation. The present study contributes to the discussion concerning both the 'teachability' of dictionary use strategies and skills and the effective forms of intervention programs raising dictionary use awareness and culture.

Keywords Dictionary use strategies; explicit and integrated intervention program; dictionary culture; pedagogical lexicography

References

Chadjipapa, E./Gavriilidou, Z./Markos, A./Mylonopoulos, A. (2020): The effect of gender and educational level on dictionary use strategies adopted by upper-elementary and lower-secondary students attending Greek Schools. In: International Journal of Lexicography 33 (4), pp. 443–462.

Gavriilidou, Z. (2010): Profiling Greek adult dictionary users. In: Studies in Greek Linguistics 31, pp. 166–172.

Gavriilidou, Z. (2011): Users' abilities and performance in dictionary look up. In: Lavidas, N. (ed.): Selected papers of the 20th International Symposium of Theoretical and Applied Linguistics. Thessaloniki, pp. 41–51.

Gavriilidou, Z. (2013): Development and validation of the Strategy Inventory for Dictionary Use (S.I.D.U). In: International Journal of Lexicography 22 (2), pp. 135–154.

Gavriilidou, Z./Konstantinidou, E. (2021): The design of an explicit and integrated intervention program for pupils aged 10–12 with the aim to promote dictionary culture and strategies. In: Gavriilidou, Z./Mitis, L./Kiosses, S. (eds.): Proceedings of the XIX Euralex Congress: Lexicography for Inclusion. Vol. 2. Komotini, pp. 735–745. https://euralex2020.gr/wp-content/uploads/2021/09/Pages-from-EURALEX2021_ProceedingsBook-Vol2-p735-745.pdf.

Gavriilidou, Z./Mavrommatidou, S./Markos, A. (2020): The effect of gender, age and career orientation on digital dictionary use strategies. In: International Journal of Research Studies in Education 9 (6), pp. 63–76. https://doi.org/10.5861/ijrse.2020.5046.

Prichard, C. (2008): Evaluating L2 readers' vocabulary strategies and dictionary use. In: Reading in a Foreign Language 20 (2), pp. 216–231.

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Contact information

Zoe Gavriilidou
zoegab@otenet.gr
Greece Democritus University of Thrace

Evi Konstandinidou
evi1990@hotmail.gr
Greece Democritus University of Thrace

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Theresa Kruse/Ulrich Heid

LEARNING FROM STUDENTS

On the design and usability of an e-dictionary of mathematical graph theory

Abstract We created a prototype of an electronic dictionary for the mathematical domain of graph theory. We evaluate our prototype and compare its effectiveness in task-based tests with that of Wikipedia. Our dictionary is based on a corpus; the terms and their definitions were automatically extracted and annotated by experts (cf. Kruse/Heid 2020). The dictionary is bilingual, covering German and English; it gives equivalents, definitions and semantically related terms. For the implementation of the dictionary, we used LexO (Bellandi et al. 2017). The target group of the dictionary are students of mathematics who attend lectures in German and work with English resources. We carried out tests to understand which items the students search for when they work on graph-theoretical tasks. We ran the same test twice, with comparable student groups, either allowing Wikipedia as an information source or our dictionary. The dictionary seems to be especially helpful for students who already have a vague idea of a term because they can use the resource to check if their idea is right.

Keywords Dictionary use; specialised languages; terminology; terminography; mathematics

References

Bellandi, A. et al. (2017): Developing LexO: a collaborative editor of multilingual lexica and termino-ontological resources in the humanities. In: Proceedings of Language, Ontology, Terminology and Knowledge Structures Workshop (LOTKS 2017). Montpellier, France. https://aclanthology.org/W17-7010.

Kruse, T./Heid, U. (2020): Lemma selection and microstructure: definitions and semantic relations of a domain-specific e-dictionary of the mathematical domain of graph theory. In: Gavriilidou, Zoe/Mitsiaki, Maria/Fliatouras, Asimakis (eds.): Lexicography for Inclusion: Proceedings of the 19th EURALEX International Congress, 7–9 September 2021. Alexandroupolis. Volume 1. Alexandroupolis, pp. 227–233.

Contact information

Theresa Kruse
University of Hildesheim
kruset@uni-hildesheim.de

Ulrich Heid
University of Hildesheim
heidul@uni-hildesheim.de

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Sascha Wolfer/Robert Lew

PREDICTING ENGLISH WIKTIONARY CONSULTATIONS

Keywords Wiktionary; English; corpora; frequency; prevalence; age of acquisition; polysemy

Dictionaries have been part and parcel of literate societies for many centuries. They assist in communication, particularly across different languages, to aid in understanding, creating, and translating texts. Communication problems arise whenever a native speaker of one language comes into contact with a speaker of another language. At the same time, English has established itself as a *lingua franca* of international communication. This marked tendency gives lexicography of English a particular significance, as English dictionaries are used intensively and extensively by huge numbers of people worldwide.

In doing so, users make choices about which words to look up, and our aim is to identify the lexical variables that affect the likelihood of those choices by using the log files of a popular crowd-sourced dictionary: the English Wiktionary. The choice of the English Wiktionary is motivated by the availability and size of the log files. While not seeking a single lexical processing or representation model, we are interested in what drives people's decisions to look up a specific word in terms of language experience. We are contributing towards at least two research questions: 1) What makes people interested in specific words, what prompts them to seek information on these words in lexical resources? 2) Can we formulate guidelines for dictionary compilation by offering empirically-based quantifiable measures of which specific words are more likely to be sought by users, so lexicographic work can prioritize these words, and words exhibiting similar characteristics? We note that the specific look-up context as well as idiosyncratic user characteristics cannot be known at the stage of lexicographic design, and so our approach also ignores these factors.

One factor that is already known to guide people's look-up behaviour is corpus-based lexical frequency. While, quite surprisingly, the positive relationship between dictionary look-up and corpus frequency did not turn out to be apparent at all in initial studies (de Schryver/Joffe 2004; de Schryver et al. 2006; Verlinde/Binon 2010), it has since been established empirically with some confidence (Koplenig/Meyer/Müller-Spitzer 2014; Müller-Spitzer/Wolfer/Koplenig 2015; de Schryver/Wolfer/Lew 2019). However, we see a clear advantage in including further variables. Metrics reflecting other properties of words (some of them closely related – but not identical – to corpus frequency) can help us understand better the effect of corpus frequency and the relationships between different variables predicting look-up behaviour. As a first step, we will consider word prevalence, age of acquisition, and number of senses (or degree of polysemy) of the headword.

The prevalence of a word is the extent to which it is known amongst the native-speaking population. It stands to reason that words which occur with relatively higher frequency in texts and discourse should be more likely to be known by a large proportion of the speakers (Weizman/Snow 2001; Longobardi et al. 2015). However, it remains to be seen whether – with the effect of frequency controlled for – word prevalence still is a relevant predictor of look-up frequency, and if so, in which direction.

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/
Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of
Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Age of acquisition is the age at which a word is, on average, acquired by native speakers in the process of (naturalistic) L1 acquisition. One might expect that this could play a role in how words acquired earlier, possibly being more deeply entrenched in the mental lexicon, get to be looked up. Age of acquisition has been found to have important and long-lasting effects on language behaviour (Ellis/Lambon Ralph 2000; Garlock/Walley/Metsala 2001; Juhasz 2005; Kuperman/Stadthagen-Gonzalez/Brysbaert 2012).

The concept of word sense is not without problems (Kilgarriff 1997; Hanks 2000) and there has been a long-drawn-out debate about the boundaries between polysemy and homonymy. To steer clear of the essentialist debate of whether words 'have' senses, we adopt a pragmatic approach of considering *lexicographic* senses: the separate blocks of meaning description as given in a dictionary. Degree of polysemy is, then, operationalized as the number of dictionary senses in the English Wiktionary itself. We have known for about 70 years (Zipf 1949) that the more frequent words tend to have more senses. However, the degree of polysemy may hold predictive potential above and beyond that of mere word frequency (Müller-Spitzer/Wolfer/Koplenig 2015).

Lexical frequency, prevalence, age of acquisition, and degree of polysemy will obviously not explain all of the variance in our model. As in any statistical model, there will inevitably remain unexplained variation represented by *residual* or *error* variance. Detailed investigation of this residual variance (complemented also with a more qualitative perspective on the observed data), additional factors might have to be brought into the picture to more fully account for look-up behaviour.

With our analyses, we hope to provide more information on the lexical variables that affect look-up behaviour – apart from mere corpus frequency. At the same time, we will try to shed some light on (groups of) headwords that might not follow the overall pattern of the data. Such outliers could point to the fact that some additional variables (or interactions between variables) have to be taken into consideration to broaden our understanding of how people use dictionaries.

References

de Schryver, G.-M. et al. (2006): Do dictionary users really look up frequent words? – On the overestimation of the value of corpus-based lexicography. In: Lexikos 16, pp. 67–83.

de Schryver, G.-M./Joffe, D. (2004): On how electronic dictionaries are really used. In: Williams, G./Vessier, S. (eds.): Proceedings of the Eleventh EURALEX International Congress (EURALEX 2004), Lorient, France, July 6–10, 2004, Vol. 1. Lorient, pp. 187–196.

de Schryver, G.-M./Wolfer, S./Lew, R. (2019): The relationship between dictionary look-up frequency and corpus frequency revisited: a log-file analysis of a decade of user interaction with a Swahili-English dictionary. In: GEMA Online Journal of Language Studies 19 (4), pp. 1–27. http://doi.org/10.17576/gema-2019-1904-01.

Ellis, A. W./Lambon Ralph, M. A. (2000): Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: insights from connectionist networks. In: Journal of Experimental Psychology: Learning Memory and Cognition 26 (5), pp. 1103–1123. http://doi.org/10.1037/0278-7393.26.5.1103.

Garlock, V. M./Walley, A. C./Metsala, J. L. (2001): Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. In: Journal of Memory and Language 45 (3), pp. 468–492. http://doi.org/10.1006/jmla.2000.2784.

Hanks, P. (2000): Do word meanings exist? In: Computers and the Humanities 34 (1–2), pp. 205–215.

Juhasz, B. J. (2005): Age-of-acquisition effects in word and picture identification. In: Psychological Bulletin 131 (5), pp. 684–712. http://doi.org/10.1037/0033-2909.131.5.684.

Kilgarriff, A. (1997): I don't believe in word senses. In: Computers and the Humanities 31 (2), pp. 91–113.

Koplenig, A./Meyer, P./Müller-Spitzer, C. (2014): Dictionary users do look up frequent words. A log file analysis. In: Müller-Spitzer, C. (ed.): Using online dictionaries. (= Lexicographica Series Maior 145). Berlin, pp. 229–249.

Kuperman, V./Stadthagen-Gonzalez, H./Brysbaert, M. (2012): Age-of-acquisition ratings for 30,000 English words. In: Behavior Research Methods 44 (4), pp. 978–990. http://doi.org/10.3758/s13428-012-0210-4.

Longobardi, E. et al. (2015): Children's acquisition of nouns and verbs in Italian: contrasting the roles of frequency and positional salience in maternal language. In: Journal of Child Language 42 (1), pp. 95–121. http://doi.org/10.1017/S0305000913000597.

Müller-Spitzer, C./Wolfer, S./Koplenig, A. (2015): Observing online dictionary users: studies using Wiktionary log files. In: International Journal of Lexicography 28, pp. 1–26. http://doi.org/10.1093/ijl/ecu029.

Verlinde, S./Binon, J. (2010): Monitoring dictionary use in the electronic age. In: Dykstra, A./Schoonheim, T. (eds.): Proceedings of the XIV Euralex International Congress. Ljouwert, pp. 1144–1151.

Weizman, Z. O./Snow, C. E. (2001): Lexical input as related to children's vocabulary acquisition: effects of sophisticated exposure and support for meaning. In: Developmental Psychology 37 (2), pp. 265–279. http://doi.org/10.1037/0012-1649.37.2.265.

Zipf, G. K. (1949): Human behavior and the principle of least effort. Cambridge, MA.

Contact information

Sascha Wolfer
Leibniz-Institut für Deutsche Sprache
wolfer@ids-mannheim.de

Robert Lew
Adam Mickiewicz University
rlew@amu.edu.pl

Manuel Raaf

EVALUATION DES USER-CENTERED DESIGNS EINES SPRACHINFORMATIONSSYSTEMS

Planung, Durchführung und Ergebnisse einer Benutzerumfrage zu Usability und User Experience

Keywords Nutzerforschung; User-Centered Design; Usability; User Experience; Wörterbuchbenutzung

Bei der Entwicklung einer Webseite muss man sich fragen, wie das Design bestmöglich für die anvisierte(n) Zielgruppe(n) gestaltet werden muss, damit diese den Internetauftritt nicht direkt beim ersten Besuch frustriert verlassen und nicht wiederkehren – unabhängig davon, ob man nun im wissenschaftlichen Umfeld tätig ist oder im privatwirtschaftlichen. Eine erfolgreiche und zufriedenstellende Nutzung einer Webseite ist ohne solch ein Design – das weit mehr ist, als nur ein Layout – i. d. R. nicht möglich.

Im Projekt „Bayerns Dialekte Online“ wird der Inhalt des Bayerischen Wörterbuchs, des Fränkischen Wörterbuchs sowie des Dialektologischen Informationssystems von Bayerisch-Schwaben vereint. Das Projekt adressiert dabei explizit, jedoch nicht exklusiv die interessierte Öffentlichkeit. Zunächst wird einleitend skizziert, warum und wie das sogenannte User-Centered Design (UCD), das potenzielle Nutzergruppen im Fokus hat, erarbeitet und implementiert wurde. Auf die wissenschaftliche Basis wird ebenfalls am Rande eingegangen, zumal diese bisher kaum Beachtung in den Geisteswissenschaften fand (siehe auch nächster Absatz). Im Hauptteil des Vortrags wird illustriert, dass die Strategie des UCD zielführend war. Hierfür werden die Nutzerumfragen vorgestellt, die im Winter 2021/2022 via Online-Fragebogen und persönlicher, pandemiebedingt aber digital stattgefundener Interviews durchgeführt wurden. Die jeweilige Methodik wird erläutert, bevor im Anschluss detailliert das positive Gesamtergebnis sowie Optimierungsbedarfe, die sich aus der Umfrage ergeben, präsentiert werden. Eine Empfehlung dazu, wie und aus welchen Gründen die gewählten Methoden verbessert werden können oder gar müssen, folgt abschließend ebenso wie ein Fazit hinsichtlich der Methodiken selbst: So stellte sich zum einen u. a. heraus, dass viele Nutzer:innen offenbar vom Umfang abgeschreckt waren und ein kleinerer Rahmen vermutlich mehr Teilnahmen hervorgebracht hätte. Zum anderen zeigte sich aber auch ganz klar, dass eine Handvoll Interviews deutlich wertvollere Einblicke liefern als hunderte Fragebögen und Interviews somit generell zu bevorzugen sind, um die Gebrauchstauglichkeit und Nutzerzufriedenheit eines Online-Wörterbuchs zu erfragen.

Die Notwendigkeit der Durchführung eigener Erhebungen soll im gesamten Vortrag unterstrichen werden, da es sich bei einer Arbeit dieser Art um ein theoretisches sowie mehrheitlich gar praktisches Desiderat handelt: Zur Usability (Benutzbarkeit, Benutzerfreundlichkeit, Gebrauchstauglichkeit) und zur User Experience (Benutzerzufriedenheit, Nutzungserfahrung) liefern Kognitions- und Designwissenschaften mit ihrer Fachliteratur zwar seit nunmehr vier Jahrzehnten reichlich Informationen, doch beziehen sich diese nicht ausschließlich auf Webseiten und schon gar nicht auf geisteswissenschaftliche Projekte wie z. B. Sprachinformationssysteme oder einzelne Online-Wörterbücher (vgl. Nielsen 1993; Nielsen/Loranger 2006; Krug 2014; Shneiderman et al. 2017; Jacobsen/Meyer 2019; Sharp/

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Preece/Rogers 2019). Sie blieben bisher zudem fast vollständig unbeachtet innerhalb der digitalen Geisteswissenschaften. Womöglich mag dies dem Umstand geschuldet sein, dass sie primär in der Informationstechnologie verortet werden und die Begrifflichkeiten zu technisch verstanden werden. Sie gehen jedoch weit darüber hinaus, wie im Vortrag kurz skizziert werden wird.

Zwar gibt es Arbeiten zur Benutzung von elektronischen Wörterbüchern, jedoch „fehlt eine systematische Grundlagenforschung zur Internetlexikographie“ (Schierholz 2019, S. 167) und vorhandene Studien beziehen sich fast ausschließlich auf die Benutzung (Lew/de Schryver 2014; Lew 2015; Müller-Spitzer 2016; Kosem et al. 2018), selten auf die Benutzbarkeit oder Nutzerzufriedenheit (Klosa/Koplenig/Töpel 2011; Bank 2012; Heid/Zimmermann 2012; Müller-Spitzer/Koplenig/Töpel 2015). Sie gehen also mehr der Frage nach, warum und von wem ein elektronisches Wörterbuch genutzt wird, weniger jedoch der, ob z. B. das Layout intuitiv gestaltet ist oder die Funktionen zufriedenstellend sind. Des Weiteren wurden bisher oft Log-Dateien zur Auswertung verwendet oder projektspezifische Fragestellungen genutzt, allerdings keine Nutzerbefragungen durchgeführt, die auf reliablen Fragebögen basieren (vgl. Thomaschewski/Hinderks/Schrepp 2018). Darüber hinaus sind vorhandene Studien aufgrund der starken Divergenz der untersuchten Ressourcen nur schwer miteinander zu vergleichen.

Dabei stellt der Themenkomplex Usability und User Experience doch gerade für die Digital Humanities eine besondere Herausforderung dar, da zunehmend nicht nur Fachwissenschaftler:innen adressiert werden sollen, sondern auch die interessierte Öffentlichkeit. Somit muss hier eine besondere Balance gehalten werden zwischen granularen Suchoptionen für Fachleute auf der einen und dem niedrigschwelligen Zugang für Lai:innen auf der anderen Seite. Die Parameter, die hierfür vonnöten sind, werden am besten durch Nutzerumfragen eruiert, da letztlich die Masse der Benutzer:innen eine verlässlichere, unvoreingenommenere Meinung zu verschiedenen Fragen liefern kann als Projektmitarbeiter:innen, denen dazu schlichtweg die Distanz fehlt. Im besonderen Maße trifft dies auf wenig computeraffine Lai:innen zu, die im Falle des Projekts „Bayern Dialekte Online“ eine wichtige Nutzergruppe darstellen.

Literatur

Bank, C. (2012): Die Usability von Online-Wörterbüchern und elektronischen Sprachportalen. In: Information – Wissenschaft & Praxis 63 (6), S. 345–360. http://doi.org/10.1515/iwp-2012-0069.

Heid, U./Zimmermann, J. T. (2012): Usability testing as a tool for e-dictionary design: collocations as a case in point. In: Fjeld, R. V./Torjusen, J. M. (Hg.): Proceedings of the 15th EURALEX International Congress. Oslo, S. 661–671.

Jacobsen, J./Meyer, L. (2019): Praxisbuch Usability und UX: Was jeder wissen sollte, der Websites und Apps entwickelt. Bonn.

Klosa, A./Koplenig, A./Töpel, A. (2011): Benutzerwünsche und Meinungen zu einer optimierten Wörterbuchpräsentation: Ergebnisse einer Onlinebefragung zu elexiko. (= OPAL – Online publizierte Arbeiten zur Linguistik 3/2011). Mannheim. http://pub.ids-mannheim.de/laufend/opal/pdf/opal2011-3.pdf (Stand: 23.3.2022).

Kosem, I./Lew, R./Müller-Spitzer, C./Ribeiro Silveira, M./Wolfer, S. (2018): The image of the monolingual dictionary across Europe. Results of the European survey of dictionary use and culture. In: International Journal of Lexicography 32 (1), S. 1–23. http://doi.org/10.1093/ijl/ecy022.

Krug, S. (2014): Don't make me think! Web & Mobile Usability – das intuitive Web. 3. Auflage. Heidelberg.

Lew, R. (2015): Research into the use of online dictionaries. In: International Journal of Lexicography 28 (2), S. 232–253. http://doi.org/10.1093/ijl/ecv010.

Lew, R./de Schryver, G.-M. (2014): Dictionary users in the digital revolution. In: International Journal of Lexicography 27 (4), S. 341–359. http://doi.org/10.1093/ijl/ecu011.

Müller-Spitzer, C. (2016): Wörterbuchbenutzungsforschung. In: Klosa, A./Müller-Spitzer, C. (Hg.): Internetlexikografie. Ein Kompendium. Berlin, S. 291–342. http://doi.org/10.1515/9783050095615-010.

Müller-Spitzer, C./Koplenig, A./Töpel, A. (2015): What makes a good online dictionary? Empirical insights from an interdisciplinary research project. Mannheim. [Herausgegeben von I. Kosem/ K. Kosem].

Nielsen, J. (1993): Usability engineering. San Diego.

Nielsen, J./Loranger, H. (2006): Web usability. München. [Übersetzt von I. Kommer/C. Kommer].

Schierholz, S. J. (2019): Brauchen wir noch Wörterbücher? Ja! Aber welche? In: Eichinger, L. M./ Plewnia, A. (Hg.): Neues vom heutigen Deutsch. Berlin/Boston, S. 163–198. http://doi.org/10.1515/9783110622591-009.

Sharp, H./Preece, J./Rogers, Y. (2019): Interaction design: beyond human-computer interaction. 5. Auflage. Indianapolis.

Shneiderman, B./Plaisant, C./Cohen, M. S./Jacobs, S. M./Elmqvist, N. (2017): Designing the user interface: strategies for effective human-computer interaction. 6. Auflage. Boston u. a.

Thomaschewski, J./Hinderks, A./Schrepp, M. (2018): Welcher UX-Fragebogen passt zu meinem Produkt? In: Hess, S./Fischer, H. (Hg.): Mensch und Computer 2018 – Usability Professionals. Bonn, S. 437–446. http://doi.org/10.18420/muc2018-up-0150.

Kontaktinformationen

Manuel Raaf
Bayerische Akademie der Wissenschaften
raaf@badw.de

Geraint Paul Rees

ONLINE DICTIONARIES AND ACCESSIBILITY FOR PEOPLE WITH VISUAL IMPAIRMENTS

Keywords Accessibility; online dictionaries; visual impairments

At its core, much lexicographic research is concerned with the accessibility of dictionaries. In general, this interest is apparent in research on look-up patterns (Laufer/Kimmel 1997), signposting (DeCesaris 2012), and other elements of microstructure. As regards digital dictionaries, concern for accessibility is evident in the use of novel methods, for example, eye-tracking (Lew/Grzelak/Leszkowicz 2013) and the investigation of phenomena such as the effect of advertisements in online dictionaries on users (Dziemianko 2020). For many users, this has led to improvements in accessibility. There is, however, a significant minority of users for whom accessing dictionaries still poses a significant problem.

Globally, there are an estimated 285-million people with visual impairments. This group is particularly at risk of social exclusion (WHO no date). This is evident in high rates of unemployment and underemployment among this cohort (National Federation of the Blind no date). Recent years have brought greater emphasis on making websites more accessible for this group (Barreto/Hollier 2019). It is against this backdrop that this study evaluates three popular online dictionary websites in terms of their accessibility for people with visual impairments. This is undertaken with the aim of highlighting good practice and offering suggestions for improvement. The goal of making online dictionaries more accessible for people with visual impairments needs no further justification. However, its importance is clearer still when one considers that in addition to their role in recording language and resolving doubts, dictionaries also play a role in marking out socio-cultural identities (Lew 2014). From this perspective, accessible dictionaries may play a role in promoting the social inclusion of this oft-marginalised group.

The dictionary websites evaluated are the *collinsdictionary.com* portal (CDP); *dle.rae.es*, the online version of the *Real Academic Española: Diccionario de la lengua Española* (DLE); and the *merriam-webster.com Dictionary* (MWD). These dictionaries frequently rank among the most visited dictionary websites.

	Language	**Location**	**Type**	**Multimedia**
CPD	English	UK/USA	General, bilingual*, MLD, specialist	Audio, photographs, video
DLE	Spanish	Spain	General	None
MWD	English	USA	General	Audio

Table 1: Characteristics of the dictionary websites sampled
(* The bilingual functions are not included in the evaluation)

As Table 1 shows, the sample covers several monolingual dictionary types and contains dictionary websites with features which may influence accessibility. The relatively high rate of employment and standard of living of people with visual impairments in Spain (Suther-

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

land-Meier 2015) and the inclusion of a Spanish dictionary published by the *Real Academic Española*, an organisation supported by the Spanish state, permits speculation about how differences in website accessibility might relate to the degree of social inclusion of people with visual impairments. Websites with a range of multimedia enhancements are included as these features may influence website accessibility.

In an attempt to replicate the way users typically navigate dictionary websites, this study evaluates a structured sample of webpages. For each website the sample includes the landing page, pages pertaining to a short monosemic entry, a long polysemic entry, an entry with an image, and an entry with a video. In line with common practice in website accessibility evaluations, a two-stage method is employed. First, the extent to which the websites comply with the *Web Content Accessibility Guidelines* (W3C no date) is evaluated using three automatic tools (Abascal/Arrue/Valencia 2019). This gives a quantitative summary of those guidelines which have been met and those which have not. The second manual analysis stage involves navigating through the dictionary website as a user might. This is done three times, once without assistive technologies, once with screen reading software which reads the screen using a computer synthesised voice, and once with magnifying software.

The websites examined present accessibility challenges for people with visual impairments. Many of these challenges are not related to the core lexicographic data, but ancillary elements such as word-of-the-day features, mailing list sign-up forms, advertisements, and other promotional materials which lack labels for screen readers and other assistive technologies. As far as the dictionary entries are concerned, a lack of contrast between foreground and background colours is noted for microstructural elements such as usage and world class labels. Fortunately, many of these issues can be easily remedied.

It is hoped that these findings will not only help people with visual impairments gain improved access to online dictionaries, but also provide developers of digital dictionaries with practical advice for making resources more accessible to people with visual impairments.

References

Abascal, J./Arrue, M./Valencia, X. (2019): Tools for web accessibility evaluation. In: Yesilada, Y./Harper, S. (eds.): Web accessibility: a foundation for research. 2nd edition. London, pp. 479–503.

Barreto, A./Hollier, S. (2019): Visual disabilities. In: Yesilada, Y./Harper, S. (eds.): Web accessibility: a foundation for research. 2nd edition. London, pp. 3–17.

DeCesaris, J. (2012): On the nature of signposts. In: Fjeld, R. V./Torjusen, J. M. (eds.): Proceedings of the 15th EURALEX International Congress. Oslo, pp. 532–540.

Dziemianko, A. (2020): Smart advertising and online dictionary usefulness. In: International Journal of Lexicography 33 (4), pp. 377–403.

Laufer, B./Kimmel, M. (1997): Bilingualised dictionaries: how learners really use them. In: System 25 (3), pp. 361–369.

Lew, R. (2014): Dictionaries and their users. In: Hanks, P./de Schryver, G.-M. (eds.): International handbook of modern lexis and lexicography. Berlin/Heidelberg, pp. 1–9.

Lew, R./Grzelak, M./Leszkowicz, M. (2013): How dictionary users choose senses in bilingual dictionary entries: An eye-tracking study. In: Lexikos 23 (1), pp. 228–254.

National Federation of the Blind (no date): Blindness statistics. National Federation of the Blind. https://nfb.org/resources/blindness-statistics (last access: 25-03-2022).

Sutherland-Meier, M. (2015): Toward a history of the blind in Spain. In: Disability Studies Quarterly 35 (4).

W3C (no date): Web content accessibility guidelines (WCAG) overview. Web Accessibility Initiative (WAI). https://www.w3.org/WAI/standards-guidelines/wcag/ (last access: 15-11-2021).

WHO (no date): WHO | 10 facts about blindness and visual impairment. https://www.who.int/features/factfiles/blindness/blindness_facts/en/ (last access: 15-11-2021).

Contact information

Geraint Paul Rees
Universitat Rovira i Virgili
geraintpaul.rees@urv.cat

Acknowledgements

This research was made possible by the kind support of the BALEAP Funding Stream 2021: *Social Justice in EAP*.

Silga Sviķe

SURVEY ANALYSIS OF DICTIONARY-USING SKILLS AND HABITS AMONG TRANSLATION STUDENTS

Abstract The paper presents the results of empirical research conducted with students from the Faculty of Translation studies of Ventspils University of Applied Sciences (VUAS) in Latvia. The study investigates the habits and practices concerning the use of dictionaries on the part of translation students, as well as types of dictionaries used, frequency of use, etc. The study also presents an insight into the evaluation of the usefulness of dictionaries by Latvian students. The research describes the advantages and disadvantages of dictionaries used by the respondents, the importance of the preface and the explanation of the terms and abbreviations used in dictionaries. The research conducted, as well as the insights, results and recommendations presented, will be relevant for the lexicographic community, as it reflects the experience of one Latvian University to improve the teaching of dictionary use and lexicographic culture in this country and to complement dictionary use research with the Latvian experience.

Keywords Dictionaries; dictionary use; translators; translation studies; translation tools, survey

Contact information

Silga Sviķe
Ventspils University of Applied Sciences
silga.svike@venta.lv

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Carole Tiberius / Jelena Kallas / Svetla Koeva / Margit Langemets / Iztok Kosem

AN INSIGHT INTO LEXICOGRAPHIC PRACTICES IN EUROPE

Results of the extended ELEXIS survey on user needs

Abstract the paper presents the results of a survey on lexicographic practices and lexicographers' needs across Europe that was conducted in the context of the Horizon 2020 project European Lexicographic Infrastructure (ELEXIS) among the observer institutions of the project. The survey is a revised and upgraded version of the survey which was originally conducted among ELEXIS lexicographic partner institutions in 2018 (Kallas et al. 2019). The main goal of this new survey was to complement the data from the ELEXIS lexicographic partner institutions in order to get a more complete picture of lexicographic practices both for born-digital and retro-digitised resources in Europe. The results offer a detailed insight into many aspects of the lexicographic process at European institutions, such as funding, training, staff, lexicographic expertise, software and tools. In addition, the survey reflects on current trends in lexicography and reveals what institutions see as the most important emerging trends that will affect lexicography in the short-term and long-term future. Overall, the results provide valuable input informing the development of tools, resources, guidelines and training materials within ELEXIS.

Keywords E-lexicography; lexicographic practices; lexicographers' needs; survey; ELEXIS

Reference

Kallas, J./Koeva, S./Kosem, I./Langemets, M./Tiberius, C. (2019): ELEXIS deliverable 1.1 Lexicographic Practices in Europe: A Survey of User Needs. https://elex.is/wp-content/uploads/2020/06/Revised-ELEXIS_D1.1_Lexicographic_Practices_in_Europe_A_Survey_of_User_Needs.pdf (last access: 25-03-2022).

Contact information

Carole Tiberius
Instituut voor de Nederlandse Taal
carole.tiberius@ivdnt.org

Jelena Kallas
Institute of the Estonian Language
Jelena.kallas@eki.ee

Svetla Koeva
Institute for Bulgarian Language
svetla@dcl.bas.bg

Margit Langements
Institute of the Estonian Language
margit.langemets@eki.ee

Iztok Kosem
Jožef Stefan Institute
iztok.kosem@ijs.si

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Agnes Wigestrand Hoftun

CONSULTATION BEHAVIOR IN L1 ERROR CORRECTION

An exploratory study on the use of online resources in the Norwegian context

Abstract This think-aloud study charts the use of online resources by five final-year MA students in a Nordic languages and literacy program based on the analysis of screen and audio recordings of an error-correction task. The article briefly presents some linguistic features of Norwegian Nynorsk that are not common in the context of other European languages, that is, norm optionality with regards to inflection and spelling. While performing the task, the participants were allowed to use all digital aids. This article examines their resource consultation behavior, and it makes use of Laporte/Gilquin's (2018) annotation protocol. The following research questions are posed: What online resources are used by the students? What characterizes the use? Are online resources helpful? This study provides new insights into an as yet little explored topic within the Norwegian context. The findings demonstrate that the participants relied heavily on the official monolingual dictionary Nynorskordboka. Indeed, the dictionary was helpful in the vast majority of the searches, either resulting in error improvement or the validation of a word; that is, many of the searches considered correct words. The findings suggest severe norm insecurity and emphasize the need to improve norm knowledge and metalinguistic knowledge as prerequisites for better utilization of aids. It is also suggested to include necessary information on norm optionality and other commonly queried issues in the dictionary architecture.

Keywords Consultation behavior; L1 error correction; dictionary use; online resources; Norwegian Nynorsk

Reference

Laporte, S./Gilquin, G. (2018): Annotating the use of online writing resources in a video corpus of written process data in ELAN. Annotation manual version 1.1. http://hdl.handle.net/2078.1/204351.

Contact information

Agnes Wigestrand Hoftun
University of Stavanger, Norway
agnes.w.hoftun@uis.no

Dictionary Projects

XX EURALEX INTERNATIONAL CONGRESS
DICTIONARIES AND SOCIETY
12-16 July 2022, Mannheim, Germany

Hauke Bartels

THE LONG ROAD TO A HISTORICAL DICTIONARY OF LOWER SORBIAN

Towards a lexical information system

Abstract The Sorbian Institute has been taking preparatory steps for a historical-documentary vocabulary information system for Lower Sorbian for about 10 years. To this end, the entire extant written material (16th–21st centuries) of this strongly endangered European minority language is to be systematically evaluated. An attempt made a few years ago to organise and finance the project as a long-term scientific project was not successful in the end. Therefore, it can only be advanced step by step and via some detours. The article informs about the interim status of the project, especially with respect to the creation of a reliable database.

Keywords Lower Sorbian; historical lexicography; minority language; e-lexicography; lexical information system; text corpus; Sorbian institute; language portal

Contact information

Hauke Bartels
Sorbisches Institut
hauke.bartels@serbski-institut.de

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Hanno Biber

"BLOODY WORD RIPPING" – PRACTICAL AND THEORETICAL PROSPECTS OF A CORPUS-BASED LEXICOGRAPHIC EXPLORATION OF THE TEXTS BY THOMAS BERNHARD

Keywords Digital literary studies; text lexicography; author-based dictionaries; corpus linguistics

In the following abstract a research concept will be presented, in which a text corpus of the literary works by Thomas Bernhard (1931–1989) will be used for creating a lexicographic description of the lexical units in his texts to conceptualise the practical and theoretical implications and the prospects for creating a digital author-based dictionary. The author is renowned for his literary work, his novels, novellas, short stories, poems, letters, interviews, micro dramas, dramas, whether tragedies or comedies, which have been translated from the German original into many different languages, and are to be viewed in the historical context of coming to terms with the persistent presence of the Nazi-legacy in Germany and Austria after 1945. The lexicographic description of his language can be used as a paradigm for the description of linguistic elements in actual use in literature. And, it is possible to regard the digital text corpus of his works as an example for a research exploration of how to apply a lexicographic concept based upon corpus-generated data for the purpose of dictionary making and creating a special author-based dictionary out of a text corpus. Such a text corpus of all his works will have to be established as a research resource and taken as a source for the purpose of making an author-based dictionary by giving lexicographic descriptions of the lexical entities, single words or multi word units, used in his texts (Bernhard 2003–2015). The implications of the lexicographic endeavour will have to be assessed to consider the prospects for creating a digital author-based dictionary and its methodological background. The dimensions and constraints of author dictionaries have been laid out (cf. Karpova 2011), as have some of the consequences of this approach for the lexicography of literature (cf. Lobenstein-Reichmann 2016). And the possible options for this method are described in the research field of the intersection of corpus linguistics and literary analysis (cf. Fischer-Starcke 2010) and for the area of "corpus stylistics" (cf. Mahlberg 2013). Following already existing research paths, a projected dictionary of a wider scope would require a concise digital lexicographic documentation of all the lexical entities used by one author and thus provide a large resource also for the research on individual language use, whereby a framework in accordance with "lexical analysis" (cf. Hanks 2013) would be suitable. Such a project can be seen as an actual and vivid process, which the quotation and opening motto refers to by "bloody word ripping" that reads in the full translation from its original version of "Wörterherausreißen blutig" (Bernhard 2003, p. 146) as: "He rips the words out of himself as from a swamp. This violent ripping out of words leaves him dripping with blood." (Bernhard 2006, p. 126). This passage from a novel about the personal individuality of the artist can also be read as an answer to the question posed by the fictional narrator at the beginning of the chapter: "Can it still be described as language?" (Bernhard 2006, p. 126). For example, an idiomatic dictionary has to be considered prototypical as a model how to explicate the use of multi word units in a text dictionary (cf. Welzig 1999) built on the basis of and

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

leading to a large text corpus (cf. Biber 2007), in which corpus linguistic searches are possible for lexical explorations (cf. Biber 2015). The descriptions of the lexical entries in a dictionary of literary language use can be pictured as combining literary studies with linguistics, for which the context of each lexical entity is crucial and would provide knowledge about the actual linguistic performance in a literary text. The lexicographic description of such a proposed author-based dictionary would provide data useful also for translators, interpreters, dramatic or narrative performers of the texts of an author famous for his ironic style and narrative compositions, in which author-specific creative compounds, collocations and other lexical formations can be found and be regarded as possible answers given to this research question by means of a digital text corpus together with digital lexicography thereby creating an author-based dictionary of Thomas Bernhard.

References

Bernhard, T. (2006): Frost. Translated from the German by Hofmann, M. New York.

Bernhard, T. (2003): Werke. Vol. 1: Frost. Edited by Huber, M./Schmidt-Dengler, W. Frankfurt a. M.

Bernhard, T. (2003–2015): Werke. Edited by Huber, M./Schmidt-Dengler, W. 22 volumes. Berlin/Frankfurt a. M.

Biber, H. et al. (eds.) (2007): AAC-Austrian Academy Corpus: Die Fackel.
https://fackel.oeaw.ac.at (last access: 21-03-2022).

Biber, H. (2015): AAC-Fackel. Das Beispiel einer digitalen Musteredition. In: Baum, C./Stäcker, T. (eds.): Grenzen und Möglichkeiten der Digital Humanities. Sonderband 1 (2015) der Zeitschrift für digitale Geisteswissenschaften.
https://www.zfdg.de/sb001_019 (last access: 21-03-2022),
DOI: 10.17175/sb001_019.

Fischer-Starcke, B. (2010): Corpus linguistics in literary analysis: Jane Austen and her contemporaries. London.

Hanks, P. (2013): Lexical analysis: norms and exploitations. Cambridge, MA.

Karpova, O. M. (2011): English author dictionaries (the XVIth – the XXIst cc.). Newcastle upon Tyne.

Lobenstein-Reichmann, A. (2016): Historischer Wortschatz: Text- und Autorenwörterbücher. In: Lobenstein-Reichmann, A./Müller, P. O. (eds.): Historische Lexikographie zwischen Tradition und Innovation. Berlin, pp. 77–100.

Mahlberg, M. (2013): Corpus stylistics and Dickens's fiction. London.

Welzig, W. et al. (eds.) (1999): Wörterbuch der Redensarten zu der von Karl Kraus 1899–1936 herausgegebenen Zeitschrift 'Die Fackel'. Vienna.

Contact information

Hanno Biber
Austrian Academy of Sciences
Hanno.Biber@oeaw.ac.at

Isidora Despotidou/Zoe Gavriilidou

AN ONLINE SCHOOL DICTIONARY IN GREEK SIGN LANGUAGE FOR SENIOR ELEMENTARY PUPILS

Keywords Specialized lexicography; bilingual lexicography; sign language lexicography; Greek Sign Language

Sign language lexicography is a comparatively new special field of lexicography which still remains unexplored due to the nature and characteristics of sign languages themselves (McKee/Vale 2017). As a result, the design and creation of sign language dictionaries is not very common mainly due to linguistic, financial and social reasons (Vacalopoulou 2020). Not many dictionaries have been created after the compilation of the *Dictionary of American Sign Language on Linguistic Principles* (Stokoe/Casterline/Croneberg 1965) which is considered to be the first such dictionary, while such resources are extremely limited for Greek Sign Language (GSL). This causes accessibility issues for children having GSL as a mother tongue (L1).

Our aim is to offer the organization plan and the dictionary conceptualization plan of a specialized online bilingual (Greek-GSL) school dictionary in the GSL targeted to senior elementary children belonging to the deaf community in Greece. Our genuine purpose stems from the need to address the lack of specialized resources for children with GSL as L1 Greece.

The dictionary in GSL can be found at the following link http://synmorphose.gr/index.php/en/ and includes so far 200 videotaped entries; the headword selection was based on the paper school dictionary entitled Το Λεξικό μας (Our dictionary) which is available for pupils aged 9–12 attending schools in Greece.

The microstructure of the dictionary includes the videotaped definitions of the words in Greek Sign Language followed by a list of synonyms and antonyms for each entry. We opted for a representation of signs using word glosses (Miller 2006). We offer the basic characteristics of what a Sign Language is with special reference to GSL. Then we thoroughly describe the compilation procedure, the lexicographic functions of the dictionary and we make a detailed demonstration of the microstructure and entry characteristics of the dictionary. Finally we highlight the main challenges encountered during dictionary compilation. The availability of more signed resources like the one presented here will result in equality and inclusion of the deaf community in Greek education.

References

McKee, R./Vale, M. (2017): Sign language lexicography. In: Hanks, P./de Schryver, G.-M. (eds.): International handbook of modern lexis and lexicography. https://www.semanticscholar.org/paper/Sign-language-lexicography-Mckee-Vale/e3d1d3a158ec1fd1652fa6462129616fd1baf86a (last access: 27-04-2020).

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Miller, C. (2006): Sign language: transcription, notation, and writing. In: Brown, K. (ed.): Encyclopedia of language and linguistics. 2nd edition. Amsterdam, pp. 328–338. https://www.sciencedirect.com/science/article/pii/B008044854200242X (last access: 27-04-2022).

Stokoe, W. C./Casterline, D. C./Croneberg, C. G. (1965): Dictionary of American Sign Language on linguistic principles. Washington, DC.

Vacalopoulou, A. (2020): Sign language corpora and dictionaries: a multidimensional challenge. In: Gavriilidou, Z./Mitsiaki, M./Fliatouras, A. (eds.): XIX Euralex Proceedings. Lexicography for Inclusion. Volume 1, pp. 427–434.

Contact information

Zoe Gavriilidou
Democritus University of Thrace
zoegab@otenet.gr

Isidora Despotidou
Democritus University of Trace
Isidora.despotidou@gmail.com

Carolina Flinz/Laura Giacomini/Weronika Szemińska

TERMIKNOWLEDGE: EIN EINBLICK IN DIE DATENBESCHAFFUNG UND DATENAUFBEREITUNG EINES ONLINE-FACHWÖRTERBUCHS ZUM THEMA COVID-19

Keywords Lexicographic process; LSP dictionary; online dictionary

Die *TermiKnowledge Multilingual Knowledge Base* ist ein Projekt zur Erstellung eines mehrsprachigen korpusbasierten Online-Fachwörterbuchs zum COVID-19-Diskurs, welches sich aktuell im Aufbau befindet (vgl. Storrer 2001).

Vier Universitäten (Universität Warschau, Karls-Universität in Prag, Universität Heidelberg, Universität Mailand) haben sich innerhalb eines 4EU+ Educational Project mit dem Titel „Knowledge through Terminology – From Multilingual Data to Domain-specific Knowledge via Terminological Resources“, abgekürzt *TermiKnowledge*, das Ziel gesetzt, Studierenden von BA- und MA-Studiengängen (Sprach- und Translationswissenschaft sowie Terminologie) Basiskompetenzen zur korpusgestützten und korpusbasierten Fachlexikographie (vgl. Lemnitzer/Zinsmeister 2015, S. 34–37) zu vermitteln, die sie dann zur Erstellung des anvisierten Internetfachwörterbuches (zur Internetlexikographie vgl. Klosa/Müller-Spitzer 2016) unter Betreuung von Fachexpertinnen und Fachexperten einsetzen können.[1] Das Wörterbuch richtet sich an Studierende der Sprach- und Translationswissenschaft, der Terminologie sowie an sonstige Interessierte an diesem Fachdiskurs.

Die Phasen des lexikographischen Prozesses sind von einem Zusammenfließen und Überlappen der Phasen wie in einem Kreis charakterisiert (vgl. Klosa 2013): In der Vorbereitungsphase wurde das Wörterbuch inhaltlich konzipiert und computertechnische Möglichkeiten wurden exploriert. Nach Abwägung unterschiedlicher Möglichkeiten wurde ein Mediawiki-System ausgewählt (vgl. dazu auch Flinz 2018).

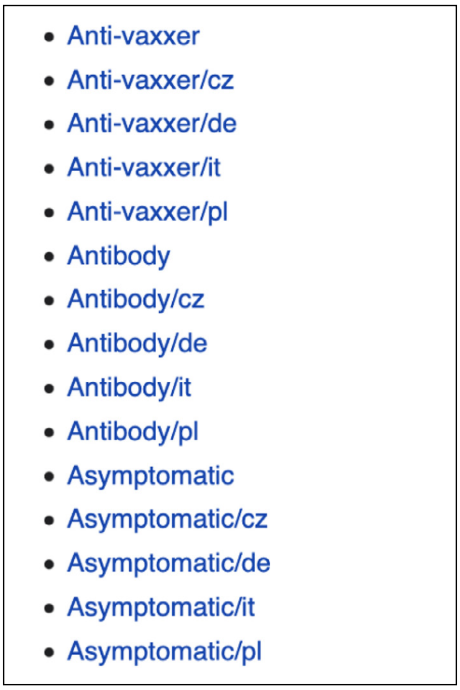
In der Phase der Datenbeschaffung wurden die Quellen für die Wörterbuchbasis zusammengestellt. Es handelte sich um *ad hoc* erstellte Vergleichskorpora in deutscher, englischer, italienischer, polnischer und tschechischer Sprache, die aus unterschiedlichen Textsorten kompiliert wurden: normativen Texten (u.a. rechtlichen und medizinischen Leitlinien), wissenschaftlichen Texten (Aufsätzen), Presse-Texten aus allgemeinen Zeitungen sowie Online-Kommentaren zu Presse-Texten. Die Korpora wurden mit der Korpusplattform Sketch Engine aufgebaut und analysiert (vgl. Kilgarriff et al. 2004).

[1] Das Team bestand aus 7 Dozierenden und mehr als 30 Studierenden (7–8 pro Universität und 1 TutorIn pro Universität). Die Arbeit fand online statt; der Kurs (6 cfu) wurde dann jedoch im Frühjahr 2022 mit einem abschließenden Präsenzmeeting in Heidelberg beendet. Die Studierenden haben nach einer theoretischen Einführung in unterschiedlichen Gruppenkonstellationen gearbeitet. In der ersten Arbeitsphase (Erstellung der Korpora und der Lemmalisten) wurden die Gruppen je nach Art von Korpora zusammengestellt; in der zweiten Phase (Herausfilterung der mikrostrukturellen Angaben) erfolgte die Zusammenführung der Studierenden in Gruppen nach Sprachen.

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

In der Phase der Datenaufbereitung wurden in jeder Sprache die provisorischen Lemmakandidatenlisten erstellt. Diese sind aus dem Zusammenspiel von zwei Arbeitsschritten entstanden: Extrahierung von Frequenzlisten (absoluter und relativer Häufigkeit) aus den unterschiedlichen Fokuskorpora und von Keywordlisten (einzelner Keywords und Mehrwortverbindungen) auf der Basis der in Sketch Engine integrierten Referenzkorpora (Webkorpora). Die provisorischen Stichwortlisten (ca. 57 Lemmata) wurden anschließend untereinander verglichen, um Äquivalenzbeziehungen zu identifizieren (vgl. Flinz/Perkuhn 2018; Szemińska/Więch 2019): z.B. *infection rate* (en) – *Infektionsrate* (de) – *tasso di infezione* (it) – *šíření nákazy* (cz) – *wskaźnik zakażeń* (pl).

Die definitive Lemmaliste (38 Stichwörter), die Entsprechungen in allen involvierten Sprachen hat, wurde dann online gestellt (vgl. Abb. 1):[2]

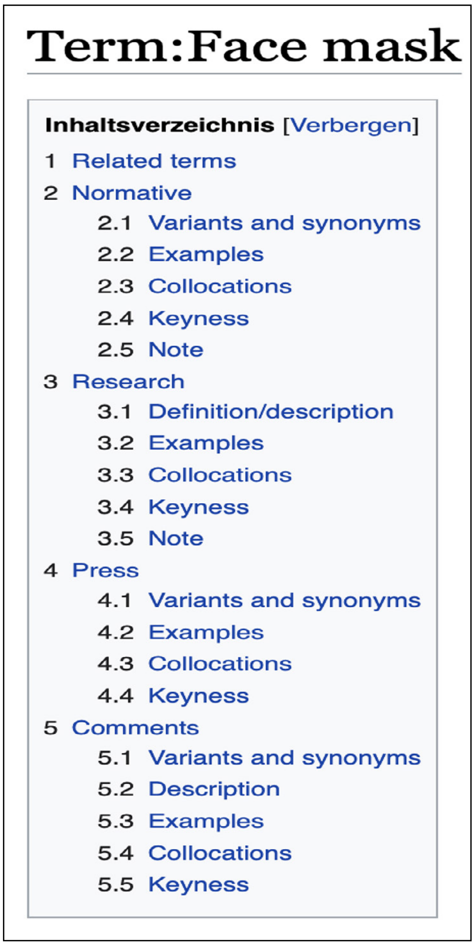


- Anti-vaxxer
- Anti-vaxxer/cz
- Anti-vaxxer/de
- Anti-vaxxer/it
- Anti-vaxxer/pl
- Antibody
- Antibody/cz
- Antibody/de
- Antibody/it
- Antibody/pl
- Asymptomatic
- Asymptomatic/cz
- Asymptomatic/de
- Asymptomatic/it
- Asymptomatic/pl

Abb. 1: Screenshot der Lemmaliste

Anschließend wurden die Listen validiert und die lexikographischen Daten unter Berücksichtigung der mikrostrukturellen Angaben (u. a. des Formkommentars und des semantischen Kommentars) in einer Datenbank eingetragen. Für jedes Stichwort wurden folgende Informationen aus den Korpora herausgefiltert: semantisch nahe Wörter (auch auf der Basis des Kookkurrenzprofils), orthographische Varianten, Synonyme, Definitionen, Beispiele, Kollokationen und Angaben zur *Keyness*. Wenn möglich wurden die Angaben nach Korpustyp unterschieden, um auch die intralinguale Variation innerhalb unterschiedlicher Textsorten abzubilden (Abb. 2).

[2] Vgl. https://terminology.mimuw.edu.pl/index.php?title=Main_Page. Es gab auch Termini, die keine Entsprechung in einer Sprache oder in einer bestimmten Domäne hatten: In diesem Fall wurde ein Hinweis dazu gegeben.



Term:Face mask

Inhaltsverzeichnis [Verbergen]

1 Related terms
2 Normative
  2.1 Variants and synonyms
  2.2 Examples
  2.3 Collocations
  2.4 Keyness
  2.5 Note
3 Research
  3.1 Definition/description
  3.2 Examples
  3.3 Collocations
  3.4 Keyness
  3.5 Note
4 Press
  4.1 Variants and synonyms
  4.2 Examples
  4.3 Collocations
  4.4 Keyness
5 Comments
  5.1 Variants and synonyms
  5.2 Description
  5.3 Examples
  5.4 Collocations
  5.5 Keyness

Abb. 2: Screenshot der Struktur des Eintrags face mask

Das mehrsprachige Wörterbuch wurde bereits online gestellt, auch wenn die Umtexte noch im Aufbau sind.

Die Studierenden haben abschließend die eigene praktische lexikographische Arbeit sowie den theoretischen Hintergrund reflektiert und bewertet: Vor- und Nachteile wurden diskutiert. Dieser letzte Arbeitsschritt war für das Team von großer Relevanz, da Konsequenzen für das bereits gestartete neue *Termiknowledge*-Projekt gezogen werden konnten.

Literatur

Flinz C. (2018): Der lexikographische Prozess bei Tourlex (ein deutsch-italienisches Fachwörterbuch zur Tourismussprache) für italienische DaF-Lerner. In: Klosa, A./Storrer, A./Taborek, J. (Hg.): Internetlexikographie und Sprachvermittlung. Jahrbuch Lexicographica. Berlin, S. 9–36.

Flinz, C./Perkuhn, R. (2018): Wortschatz und Kollokationen in ‚Allgemeine Reisebedingungen'. Eine intralinguale und interlinguale Studie. In: Krek, S. et al. (Hg.): Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Context. Ljubljana, S. 959–967. https://euralex.org/publications/wortschatz-und-kollokationen-in-allgemeine-reisebedingungen-eine-intralinguale-und-interlinguale-studie-zum-fachsprachlich-lexikographischen-projekt-tourlex/ (last access: 25-03-2022).

Kilgariff, A. et al. (2004): The Sketch Engine. In: Williams, G./Vessier, S. (eds.): Proceedings of the 11th Euralex International Congress, Lorient, France, July 6–10. Bd. 1. Lorient, S. 105–115.

Klosa, A. (2013): The lexicographical process (with special focus on online dictionaries). In: Gouws, R. et al. (eds.): Dictionaries. An international encyclopedia of lexicography. Supplementary volume: Recent developments with focus on electronic and computational lexicography. Berlin/Boston, S. 517–524.

Klosa, A./Müller-Spitzer, C. (Hg.) (2016): Internetlexikographie. Ein Kompendium. Berlin/New York.

Lemnitzer, L./Zinsmeister, H. (2015): Korpuslinguistik. Eine Einführung. Tübingen.

Storrer, A. (2001): Digitale Wörterbücher als Hypertexte: Zur Nutzung des Hypertextkonzepts in der Lexikographie. In: Lemberg, I./Schröder, B./Storrer, A. (Hg.): Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher. Tübingen, S. 53–69.

Szemińska, W./Więch, A. (2019): eLex2. A prototype electronic dictionary application for legal translators. In: Terminology 25 (2), S. 198–221.

Kontaktinformationen

Carolina Flinz
Università degli Studi di Milano
carolina.flinz@unimi.it

Laura Giacomini
Universität Heidelberg
laura.giacomini@iued.uni-heidelberg.de

Weronika Szemińska
Universität Warschau
w.szeminska@uw.edu.pl

Polona Gantar/Simon Krek

CREATING THE LEXICON OF MULTI-WORD EXPRESSIONS FOR SLOVENE

Methodology and structure

Abstract This paper describes a method for automatic identification of sentences in the Gigafida corpus containing multi-word expressions (MWEs) from the list of 5,242 phraseological units, which was developed on the basis of several existing open-access lexical resources for Slovene. The method is based on a definition of MWEs which includes information on two levels of corpus annotation: syntax (dependency parsing) and morphology (POS tagging), together with some additional statistical parameters. The resulting lexicon contains 12,358 sentences containing MWEs extracted from the corpus. The extracted sentences were analysed from the lexicographic point of view with the aim of establishing canonical forms of MWEs and semantic relations between them in terms of variation, synonymy, and antonymy.

Keywords Identification of multi-word expressions; multi-word expressions lexicon; canonical form of multi-word expressions; Slovene; digital dictionary database

Contact Information

Polona Gantar
University of Ljubljana
apolonija.gantar@guest.arnes.si

Simon Krek
Jožef Stefan Institute
simon.krek@guest.arnes.si

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Zoe Gavriilidou/Apostolos Garoufos

THE LEXICOGRAPHIC PROTOCOL OF Mikaela_Lex: A FREE ONLINE SCHOOL DICTIONARY OF GREEK ACCESSIBLE FOR VISUALLY-IMPAIRED SENIOR ELEMENTARY CHILDREN

Abstract The purpose of this paper is to present the lexicographic protocol and to report on the progress of compilation of Mikaela_Lex, which is a Greek, free online monolingual school dictionary for upper elementary students with visual impairments including 4,000 lemmata. The dictionary is equipped with new digital tools, such as the "Braille-system-keyboard, a "speech-to-text" tool, a "text-to-speech" tool and also a qwerty accessibility for visually non-impaired students.

Keywords Inclusive lexicography; blindness; visually impaired children; pedagogical lexicography Greek

Contact information

Gavriilidou Zoe
Democritus University of Thrace
zoegab@otenet.gr

Garoufos Apostolos
Democritus University of Thrace
agaroufo@helit.duth.gr

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Vanessa González Ribao

FACHLEXIKOGRAFIE IN DIGITALEM ZEITALTER

Ein metalexikografisches Forschungsprojekt

Abstract This paper presents the methodology of a research project on the use of specialised German dictionaries. A mixed-methods research approach will help to answer the following main questions, concerning the lexicographic presentation of the data on the one hand and the data collection on the other hand: How do different systems of data organization and presentation affect the likelihood that users will correctly find and select the data they look up? And does the probability of success increase if users are familiar with the system? Which advantages and disadvantages do lexicographers and specialised languages experts see in using quantitative methods to extract terms? And are these methods accepted and considered reliable by the user community?

Keywords Specialised lexicography for the German language; user research; mixed methods

Contact information

Vanessa González Ribao
Postdoc-Stipendiatin der Fritz-Thyssen-Stiftung
vanessina_gr@hotmail.com

Peter Meyer

LEHNWORTPORTAL DEUTSCH: A NEW ARCHITECTURE FOR RESOURCES ON LEXICAL BORROWINGS

Abstract This paper presents the *Lehnwortportal Deutsch*, a new, freely accessible publication platform for resources on German lexical borrowings in other languages, to be launched in the second half of 2022. The system will host digital-native sources as well as existing, digitized paper dictionaries on loanwords, initially for some 15 recipient languages. All resources remain accessible as individual standalone dictionaries; in addition, data on words (etyma, loanwords etc.) together with their senses and relations to each other is represented as a cross-resource network in a graph database, with careful distinction between information present in the original sources and the curated portal network data resulting from matching and merging information on, e. g., lexical units appearing in multiple dictionaries. Special tooling is available for manually creating graphs from dictionary entries during digitization and for editing and augmenting the graph database. The user interface allows users to browse individual dictionaries, navigate through the underlying graph and 'click together' complex queries on borrowing constellations in the graph in an intuitive way. The web application will be available as open source.

Keywords Multilingual lexicography; lexical borrowings; graph database

Contact information

Peter Meyer
Leibniz-Institut für Deutsche Sprache
meyer@ids-mannheim.de

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Iryna Ostapova/Volodymyr Shyrokov/Yevhen Kupriianov/
Mykyta Yablochkov

ETYMOLOGICAL DICTIONARY IN DIGITAL ENVIRONMENT

Abstract The digital environment represents a qualitatively new level of service for research work with linguistic information presented in dictionary form. And first of all, this applies to index systems. By dictionary indexing we mean a set of formalized rules and procedures, on the basis of which it is possible to obtain information about certain linguistic facts recorded in the dictionary. These rules are implemented in the form of user interfaces. However, one should take into account the fact that the effectiveness of automatic construction of index schemes for a digital dictionary is possible only in a sufficiently formalized environment. This article describes the method and technology of indexing the Etymological Dictionary of the Ukrainian Language (EDUL). For the language indexing of the dictionary, a special computer instrumental system (VLL – virtual lexicographic laboratory) was developed, and adapted to the structure of the EDUL and focused on the creation of indexes in automatic mode. The digital implementation of the EDUL made it possible to access the entire corpus of the dictionary text regardless of the time of publication of the corresponding volume and opened up opportunities for various digital interpretations of etymological information.

Keywords Ukrainian language; etymology; formal model; lexicographical system; etymological data base; index

Contact information

Iryna Ostapova
Ukrainian Lingua-Information Fund of National Academy of Sciences of Ukraine
irinaostapova@gmail.com

Volodymyr Shyrokov
Ukrainian Lingua-Information Fund of National Academy of Sciences of Ukraine
Vshirokov48@gmail.com

Yevhen Kupriianov
National Technical University "Kharkiv Polytechnic Institute"
eugeniokuprianov@gmail.com

Mykyta Yablochkov
Ukrainian Lingua-Information Fund of National Academy of Sciences of Ukraine
gezartos@gmail.com

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Anna Pavlova

MEHRSPRACHIGE DATENBANK DER PHRASEM-KONSTRUKTIONEN

Abstract The paper describes an online German-Russian database for phraseological constructions (PhC), or syntactic idioms. It is a linguistic phenomenon representing a stable multi-word form that usually contains some auxiliary words ("anchors") and partially opens up empty spaces ("slots") which are filled directly in spoken language by various lexemes or combinations of lexemes ("fillers", or "slot fillers"). Linguists from several German institutions are currently working on the database. The PhCs selected for the database have to meet special criteria. The database is a manual that combines scientific descriptions, a thesaurus and a bilingual dictionary. The database is designed as an active aid for text production in the respective foreign language; it is also a manual for language researchers and for translators. Apart from that, it can serve as a basis for extensions for other language pairs. The aim of the project is to record and to describe 300 PhC before the database is published. Our objective is to enable foreign language learners to use the syntactic idioms correctly in the texts they produce rather than create a big-sized database. The paper describes some issues related to the creation of the database, namely objectives and target groups, material and methods, microstructure of the database article and some others.

Keywords Phraseme constructions; syntactic phrasemes; syntactic idioms; online database; bilingual dictionary

Contact information

Anna Pavlova
JGU Mainz, FTSK Germersheim
pavloan@uni-mainz.de

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Ralf Plate

WORD FAMILIES IN DIACHRONY

An epoch-spanning structure for the word families of Older German

Abstract The 'Word Families in Diachrony' project (WoDia), for which a funding application to the DFG is in preparation, aims to provide a database-driven online research environment that will enable processes of change in the entire historical vocabulary of German to be investigated by focusing on the changes in word families and the individual means of word formation. WoDia will embed the vocabularies of Old High German (OHG), Middle High German (MHG), Old Saxon (OS), and Middle Low German (MLG) in a database, resulting in a word-family structure for High and Low German from the beginnings up to the 15th century (for High German) and up to the 17th century (for Low German). The basis of the vocabulary is provided by reference dictionaries of the four historical varieties, whereas the word families' historical structure is based on the word-family dictionary of OHG by Jochen Splett (1992). Each lemma in the database will be assigned, where appropriate, to a word family. The individual word-formation elements and the word-formation hierarchy will be mapped in a structural formula. The etymologically corresponding lemmas and word families of the different periods/varieties of older German will be linked so that an analysis across the varieties will also be possible. The annotations of word families in the database (e.g., relating to word structure) will be supplemented by linking their lemmas to the online dictionaries and to the reference corpora of Old German (OS and OHG), MHG, and MLG.

Keywords Older German (OHG, MHG, OS, MLG); word family database; historical word formation of german

Contact Information

Ralf Plate
Akademie der Wissenschaften und der Literatur | Mainz, Mittelhochdeutsches Wörterbuch,
Arbeitsstelle an der Universität Trier
plate@uni-trier.de

Kyriaki Salveridou / Zoe Gavriilidou

COMPILATION OF AN ANCIENT GREEK – MODERN GREEK ONLINE THESAURUS FOR TEACHING PURPOSES: MICROSTRUCTURE AND MACROSTRUCTURE

Abstract To effectively design online tools and develop sophisticated programs, for the teaching of Ancient Greek language, there is a clear need for lexical resources that provide semantic links with Modern Greek. This paper proposes a microstructure for an online Ancient Greek to Modern Greek thesaurus (AMGthes) that serves educational purposes. The terms of this bilingual thesaurus have been selected from reference Ancient Greek texts, taught and studied during lower and upper secondary education in Greece. The main objective here is to build a semantic map that helps students find relevant and semantically related terms (synonyms and antonyms) in Ancient Greek, and then provide a rich set of suitable translations and definitions in Modern Greek. Designed to be an online resource, the thesaurus is being developed using web technologies, and thus will be available to every school and university student that pursues a degree in digital humanities.

Keywords Online thesaurus; bilingual thesaurus; Ancient Greek; pedagogical lexicography

Contact information

Kyriaki Salveridou
Democritus University of Thrace
ksalveri@helit.duth.gr

Zoe Gavriilidou
Democritus University of Thrace
zgabriil@helit.duth.gr

This page is part of the publication: Klosa-Kückelhaus, Annette / Engelberg, Stefan / Möhrs, Christine / Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Bilingual Dictionaries

Voula Giouli/Anna Vacalopoulou/
Nikos Sidiropoulos/Christina Flouda/Athanasios Doupas/
Gregory Stainhaouer

FROM MYTHOS TO LOGOS: A BILINGUAL THESAURUS TAILORED TO MEET USERS' NEEDS WITHIN THE ECOSYSTEM OF CULTURAL TOURISM

Abstract Thesauri have long been recognized as valuable structured resources aiding Information Retrieval systems. A thesaurus provides a precise and controlled vocabulary which serves to coordinate data indexing and retrieval. The paper presents a bilingual Greek and English specialized thesaurus that is being developed as the backbone of a platform aimed at enhancing and enriching the cultural experiences of visitors in Eastern Macedonia and Thrace, Greece. The cultural component of the intended platform comprises textual data, images of artifacts and living entities (animals and plants in the area), as well as audio and video. The thesaurus covers the domains of Archaeology, Literature, Mythology, and Travel; therefore, it can be viewed as a set of inter-linked thesauri. Where applicable, terms and names in the database are also geo-referenced.

Keywords Thesaurus; cultural heritage resources; content management platform; controlled vocabularies; bilingual resources

Contact information

Voula Giouli
Institute for Language and Speech Processing, ATHENA RC
voula@athenarc.gr

Anna Vacalopoulou
Institute for Language and Speech Processing, ATHENA RC
avacalop@athenarc.gr

Nikos Sidiropoulos
Institute for Language and Speech Processing, ATHENA RC
nsidir@athenarc.gr

Christina Flouda
Institute for Language and Speech Processing, ATHENA RC
cflouda@athenarc.gr

Athanasios Doupas
Institute for Language and Speech Processing, ATHENA RC
adoupas@athenarc.gr

Gregory Stainhaouer
Institute for Language and Speech Processing, ATHENA RC
stein@athenarc.gr

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/
Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of
Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Iztok Kosem

THE COMPREHENSIVE SLOVENIAN-HUNGARIAN DICTIONARY: BILINGUAL LEXICOGRAPHY MEETS MONOLINGUAL LEXICOGRAPHY

Keywords Bilingual dictionary; Digital Dictionary Database; Slovenian; semantic concepts; database model

The last decade has been relatively quiet for bilingual lexicography in Slovenia, with only few bilingual dictionaries being published. The methodology has not changed much either, with the usual approach including the use of an existing bilingual dictionary database, and replacing the target language with new language translations. However, even this has become a problem, particularly when compiling a dictionary of Slovene to another target language, mainly due to the fact that there is currently no dictionary, bilingual or monolingual, that would satisfactorily describe contemporary Slovene.

In recent years, various developments in Slovenian and international lexicography have made it possible to start addressing this gap in Slovenian bilingual lexicography. Among the most important were the initiatives by linguists and politicians to start the work on a new Slovenian-Hungarian dictionary, an important education and cultural resource for the two neighbouring countries. The main result of the initiatives was a project funded by the Slovenian Research Agency and the Ministry of Education, Science and Sport of the Republic of Slovenia focussed on producing a concept for a Slovenian-Hungarian dictionary (Kosem et al. 2018b). The project provided a unique opportunity to thoroughly investigate the state-of-the-art in bilingual lexicography, and design and test the best methodology for the compilation of not only this particular dictionary, but any others that may benefit from its data. After the project ended, the work on the Comprehensive Slovenian-Hungarian Dictionary began in 2018 at the Centre for Language Resources and Technologies of the University of Ljubljana (CJVT UL). The dictionary compilation is co-funded by the Slovenian Research Agency (ARRS) through CJVT-dedicated funds as part of the Network of Infrastructure Centres at the University of Ljubljana.

In parallel, CJVT UL has been compiling various other dictionaries, e.g. the Collocations Dictionary of Modern Slovene (Kosem et al. 2018a) and the Thesaurus of Modern Slovene (Arhar et al. 2018). Moreover, CJVT UL has started developing a Digital Dictionary Database, which aims to become a one for-all database for the Slovenian language, to be used both in the compilation of language resources and for natural language processing tasks. The plans for the database have been described in detail in Klemenc et al. (2017). The main aspect of the Digital Dictionary Database relevant for this paper is that it intends to use the same concepts, which translate into dictionary senses, for all the resources coming out of the database. So any resource being compiled would use the senses of existing lexical units found in the database, or contribute new senses for lexical units not yet analysed. In terms of the Comprehensive Slovenian-Hungarian Dictionary, this meant that the workflow was split into two major stages: in the first stage the (monolingual) entries are compiled com-

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

pletely from scratch, with a great amount of contextual information (collocations, examples etc.) being provided. In the second stage the translation into Hungarian is conducted, with the lexicographers-translators translating the senses and selected parts of contextual information. By using this approach, we are excluding target language bias (i. e. providing only contextual data of contrastive interest for a particular language pair) and creating a database that can be used in the compilation of language resources of other Slovenian-foreign language pairs.

In October 2021, version 1.0 of the Comprehensive Slovenian-Hungarian Dictionary (Kosem et al. 2021a) was published, containing 10,946 headwords, 33,298 translations, 15,265 collocations and other word combinations, and 2,416 examples. The Comprehensive Slovenian-Hungarian dictionary is a growing dictionary, which means that new headwords will be added in regular intervals. Importantly, the dictionary database is available under CC BY-SA 4.0 license in the CLARIN.SI repository (Kosem et al. 2021b). In our presentation, we will discuss the lexicographic workflow, the steps and recording of data relevant for other bilingual and monolingual dictionaries, and some of the challenges of adopting the same sense structure for all the dictionaries in the Digital Dictionary Database. We will also give a short demonstration of the dictionary interface, and present future plans, especially those related to improving the efficiency of lexicographic workflow.

References

Arhar Holdt, Š./Čibej, J./Dobrovoljc, K./Gantar, P./Gorjanc, V./Klemenc, B./Kosem, I./Krek, S./Laskowski, C./Robnik Šikonja, M. (2018): Thesaurus of modern Slovene: by the community for the community. In: Čibej, J./Gorjanc, V./Kosem, I./Krek, S. (eds.): Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts. Ljubljana, pp. 401–410. https://e-knjige.ff.uni-lj.si/znanstvena-zalozba/catalog/view/118/211/3000-1 (last access: 25-04-2022).

Klemenc, B./Robnik-Šikonja, M./Fürst, L./Bohak, C./Krek, S. (2017): Technological design of a state-of-the-art digital dictionary. In: Gorjanc, V./Gantar, P./Kosem, I./Krek, S. (eds): Dictionary of modern Slovene: problems and solutions. Ljubljana, pp. 10–22.

Kosem, I./Krek, S./Gantar, P./Arhar Holdt, Š./Čibej, J./Laskowski, C. (2018a): Collocations dictionary of modern Slovene. In: Čibej, J./Gorjanc, V./Kosem, I./Krek, S. (eds.): Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts. Ljubljana, pp. 989–997. https://e-knjige.ff.uni-lj.si/znanstvena-zalozba/catalog/view/118/211/3000-1 (last access: 25-04-2022).

Kosem, I./Bálint Čeh, J./Gorjanc, V./Kolláth, A./Kovács, A./Krek, S./Novak-Lukanovič, S./Rudaš, J. (2018b): Osnutek koncepta novega velikega slovensko-madžarskega slovarja. Ljubljana. https://www.cjvt.si/komass/wp-content/uploads/sites/17/2020/08/Osnutek-koncepta-VSMS-v1-1.pdf (last access: 25-04-2022).

Kosem, I. et al. (2021a): Comprehensive Slovenian-Hungarian Dictionary. Ljubljana.

Kosem, I. et al. (2021b): Comprehensive Slovenian-Hungarian Dictionary 1.0, Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1453 (last access: 25-04-2022).

Contact information

Iztok Kosem
Faculty of Arts, University of Ljubljana
iztok.kosem@cjvt.si

Anke Müller/Gabriele Langer/Felicitas Otte/Sabrina Wähl

CREATING A DICTIONARY OF A SIGNED MINORITY LANGUAGE

A bilingualized monolingual dictionary of German Sign Language

Abstract Lexicographers working with minority languages face many challenges. When the language in question is also a sign language, circumstances specific to the visual-spatial modality have to be taken into consideration as well. In this paper, we aim to show and discuss which challenges we encounter while compiling the Digitales Wörterbuch der Deutschen Gebärdensprache (DW-DGS), the first corpus-based dictionary of German Sign Language (DGS). Some parallel the challenges minority language lexicographers of spoken languages encounter, e. g. few resources, no written tradition, and having to create one dictionary for all potential user groups, while others are specific to sign languages, e. g. representation of visual-spatial language and creating access structures for the dictionary.

Abstract Sign language dictionary; minority language; bilingualized dictionary

Contact Information

Anke Müller
University of Hamburg
anke.mueller@uni-hamburg.de

Gabriele Langer
University of Hamburg
gabriele.langer@uni-hamburg.de

Felicitas Otte
University of Hamburg
Felicitas.Otte@uni-hamburg.de

Sabrina Wähl
University of Hamburg
sabrina.waehl@uni-hamburg.de

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Specialised Dictionaries

XX EURALEX INTERNATIONAL CONGRESS
DICTIONARIES AND SOCIETY
12-16 July 2022, Mannheim, Germany

Maria Aldea

BIEN ÉCRIRE, BIEN PARLER AU XIX^{E} SIÈCLE. LE RÔLE DU DICTIONNAIRE DANS L'APPRENTISSAGE DE LA LANGUE MATERNELLE: LE CAS DU ROUMAIN

Abstract In this paper, the author studies the role of the dictionary in the first language acquisition, highlighting its didactic value. Based on two Romanian lexicographical works of the 19th century, *Lexiconul de la Buda* (Buda, 1825) [*the Lexicon of Buda*] et *Vocabularu romano-francesu* (Bucarest, 1870) [*the Romanian-French Vocabulary*], the author analyses the normative information recorded in the articles in order to observe which level of language (i. e. phonetical, morphological, syntactical and lexical) is concerned. Such an approach allows to distinguish between the possible changings both at the level of the perception or at the grammatical, lexical and semantical description, i. e. the settlement of the word in the first language, and at a technical level, i. e. the making of article and of dictionary.

Keywords First language acquisition; dictionary; linguistic norm; cultural norm; Lexiconul de la Buda; Vocabularu romano-francesu

Contact information

Maria Aldea
Université Babeş-Bolyai de Cluj-Napoca
maria.aldea@ubbcluj.ro

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Harald Bichlmeier

ALMANCA TUHFE / DEUTSCHES GESCHENK (1916) oder: Wie schreibt man deutsch mit arabischen Buchstaben?

Abstract Versified dictionaries are bilingual/multilingual glossaries written in verse form to teach essential words in any foreign language. In Islamic culture, versified dictionaries were produced to teach the Arabic language to the young generations of Muslim communities not native in Arabic. In the course of time, many bilingual/multilingual versified dictionaries were written in different languages throughout the Islamic world.

The focus of this study is on the Turkish-German versified dictionary titled *Almanca Tuhfe / Deutsches Geschenk* [German Gift], published by Dr. Sherefeddin Pasha in Istanbul in 1916. This dictionary is the only dictionary in verse ever written combining these two languages. Moreover the dictionary is one of the few texts containing German words written in Arabic letters (applying Ottoman spelling conventions). The study concentrates on the way German words are spelled and tries to find out, whether Sherefeddin Pasha applied something like fixed rules to write the German lexemes.

Keywords Osmanische Wörterbücher; Osmanisch-deutsche Wörterbücher; Wörterbücher in Versen; Ajamiado-Literatur; arabische Schrift; osmanisch-türkische Orthographie

Kontaktinformationen

Harald Bichlmeier
Sächsische Akademie der Wissenschaften zu Leipzig,
Arbeitsstelle Jena: Etymologisches Wörterbuch des Althochdeutschen
harald.bichlmeier@uni-jena.de

This page is part of the publication: Klosa-Kückelhaus, Annette / Engelberg, Stefan / Möhrs, Christine / Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Walter Amaru Flores Flores/Daniel Kroiß

DAS DIGITALE FAMILIENNAMENWÖRTERBUCH DEUTSCHLANDS (DFD)

Keywords Namenforschung; Familiennamen; Sprachgeschichte; Kulturgeschichte; Sprachkontakt; Migration; Kartierung; Visualisierung; Etymologie; Digitalität; Vernetzung

Kaum ein lexikografisches Projekt gibt mehr Auskunft über die mittelalterliche Lebenswelt als das Digitale Familiennamenwörterbuch Deutschlands (DFD). Die deutschen Familiennamen reichen bis ins späte Mittelalter zurück und spiegeln als sprach- und kulturhistorische Zeugen nicht nur die Spezialisierung der Berufe, z.B. des Müllers (*Weiß-*, *Haber-*, *Hopfenmüller*; *Oel-*, *Pulvermüller*) oder die Produktvielfalt der Nahrungsmittel (*Krapf*, *Küchle*, *Pfannkuche*) und Kleidung (*Bundschuh*, *Hornschuh*), sondern auch Normvorstellungen (*Unverricht*, *Unglaube*, *Trinkaus*) sowie (volkstümliche) Bräuche (*Maibaum*, *Palmetag* 'Palmsonntag'). Die Verbreitung der Namen kann darüber hinaus Auskunft über Migrations- und Wanderungsbewegungen geben, z.B. haben der Zustrom und die Einwanderung von Neusiedlern nach dem Dreißigjährigen Krieg (v.a. Handwerker und Bauern aus Frankreich, Tirol, Schweiz usw.), zur Zeit der Industrialisierung (Bergleute aus Polen) sowie aufgrund der Vertreibungen nach dem Zweiten Weltkrieg nachweislich Spuren im Namenbestand hinterlassen. Auch moderne Migration wird erfasst, etwa in der Form unterschiedlicher Zuwanderungsmuster, die sich in der zweiten Hälfte des 20. Jahrhunderts aus den Anwerbeabkommen mit verschiedenen Ländern (v.a. Italien, Türkei) ergeben haben.

In der Mainzer Akademie der Wissenschaften und der Literatur entsteht derzeit (in Kooperation mit der Johannes Gutenberg-Universität Mainz und der Technischen Universität Darmstadt) ein zentrales, leicht zugängliches und gut durchsuchbares Wörterbuch, das diesen einmaligen Wissensspeicher für ein Fachpublikum sowie die breite Öffentlichkeit zugänglich macht. Auf Grundlage der Festnetzanschlüsse der Deutschen Telekom von 2005 wird systematisch ein Grundbestand von ca. 200.000 Familiennamen erfasst, kartiert und etymologisiert. Berücksichtigt werden Namen ab 10 Telefonanschlüssen inklusive fremdsprachiger Namen. Insgesamt gibt es in Deutschland aktuell ca. 850.000 unterschiedliche Familiennamen; im Vergleich dazu enthalten die derzeit vorhandenen Familiennamenlexika gerade einmal rund 70.000 verschiedene Namen – weniger als 10% der Gesamtmenge.

In diesem Beitrag werden zum einen die oben genannten sprach- und kulturhistorischen Aspekte fokussiert, z.B. wie die Stabilität der Familiennamen als Quelle und Wissensspeicher für sprachhistorische Forschungen genutzt werden kann, denn über Jahrhunderte hinweg hat sich die grundlegende Landschaft der Familiennamen nur wenig verändert: Nach wie vor finden wir Namentypen wie *Petersen* und *Hansen* vor allem im Norden, *Häberle* im Schwäbischen und *Mayr* im Südosten. Andererseits bieten sich die Namen dazu an, Migrations- und Wanderungsbewegungen zu beobachten. So lassen eingedeutschte Namen wie *Pospischil*, *Pospischill* und *Pospischiel* kaum noch erahnen, dass sie auf Namen tschechischen Ursprungs als Ergebnis jahrhundertelangen deutsch-böhmischen Sprachkontakts zurückgehen. Des Weiteren wird näher vorgestellt, wie das Projekt die Vorteile einer digitalen Publikation nutzt: Diese bietet erweiterte Suchmöglichkeiten, erlaubt Vernetzung mit anderen Projekten und Datenquellen und ermöglicht die Erstellung dynamischer (Geo-)Visualisierungen.

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Literatur

Digitales Familiennamenwörterbuch Deutschlands (DFD): https://www.namenforschung.net/dfd/woerterbuch/liste/ (Stand: 14.4.2022).

Kontaktinformationen

Walter Amaru Flores Flores
Akademie der Wissenschaften und der Literatur Mainz
flores@uni-mainz.de

Daniel Kroiss
Akademie der Wissenschaften und der Literatur Mainz
daniel.kroiss@uni-mainz.de

Dominika Kováříková/Michal Škrabal

THE DICTIONARY OF CZECH CORE ACADEMIC VOCABULARY

Keywords Academic word list, core academic vocabulary, modularity, Akalex

Over the past two decades, several lists of academic words and academic phrases emerged, often focused on L2 teaching or undergraduate students of academic writing classes (Coxhead 2000; Paquot 2010; Gardner/Davies 2014; Morley 2014). Most of the lists are focused on English as a lingua franca of science and research but with time, lists for other languages, such as Portuguese (Baptista et al. 2010), Swedish (Carlund et al. 2012) or Czech (Kováříková/Kovářík 2021), have been created too.

Academic word list by itself is a powerful tool for both teachers and students, but it can be further enhanced by additional information about the headword and its context as offered by a dictionary. Currently, an online dictionary of core Czech academic vocabulary is being developed, which will contain modules with information relevant to various target groups: undergraduate students (and possibly high school students), philology students, academic writers (professional and in training), and university students of Czech as a second language. Depending on the erudition level, on the field of study, or the specific task, the user will be able to adjust the content of the dictionary by choosing which of the modules will be displayed.

The following information will be included in the modules: 1. frequency information, 2. link to the corpus concordance substituting exemplification, 3. meaning(s) in academic texts, 4. etymology (if relevant, especially for loanwords), 5. common academic collocations including combinations with typical function words, linked to the relevant corpus concordance, 6. synonyms typical for academic texts, and opposites (if relevant), 7. derived words typical for academic texts, 8. translation equivalents in English, 9. translation equivalents in other languages.

Since this is quite an ambitious task, we decided to focus on one specific type of user in this study and elaborate only the modules that we expect would be chosen by an undergraduate student in an academic writing class. Among the modules relevant to this lower-level user are: frequency information, link to corpus concordance, meaning definition(s), etymology, common collocations, and synonyms.

A detailed analysis of several headwords of various word classes focuses on the meaning description module and on the procedure of finding typical collocations and appropriate synonyms. We discuss the benefits and drawbacks of two types of definitions for this specific dictionary (Aristotelian genus-differentia definition vs. full sentence as provided in Collins COBUILD dictionaries). Further, we discuss the possibility of using existing resources for creating the definitions of meaning, such as monolingual and bilingual dictionaries (e.g. Kraus et al. 1995; Sinclair et al. 1998), and the Frequency dictionary of Czech (Čermák/Křen 2010) which can serve as a defining vocabulary.

The dictionary is based on the Czech list of academic words and phrases that has been published recently as a part of an online application Akalex (Kováříková/Kovářík 2021). The list

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

contains approximately 1,000 single-word and multi-word expressions, and it is based on frequency and distribution criteria similar in some respects to other academic word lists. The material for the academic word list is data from two representative corpora of contemporary written Czech SYN2015 and SYN2020. The words and multi-word units that have been included in Akalex are significantly more common in academic than non-academic texts, they are relatively frequent in academic texts, and are attested and evenly distributed in at least 21 of 24 academic disciplines available in the corpus material. These relatively simple criteria produced outstanding and convincing results comparable to other lists of academic words in size as well as in content (namely the Academic Keyword List by Paquot 2010).

For compiling the dictionary, we utilize some of the online corpus tools available at the Czech National Corpus web page (www.korpus.cz). Apart from the corpus manager KonText (Machálek 2014), which is a primary tool for examining the context of the lemma and for finding collocations, we use the database of translation equivalents Treq (Vavřín/Rosen 2015). Treq can be used not only to search for relevant equivalents in English and other languages but also for finding synonyms through the translation of various equivalents (as suggested by Čibej/Holdt, 2019). Another application, Word at a Glance (Machálek 2020), provides a basic overview of the searched word including collocations and similarly used words.

References

Baptista, J./Costa, N./Guerra, J./Zampieri, M. (2010): P-AWL: Academic Word List for Portuguese. In: Proceedings of Computational Processing of the Portuguese Language, PROPOR 2010, Porto Alegre, Brazil, pp. 120–123.

Carlund, C./Jansson, H./Johansson Kokkinakis, S./Prentic, J./Ribeck, J. (2012): An academic word list for Swedish – a support for language learners in higher education. In: Proceedings of the SLTC 2012 workshop on NLP for CALL, pp. 20–27.

Čermák, F./Křen, M. (2010): Frequency dictionary of Czech: core vocabulary for learners. London.

Čibej, J./Holdt, Š. A. (2019): Repel the syntruders! A crowdsourcing cleanup of the Thesaurus of Modern Slovene. In: Kosem, I. et al. (eds.): Electronic Lexicography in the 21st Century. Proceedings of the eLex 2019 Conference, 1–3 October 2019, Sintra, Portugal. Brno: Lexical Computing CZ, pp. 338–356.

Coxhead, A. (2000): A new academic word list. In: TESOL Quarterly 34 (2), pp. 213–238.

Gardner, D./Davies, M. (2014): A new academic vocabulary list. In: Applied Linguistics 35 (3), pp. 305–327.

Kováříková, D./Kovářík, O. (2021): Akalex: Czech Academic Word List. Prague: Institute of the Czech National Corpus. www.korpus.cz/akalex (last access: 10-01-2022).

Petráčková, V./Kraus, J. (1995): Akademický slovník cizích slov. Prague.

Křen, M. et al. (2015): SYN2015 – representative corpus of contemporary written Czech. Prague: Institute of the Czech National Corpus. www.korpus.cz (last access: 10-01-2022).

Křen, M. et al. (2020): SYN2020 – representative corpus of contemporary written Czech. Prague: Institute of the Czech National Corpus. www.korpus.cz (last access: 10-01-2022).

Machálek, T. (2014): KonText – corpus query interface. Prague: Institute of the Czech National Corpus. kontext.korpus.cz (last access: 10-01-2022).

Machálek, T. (2020): Word at a glance: modular word profile aggregator. In: Calzolari, N. et al. (eds.): Proceedings of the 12th Language Resources and Evaluation Conference. Marseille, pp. 7011–7016.

Morley, J. (2014): Academic phrasebank: a compendium of commonly used phrasal elements in academic English in PDF format. Manchester.

Paquot, M. (2010): Academic vocabulary in learner writing: from extraction to analysis. London/New York.

Sinclair, J. et al. (1998): Anglicko-český výkladový slovník. Prague: Nakladatelství Lidové noviny.

Vavřín, M./Rosen, A. (2015): Treq – database of translation equivalents. FF UK. Prague. https://treq.korpus.cz/ (last access: 10-01-2022).

Contact information

Dominika Kovaříková
Institute of the Czech National Corpus, Charles University, Prague
dominika.kovarikova@ff.cuni.cz

Michal Škrabal
Institute of the Czech National Corpus, Charles University, Prague
michal.skrabal@ff.cuni.cz

Acknowledgements

The paper was supported by the European Regional Development Fund-Project “Creativity and Adaptability as Conditions of the Success of Europe in an Interrelated World” (No. CZ.02.1.01/0.0/0.0/16_019/0000734).

Lorna Morris

THE TREATMENT OF HUMAN REPRODUCTIVE ORGANS IN SCHOOL DICTIONARIES, WITH RECOMMENDATIONS FOR SOUTH AFRICAN PRIMARY SCHOOL DICTIONARIES

Keywords School dictionaries; illustrations; taboo topics; illustrations in school dictionaries

This paper is a pilot study that investigates options for treating human reproductive organs in primary school dictionaries in South Africa, with particular emphasis on the illustrations. This study is a response to concern by some Grade 5 and 6 learners that younger children would be exposed to inappropriate illustrations in school dictionaries.

This paper is placed in the South African context and shows how this is a complex and relevant topic in South Africa, due to the different cultures that are represented in each classroom.

The study shows how school dictionaries, both print and online, currently treat human reproductive organs, and presents examples of entries from school dictionaries. It also presents examples of other organs in the human body as comparison. This article investigates whether the reproductive organs should be treated differently to other organs. The eight dictionaries considered are the *Longman South African School Dictionary*, *Collins New School Dictionary 2e*, *Pharos English Dictionary for South African Schools*, *Illustrated School Dictionary for Southern Africa*, *Oxford South African Illustrated School Dictionary*, *Oxford South African School Dictionary 4e*, *Britannica Kids* (online), and *Word Explorer Children's Dictionary* (online).

The literature examined will comprise lexicographic theory on illustrations in dictionaries and the treatment of taboo topics in dictionaries. Literature on the following aspects is also discussed: cultural aspects of sex education in southern Africa, and sex education in primary schools globally.

The study includes questionnaires completed by primary school teachers and parents to establish their attitudes on whether reproductive organs should be included in primary school dictionaries, and, if so, how they should be treated. Some examples of different options for illustrations are also included. The teachers and parents were given different questionnaires, but they included the same options for illustration preferences. The options were: an anatomical drawing, such as a cross section showing internal and external parts; a "coy" illustration that suggests more than shows, such as a child looking down their pants; an illustration hidden behind a click; a full (naked) child's body, with everything labelled – arm, leg, penis; and a clear, non-nonsense picture of the body part in question. The teachers' preference was an anatomical drawing, such as one would find in a science textbook, while the parents preferred the diagram of a full body with all the parts labelled. The majority of parents stated their preference to include these terms in a primary school dictionary, and to treat them the same as other organs and body parts.

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

This study is a pilot study due to the small sample of parents and teachers surveyed and it will discuss how a larger study can be conducted. The questionnaires show an overwhelming preference to include terms relating to sexuality, but the literature shows significant cultural reservations to these terms being used at school, and especially at primary school.

This presentation will show why it is important to treat these terms in a school dictionary in a clear and unambiguous way, despite this causing potential discomfort to some users. Further research is required in this area, largely because there is such a lack of user research in school dictionaries. Further research is also required with a statistically significant sample size of teachers and parents from different demographic groups. This study could simply be replicated on a larger scale.

The presentation will conclude with recommendations for the treatment of human reproductive organs in primary school dictionaries, as well as recommendations for further research in this area.

References

Bullon, S. (2007): Longman South African school dictionary. England: Pearson Education.

Cullen, K. (2002): Collins new school dictionary. 2nd edition. Glasgow: HarperCollins.

De Kock, C. (2014): Pharos English dictionary for South African schools. Cape Town: Pharos Dictionaries.

Dictionary Unit for South African English (1999): Illustrated school dictionary for Southern Africa. Cape Town: Francolin Publishers.

Encyclopaedia Britannica (2022): Britannica kids dictionary. https://kids.britannica.com/kids/browse/dictionary (last access: 22-02-2022).

Hiles, L. (2008): Oxford South African illustrated school dictionary. Cape Town: Oxford University Press.

Reynolds, M. J. (2019): Oxford South African school dictionary. 4th edition. Cape Town: Oxford University Press.

Wordsmyth (2022): Word Explorer Children's Dictionary. Available at: https://kids.wordsmyth.net/we/ (last access: 22-02-2022).

Contact information

Lorna Morris
Stellenbosch University
lorna@lemma.co.za

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

María Pozzi

DESIGN OF A DICTIONARY TO HELP SCHOOL CHILDREN TO UNDERSTAND BASIC MATHEMATICAL CONCEPTS

Abstract This paper presents the decisions behind the design of a maths dictionary for primary school children. We are aware that there has been a considerable problem regarding Mexican children's performance in maths dragging on for a long time, and far from getting better, it is getting worse. One of the probable causes seems to be the lack of coordination between maths textbooks and teaching methods. Most maths textbooks used in primary schools include lots of activities and problem-solving techniques, but hardly any conceptual information in the form of definitions or explanations. Consequently, many children learn to do things, but have difficulty understanding mathematical concepts and applying them in different contexts. To help solve this problem, at least partially, the project of the dictionary was launched aiming at helping children to grasp and understand maths concepts learned during those first six years of their formal education. The dictionary is a corpus-based terminographical product whose macrostructure, microstructure, typography, and additional information were specifically designed to help children understand mathematical concepts.

Keywords children's specialised lexicography; corpus-based terminography; mathematical terms; children's vocabulary; conceptualisation

Contact information

María Pozzi
El Colegio de México
pozzi@colmex.mx

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Stefan J. Schierholz/Monika Bielinska/
Maria José Domínguez Vázquez/Rufus H. Gouws/
Martina Nied Curcio

THE EMLex DICTIONARY OF LEXICOGRAPHY (EMLexDictoL)

Abstract The EMLex Dictionary of Lexicography (= EMLexDictoL) is a plurilingual subject field dictionary (in German, English, Afrikaans, Galician, Italian, Polish and Spanish) that contains the basic subject field terminology of lexicography and dictionary research, in which the dictionary article texts are presented in a sophisticated but comprehensible form. The articles are supplemented by a complex cross-referencing system and the current subject field literature of the respective national languages. Following the lemma position, the dictionary articles contain items regarding morphology, synonymy, the position of the definiens, additional explanations, the cross-reference position, the position for literature, the equivalent terms in the other six languages of the dictionary as well as the names of the authors.

Keywords Special field lexicography; multilingual dictionary; EMLex

Contact information

Stefan J. Schierholz
Friedrich-Alexander-Universität Erlangen
Stefan.Schierholz@fau.de

Monika Bielinska
Uniwersytet Śląski
monika.bielinska@us.edu.pl

Maria José Domínguez Vázquez
Universidade Santiago de Compostela
majo.dominguez@usc.es

Rufus H. Gouws
Stellenbosch University
rhg@sun.ac.za

Martina Nied Curcio
Università degli Studi Roma Tre
martina.nied@uniroma3.it

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Historical Lexicography: German

XX EURALEX INTERNATIONAL CONGRESS
DICTIONARIES AND SOCIETY
12-16 July 2022, Mannheim, Germany

Andreas Deutsch

FREMDWÖRTER IM DEUTSCHEN RECHTSWÖRTERBUCH (DRW)

Keywords Belegauswahl; Bedeutungserklärung; Codeswitching; Fachterminologie; Fremdwörter; Rechtssprache; Rechtswörterbuch; Rezeption des römischen Rechts

Fremdwörter spielen in der Rechtssprache – wie in vielen Fachsprachen – bis heute eine große Rolle. Neben lateinischen Fachbegriffen haben etwa auch ursprünglich griechische, französische und italienische Wörter in die deutsche Rechtsterminologie Eingang gefunden. Einen gegenüber dem modernen Recht noch weit größeren Schatz an Fremdwörtern enthält die historische Rechtssprache. Viele der historischen Rechtswörter sind dem heutigen Lesepublikum allerdings aufgrund der herausgebildeten Spezialbedeutungen gänzlich unverständlich.

Als wichtigstes Hilfsmittel zum Verständnis der historischen deutschen Rechtssprache darf das „Deutsche Rechtswörterbuch" (DRW) gelten. Das an der Heidelberger Akademie der Wissenschaften bearbeitete Großwörterbuch erläutert den rechtlich relevanten deutschen und westgermanischen Wortschatz vom Beginn der schriftlichen Aufzeichnung in der Spätantike bis ins 19. Jahrhundert. Seit 1897 wird an diesem Mammutvorhaben gearbeitet. 100.000 Wortartikel sind mittlerweile publiziert. Sie umfassen die Buchstabenbereiche von „A" bis zum Beginn von „T". Derzeit wird am 14. Wörterbuchband gearbeitet, jährlich werden rund 1.000 neue Wörterbuchartikel fertiggestellt. Das DRW ist online frei zugänglich (www.deutsches-rechtswoerterbuch.de).

In der Anfangsphase des Wörterbuchprojekts wurden die Fremdwörter allerdings bewusst ausgeklammert, da die sprachlichen und rechtlichen Verflechtungen des westgermanischen Sprachraums im Fokus des Interesses standen. Diese Einschätzung änderte sich im Zuge einer konzeptionellen Reform 1971: Seither sollen auch Fremdwörter ins DRW aufgenommen werden. Da die Quellenexzerpte aus der Anfangsphase des Projekts den Fremdwortschatz jedoch nicht berücksichtigten, konnte diese neue Aufgabe von den Wörterbuchbearbeitern lange Zeit nur eingeschränkt erfüllt werden. Mittlerweile stehen der Forschungsstelle aber umfangreiche elektronische rechtssprachliche Korpora zur Verfügung, die den Zugang auch zum Fremdwortschatz ermöglichen. Besonders wichtig ist hierbei die elektronische Edition DRQEdit (online frei zugänglich: www.drqedit.de), in welcher die wichtigsten bis 1600 gedruckten deutschsprachigen Rechtstexte als Volltexte verfügbar gemacht sind. DRQEdit erfasst somit insbesondere den Wortschatz der vor rund 500 Jahren stattfindenden Rezeption des römischen Rechts in Deutschland, im Zuge derer sehr zahlreiche lateinische Fremdwörter in die deutsche Rechtssprache Eingang fanden.

Viele dieser Fremdwörter stellen für die lexikographische Arbeit eine besondere Herausforderung dar. Oft stellt sich die Frage, ob der Textverfasser vor 500 Jahren das betreffende Wort (bereits) als ein deutsches Wort ansah oder nur einen lateinischen Begriff in seine Ausführungen eingeschoben hat (sog. Codeswitching). Nur im ersten Fall darf das Wort ins DRW aufgenommen werden. Sehr häufig sind die neuen Fremdwörter zudem nicht Teil der Allgemeinsprache geworden oder haben besondere rechtliche Bedeutungen ausgebildet,

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

wodurch auch die Worterklärung besondere Sorgfalt verlangt. Der Vortrag will aufzeigen, wie bei der Wörterbuchbearbeitung auf diese besonderen Herausforderungen eingegangen wird.

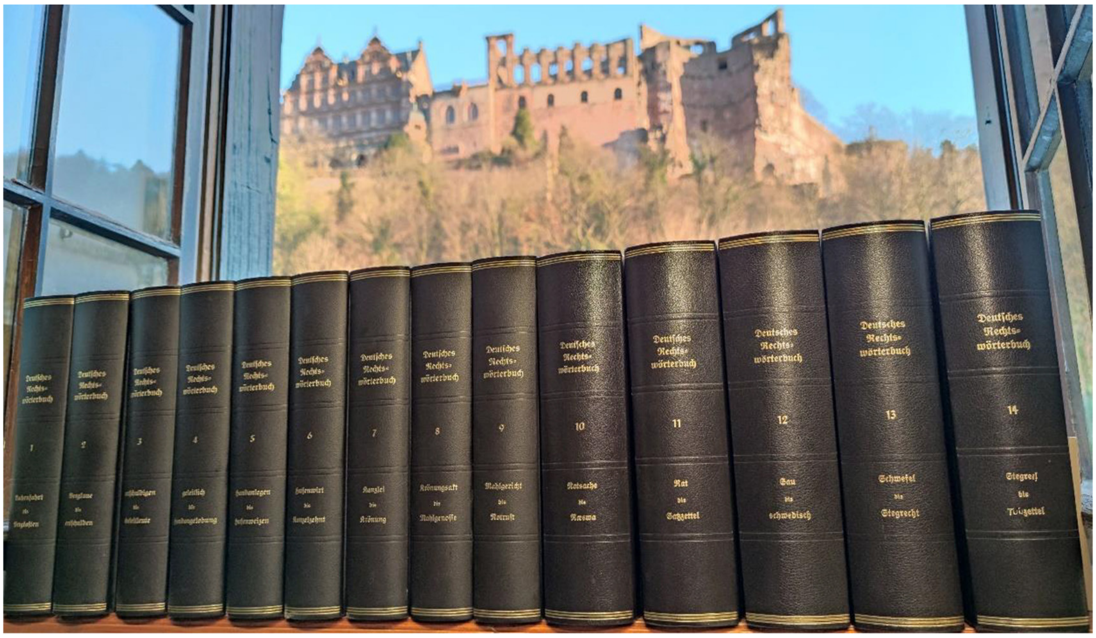
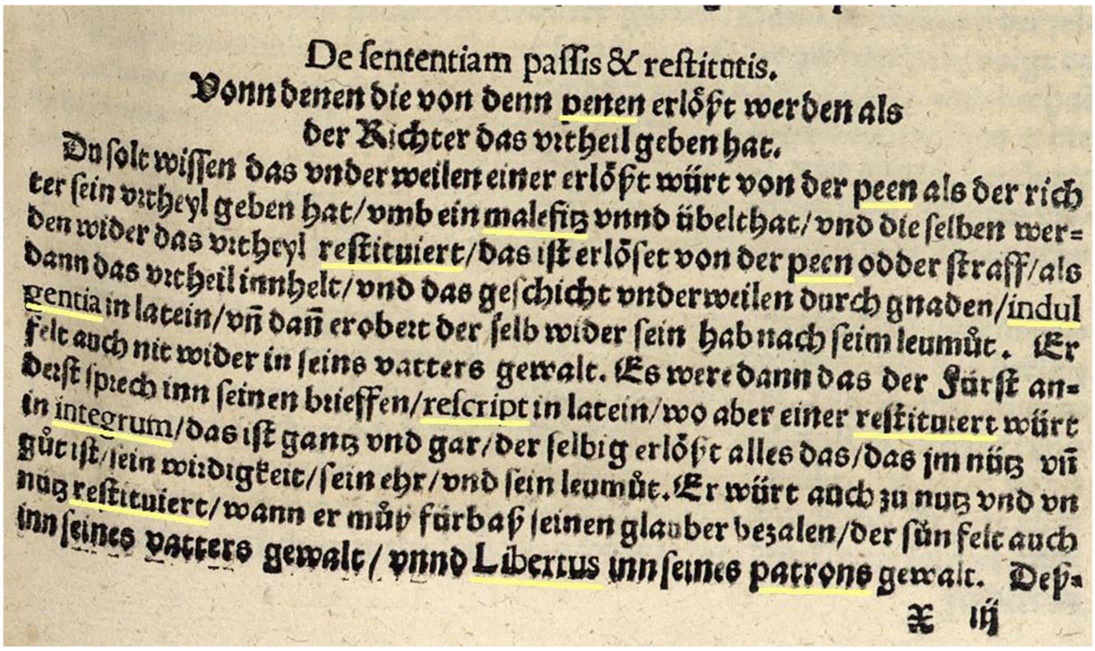


Abb. 1: Das Deutsche Rechtswörterbuch – 100.000 Wortartikel aus den Buchstabenbereichen A bis T sind bislang fertig



De ſententiam paſſis & reſtitutis.
Vonn denen die von denn penen erlöſt werden als der Richter das vrtheil geben hat.

Du ſolt wiſſen das vnderweilen einer erlöſt würt von der peen als der richter ſein vrtheyl geben hat/vmb ein malefitz vnnd übelthat/vnd die ſelben werden wider das vrtheyl reſtituiert/das iſt erlöſet von der peen odder ſtraff/als dann das vrtheil innhelt/vnd das geſchicht vnderweilen durch gnaden/indulgentia in latein/vñ dañ erobert der ſelb wider ſein hab nach ſeim leumůt. Er felt auch nit wider in ſeins vatters gewalt. Es were dann das der Fürſt anderſt ſprech inn ſeinen brieffen/reſcript in latein/wo aber einer reſtituiert würt in integrum/das iſt gantz vnd gar/der ſelbig erlöſt alles das/das im nütz vñ gůt iſt/ſein wirdigkeit/ſein ehr/vnd ſein leumůt. Er würt auch zu nutz vnd vnnütz reſtituiert/wann er můß fürbaß ſeinen glauber bezalen/der ſün felt auch inn ſeines vatters gewalt/vnnd Libertus inn ſeines patrons gewalt. Deſ=

X iij

Abb. 2: Oft ist es schwer, in Rechtstexten der Frühneuzeit zwischen Codeswitching und Fremdwortgebrauch zu unterscheiden. Beispiel aus: Justin Gobler, Gerichtlicher Prozeß (Frankfurt a. M. 1536), Bl. 123r

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Wortanfang	Zahl der Fremdwörter im DRW	Beispiele
sub-	*88*	*subarrendieren, subarrhieren, Subhypothek, subreptiv*
suk-	17	*sukkumbieren, Sukzentor, Sukzessionsgerechtigkeit*
summ-	30	*Summari' appellationklage, summarie, Summbrief*
sup-/super-	59	*Supan, Superintendentur, Superkargo, Supputation*
sus-	12	*suspektieren, Suspensor, Suspition*
syn-	18	*Syndikant, Syndikator, synodieren*
tab-	47	*Tabakrobot, Tabellier, Tabellionat, tabulieren*

Tab. 1: Allein im neuesten DRW-Doppelheft finden sich über 250 Fremdwörter erklärt. Viele sind in anderen Wörterbüchern nicht oder nur mit anderen Bedeutungen gebucht

Literatur

Bedenbender, A. (2014): Das Deutsche Rechtswörterbuch im Netz. In: Abel, A./Lemnitzer, L. (Hg.): Vernetzungsstrategien, Zugriffsstrukturen und automatisch ermittelte Angaben in Internetwörterbüchern. (= OPAL – Online publizierte Arbeiten zur Linguistik 2/2014). Mannheim, S. 22–28. http://pub.ids-mannheim.de/laufend/opal/opal14-2.html (Stand: 21.3.2022).

Considine, J. (2016): Historical dictionaries. In: Durkin, P. (Hg.): The Oxford handbook of lexicography. Oxford, S. 163–175.

Deutsch, A. (2019): Das Deutsche Rechtswörterbuch – ein Fachwörterbuch zwischen Recht, Sprache und Geschichte. In: Harm, V./Lobenstein-Reichmann, A./Diehl, G. (Hg.): Wortwelten: Lexikographie, Historische Semantik und Kulturwissenschaft. Berlin/Boston, S. 97–112.

Deutsch, A. (2021): „Als wolte ich in amplissima illa materia … ein Tractat beschreiben" – Zur Rolle von Codeswitching in Rechtsbüchern aus der Rezeptionszeit des römischen Rechts. In: Glaser E./Prinz, M./Ptashnyk, S. (Hg.): Historisches Codeswitching mit Deutsch. Berlin/Boston, S. 91–112.

Deutsch, A. (2022): 125 Jahre Deutsches Rechtswörterbuch: Bundesverfassungsgerichtspräsident Harbarth gratuliert, In: H-Soz-Kult, 10.01.2022. www.hsozkult.de/news/id/news-115018 (Stand: 5.5.2022).

Deutsch, A. (2011): The „Dictionary of Historical German Legal Terms" and its European concept. In: Oxford University Research Archive.
http://ora.ox.ac.uk/objects/uuid:ef5d07d3-77fc-4f07-b13f-d4c24b4d1848 (Stand: 21.3.2022).

Deutsches Rechtswörterbuch (1912 ff.): Wörterbuch der älteren deutschen Rechtssprache. Hrsg. von der Heidelberger Akademie der Wissenschaften. Weimar.

DRQEdit: Deutschsprachige Rechtsquellen in digitaler Edition. http://drqedit.de (Stand: 21.3.2022).

Kontaktinformationen

Andreas Deutsch
Leiter der Forschungsstelle Deutsches Rechtswörterbuch
Heidelberger Akademie der Wissenschaften
drw@hadw-bw.de

Volker Harm

WORTGESCHICHTE DIGITAL: A HISTORICAL DICTIONARY OF NEW HIGH GERMAN

Abstract *Wortgeschichte digital* (‘digital word history’) is a new historical dictionary of New High German, the most recent period of German reaching from approximately 1600 AD up to the present. By contrast to many historical dictionaries, *Wortgeschichte digital* has a narrated text – a “word history” – at the core of its entries. The motivation for choosing this format rather than traditional microstructures is briefly outlined. Special emphasis it put on the way these word histories interact with other components of the dictionary, notably with the quotation section. As *Wortgeschichte digital* is an online-only project, visualizations play an important role for the design of the dictionary. Two examples are presented: first, the “quotation navigator” which is relevant for the microstructure of the entries, and, second, a timeline (“Zeitstrahl”) which is part of the macrostructure as it gives access to the lemma inventory from a diachronic point of view.

Keywords Historical lexicography; word history; quotations; visualizations

Contact information

Volker Harm
Zentrum für digitale Lexikographie der deutschen Sprache (ZDL), Akademie der Wissenschaften zu Göttingen
vharm@uni-goettingen.de

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Andrea Moshövel

SKATOLOGISCHER WORTSCHATZ IM FRÜHNEUHOCHDEUTSCHEN ALS KULTURGESCHICHTLICHE UND LEXIKOGRAPHISCHE HERAUSFORDERUNG

Abstract This paper deals with the lexicographic treatment of the evidently plenty and pervasive scatological vocabulary, that is vocabulary concerning the process and products of bodily excretion (especially feces), in the synchronical Early New High German Dictionary (FWB = Frühneuhochdeutsches Wörterbuch) from a dictionary user's view. Initially, different cultural concepts of scatology by Norbert Elias, Michail Bachtin and Mary Douglas among others and the term taboo are reflected. Subsequently, selected lexical items such as words with a primary scatological meaning (e.g. *drek*, *kot*, *scheisse*), concealing expressions (euphemisms, periphrases, metaphors, e.g. *sitzen*, *seine notdurft tun, bauernveiel*), and certain aspects within the polysemy of the verb *scheissen* are discussed, the latter on the one hand referring to a physical process with uncontrollable aspects and on the other hand denoting a deliberate action and functionalized as a fighting word during the reformation. Focussing on different positions of lexicographical information within the microstructure of the FWB, the surveillance shows that in a synchronical perspective Early New High German scatological vocabulary is a heterogeneous and complex phenomenon due to speaker, context and respectively semantic and pragmatic purposes.

Keywords Scatological vocabulary; Early New High German; Early New High German Dictionary (FWB); historical lexicography; historical lexicology; cultural history

Reference

FWB (1986 ff.): Goebel, U./Lobenstein-Reichmann, A./Reichmann, O. (eds.) (1986 ff.): Frühneuhochdeutsches Wörterbuch. Seit 2013 im Auftrag der Akademie der Wissenschaften zu Göttingen. Berlin et al. Online: https://fwb-online.de/ (last access: 23-03-2022).

Kontaktinformationen

Andrea Moshövel
Akademie der Wissenschaften zu Göttingen, Arbeitsstelle Frühneuhochdeutsches Wörterbuch (FWB)
amoshoe@gwdg.de

Historical Lexicography: Romance and Other Languages

XX EURALEX INTERNATIONAL CONGRESS
DICTIONARIES AND SOCIETY
12-16 July 2022, Mannheim, Germany

Maria Arapopoulou/Georgios Kalafikis/Dimitra Karamitsou/
Efstratios Sarischoulis/Sotiris Tselikas

"VOCABULA GRAMMATICA": THREADING A DIGITAL ARIADNE'S STRING IN THE LABYRINTH OF ANCIENT GREEK SCHOLARSHIP

Abstract An ongoing academic and research program, the "Vocabula Grammatica" lexicon, implemented by the Centre for the Greek Language (Thessaloniki, Greece), aims at lemmatizing all the philological, grammatical, rhetorical, and metrical terms in the written texts of scholars (philologists and scholiasts) who curated the ancient Greek literature from the beginning of the Hellenistic period (4th/3rd c. BC) until the end of the Byzantine era (15th c. AD). In particular, it aspires to fill serious gaps (a) in the study of ancient Greek scholarship and (b) in the lexicography of the ancient Greek language and literature. By providing specific examples, we will highlight the typical and methodological features of the forthcoming dictionary.

Keywords Humanities; digital lexicography; specialized dictionary; Ancient Greek language; Ancient Greek scholarship

Contact information

Maria Arapopoulou
Centre for the Greek Language, Thessaloniki, Greece
marapopoulou@gmail.com

Georgios Kalafikis
Centre for the Greek Language, Thessaloniki, Greece
gkalafikis@gmail.com

Dimitra Karamitsou
Centre for the Greek Language, Thessaloniki, Greece
dimika91@gmail.com

Efstratios Sarischoulis
Centre for the Greek Language, Thessaloniki, Greece
efstratios.sarischoulis@gmail.com

Sotiris Tselikas
Centre for the Greek Language, Thessaloniki, Greece
stselik@gmail.com

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Anaïs Chambat

LA LIGNÉE « CAPURON-NYSTEN-LITTRÉ » ENTRE RUPTURES ET CONTINUITÉS DOCTRINALES

Abstract This article aims to show the influence of doctrines in the medical lexicographers choices, with the Capuron-Nysten-Littré lineage as a case study. Indeed, the *Dictionnaire de médecine* has been crossed by several schools of thought such as spiritualism and positivism. While lexical continuity may seem self-evident due to the nature of the work, thus reducing the reprint to a simple lexical increase, this process introduces neologisms and deletions, all can be considered in their effects by using text statistics and factorial analysis.

Keywords History of medical science dictionaries; 19th–20th centuries; medical lexicographers; spiritualism; positivism; text statistics; Joseph Capuron; Pierre-Hubert Nysten; Émile Littré; Charles Robin; Augustin Gilbert; Medica

Contact

Anaïs Chambat
Université Paris Cité
Direction générale déléguée des bibliothèques et musées (DGDBM)
Bibliothèque interuniversitaire de santé – pôle médecine
ana.chambat@gmail.com

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

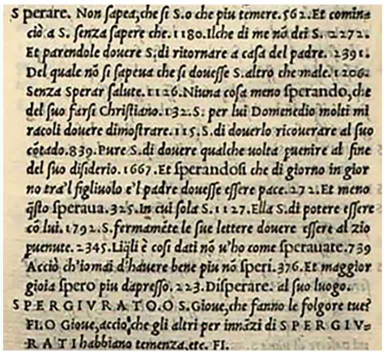
Sarah Mantegna/Carla Marello

THE MULTILINGUAL APPENDIX OF *LE RICCHEZZE DELLA LINGUA VOLGARE* (1543) BY FRANCESCO ALUNNO

A lexicographer's "service list" and an intercomprehension tool

Keywords Decameron concordance; multilingual appendix; intercomprehension

Francesco del Bailo (1474–1556), alias Francesco Alunno, was a humanist: born in Ferrara, he spent most of his life between Udine and Venice. Alunno is best known as author of *La fabrica del mondo* 'The texture of the world' (1548), which was considered the first onomasiological dictionary in Italian. Here, we try to stress the value of *Le Ricchezze della lingua volgare* as an accurate concordance of Boccaccio and focus on its multilingual appendix.

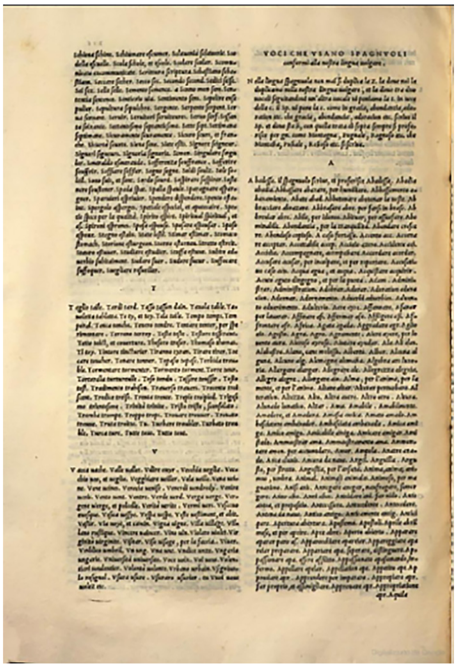


S perare. Non ſapea,che ſi S.o che piu temere.562.Et comin-
ciò a S. ſenza ſapere che.1180.Ilche di me nō dei S. 2272.
Et parendole douere S.di ritornare a caſa del padre. 2391.
Del quale nō ſi ſapeua che ſi doueſſe S.altro che male.1206.
Senza Sperar ſalute.1126.Niuna coſa meno ſperando,che
del ſuo farſi Chriſtiano.132.S. per lui Domenedio molti mi
racoli douere dimoſtrare.115.S.di douerlo ricouerare al ſuo
cōtado.839.Pure S.di douere qualche uolta ꝑuenire al fine
del ſuo diſiderio.1667.Et ſperandoſi che di giorno in gior
no tra'l figliuolo e'l padre doueſſe eſſere pace.272.Et meno
q̄ſto ſperaua.325.In cui ſola S.1127.Ella S.di potere eſſere
cō lui.1792.S.fermamēte le ſue lettere douere eſſere al zio
ꝑuenute.2345.Li q̄li è coſi dati nō u'ho come ſperauate.739
Acciò ch'io mai d'hauere bene piu nō ſperi.376.Et maggior
gioia ſpero piu dapreſſo.223.Diſperare. al ſuo luogo.
SPERGIVRATO.O S.Gioue,che fanno le folgore tue?
FI.O Gioue,acciò che gli altri per innāzi di SPERGIV-
RATI habbiano temenza,etc. FI.

Fig. 1: Entry *sperare* 'to hope' (Alunno 1543, p. 171)

At page 3 recto Alunno explains his lemmatisation rules and his nested microstructure for the *Decameron* concordances: derived forms are in the microstructure of their "primitives"; conjugated verb forms are given under an entry in the infinitive; contexts both for the singular and for the plural are grouped under an entry in the singular. An example of some of these details is visible in the microstructure of *sperare* 'to hope' in Figure 1: the forms of the verb *sperando*, *sperandosi*, *sperava*, *speravate, spero* are in the same font as the entry *sperare* and assume the role of subentries.

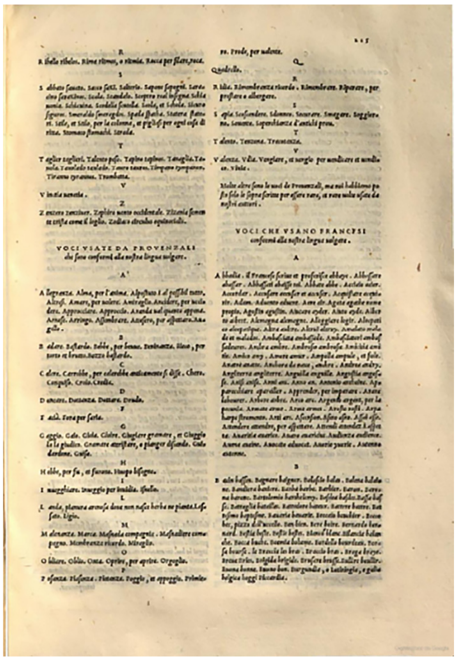
This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.



VOCI CHE VSANO SPAGNVOLI
conformi alla noſtra lingua volgare.

Fig. 2: List of Spanish words (Alunno 1543, p. 215)

The lists of "words used by Latins, Greeks, Provençals, Frenchmen, Spaniards, Germans, Englishmen, Goths and other nations of which we list only those which best conform to our vernacular Italian and were used by Boccaccio, Dante, and Petrarca" cover about twenty pages in two columns. Alunno seems to adopt an Italian-centric perspective and a synchronic approach in compiling lists of Italian words which correspond to remarkably similar lexical items in Latin, Greek, Spanish, French, German and English. Alunno lists corresponding vernacular words for Romance languages, while for non-Romance languages he lists borrowings of Latin origin.



VOCI VSATE DA PROVENZALI
che ſono conformi alla noſtra lingua volgare.

VOCI CHE VSANO FRANCESI
conformi alla noſtra lingua volgare.

Fig. 3: List of French words (Alunno 1543, p. 217)

If we consider the way in which Alunno reports Spanish and French words we find similarities with Nebrija's *Spanish and Latin dictionary* (1492) or Estienne's *French and Latin dictionary* (1538). As for Latin and Greek, an Aldine edition of *Calepinus* had just appeared in 1542 by the same printer of Alunno (Manutius). Alunno might very well have adopted as a source the above-mentioned bilingual Latin-Vernacular: Latin was the right bridge language to easily find words for his lists. For non-Romance languages the *Quinque* (or *sex*) *linguarum* lists are a more probable source, but we cannot exclude some oral informants.

Since *Le ricchezze della lingua volgare* is a single author's dictionary, microstructures gather all the contexts in which Boccaccio used the word corresponding to the entry. Alunno does not provide a definition and does not suggest an etymology. Today, however, we can appreciate the multilingual appendix not only as a curiosity, but as a sort of forerunning experiment of intercomprehension, in a broad sense.

Alunno was aware of some spelling regularities in graphic differences between Italian and Spanish and he reports them at the beginning of the list. We contrastively analysed these differences with the other word lists in other languages through the intercomprehensive principle of the "seven sieves" of the EuroCom project (Meissner et al. 2004) and if we consider the lists consulted in connection with the concordance, we also have contexts. Those who study intercomprehension know that context is one of the basic elements that enable the use of different strategies to understand language and texts. These considerations encourage us to figure the use of Alunno's lists within the frame of the concordance at his time and that of similar lists supported by a corpus today. It is important to stress that Alunno's lists can be seen as an example of a primitive attempt to make intercomprehension possible as early as the XVI century.

The work of a lexicographer like Alunno embodies an approach which, if considered in scientific terms, is neither top-down like Bach et al. (2008) nor bottom-up. Although he lacked a strong philological awareness, he cannot be said to be naïve. His point of view was closer to that of a present-day cultivated practitioner of intercomprehension.

References

Alunno, F. (1543): Le ricchezze della lingua volgare sopra il Boccaccio Figli di Aldo Manuzio, in Vinegia. 2nd edition 1551. https://books.google.it/books?hl=it&id=mr9KAAAAcAAJ&q=zafferano#v=onepage&q&f=true (last access: 06-03-2022).

Bach, S./Brunet, J./Mastrelli, C. A. (2008): Quadrivio romanzo: dall'italiano al francese, allo spagnolo, al portoghese. Florence.

Boccaccio, G. (1526): Il Decamerone: novamente stampato da Sabbio. Vineggia.

Estienne, R. (1539): Dictionnaire français latin Paris de l'imprimerie de Robert Estienne.

Meissner, F.-J./Meissner, C./Klein, H. G./Stegmann, T. D. (2004): EuroComRom – Les sept tamis: lire les langues romanes dès le départ. Aachen.

Nebrija, A. de (1492): Dictionarium latino-hispanicum. [Juan de Porras (ed.)]. Salamanca.

Quinque linguarum utilissimus Vocabulista Latine, Italice, Gallice, Hyspane et Alemanice. Vocabulista de le cinque lingue ... Augsburg: Philipp Ulhart, 1533 Venetiis, Marchio Sessa 1537, 1538.

Sex linguarum, latinae, gallicae, hispanicae, italicae, anglicae et teutonice dilucidissimus dictionarius: Vocabulaire de six languages, latin, francoys, espagniol, italien, anglois et aleman 1541 Venetiis Marchio Sessa 1.

Contact information

Sarah Mantegna
Univerisité Savoie Mont-Blanc
sarah.mantegna@univ-savoie.fr

Carla Marello
Università di Torino
carla.marello@unito.it

Mihai-Alex Moruz/Mădălina Ungureanu

17TH-CENTURY ROMANIAN LEXICAL RESOURCES AND THEIR INFLUENCE ON ROMANIAN WRITTEN TRADITION

Abstract This paper focusses on the first Slavonic-Romanian lexicons, compiled in the second half of the 17th century and their use(rs), proposing a method of investigating the manner in which lexical information available in the above corpus relates, if at all, to the vocabulary of texts from the same period. We chose to investigate their relation to an anonymous Old Testament translation made from Church Slavonic, also from the second half of the 17th century, which was supposed to be produced in the same geographical area, in the same Church Slavonic school or even by the same author as the lexicons. After applying a lemmatizer on both the Biblical text (Books of Genesis and Daniel) and the Romanian material from the lexicons, we analyse the results and double the statistical analysis with a series of case studies, focusing on some common lexemes that might be an indicator of the relatedness of the texts. Even if the analysis points out that the lexicons might not have been compiled as a tool for the translation of religious texts, it proves to be a useful method that reveals interesting data and provides the basis for more extensive approaches.

Keywords 17th century; Church Slavonic; bilingual dictionaries; Old Testament; Old Romanian

Contact information

Mihai-Alex Moruz
Faculty of Computer Science, “Alexandru Ioan Cuza” University, Iași
mmoruz@info.uaic.ro

Mădălina Ungureanu
Institute of Interdisciplinary Research, Department of Social Sciences and Humanities, “Alexandru Ioan Cuza” University, Iași
madandronic@gmail.com

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Clarissa Stincone

USAGE LABELS IN BASNAGE'S *DICTIONNAIRE UNIVERSEL* (1701)

Abstract Basnage's revision (1701) of Furetière's *Dictionnaire universel* is profoundly different from Furetière's work in several regards. One of the most noticeable features of the dictionary lies in his increased use of usage labels. Although Furetière already made use of usage labels (see Rey 1990), Basnage gives them a prominent role. As he states in the preface to his edition, a dictionary that aspires to the title of "universal" should teach how to speak *in a polite way* ("poliment"), *right* ("juste") and making use of specific terminology for each art. He specifies, lemma by lemma, the diaphasic dimension by indicating the word's register and context of use, the diastratic one by noting the differences in the use of the language within the social strata, the diachronic evolution by indicating both archaisms and neologisms, the diamesic aspect by highlighting the gaps between oral and written language, the diatopic one by specifying either foreign borrowings or regionalisms.

After extracting the entries containing formulas such as "ce mot est …", "ce terme est …" and similar ones, we compare the number of entries and the type of information provided by the two lexicographers[1]. In this paper, we will focus on Basnage's innovative contribution. Furthermore, we will try to identify the lexicographer's sources, i. e. we will try to establish on which grammars, collections of linguistic remarks or contemporary dictionaries Basnage relies his judgements.

Keywords Historical lexicography; *Dictionnaire universel*; Basnage de Beauval; 17th century; usage labels

References

Basnage de Beauval, H. (1701): Dictionnaire universel. La Haye et Rotterdam.

Rey, A. (1990): Le marques d'usage et leur mise en place dans les dictionnaires du XVII^e^ siècle : le cas Furetière. In: Lexique 9/ Les marques d'usage dans les dictionnaires (XVII^e^, XVIII^e^ siècles). Lille, pp. 17–29.

Contact information

Clarissa Stincone
Université Sorbonne Nouvelle – Paris 3
clarissastincone@live.it

[1] The .txt files digitised with Transkribus and subsequently analysed with BBEdit are flawed (incorrectly separated words, confused or missing letters, etc.). It is possible that some usage labels may have escaped analysis.

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Marija Žarković

THE LEGAL LEXICON IN THE FIRST DICTIONARY OF THE SPANISH ROYAL ACADEMY (1726–1739)

The concept of the judge

Abstract This paper consists of a short analysis of the sources and the treatment of the legal lexicon in the first dictionary published by the Spanish Royal Academy (1726–1739), followed by a longer commentary on the representation and the treatment of the concept of judge, in which the reflection of the extra-linguistic factors in the definitions stands in focus. The results highlight the relevance of the legal context of that era for the treatment of the lexicon related to the legal domain, but they also demonstrate the pattern in which the lexicographic data displays peculiarities of legal matters.

Keywords History of lexicography; legal lexicon; Spanish lexicography; Spanish Royal Academy

Contact information

Marija Žarković
Universitat Autònoma de Barcelona
marija.zarkovic@e-campus.uab.cat

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

(Historical) Lexicology

XX EURALEX INTERNATIONAL CONGRESS
DICTIONARIES AND SOCIETY
12-16 July 2022, Mannheim, Germany

Rita Calabrese/Katherine E. Russo

METAPHORICAL CONSTRUCTIONS IN INDIAN ENGLISH AND AUSTRALIAN ABORIGINAL ENGLISH

From compositionality to grammaticalization

Keywords Verbal MWCs; variability; World Englishes; idiomaticity

The study of verbal multi-word constructions (MWCs) presents intriguing theoretical and methodological challenges which touch on both the relation between morphology and syntax, more generally on the architecture of grammar (Los et al. 2012). Verbal MWCs occur in most Germanic languages and are generally referred to in the literature as verb-particle constructions, phrasal verbs, separable (complex) verbs or particle verbs occurring in transitive, intransitive or more complex variants (Dehé 2002). They represent a highly innovative area in any variety of English (Baugh/Cable 2002) and the divergence of these constructions in World Englishes is highly significant. The present research focuses on verbal MWCs possessing semantic opacity, with a specific focus on the possible development of figurative and non-compositional meaning (Brinton/Closs Traugott 2005; Rodriguez-Puente 2012, 2021) in two varieties of English, namely Indian English and Aboriginal English. Such constructions develop through gradual stages of semantic reinterpretation from semantically transparent according to the 'Principle of Compositionality' (Szabò 2012) to idiomatic constructions via metaphors (Kovàcs 2007). This process is also reflected in both the syntactic structure of particle verbs and in the semantic contribution of the accompanying particles to the meaning of the whole combination. For example, the aspectual or telic function implied in these verb+particle combinations have led some authors to analyse them as lexicalisation of the functional category 'telicity' (Dehé 2002).

The study aims to address the major issue concerning whether phrasal verbs with figurative and non-compositional meaning follow a common underlying phenomenon across the two varieties under investigation. Corpus-based methodology and linguistic diagnostics have been matched to 1) uncover potential differences and shared features between Indian and Aboriginal English in multi-word verb constructions 2) establish the extent to which the frequency of such features across time may contribute to measuring distance and similarity between the two lects.

In order to test the aforementioned theoretical and methodological issues, the research draws on data from two annotated corpora, the Diachronic Corpus of Indian English (1,000,000 tokens) and the Diachronic Corpus of Aboriginal English (948,000 tokens), specifically compiled to represent different dimensions of linguistic variability over a period of about 150 years (1833–2013). The two corpora have been designed on the model of available multi-genre corpora like ICE-IND and ICE-AUS to provide a comparable, balanced configuration to the two corpora. As major changes in a language are assumed to come from the spoken language, speech-related written genres such as witness depositions (containing direct speech) were included for the analysis (Culpeper/Kyto 2010) and then compared to

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

selected proceedings from the *Old Bailey Corpus* dating back to the same time span. Other sections of the corpora were selected and compared to similar samples from the *BNC* and ARCHER corpora. To test whether any of attested combinations in the 1830s had followed a gradual path toward nativisation in Indian English and Aboriginal English, the comparative approach of using each sub-corpus compiled for each decade side-by-side was employed to describe the development and use of those combinations.

Early results have shown that similar lexico-grammar patterns are attested in both DiCIE and DiCAbE. The data confirms previous findings (Nihalani/Ray/Tongue 2005; Malcolm 2013) showing similar patterns of variation in the two varieties. The occurrences of verbal MWC variation identified in DiCIE and DiCAbE were also analysed from a semantic perspective to investigate whether they developed a figurative meaning (Rodriguez-Puente 2021) over time. While similar converging co-occurrences were relatively scarce due to different contact ecologies, from a diachronic perspective it is remarkable that the overlap in Indian-specific and Aboriginal-specific occurrences emerges with greater frequency up until the year 1938. A possible explanation may be that commonalities were due to the fact that in their first phases of evolution they followed a similar path still linked to the common lexifier but after that date the two varieties underwent a further phase of nativisation (Schneider 2007). In other words, the comparison of data from different time periods suggests that the creation of new combinatory lexico-grammar patterns is an overall phenomenon occurring regularly and steadily over time, but that in both varieties it was at first influenced by their supposed similar lexifiers and later gave rise to divergence.

References

Baugh, A. C./Cable, T. (2002): A history of the English language. 5th edition. London.

Brinton, L./Closs Traugott, E. (2005): Lexicalization and language change. Cambridge.

Culpeper, J./Kyto, M. (2010): Early Modern English dialogues. Spoken interaction as writing. Cambridge.

Dehé, N. (2002): Particle verbs in English. Syntax, information structure and intonation. Amsterdam.

Kovàcs, E. (2007): Reflections on English phrasal verbs. In: Publicationes Universitatis Miskolcinensis XII (2), pp. 5–18.

Los, B./Blom C./Boogij, G./Elenbaas, M./Van Kemenade, A. (eds.) (2012): Morphosyntactic change: a comparative study of particles and prefixes. Cambridge.

Malcolm, I. (2013): Aboriginal English: some grammatical features and their implications. Australian Review of Applied Linguistics 36 (3), pp. 267–284.

Nihalani, P./Ray, K./Tongue, P. H. (2005): Indian and British English. A handbook of usage and pronunciation. 2nd edition. Oxford.

Rodrìguez-Puente, P. (2021): The English phrasal verb, 1650 – Present. History, stylistic drift, and lexicalisation. Cambridge.

Schneider, E. W. (2007): Postcolonial English: varieties around the world. Cambridge.

Contact information

Rita Calabrese
University of Salerno
rcalabrese@unisa.it

Katherine E. Russo
Universtà degli Studi di Napoli "L'Orientale"
kerusso@unior.it

Zoe Gavriilidou/Asimakis Fliatouras/Elina Chadjipapa

ARABIC LOANWORDS IN GREEK

Keywords Loanwords; Arabic; Greek language

Borrowing is a linguistic phenomenon that emerges in situations of linguistic interaction and language contact. The transfer of lexical elements between languages induces language changes and reveals the relations between people and their stereotypes (Anastassiadis 1994; Gavriilidou 2018). Borrowing can be found as a result of dominance of one language to another, immigration, etc. and it has been intensively studied throughout the last decades (Thomason/Kaufman 1992; Thomason 2001; Haspelmath 2008; Haspelmath/Tadmor 2009).

There is a large amount of previous research on lexical borrowing from English, French or Russian into Greek (Contossopoulos 1978; Apostolou-Panara 1991; Anastassiadis 1994; Gavriilidou 2018), however only few papers (Chatzisavvidis 2015) have occasionally studied borrowings from Arabic into Greek, even though Arabic is a language with an important cultural impact in languages such as Spanish, English, or French and despite its long historical relation with Greek.

Addressing this gap, the aim of this presentation is to investigate Arabic borrowings that entered the Greek lexicon by considering the factors involved in borrowing and accounting the reasons for social and attitudinal influence. Our purpose is also to investigate how these borrowings are included in two major dictionaries of Modern Greek (MG), the Dictionary of Standard Modern Greek (DSMG) (1998) and the User's Dictionary of the Academy of Greece (UD) (2014).

The study is based on 324 loanwords from Arabic into Greek extracted from the online version of DSMG (http://www.greek-language.gr/greekLang/modern_greek/tools/lexica/triantafyllides/) through advanced searches, classified according to Haugen (1950) typology and further analyzed phonologically, morphologically and semantically.

It is shown that Arabic words have been introduced into Greek as direct loans mainly during the Byzantine period, e. g. βερίκοκο< berkuk ‘apricot’, γάιδαρος< gadar, gaidar ‘donkey’ (Browning 2008), many of which have not survived in Modern Greek (e. g. αζάπης ‘Turkish soldier’), or as indirect loans with the intermediate of Latin, Turkish or French in later years (e. g. χαρέμι< ‘harem’ <’ haram’, French e. g. γάζα< ‘gaz(e)’ < ‘Gazza’). Fliatouras (2020) has shown that 5% of direct loans in Modern Greek are of Arabic origin and that Arabic is the most productive category of indirect loans, especially with the intermediate of Turkish. Moreover, many direct loans can be found in Modern Greek dialects, especially in Cretan, e. g. κερεβίζι ‘celery’.

In the first part of the paper, we consider the contact situations that led to borrowing from Arabic into Greek and the reasons of borrowing. Then we critically compare how these borrowings are recorded in DSMG and UD for shedding light to the inconsistencies concerning the etymology of these entries and we discuss the important role of Greek dialects such as Cretan in Arabic loanword preservation in Greek. Next, we offer a morphophonological analysis of borrowings. The semantic fields in which the borrowings belong to are also studied. The semantic field list of Haspelmath/Tadmor (2009), used in the Loanword

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Typology Project, including 24 fields (e. g. physical world, kinship, animals, food and drink, social and political relations, etc.) is adopted here, in order to provide data comparable with findings from other studies which used the same list. Finally, in the last section we offer a lexicographic proposal for the treatment of such lexical units in Greek dictionaries.

References

Anastassiadis, A. (1994): Neological borrowing in Modern Greek [In Greek]. Thessaloniki.

Apostolou-Panara, A. M. (1991): English loanwords in Modern Greek: an overview. In: Terminologie et Traduction 1, pp. 45–58.

Browning, R. (2008): The Medieval and Modern Greek language. Athens.

Chatzisavvidis, S. (2015): The Greek between east and west: loans from eastern and western languages. Theory and research in education. Dedicated volume to Sofronios Chatzisavvidis. Patra, pp. 153–163. http://periodiko.inpatra.gr/issue/issue5/issue5.pdf (last access: 07-09-2021).

Contossopoulos, N. (1978): L'influence du français sur le grec, emprunts lexicaux et calques phraséologiques. Athènes.

Fliatouras, A. (2020): How many and which are the loans in Modern Greek? A first approach. In: Glossologia 29, pp. 21–44.

Gavriilidou, Z. (2018): Russian borrowings in Greek and their presence in two Greek dictionaries. In: Cibej, J./Gorjang, V./Kozem, I./Krek, S. (eds.): Proceedings of the XVIII Euralex International Congress, Lexicography in Global Contexts. Ljubljana, pp. 297–308.

Haspelmath, M. (2008): Loanword typology: steps toward a systematic cross-linguistic study of lexical borrowability. In: Stolz, T./Bakker, D./Salas Palomo, R. (eds.): Aspects of language contact: new theoretical, methodological and empirical findings with special focus on Romancisation processes. Berlin/New York, pp. 43–62.

Haspelmath, M./Tadmor, U. (2009): Loanwords in the world's languages. The Hague.

Haugen, E. (1950): The analysis of linguistic borrowing. In: Language 26, pp. 210–231.

Thomason, S. (2001): Language contact. Washington.

Thomason, S./Kaufman, T. (1992): Language contact, creolization and genetic linguistics. Oakland.

Contact information

Zoe Gavriilidou
Democritus University of Thrace
zoegab@otenet.gr

Asimakis Fliatouras
Democritus University of Thrace
afliatouras@yahoo.com

Elina Chadjipapa
Democritus University of Thrace
elinaxp@hotmail.com

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Ellert Thor Johannsson

OLD WORDS AND OBSOLETE MEANINGS IN MODERN ICELANDIC

Abstract This paper examines a certain subset of the vocabulary of Modern Icelandic, namely those words that are labelled as ‘ancient’ in the *Dictionary of Contemporary Icelandic* (DCI). The words were analysed and grouped into two main categories, 1) Words with only ‘ancient’ sense(s) and 2) words that have modern as well as an obsolete older sense. Several subgroups were identified as well as some lexical characteristics. The words in question were then analysed in two other sources, the *Dictionary of Old Norse Prose* (ONP) and the *Icelandic Gigaword Corpus* (IGC). The results show that the words belong to several semantic domains that reflect the types of texts that have survived until modern times. Most of the words are robustly attested in Old Norse sources, although there are a few exceptions. Large majority of the words can be found in Modern Icelandic texts, but to a varying degree. Limits of the corpus material makes it difficult to analyse some of the words. The result indicate that the words labelled ‘ancient’ can be divided into three main groups: a) words that are poorly attested and should perhaps not be included in the lexicographic description of Modern Icelandic; b) words that are likely to occur sometimes in Modern Icelandic; c) words that function as other inherited Old Norse words and perhaps do not require a special label or should have an additional sense in the DCI.

Keywords Modern Icelandic, Old Norse, historical lexicology

Contact information

Ellert Thor Johannsson
The Arni Magnusson Institute for Icelandic Studies
Laugavegi 13, IS-105, Reykjavík, Iceland
etj@hi.is

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Claudia Lauer/Birgit Herbers

HAPAX LEGOMENA IN DER DEUTSCHSPRACHIGEN LITERATUR DES MITTELALTERS. BEDINGUNGEN, VERFAHREN UND BEDEUTUNGEN

Ein Projektbericht

Keywords *Hapax legomenon*; Historische Semantik; Deutschsprachige Literatur des Mittelalters (12.–14. Jahrhundert)

Das avisierte DFG-Projekt „Wort(er)findungen. Die Kunst und Semantik sprachlicher Einmalbildungen in der weltlichen Literatur des Mittelalters (12.–14. Jahrhundert)“ befasst sich mit einer besonderen Gruppe von Wörtern, die in einem bestimmten Textkorpus oder einer bestimmten Zeit nur ein einziges Mal belegt sind: sog. *Hapax legomena.* Derartige Einmalbildungen lassen sich nicht nur regelmäßig in der deutschsprachigen Literatur des Mittelalters greifen. Sie besitzen dabei oft auch spezielle semantische Eigenschaften und poetisch-rhetorische Funktionen, die sie in die Nähe von sog. Ad-hoc-Bildungen und situativen Wortneuschöpfungen (vgl. Hohenhaus 1996) rücken.

In der germanistisch-mediävistischen Forschung werden diese Wörter überwiegend nur als Gruppe benannt (vgl. Harm 2015, S. 119), auch existieren nurmehr punktuell Beiträge zu Einzelwörtern (vgl. Lühr 1990; Steinmetz 2000), d. h. eine systematische Erfassung und profunde Bedeutungserschließung dieser ‚einmaligen Worte‘, die über das hinausgeht, was ein lexikographisches Werk bieten kann, steht weitgehend aus. Folgerichtig reagiert das Projekt auf dieses Desiderat: Ausgehend von einem eigenen sprach- und literaturwissenschaftlichen Ansatz und einem extra entwickelten methodischen Zugriff zielt das Projekt in einem neuartigen interdisziplinären Verbund von Lexikographie, mediävistischer Sprach- und Literaturwissenschaft sowie Philologie und Digital Humanities erstmals überhaupt auf die Erschließung eines breiten Datenbestands sprachlicher Einmalbildungen in der weltlichen Literatur des Mittelalters (12.–14. Jahrhundert), die nicht nur im Rahmen einer öffentlich zugänglichen lexikalisch-semantischen Datenbank gesammelt, sondern auch sprach- und literaturwissenschaftlich analysiert sowie lexikonartig aufbereitet werden. Ein wesentlicher Grundbaustein der avisierten Online-Datenbank ‚Hapaxonwoerterbuch‘ sind dabei sog. ‚Wortvisitenkarten‘, die – vergleichbar einem Wörterbuch-/Lexikon-Artikel – die wichtigsten sprach- und literaturwissenschaftlichen Angaben der jeweiligen Einmalbildung beinhalten: zu Werk und Autor, literarischem Kontext, sprachlichen Verfahren der Wortbildung, Übersetzungsmöglichkeiten und literarische(n) Bedeutung(en) innerhalb des Werkes. Diese ‚Wortvisitenkarten‘ sollen nicht nur aufgerufen und wie in einer Art Lexikon jeweils Auskunft über die Bedingungen, Bedeutungen und Funktionen der sprachlichen Einmalbildung geben. Ziel ist es, die Inhalte dieser Karten auch untereinander zu verknüpfen, um dynamische Suchabfragen zu ermöglichen (Autor, Werk, Zeitraum, Wortbildungsmuster etc.) und so weitere Ergebnisse und Erkenntnisgewinne für Nutzer:innen zu generieren.

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Das Vorhaben besitzt so aufs Ganze gesehen den Status eines wissenschaftlichen Grundlagenprojekts und lässt in dreifacher Hinsicht innovative Ergebnisse und Erkenntnisgewinne erwarten: Erstens wird ein valider Datenbestand mittelhochdeutscher Einmalbildungen erhoben und online zugänglich gemacht, zweitens eröffnet die interdisziplinäre Auswertung und lexikonartige Aufbereitung des Datenmaterials neue sprach- und literaturwissenschaftliche Verständniswege zur Kunst und Semantik der deutschsprachigen Dichtung des Mittelalters, und drittens kann das Projekt damit insgesamt neue Einblicke in die sprachliche und literarische Kreativität und Innovationskraft der deutschsprachigen Literatur des Mittelalters geben, die nicht zuletzt auch Brücken zur modernen Sprache, Kunst und Kultur von Worterfindungen schlagen, wie sie heute für jede:n Sprachbenutzer:in literarisch, aber auch im gesellschaftlichen Alltag greifbar sind (vgl. Neologismenwörterbuch 2021).

Literatur

Harm, V. (2015): Einführung in die Lexikologie. Darmstadt.

Hohenhaus, P. (1996): Ad-hoc-Wortbildung. Terminologie, Typologie und Theorie kreativer Wortbildung im Englischen. Frankfurt a. M.

Lühr, R. (1990): Hapax legomena in der althochdeutschen Glossenüberlieferung. In: Sprachwissenschaft 15, S. 164–183.

Neologismenwörterbuch (2021): https://www.owid.de/docs/neo/start.jsp (Stand: 21.03.2022).

Steinmetz, R.-H. (2000): Tristans *erbeminne*. Versuch über vier *hapax legomena* bei Gottfried von Straßburg. In: Zeitschrift für deutsches Altertum und deutsche Literatur (ZfdA) 129, S. 388–408.

Kontaktinformationen

Claudia Lauer
Johannes Gutenberg-Universität Mainz
lauercl@uni-mainz.de

Birgit Herbers
Johannes Gutenberg-Universität Mainz/
Akademie der Wissenschaften und der Literatur Mainz
herbers@uni-mainz.de

Geda Paulsen/Ene Vainik/Maria Tuulik/Ahti Lohk

THE MORPHOSYNTACTIC PROFILE OF PROTOTYPICAL ADJECTIVES IN ESTONIAN

Keywords Lexical categories; morphosyntax; lexicography; language technology; Estonian

The current direction in Estonian lexicography is a unification of lexical resources (dictionaries and term bases) into a central superdictionary, the EKI Combined Dictionary (CombiDic), supported by the dictionary writing system Ekilex. The lexicographic work is moving towards a higher degree of automation and processing of corpora (Koppel et al. 2019; Tavast et al. 2020) The lexical database of Ekilex includes automatically generated lists over dictionary entry candidates, requiring assessment of their degree of lexicalisation. An urgent lexicographic issue is providing the underspecified Ekilex entries with PoS tags and assessing the candidates for their potential status as a lexical entry. Today, 72% of the total number of the public CombiDic headwords miss the PoS tag.

In this study, we focus on one of the most ambiguous parts of speech posing categorisation problems for the lexicographers, the adjective (Paulsen et al. 2019, p. 327). To clarify this issue, we aim to develop a multi-parameter solution for determining the relative adjectiveness of a word or a word form, e. g., the adjectivizing participles or nominals (for the border areas of adjectives with other lexical classes in Estonian, see Vainik/Paulsen/Lohk 2020).

In our vision, the properties of a concrete word can be compared with the profile of a typical adjective, using the profile as a similarity measure. We have established and tested the characteristic attributes of the adjective in a previous study (Tuulik et al. in press), using six morphosyntactic parameters detectable in the corpus. The result of the experiment was that the tested parameters were, to different degrees, able to differentiate adjectival morphosyntactic behaviour.

To establish the exact boundaries of the profile of prototypical adjectives, the parameters should be tested on a larger sample of adjectives that represent the best examples of its category. Our aim in the present study is to find the representative profile of the prototypical adjective based on a larger sample of predefined adjectives and setting the threshold value for classifying the questionable word form as an adjective. The normal distribution of the parameter values will be used to distinguish adjectives from other words and used as a comparison to the corresponding values of the unclear cases.

The basis for the analysis is the largest corpus of contemporary Estonian, the Estonian National Corpus 2019 with 1.5 billion words. The corpus is lemmatised, tagged and disambiguated with the EstNLTKv.1.6 toolkit (Laur et al. 2020). The selection of test words contains 100 words extracted by random sampling from the 554 most central Estonian adjectives included in the Basic Estonian Dictionary (Kallas et al. 2014). The Euclidean distance analysis calculated from the profile of the prototypical adjective is the measure of a word form's similarity vs. difference according to its behaviour in the corpus.

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

References

Laur, S./Orasmaa, S./Särg, D./Tammo, P. (2020): EstNLTK 1.6: Remastered Estonian NLP Pipeline. In: Proceedings of the 12th Language Resources and Evaluation Conference. Marseille, pp. 7152–7160.

Ekilex (2022): https://ekilex.eki.ee/ (last access: 20-03-2022).

Estonian National Corpus 2019 = Koppel, K./Kallas, J. (2020): Eesti keele ühendkorpus 2019. https://doi.org/10.15155/3-00-0000-0000-0000-08565L.

CombiDic = Hein, I./Kallas, J./Kiisla, O./Koppel, K./Langemets, M./Leemets T./Melts, M./Mäearu, S./Paet, T./Päll, P./Raadik, M./Tiits, M./Tsepelina, K./Tuulik, M./Uibo, U./Valdre, T./Viks, Ü./Voll, P. (2020): The EKI Combined Dictionary. Institute of the Estonian Language. https://sonaveeb.ee (last access: 20-03-2022).

Kallas, J./Tiits, M./Tuulik, M./Koppel, K./Jürviste, M. (2014): Eesti keele põhisõnavara sõnastik [The Basic Estonian Dictionary]. Tallinn.

Koppel, K./Tavast, A./Langemets, M./Kallas, J. (2019): Aggregating dictionaries into the language portal Sõnaveeb: issues with and without a solution. In: Kosem, I./Zingano Kuhn, T./Correia, M./Ferreria, J. P./Jansen, M./Pereira, I./Kallas, J./Jakubíček, M./Krek, S./Tiberius, C. (eds.): Proceedings of the eLex 2019 Conference. 1–3 October 2019, Sintra, Portugal. Brno, pp. 434–452.

Paulsen, G./Vainik, E./Tuulik, M./Lohk, A. (2019): The lexicographer's voice: word classes in the digital era. In: Kosem, I./Zingano Kuhn, T./Correia, M./Ferreria, J. P./Jansen, M./Pereira, I./Kallas, J./Jakubíček, M./Krek, S./Tiberius, C. (eds.): Proceedings of the eLex 2019 Conference. 1–3 October 2019, Sintra, Portugal. Brno, pp. 434–452.

Vainik, E./Paulsen, G./Lohk, A. (2020): A typology of lexical ambiforms in Estonian. In: Gavriilidou, Z./Mitsiako, M./Fliatouras, A. (eds.): Lexicography for Inclusion. Proceedings of the 19th EURALEX Congress, 7-9 September 2021, Alexandroupolis. Volume 1. Alexandroupolis, pp. 119–130.

Tuulik, M./Vainik, E./Lohk, A./Paulsen, G. (in press): Kuidas ära tunda adjektiivi? Korpuskäitumise mustrite analüüs [How to recognize adjectives? An analysis of corpus patterns]. The Estonian Papers in Applied Linguistics.

Tavast A./Koppel, K./Langemets, M./Kallas, J. (2020): Towards the Superdictionary: Layers, Tools and Unidirectional Meaning Relations. In: Gavriilidou, Z./Mitsiako, M./Fliatouras, A. (eds.): Proceedings of XIX EURALEX Congress: Lexicography for Inclusion, Volume 1., Greece, pp. 215–223.

Contact information

Geda Paulsen
Institute for the Estonian Language/Uppsala University
geda.paulsen@eki.ee

Ene Vainik
Institute for the Estonian Language
ene.vainik@eki.ee

Maria Tuulik
Institute for the Estonian Language
maria.tuulik@eki.ee

Ahti Lohk
Institute for the Estonian Language
ahti.lohk@eki.ee

Acknowledgements

This work is supported by Estonian Research Council grant PSG227.

Pius ten Hacken/Renáta Panocová

THE ETYMOLOGY OF INTERNATIONALISMS

Evidence from German and Slovak

Abstract In the etymological information for a word in a dictionary, the first question to be answered is whether the word is a borrowing or the result of word formation. Here, we consider this question for internationalisms ending in *-ation* in German and in *-ácia* in Slovak. In German, *-ation* is a suffix that attaches to verbs in *-ieren*. For these verbs, it is in competition with *-ung*. In Slovak, *-ácia* is a suffix that attaches to bases of Latin or Greek origin. The corresponding verbs are often backformations. Most Slovak verbs also have a nominalization in *-nie*. In order to investigate to what extent the nouns in *-ation* or *-ácia* are borrowings or derived from the corresponding verbs in German and Slovak, we took a random sample of English nouns in *-ation* for which OED gives a corresponding verb. For this sample, we checked whether the cognate noun in *-ation* or *-ácia* is attested in standard dictionaries and in corpora. Then we did the same for the corresponding verbs and the nouns in *-ung* or *-nie*. Finally, we checked the frequency of these words in DeReKo for German and SNK for Slovak. On this basis, we found evidence that *-ation* in German has a slightly different status to *-ácia* in Slovak. This status affects the relationship to the corresponding verbs and to the nouns in *-ung* or *-nie*. Such generalizations are important as background information for specifying etymological information in dictionaries, especially for languages where first attestations dates are not readily available.

Keywords Borrowing; word formation; reanalysis

Contact information

Pius ten Hacken
Leopold-Franzens-Universität Innsbruck
pius.ten-hacken@uibk.ac.at

Renáta Panocová
Pavol Jozef Šafárik University in Košice
renata.panocova@upjs.sk

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Neologisms and Lexicography

XX EURALEX INTERNATIONAL CONGRESS
DICTIONARIES AND SOCIETY
12-16 July 2022, Mannheim, Germany

Ieda Maria Alves/Bruno Maroneze

FROM SOCIETY TO NEOLOGY AND LEXICOGRAPHY

Relationships between morphology and dictionaries

Abstract This paper aims at verifying if the most important online Brazilian Portuguese dictionaries include some of the neologisms identified in texts published in the 1990s to 2000s, formed with the elements *ciber-*, *e-*, *bio-*, *eco-* and *narco,* which we refer to as *fractomorphemes* / *fracto-morphèmes.* Three online dictionaries were analyzed (*Aulete*, *Houaiss* and *Michaelis*), as well as *Vocabulário Ortográfico da Língua Portuguesa* (*VOLP*). We were able to conclude that all three dictionaries and VOLP include neologisms with these elements; Michaelis and VOLP do not include separate entries for bound morphemes, whereas Houaiss includes entries for all of them and Aulete includes entries for *bio-*, *eco-* and *narco-*. Aulete also describes the neological meaning of *eco-* and *narco-*, whereas Houaiss does not.

Keywords Fracto-morphèmes; neologisms in Brazilian Portuguese; Brazilian Portuguese dictionaries

Contact information

Ieda Maria Alves
Universidade de São Paulo
iemalves@usp.br

Bruno Maroneze
Universidade de São Paulo
maronezebruno@yahoo.com.br

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Mikyung Baek/Jinsan An/Yelin Go

A DISTRIBUTIONAL APPROACH TO KOREAN SEMANTIC NEOLOGISMS

Identifying their first occurrences and investigating their spread

Keywords Semantic neologisms; distributional approach; first occurrence; semantic change

Diachronic corpora have enabled extensive research on semantic change (Gulordava/Baroni 2011; Kim/ et al. 2014; Hamilton/Leskovec/Jurafsky 2016), from which both theoretical linguistics and applied linguistics can benefit. In lexicography in particular, the description of new meanings proves crucial not only for communication, but also language teaching and translation, including automatic translation. In neologism research, words that display a shift in meaning are referred to as semantic neologisms and may encompass various types of semantic changes. Broadly speaking, these may include metaphorical and metonymic uses, as well as widening and narrowing of meanings (Ullman 1957; Boussidian 2013; Nam/Lee/Choi 2018). While Nam/Lee/Choi (2018) also add the time of first occurrence as an important criterion for determining semantic neologisms, a satisfying systematic methodology for identifying semantic neologisms is yet to be developed. Thus, this research seeks to overcome this shortcoming as an empirical study proposing a methodology to identify the time of first occurrence of four semantic neologisms (Examples 1–4) as well as the period when they settled in language.

(1) 인생 *insayng* ‘lifetime’ [S1], ‘best’ [S2]

(2) 폭풍 *phokphwung* ‘storm’ [S1], ‘extreme’ [S2]

(3) 공화국 *konghwakwuk* ‘republic (as a political system)’ [S1], ‘country (of a particular type)’ [S2]

(4) 영접 *yengcep* ‘formal reception (of an official person)’ [S1], ‘welcome (of object/food)’ [S2]

In order to achieve this, we first built a set of Web-crawled corpora consisting of news articles dating from January 1990 to December 2019, each subset corresponding to one semantic neologism in (1–4). We then extracted all noun phrases that comprise the above semantic neologisms either as a modifier or as a modificand. As this included both noun phrases where the word is used in its original meaning and noun phrases where it is used with a new meaning, we sorted out the noun phrases regarded as semantic neologisms by all three authors for analysis. In the next step, we applied the main two methodologies used in semantic change research, that is, collocation analysis (within a left/right window of four words) and word similarity analysis. The results yielded by these two methodologies only proved meaningful in the case of *phokphwung* ‘storm’ [S1], ‘extreme’ [S2], but were rather inconclusive for the others. Thus, we proposed another methodology here, which consisted of analyzing the collocate types (vs tokens) and the semantic categories of the collocates.

The collocate type analysis showed a clear pattern in all four semantic neologisms, which allowed us to determine the time of first occurrence and the establishment period, as shown in pictures 1–4.

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

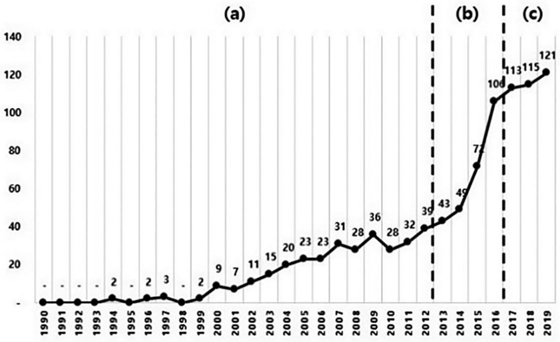


Fig. 1: Collocate type pattern for ‘*insayng* [S2] + N’

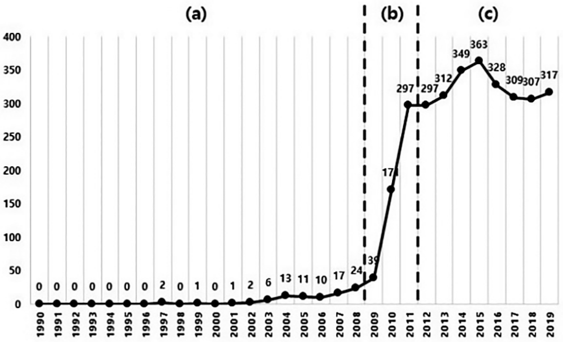


Fig. 2: Collocate type pattern for ‘*phokphwung* [S2] + N’

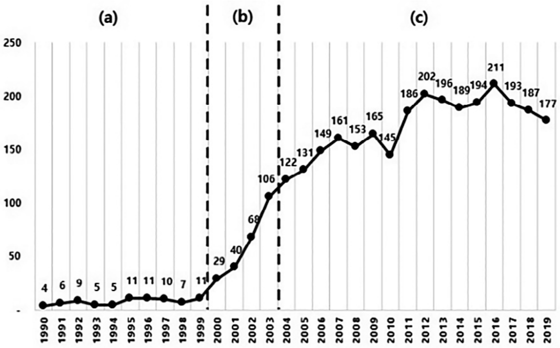


Fig. 3: Collocate type pattern for ‘N + *konghwakwuk* [S2]’

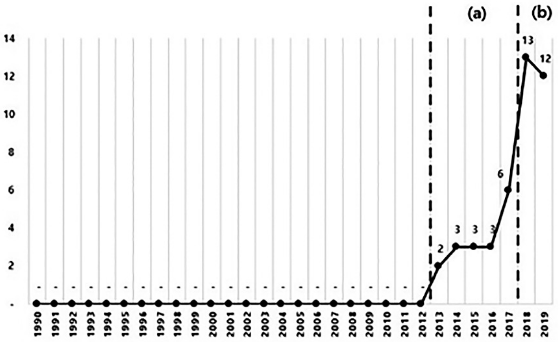


Fig. 4: Collocate type pattern for ‘N + *yengcep* [S2]’

Frequency trends can be divided in three phases (*a*, *b*, *c*). Phase *a* corresponds to a low number of collocate types, which consists of nonce expressions. Phase *b* shows a sudden and rapid diversification of collocate types, corresponding to Schmid's 'consolidation' stage (2008). Finally, the diversification of types slows down in phase *c*, which indicates that the new meaning is being established. While this methodology seems to provide systematicity, it also presents some limitations. For example, the word *seytay* 'generation' is also typically used in such constructions as 'N + *seytay*'. When analyzing the collocate types of the *seytay* noun phrases, it appears to show a similar pattern. However, these noun phrases are not indicative of a semantic shift in the word *seytay*. One explanation lies in the semantic category the collocate falls into: in the case of 'N + *seytay*', N would most likely fall in the category of 'society'. This would be in the same line with the example presented by Gulodarva/Baroni (2011) whereby the distribution of the collocates for *sleep* may have changed (e. g., deep sleep (1960s); REM sleep (1990s)) but without impacting the meaning of *sleep*. However, as the semantic category analysis shows, the diversification of collocate types is concomitant with a clear semantic shift in the semantic category of the collocates, as shown in Table 1.

	[S1]	**[S2]**
insayng 'lifetime'	*insayng mokphyo* 'life goal'; *insayng keyhoyk* 'life plan'	*insayng yenghwa* 'best movie'; *insayng meynyu* 'best menu'
phokphwung 'storm'	*phokphwung kyengpo* 'storm warning'; *phokphwung yeypo* 'storm forecast'	*phokphwung sengcang* 'extreme growth'; *phokphwung cilcwu* 'extreme speeding'
konghwakwuk 'republic'	*mincwu konghwakwuk* 'democratic republic'; *cheykho konghwakwuk* 'Czech Republic'	*senghyeng konghwakwuk* 'the country of plastic surgery'; *aphathu konghwakwuk* 'the country of apartments'
yengcep 'formal reception'	*taythonglyeng yengcep* 'formal reception of the President'; *chongli yengcep* 'formal reception of the Prime Minister'	*chikhin yengcep* 'welcome of the chicken'; *sutheyikhu yengcep* 'welcome of the steak'

Table 1: Comparative table of the collocates for the semantic neologisms under study

The semantic category of the collocates for the original meaning of the words in question is rather restricted: 'life/lifestyle' for *insayng*; 'nature/weather forecast' for *phokphwung*; 'political system' for *konghwakwuk*; 'political figure' for *yengcep*. However, the collocate types have greatly diversified in terms of semantic categories in the case of semantic neologisms. Only *yengcep* shows a more restricted semantic category pattern; nonetheless, the categories of the original and new meanings are semantically different enough to consider *yengcep* as a semantic neologism.

The methodology we presented in this study showed that the distribution of the collocate types somewhat evidences semantic change. A sole quantitative analysis proved insufficient and instead, a qualitative analysis allowed us to refine our findings. Nonetheless, our study presents a number of limitations, including the scale of our data as our corpora only covers a time span of thirty years and the selection criterion for the semantic neologisms under study, which is ultimately based on our intuition of native speakers.

References

Boussidan, A. (2013): Dynamics of semantic change. PhD Dissertation. Lyon.

Gulordava, K./Baroni, M. (2011): A distributional similarity approach to the detection of semantic change in the Google Books Ngram corpus. In: Proceedings of the GEMS 2011 Workshop on Geometrical Models of Natural Language Semantics. Edinburgh, pp. 67–71.

Hamilton, W. L./Leskovec, J./Jurafsky, D. (2016): Diachronic word embeddings reveal statistical laws of semantic change. arXiv preprint arXiv:1605.09096.

Kim, Y./Chiu, Y. I./Hanaki, K./Hegde, D./Petrov, S. (2014): Temporal analysis of language through neural language models. arXiv preprint arXiv:1405.3515.

Nam, N. Y./Lee, S. J./Choi, J. (2018): Research trends and issues on semantic neology using web corpus. In: Korean Lexicography 31, pp. 55–84.

Schmid, H. (2008): New words in the mind: concept-formation and entrenchment of neologisms, In: Anglia 126 (1), pp. 1–36.

Ullmann, S. (1957): The principles of semantics. Oxford.

Contact information

Mikyung Baek
Kyungpook National University
bmg0128@hanmail.net

Jinsan An
Kyungpook National University
siveking@naver.com

Yelin Go
Kyungpook National University
goyelin08@naver.com

Jun Choi/Hae-Yun Jung

ON LOANS IN KOREAN NEW WORD FORMATION AND IN LEXICOGRAPHY

Abstract This study examines a list of 3,413 neologisms containing one or more borrowed item, which was compiled using the databases built by the Korean Neologism Investigation Project. Etymological aspects and morphological aspects are taken into consideration to show that, besides the overwhelming prevalence of English-based neologisms, particular loans from particular languages play a significant role in the prolific formation of Korean neologisms. Aspects of the lexicographic inclusion of loan-based neologisms demonstrate the need for Korean neologism and lexicography research to broaden its scopes in terms of methodology and attitudes, while also providing a glimpse of changes.

Keywords Neologisms; lexicography; loans; clipping; blending; word formation

Contact information

Jun Choi
Kyungpook National University
c-juni@hanmail.net

Hae-Yun Jung
Kyungpook National University
haeyun.jung.22@gmail.com

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Emmanuel Cartier

DIACHRONIC SEMANTIC EVOLUTION AUTOMATIC TRACKING: A PILOT STUDY IN MODERN AND CONTEMPORARY FRENCH COMBINING DEPENDENCY ANALYSIS AND CONTEXTUAL EMBEDDINGS

Keywords Semantic change; NLP; lexical diffusion; emergence; diffusion and institutionalization; dependency analysis; word embeddings; contextual embeddings

The vocabulary of any given language is continuously evolving: new form-meaning pairs are forged, some signs fall into disuse, and some form-meaning pairs evolve by adding new meanings, losing others or slightly shifting the existing ones. Tracking and describing the dynamism of the vocabulary is one of the main tasks of lexicography. The massive availability of digital or digitalized corpora gives the discipline an unprecedented material for its study, enabling to setup systems able to mine monitor corpora to update the dictionary entries, usages and meanings with specific tools detecting neology and more generally semantic change. However, if it is relatively simple to identify new linguistic signs that appear and those that are no longer used (e.g. among others Kerremans and Prokic 2018; Cartier 2019), it is much more complex to identify semantic neology as it does not manifest at the level of the form itself. Several methods have been proposed by Natural Language Processing (NLP) to try to identify these evolutions of meaning.

Lexical change detection systems have followed advances in NLP methods: after the first systems essentially based on frequency changes (for example Gulordova/Baroni 2011), systems used *word embeddings* (e.g. Kim et al. 2014) and more recently *contextual embeddings* (Hu et al. 2019; Martinc et al. 2019; Giulianelli et al. 2020). These latter systems generally proceed by grouping the contextual vector representations of the different uses into clusters of meaning, then detect changes according to different metrics (Monteirol et al. 2021). Current systems still face many limitations. Mainly, the opacity of neural models does not make it possible to characterize these evolutions, in particular it is difficult, if not impossible, to link the semantic changes to linguistic morphological, syntactic or lexico-syntactic features, or to categorize the types of changes (extension, restriction, metaphor, metonymy, etc.).

To this end, other perspectives have been proposed, based on the hypothesis that meanings are correlated with prototypical co-occurrences or collocations and that an evolution of these elements could denote an evolution of meaning. The most advanced work in this perspective was proposed by (Gries 2012). It is based on the hypothesis that meanings are correlated to prototypical lexical-syntactic patterns (*behavioral profile*). The notion was then extended to that of *dynamic behavioral profile* (Jansegers and Gries 2017) by considering that changes in patterns correlated with semantic evolutions. For example, they show the progressive grammaticalization of one meaning of the verb *sentir*, towards the discourse marker *lo siento* ('I'm sorry') through a visualization by Multidimensional Scaling Maps

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

(MDS) built from manually annotated linguistic features on contexts of the verb, and a probabilistic model. This method has the disadvantage of a manual annotation, which is time consuming and may bias the results.

In this work, we propose, within the framework of a project aiming at building a reference dataset of semantic evolution in modern and contemporary French (1800–today), from a corpus of journalistic texts from the French National Library website (1850–1940) and a contemporary newspaper corpus from the web (2014–2021) to study a sample of polysemous words (nouns and verbs). To trace their evolution, we explore a combination of the above approaches: on the one hand, *contextual embeddings*, which allow to get a very fine representation of the context and to cluster the vectorized representations of individual occurrences, thus discovering clusters of meaning and their evolution through time; as pre-trained model, we use CamemBERT (Martin et al. 2019), a state-of-the-art model for French; on the other hand, a dependency analysis by means of a state-of-the-art parser (Straka 2018), allowing not only to annotate each token with morphological but also syntactic features. The latter approach, inspired by (Jenserges and Gries 2017) does not use any manual annotation, and focus on valid lexico-syntactic patterns for the two categories: for nouns, modifier (N ADJ, N prep N), and core argumental constructions (N subject or object of a verb). For verbs, core argumental constructions (subject group, direct and indirect object, subordinate clause).

In this contribution, we will present the different phases of the project (corpora building, choice of lexemes, pre-processing and exploration web platform) and the first lessons learned.

References

Giulianelli, M./Tredici, M.D./Fernández, R. (2020): Analysing lexical semantic change with contextualised word representations. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, July 5–10, 2020. Stroudsburg, PA, pp. 3960–3973.

Gries, St. Th. (2012): Behavioral profiles: a fine-grained and quantitative approach in corpus-based lexical semantics. In: Jarema, G./Libben, G./Westbury, Ch. (eds.): Methodological and analytic frontiers in lexical research. Amsterdam, pp. 57–80.

Gulordava, K./Baroni, M. (2011): A distributional similarity approach to the detection of semantic change in the Google Books Ngram corpus. In: Proceedings of the GEMS 2011 Workshop on Geometrical Models of Natural Language Semantics. Edinburgh, pp. 67–71.

Hu, R./Li, S./Liang, S. (2019): Diachronic sense modeling with deep contextualized word embeddings: an ecological view. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, pp. 3899–3908.

Iavarone, B./Brunato, D./Dell'Orletta, F. (2021): Sentence complexity in context. In: Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, Online Event, June 10, 2021, pp. 186–199.

Jansegers, M./Gries, St. Th. (2017): Towards a dynamic behavioral profile: a diachronic study of polysemous sentir in Spanish. In: Corpus Linguistics and Linguistic Theory 16 (1), pp. 145-187.

Kim, Y./Chiu, Y.-I./Hanaki, K./Hegde, D./Petrov, S. (2014): Temporal analysis of language through neural language models. In: Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science. Baltimore, pp. 61–65.

Kosem, I./Koppel, K./Kuhn, T. Z./Michelfeit, J./Tiberius, C. (2019): Identification and automatic extraction of good dictionary examples: the case(s) of GDEX. In: International Journal of Lexicography 32, pp. 119–137.

Martin, L./Muller, B./Ortiz Suarez, P./Dupont, Y./Romary, L./Villemonte de la Clergerie, E./Seddah, D./Sagot, B. (2020): CamemBERT: a tasty French language model. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, July 5–10, 2020. Stroudsburg, PA, pp. 7203–7219.

Martinc, M./Novak, P. K./Pollak, S. (2020): Leveraging contextual embeddings for detecting diachronic semantic shift. In: Proceedings of the 12th Conference on Language Resources and Evaluation (LREC), Marseille, 11–16 May 2020, pp. 4811–4819.

Montariol, S./Martinc, M./Pivovarova, L. (2021): Scalable and interpretable semantic change detection. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, June 6–11, 2021, pp. 4642–4652.

Raganato, A./Pasini, T./Camacho-Collados, J./Pilehvar, M. T. (2020): XL-WiC: A multilingual benchmark for evaluating semantic contextualization. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, November 16–20, 2020, pp. 7193–7206.

Segonne, V./Candito, M./Crabbé, B. (2019): Using Wiktionary as a resource for WSD: the case of French verbs. In: Proceedings of the 13th International Conference on Computational Semantics, Gothenburg, Sweden, pp. 259–270.

Straka, M. (2018): UDPipe 2.0 Prototype at CoNLL 2018 UD Shared Task. In: Proceedings of CoNLL 2018: The SIGNLL Conference on Computational Natural Language Learning, Brussels, Belgium, pp. 197–207.

Contact information

Emmanuel Cartier
LIPN – RCLN UMR 7030 CNRS, University Sorbonne Paris Nord (Paris, France)
emmanuel.cartier@univ-paris13.fr

Lars Trap-Jensen/Henrik Lorentzen

RECENT NEOLOGISMS PROVOKED BY COVID-19 IN THE DANISH LANGUAGE AND IN THE DANISH DICTIONARY

Abstract Inspired by GWLN 3, we take a look at the new words, meanings, and expressions that have been created during or promoted by the COVID-19 pandemic. The pandemic provides a rare opportunity to follow the rise, spread, and integration of words and expressions in a language that may serve as an illustration of how linguistic innovation in general works. Relevant words were selected from various lists, notably monthly and annual lists of prominent words attested in the corpus of The Danish Dictionary. Analysis of these lists gives an insight into the number of words that stand out month by month and what kinds of words are involved, both in terms of morphological type and of semantic category, with special attention given to neologisms. Finally, we discuss the criteria for selecting which words to include in the dictionary. With this study, Danish is added to the list of languages covered in the GWLN series on COVID-19 neologisms.

Keywords COVID-19; detecting neologisms; corpus-based; temporal dimension; The Danish Dictionary

Contact information

Lars Trap-Jensen
Society for Danish Language and Literature
ltj@dsl.dk

Henrik Lorentzen
Society for Danish Language and Literature
hl@dsl.dk

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Gilles-Maurice de Schryver/Minah Nabirye

TOWARDS A MONITOR CORPUS FOR A BANTU LANGUAGE

A case study of neology detection in Lusoga

Abstract This paper looks at whether, after two decades of corpus building for the Bantu languages, the time is ripe to begin using monitor corpora. As a proof-of-concept, the usefulness of a Lusoga monitor corpus for lexicographic purposes, *in casu* for the detection of neologisms, both in terms of new words and new meanings, is investigated and found useful.

Keywords Monitor corpus; neology detection; new words; new meanings; Bantu; Lusoga

Contact information

Gilles-Maurice de Schryver
BantUGent – UGent Centre for Bantu Studies, Ghent University
& Department of African Languages, University of Pretoria
gillesmaurice.deschryver@UGent.be

Minah Nabirye
BantUGent – UGent Centre for Bantu Studies, Ghent University
minah.nabirye@UGent.be

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Urška Vranjek Ošlak/Helena Dobrovoljc

NEOLOGISMS IN THE LIGHT OF THE NEW SLOVENIAN NORMATIVE GUIDE

Keywords Slovenian standard language; Slovenian orthography; Slovenian neologisms; COVID-19

This contribution presents the treatment of neologisms in the standard Slovenian language as presented in the new Slovenian normative guide, i.e., the accompanying dictionary.

1. Codification of Slovenian standard language

Normative guides provide information about the acceptability of language elements for standard language use. In Slovenian, the standard language is a supra-regional idiom agreed upon by language users that has been used in the written language since the middle of the 19th century. A normative guide (*pravopis* in Slovenian) is a manual consisting of normative rules and an orthographic dictionary. It includes not only a systematic set of basic writing rules at the vowel-letter level (orthography or spelling) but also other consensual norms of the standard language, i.e., the use of lower- and upper-case letters, borrowing, one- or two-word spelling, punctuation, etc. (Dobrovoljc 2016). Along with the monolingual dictionary (*SSKJ* 1970–1991), the normative guide (*Slovenski pravopis* 2001) is one of the two main sourcebooks for the Slovenian language. The normative rules were published in 1990 and the orthographic dictionary followed in 2001.

Due to the considerate time span in which the codification process took place, several discrepancies occurred, resulting in asynchronic codification. Therefore, a new concept had to be created so that the orthographic dictionary could effectively accompany the normative rules. The new Slovenian normative guide consists of: (1) normative rules, which form the theoretical part of the normative guide (*Pravopis 8.0*), (2) an orthographic dictionary, which provides (additional) examples (*ePravopis*), and (3) orthographic categories (*Pravopisne kategorije*), a collection of comments on changes in codification; the main purpose of the latter is to ensure a transparent codification process (Dobrovoljc/Vranjek Ošlak 2021). All these resources are available online (the *Fran* portal).

2. Detecting neologisms

The new approach to the elaboration of normative rules for the Slovenian language is problem-oriented. One of the most important resources used to identify linguistic dilemmas such as neologisms is the Language Counselling Service, the central online language counselling platform for the Slovenian language (*Jezikovna svetovalnica*, managed by the Fran Ramovš Institute of the Slovenian Language at ZRC SAZU). It is used to enquire about ambiguities in the standard language, enabling researchers to identify language description gaps. The Language Counselling Service automatically creates a provisional online language manual. Like similar platforms (cf. Beneš et al. 2017), it also reflects current social circumstances. The platform is connected to the *Fran* portal (Vranjek Ošlak/Dobrovoljc 2021).

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

3. Neologisms in the new Slovenian normative guide

Below, some neologism (sub-)types are listed as they are presented in the new Slovenian normative guide, namely its accompanying dictionary:

1) new common nouns with notation and pronunciation information, as applicable:
   a) neologisms as a result of translation or adaptation of newer borrowed expressions, e.g., *fedžoja* ('feijoa'), *kovč* ('coach'), *timbilding* ('team-building');
   b) appellativization of proper names, e.g., *zika* ('Zika virus'), *marburg* ('Marburg virus');
   c) new acronyms, e.g., *COVID-19, mRNA, PCR*;
   d) transition from acronyms to ordinary words, e.g., *kovid* < *COVID-19*, *sars* < *SARS*, *mers* < *MERS*;
   e) feminine nouns, e.g., *kupka* ('female customer'), *piska* ('female writer');
   f) compounds from prepositional phrases, e.g., *nesimptomatičen* ('asymptomatic'), *anticepilec* ('male anti-vaccinationist');
   g) blended words, e.g., *hekaton* ('hackathon'), *jogalates* ('yogalates').
2) new common noun forms and compounds traditionally included in grammatical information of dictionary entries, e.g., *telešenje* ('embodiment');
3) new proper names (*Android; AstraZeneca*, *BioNTech*, *Vaxzevria; Teams*, *Zoom*);
4) new anthroponyms, e.g., *Ciudaddemexičan*/*Ciudadčan* ('male inhabitant of Mexico City'), etc.

Such neologisms can be challenging for Slovenian language users, especially when 1) the lexeme is not included in any of the existing dictionaries or other manuals; 2) research on written language usage suggests multiple variants; and 3) phonetic realisation of (borrowed) lexemes in Slovenian is difficult or has multiple variants.

References

Beneš, M./Prošek, M./Smejkalová, K./Štěpánová, V. (2017): Interaction between language users and a language consulting centre: challenges for language management theory and research. In: Fairbrother, L./Nekvapil, J./Sloboda, M. (eds.): The language management approach: a focus on research methodology. Frankfurt a.M., pp. 119–140.

Dobrovoljc, H. (2016): Povezljivost pravopisnih pravil in slovarja: sanje pravopiscev 20. stoletja. In: Erjavec, T./Fišer, D. (eds.): Proceedings of the Conference on Language Technologies & Digital Humanities 2016, Ljubljana, 29 September–1 October 2016. Ljubljana, pp. 52–57.

Dobrovoljc, H./Vranjek Ošlak, U. (2021): Codification within reach: three clickable layers of information surrounding The new Slovenian normative guide. In: Kosem, I./Cukr, M./Jakubíček, M./Kallas, J./Krek, S./Tiberius, C. (eds.): Electronic Lexicography in the 21st Century. Proceedings of the eLex 2021 Conference, virtual, 5–7 July 2021. Brno, pp. 637–652.

Fran. https://fran.si/ (last access: 16-03-2022).

Jezikovna svetovalnica. https://svetovalnica.zrc-sazu.si/ (last access: 18-03-2022).

SSKJ (1970–1991, 2014): Slovar slovenskega knjižnega jezika. Ljubljana.

Toporišič, J. (ed.) (1990–2001): Slovenski pravopis. Ljubljana.

Vranjek Ošlak, U./Dobrovoljc, H. (2021): Integration of public engagement mechanisms in an online language counselling platform. In: Pelegrín-Borondo, J./Arias-Oliva, M./Murata, K./Palma, A. M. L. (eds.): ETHICOMP 2021. Moving technology ethics at the forefront of society, organisations and governments. Logroño, pp. 381–390.

Contact information

Urška Vranjek Ošlak
ZRC SAZU, Fran Ramovš Institute of the Slovenian Language
urska.vranjek@zrc-sazu.si

Helena Dobrovoljc
ZRC SAZU, Fran Ramovš Institute of the Slovenian Language/
University of Nova Gorica, School of Humanities
helena.dobrovoljc@zrc-sazu.si

Acknowledgements

This contribution was prepared within the framework of P6-0038 programme funded by the Slovenian Research Agency (ARRS).

Phraseology & Collocations

XX EURALEX INTERNATIONAL CONGRESS
DICTIONARIES AND SOCIETY
12-16 July 2022, Mannheim, Germany

Amparo Alcina

REPRESENTING COLLOCATIONS USING ONTOLOGIES

Keywords Ontology; terminology; collocation

1. Introduction

Dictionaries, lexicographic or terminographic, sometimes collect some typical uses of some words or terms. But methodology for preparing dictionaries rarely includes the systematic detection of contexts of use. Furthermore, computer tools have not been developed to facilitate the systematic or timely collection of this type of information.

Therefore, it is of great interest to develop and apply models and structures that facilitate the collection of contextual information on terms and their formal representation, that can be useful for both humans and machines, so that terminological resources offer access to this information.

2. Databases and knowledge bases

Research on the lexicon often requires storing data and processing it in various ways in databases. These systems facilitate the work of storing, maintaining and managing linguistic data; process its ordering by different criteria; manage the filtering, consultation, modification and updating of data; and its display on the screen or its printing in various ways (Cabré Castellví 1999; Sager 1990). They are used in the investigation of the different linguistic levels (phonology, morphology, lexicon, syntax, semantics).

In the last twenty years, these databases have improved in terms of technical aspects (storage capacity, friendlier interfaces), interaction with other systems (automatic extraction of terms, assisted and automatic translation) and in terms of developing standards, such as LMF, TMF, TBX, to ensure lexical data exchange and reuse (Francopoulo (ed.) 2013; Francopoulo et al. 2007; Melby 2015).

However, sometimes, databases do not allow us to model the data in the way we would like, to develop improvements in the access and query of the data or in the organization or to diversify the presentation of the data structures. Some authors point at lexical networks to represent lexical (Lamb 1999; Moerdijk 2008; Moerdijk/Tiberius/Niestadt 2008; Polguère 2014) and terminological (Buendía Castro/Montero Martínez/Faber 2014; Faber 2012) information and relations.

The limitations that the databases present come largely from technical limitations, derived from the fact that their base structure, records and fields, generates a rigid organization of the data in the form of a table.

To achieve more flexible representations of lexical data, it is necessary to abandon the rigidity of the record structures and fields that databases currently employ and move to represent the lexical data with more flexible systems, such as knowledge bases or ontologies.

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Ontologies, through their structure of classes, properties and individuals, allow to establish data relationships that result in network structures, much closer to what the representation of the lexicon requires.

3. Representing collocations in an ontology model

In our previous research, the processing of lexical data through ontologies has allowed us to advance in the hierarchical organization of concepts and their detailed analysis in characteristics and attributes. This has been of great help in an onomasiological access to the terms, the visual representation of the structure of their relationships and the elaboration of lexicographically correct definitions (Alcina 2020; Alcina/Valero Doménech 2018).

More recently, we have added to this ontological model the ability to represent the collocations of terms. For this, we have implemented the terms of industrial ceramics and their collocations. From a theoretical point of view, the lexical functions of Explanatory and Combinatorial Lexicography (Jousse/Bouveret 2003; L'Homme 2020; Mel'čuk/Clas/Polguere 1995; Polguère 2003) have helped us to establish the binary relationships between terms.

In this ontological model, the terms have been represented as individuals, and they have been classified as class instances. Linguistic concepts (such as Term, Grammar Category, or Concept Type) are implemented as Classes. Collocations have been implemented as properties that link two terms (individuals). We can create subtypes of collocations using restrictions for these properties following the caracterizations of lexical functions in Meaning Text Theory (such as Oper, Real, etc.), or other lexical semantics theories. In Figure 1, you can see the collocations of the term *absorber* in the ceramic industry terminology.

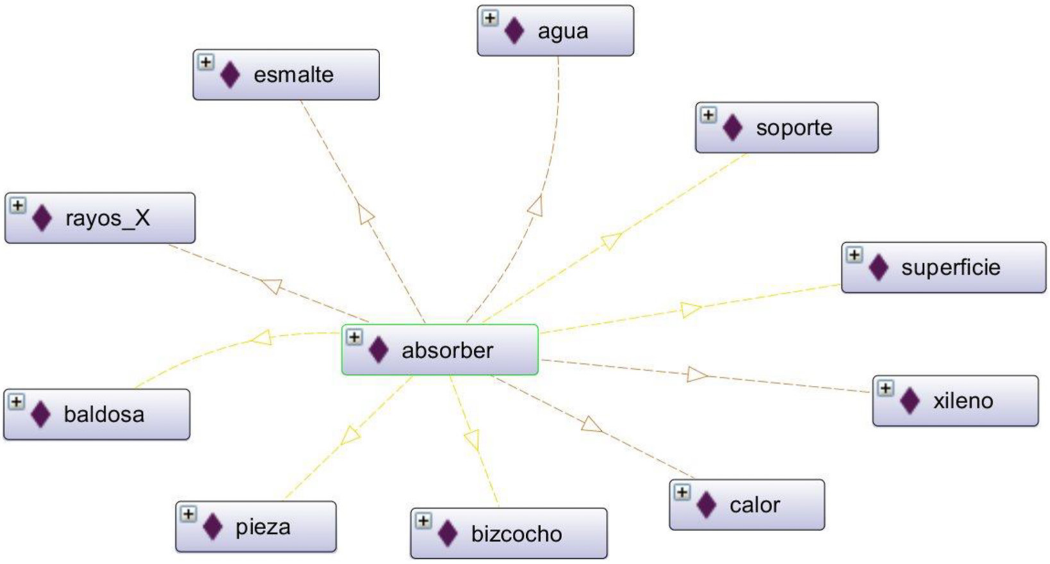


Fig. 1: Collocations of 'absorber' in Protégé

We have developed the research using the Protégé ontology editor (Musen 2015). We will present how we have used the tool: the class configuration, properties used to model the collocations of the industrial ceramics language.

References

Alcina, A. (2020): La representación de relaciones conceptuales en una ontología. In: Ibáñez Rodríguez, M. (ed.): Enotradulengua. Vino, lengua y traducción. Berlín.

Alcina, A./Valero Doménech, E. (2018): Description of the terminological concept in an ontology. In: TOTH 2017: Terminologie & Ontologie: Théories et Applications. Chambéry, pp. 161–179. http://hdl.handle.net/10234/188438 (last access: 05-04-2022).

Buendía Castro, M./Montero Martínez, S./Faber, P. (2014): Verb collocations and phraseology in EcoLexicon. In: Yearbook of Phraseology 5 (1), pp. 57–94.

Cabré Castellví, M. T. (1999): Informática y terminología. In: Blecua, J. M./Clavería, G./Sánchez, C./Torruella, J. (eds.): Nuevas tecnologías en los estudios filológicos. Barcelona.

Faber, P. (ed.) (2012): A cognitive linguistics view of terminology and specialized language. Berlin.

Francopoulo, G. (ed.) (2013): LMF lexical markup framework. London.

Francopoulo, G./Bel, N./George, M./Calzolari, N./Monachini, M./Pet, M./Soria, C. (2007): Lexical Markup Framework: ISO Standard for Semantic Information in NLP Lexicons. Tübingen.

Jousse, A. L./Bouveret, M. (2003): Lexical functions to represent derivational relations in specialized dictionaries. In: Terminology 9 (1), pp. 71–98.

L'Homme, M. C. (2020): Lexical semantics for terminology: an introduction. Amsterdam.

Lamb, S. M. (1999): Pathways of the brain: the neurocognitive basis of language. Amsterdam.

Mel'čuk, I./Clas, A./Polguere, A. (1995): Introduction à la lexicologie explicative et combinatoire. Louvaine la Neuve.

Melby, A. K. (2015): TBX: A terminology exchange format for the translation and localization industry. In: Kockaert, H. J./Steurs, F. (eds.): Handbook of terminology, pp. 393–424.

Moerdijk, F. (2008): Frames and semagrams. Meaning description in the General Dutch Dictionary. In: Bernal, E./DeCesaris, J. A. (eds.). Proceedings of the XIII EURALEX International Congress. Barcelona, pp. 561–569.

Moerdijk, F./Tiberius, C./Niestadt, J. (2008): Accessing the ANW dictionary. In: Coling 2008: Proceedings of the Workshop on Cognitive Aspects of the Lexicon (COGALEX 2008). Manchester, pp. 18–24.

Musen, M. A. (2015): The Protégé project: A look back and a look forward. In: AI Matters. Association of Computing Machinery Specific Interest Group in Artificial Intelligence 1 (4), pp. 4–12.

Polguère, A. (2003): Lexicologie et sémantique lexicale. Montreal.

Polguère, A. (2014): From writing dictionaries to weaving lexical networks. In: International Journal of Lexicography 27 (4), pp. 396–418.

Sager, J. C. (1990): A practical course in terminology processing. Amsterdam/Philadelphia.

Contact information

Amparo Alcina
TecnoLeTTra Research Group, Universitat Jaume I
alcina@uji.es

Acknowledgements

This research is part of the project: PRO-ONTODIC: Protocolos para la creación de diccionarios terminológicos basados en ontologías (Modelo ONTODIC), funded by Universitat Jaume I (Ref. UJI-B2018-65).

Valeria Caruso/Angela Caiazza

LEXICOGRAPHY AND PHRASEOLOGY OF ROMANCE LANGUAGES

The case of procomplement verbs

Keywords Multiword expressions; phraseology; Romance languages; clitic pronouns; Italian lexicography; French lexicography; Spanish lexicography

The *Grande Dizionario dell'Uso* edited by Tullio De Mauro in 1999 has brought about a revolution in Italian lexicography, especially for its focus on phraseology. Multiword expressions are listed as lemmas and a new class of verbs is introduced and given for the first time a proper metalinguistic label, i.e. *procomplement verbs* (PV). In the preface it is explained that these verbs display "a meaning of their own which either cannot be linked to the main verb or which is already steadily lexicalized: *avercela*/*avere* (En. 'bear a grudge'/'have'), *cavarsela*/*cavare* ('get by'/'pull out'), *fregarsene*/*fregare* ('not to care about'/'rub')". The clitic pronouns *la* and *ne* added to the main verbs in the examples (*avere*, *cavare*, *fregare*) are anaphoric to a hidden argument which is difficult to retrieve and causes idiomaticity. This might have inspired the label 'procomplement' used by De Mauro, with *pro-* meaning 'in place of' (see $_2$*pro-* in *Nuovo De Mauro*).

Indeed, *procomplement* verbs allow for a better characterization of the lexicalised clitics to be found in Romance languages and are considered to be typical of contemporary spoken Italian (D'Achille 2003; Russi 2008). More recently, a few types of PVs are gaining some attention in research carried out on the Spanish language (Arellano 2020; Bibis/Roberge 2004; Espinal 2009) though De Mauro's classification is still unacknowledged and more systematic analyses should be carried out (Russi 2008). In the *RAE dictionary*, for example, *pagarla* ('to pay for an offence') is given as an instance of "locuciones verbales coloquiales" ('colloquial verbal idiom') and in the *Le Grand Robert* procomplement verbs are frequently listed as example sentences. Furthermore, in bilingual lexicography reversibility from one language to another is scant and suitable morphosyntactic descriptions for foreign speakers are still missing (Viviani 2007). As an example, dictionaries should point out that PVs generally have fixed pronouns, such as in *finiscila, altrimenti le prendi!* ('stop it or you'll get a beating/spanking!'), whereas in compositional verb-clitic constructions the clitic is inflected to agree with a discourse topic: *finiscila/le, la pasta/le verdure!* (*eat the pasta/vegetables!*).

Aiming at identifying relevant features to be added as microstructural items (Wiegand/Smit 2013) in bilingual dictionaries, *procomplement* verbs belonging to three Romance languages were collected from existing lexicographic resources: 86 lexical units for French, 259 for Italian and 137 for Spanish.

The inventory also counts *procomplement verb constructions* (PVC) of the type described by Espinal (2009): "V+Cl+XP", such as Fr. *en faire de belles*, It. *combinarle*/*farle grosse*, Sp. *hacerla buena*/*bonita*/*menuda* ('to blunder'). These units undeniably represent missing fragments to be added to the phraseological spectrum, such as *procomplement* phrasal verbs (see 3, V+CL+ADV), light verbs (IT, *farla lunga*/*breve*, 'drug on'/'cut something short') or idioms with free slots (Mellado Blanco 2015, p. 113, IT. "*rimetterci X*": rimetterci i polmoni/la camicia/la pelle/le cuoia…).

Starting from De Mauro's description, the inventoried PVs were classified according to different degrees of semantic opacity. For most of the inventoried lexical units (Fr 79%; It 71%; Sp 82%) the hidden argument proves to be retrievable (1), whereas in a small group (Fr 2%; It 5%; Sp 1%) an emphatic meaning emerges (3), and all the others (Fr 20%, It 24%, Sp 18%) display full pronominal semantic bleaching and are completely opaque (2). Some of these features also have morphosyntactic correlates. Subject retrievability, for example, tends to be matched by instances of change of meaning when different clitics are added to the same base verb (1.b), by contrast, clitic variation to convey the same meaning highlights semantic opacity (2.b):

(1) Fr. *la sauter* (*la= nourriture*), 'to be on the breadline',
(1.b) Sp. *buscárse**la*** (*la plea*, 'the fight') 'to ask for trouble', and *buscárse**las*** (*Las soluciones*, 'the remedies') 'to make do',

(2) It. *farci* as in *ma ci fa o ci è?*, 'to play dumb', 'to take advantage of something',
(2.b) Fr. ***en*** *écosser*/***les*** *écosser*, 'to shell out', it. *averse**ne***/*averse**la***, take offence',

(3) It. *darci dentro*, 'knuckle down'.

Issues in machine translation will be exemplified only on purpose of soliciting more thorough investigations on this class of verbs.

References

Arellano, N. (2020): Entre la morfología y la sintaxis: una aproximación a la creación de verbos con pronombre acusativo «la». In: Forma y Función 33 (2), pp. 81–108.

Bibis, N./Roberge, Y. (2004): Marginal clitics. In: Lingua. International Review of General Linguistics 114 (8), pp. 1015–1034.

D'Achille, P. (2003): L'italiano contemporaneo. Bologna.

Espinal, M. T. (2009): Clitic incorporation and abstract semantic objects in idiomatic constructions. In: Linguistics 47 (6), pp. 1221–1271.

GRADIT (1999): Grande dizionario italiano dell'uso. Torino.

Il nuovo De Mauro (2014): https://dizionario.internazionale.it/ (last access: 20-03-2022).

Le Grand Robert (2017) : Grand Robert de la langue française. Dictionnaires Le Robert. www.lerobert.com (last access: 20-03-2022).

Mellado Blanco, C. (2015): Antiphrasis-based comparative constructional idioms in Spanish. In: Journal of Social Sciences 11 (3), pp. 111–127.

RAE dictionary (2021): Diccionario de la lengua española. 23rd ed. Real Academia Española. https://dle.rae.es (last access: 20-03-2022).

Russi, C. (2008): Italian clitics. An empirical study. Berlin.

Viviani, A. (2006): I verbi procomplementari tra grammatica e lessicografia. In: Studi di grammatica italiana 25, pp. 255–322.

Contact information

Valeria Caruso
Università degli Studi di Napoli "L'Orientale"
vcaruso@unior.it

Angela Caiazza
Università degli Studi di Napoli "L'Orientale"
a.caiazza@studenti.unior.it

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Maria Ermakova/Alexander Geyken/
Lothar Lemnitzer/Bernhard Roll

INTEGRATION OF MULTI-WORD EXPRESSIONS INTO THE DIGITAL DICTIONARY OF GERMAN LANGUAGE (DWDS)

Towards a lexicographic representation of phraseological variation

Abstract One central goal of the project 'Zentrum für digitale Lexikographie der deutschen Sprache' (Center for digital lexicography for the German Language, www.zdl.org) is to provide a corpus-based lexicographic component of common German multi-word expressions (MWE), including idioms, for DWDS (www.dwds.de), a general language dictionary of contemporary German. As a central challenge of this task, we have identified an adequate lexicographic representation of such common properties of MWE as variation and modification. To document the variation, we have developed a special entry-clustering model, which we call *hub-node entry*. This model comprises a core hub entry headed by a short nuclear form of the MWE and several node entries, which represent the most common variants in their full lexical forms.

Keywords Multi-word expressions, phraseological variation, dictionary entry structure

Contact information

Maria Ermakova
Zentrum für digitale Lexikographie der deutschen Sprache
Berlin-Brandenburgische Akademie der Wissenschaften
ermakova@bbaw.de

Alexander Geyken
Zentrum für digitale Lexikographie der deutschen Sprache
Berlin-Brandenburgische Akademie der Wissenschaften
geyken@bbaw.de

Lothar Lemnitzer
Zentrum für digitale Lexikographie der deutschen Sprache
Berlin-Brandenburgische Akademie der Wissenschaften
lemnitzer@bbaw.de

Bernhard Roll
Zentrum für digitale Lexikographie der deutschen Sprache
Berlin-Brandenburgische Akademie der Wissenschaften
roll@bbaw.de

Larysa Kovbasyuk

CORONA BEKENNT FARBE: PHRASEOLOGISCHE NEOLOGISMEN IM DEUTSCHEN UND UKRAINISCHEN AUS KULTURLINGUISTISCHER SICHT

Keywords Phraseologische Neologismen; Farbname; Kulturlinguistik

Seit März 2020 entstehen im Wortschatz sowohl der deutschen als auch der ukrainischen Sprachen durch die Corona-Krise viele Neubildungen, die ganz neue Konzepte versprachlichen und deshalb die Aufmerksamkeit auf sich ziehen (Klosa-Kückelhaus 2020; Kovbasyuk 2021). Einen besonderen Platz im neuen Corona-Wortschatz beider Sprachen haben phraseologische Neologismen (weiter nur PhN) mit Farbnamen, die aus der Sicht sowohl der Kulturlinguistik als auch der Kognitiven Linguistik erforscht werden müssen, weil nicht nur Phraseologismen selbst, sondern auch Farbnamen in jeder Sprache kulturgeprägt sind.

PhN werden definiert als „phraseologische Einheiten, durch die neue Erscheinungen oder Sachverhalte erstmals neu benannt werden“ (Kovbasyuk 2018, S. 126). Sie bestehen aus mehr als einem Wort (Burger 2015, S. 11) und werden vor allem durch Polylexikalität, Reproduzierbarkeit und (fakultativ) Idiomatizität gekennzeichnet (Burger 2015, S. 15–32).

Das deutsche Korpus entstammt dem *Neologismenwörterbuch. Neuer Wortschatz rund um die Coronapandemie* des Leibniz-Institutes für Deutsche Sprache und der Online-Zeitung *Zeit Online*. Das ukrainische Korpus ist dem Online-Wörterbuch des Gegenwartsukrainischen *Myslovo* und der Online-Zeitung *Ukrainska pravda* entnommen. Insgesamt enthält das Korpus 10 deutsche und 7 ukrainische PhN mit Farbnamen. Grundlage der Analyse ist ein Vergleichskorpus aus je 125 deutschen und ukrainischen Presseberichten.

Die Analyse der beiden Korpora zeigt, dass als Komponente der PhN im Deutschen und Ukrainischen Grundfarbnamen *gelb* (*жовтий*), *grün* (*зелений*), *rot* (*червоний*) und die Zwischenfarbe *orange* (*помаранчевий*) vorkommen. Die Farbnamenpalette im Deutschen ist im Gegensatz zum Ukrainischen breiter, in PhN sind die Grundfarbnamen *blau*, *schwarz*, *weiß* und die Zwischenfarbe *dunkelrot* vorhanden. Festzustellen ist, dass in den untersuchten PhN viel Wert auf den Symbolwert der Farben gelegt wird. Es handelt sich um das Wissen der Farbensymbolik in der eigenen Kultur, die, wie die Analyse zeigt, zusammenfällt. ROT signalisiert in beiden Kulturen Gefahr, SCHWARZ wird mit Tod assoziiert, WEIß – mit Sauberkeit und Reinheit etc. (Heller 2004, S. 51, 89, 145).

Das ausgewählte Belegmaterial beider Sprachen referiert aus kognitiv-semantischer Sicht auf folgende zwei konzeptuelle Ebenen des Weltbildes (Tab. 1):

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

MENSCHENWELT		UMWELT	
deutsch	ukrainisch	deutsch	ukrainisch
GEGENSTAND: *digitaler grüner Nachweis*; *grüner Pass*	GEGENSTAND: *жовті* (*зелені*) *Covid-паспорти*; *зелений цифровий паспорт*	ORT: *grüne Zone*; *weiße Zone*; *dunkelrote Regionen*	ORT: *зелена зона*; *помаранчева зона*
KREUZFAHRT: *blaue Reise*		ZEITRAUM: *schwarze Stufe*	

Tab. 1: Konzeptuelle Ebenen des Weltbildes

Aus der Sicht der Basisklassifikation (Burger 2015, S. 36) kommen in beiden Korpora meistens referentielle PhN mit der Struktur Adj. + (Adj.) + Nom. in Betracht (92%): *weiße Zone* ‚Ort bzw. Region, in der es (laut einer länderspezifischen Klassifikation) keine Coronainfektionen gibt'; *жовтий Covid-паспорт* ‚Covid-паспорт, що містить інформацію про часткову вакцинацію' (Covid-Pass mit der Information über die Teilimpfung).

Der Herkunft nach werden in beiden Sprachen unterschieden 1) kombinierte PhN (83%), die mindestens eine entlehnte Komponente enthalten: *orange Zone*, *зелений цифровий паспорт* und 2) deutsche bzw. ukrainische PhN (17%): *schwarze Stufe*; *зелений рівень небезпеки*.

Im Ergebnis der kontrastiven Analyse der Korpora aus der Sicht der Kulturlinguistik lässt sich feststellen, dass das deutsche Korpus der PhN reicher, bildhafter und verschiedenartiger im Gegensatz zum Ukrainischen ist. Das lässt sich damit erklären, dass 1) deutsche PhN auf mehrere Konzepte des Weltbildes referieren (siehe Tab. 1) und 2) die Anzahl der Farbnamen unterschiedlich ist.

Festzustellen ist, dass sich im Deutschen sechs PhN finden, um Orte mit verschiedenen Infektionszahlen zu bezeichnen: *weiße/grüne Zone*, *gelbe Zone*, *orange Zone*, *rote Zone*, *dunkelrote Regionen*. Im Ukrainischen lassen sich nur vier Zonen unterscheiden: *зелена зона*, *жовта зона*, *помаранчева зона*, *червона зона*.

Bemerkenswert ist die Tatsache, dass im Deutschen die Farbnamen *schwarz* und *blau* in PhN *schwarze Stufe* und *blaue Reise* vorkommen, mit deren Hilfe neue Konzepte des Alltagslebens verbalisiert werden.

Die Gegenüberstellung von PhN-Paaren ergibt zwei Hauptäquivalenztypen 1) Volläquivalenz: *orange Zone* – *помаранчева зона*; *digitaler grüner Nachweis* – *зелений цифровий паспорт* und 2) Nulläquivalenz: *blaue Reise*, *dunkelrote Regionen*; *жовтий Covid-паспорт* etc.

Literatur

Burger, H. (2015): Phraseologie. Eine Einführung am Beispiel des Deutschen. 5., neu bearbeitete Auflage. Berlin.

Heller, E. (2004): Wie die Farben wirken. Farbpsychologie – Farbsymbolik – Kreative Farbgestaltung. 12. Auflage. Hamburg.

Klosa-Kückelhaus, A. (2020): Wörter in der Coronakrise – von Social Distancing und Gabenzaun. In: SPRACHREPORT 2/2020, S. 6–8.

Kovbasyuk, L. (2018): Deutsche phraseologische Neologismen (am Beispiel der Nuller- und Zehnerjahre). In: Naukovyi visnyk Khersonskoho derzhavnoho universytetu. Seriia „Linhvistyka" 34 (2), S. 125–129.

Kovbasyuk, L. (2021): Coronapandemie-Wortschatz im Gegenwartsdeutschen und Gegenwartsukrainischen. In: Studies about Languages/Kalbų studijos 38, S. 81–98.

Myslovo (2022): http://myslovo.com (Stand: 22.02.2022).

Neologismenwörterbuch. Neuer Wortschatz rund um die Coronapandemie (2022): https://www.owid.de/docs/neo/listen/corona.jsp (Stand: 22.02.02022).

Schröter, L./Tienken, S./Ilg, Y./Scharloth, J./Bubenhofer, N. (2019): Linguistische Kulturanalyse. Berlin/Boston.

Ukrainska Pravda (2022): https://www.pravda.com.ua/ (Stand: 22.02.2022).

Zeit Online (2022): https://www.zeit.de/index (Stand: 22.02.2022).

Kontaktinformationen

Larysa Kovbasyuk
Staatliche Universität Cherson, Ukraine
LKovbasiuk@ksu.ks.ua

Semantics

XX EURALEX INTERNATIONAL CONGRESS
DICTIONARIES AND SOCIETY
12-16 July 2022, Mannheim, Germany

Paolo DiMuccio-Failla

DISAMBIGUATING WORD SENSES THROUGH SEMANTIC CONDITIONS

A project in learner's lexicography

Keywords Learner's lexicography; semantic conditions; phraseology; word sense disambiguation

We propose a novel type of learner's dictionary in the COBUILD tradition, based on the following general theoretical assumptions:

1) there are such things as basic units of word meaning, at least in the case of common words in normal every-day usage;

2) basic word senses are phraseological in nature, since they are not features of words in isolation but of their contextual patterns of typical usage – this hypothesis has been famously first formulated by John Sinclair (cf. Sinclair 1991);

3) such extended units of word meaning can be described by canonical forms of distinctive observable patterns, determined by collocation (the co-occurrence of particular words with the given word), colligation (the co-occurrence of particular grammatical patterns), semantic preference (the co-occurrence of words with particular meanings), and semantic prosody (a particular connotation of the described state of affairs or a particular attitude of the speaker, cf. definition by Louw 1993) – we call this *Sinclair's thesis* about lexical items (cf. Sinclair 1998; Sinclair 2004; Sinclair/Jones/Daley 2004).

On the other hand, we think that there is another independent aspect of word usage that contributes to distinguishing and disambiguating different senses: the *contextual scene construal* (cf. Langacker 1987), which is not always collocational in nature. For example, in disambiguating the fundamental (core) meaning of the verb *to follow*, the scene is usually so construed as to make clear that the person being followed was already in the act of going to a different place. To the best of our knowledge, no collocation corresponds to this state of affairs.

Hanks (2004) has noticed that semantic types alone do not suffice to provide clear distinctions between word senses, and even adding semantic roles to the picture does not solve the problem. In a previous publication (DiMuccio-Failla/Giacomini 2017) we pointed out one of the culprits: semantic types are not always linguistically collocated (i. e. they do not participate in collocations), even if *cognitively typical*. Take, for example, the verb *to toast*: there is no semantic type representing the statistically formed lexical set (cf. Jezek/Hanks 2010) of direct objects (*bread*, *sandwich*, *muffin*, *marshmallow*, *walnut*, ...) found in a corpus, yet it is hard to dispute that *breadstuff nut seed marshmallow*is the right semantic preference. So we all know that one usually toasts breadstuff, even if one does not usually say to toast breadstuff.

However, we are convinced that the main problem in finding the right semantic preferences is the need to add not only semantic roles but also *semantic conditions* to the picture. These conditions apply to the elements of a described scene in the *contextual scene construal*, giv-

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

ing a very robust contribution to word sense disambiguation (cf. Mennes/van der Waart van Gulik 2020 for critical discussion of state-of-the-art WSD).

Our semantic types are based on a growing ontology, linguistic in nature, very similar in flavor to WordNet but with distinctions which are obtained through phraseology and not through synsets. Patterns are identified by first determining colligations (valency distinctions as they are typically found in the dictionaries), semantic types, semantic roles, and conditions imposed on the participants in the scene. Finally, all relevant collocations are added. For the verb *agree*, for instance, one of the patterns of usage, corresponding to a single sense of the word, is the following:

to AGREE ‹WITH a gv. person› ...

■‹THAT a ct. case is given› [when knowing that this gv. person thinks so] **OR**

■‹ON a gv. SUBJECT/MATTER/TOPIC/ISSUE OR **ABOUT a pt. entity›** [when knowing this gv. person's opinion]

In this example, arguments are introduced by prepositions in bold characters and conditions are indicated in square brackets.

We tested our hypothesis by trying to match all usage examples found in selected public repositories (e. g. examples in the UK Dictionary at Lexico.com) with the previously defined patterns of usage. Our test was performed on the following verbs: one with very high semantic polysemy (*to follow*), one with medium degree polysemy (*to agree*), and verbs indicated by Jezek/Hanks (2010) as difficult to disambiguate neatly, e. g. *attend* and *finish.* We managed to attribute around 95% of the example sentences to the corresponding patterns. Our model seems to be able to resolve minimal distinctions in meaning while keeping ambiguous what is intrinsically ambiguous. In the framework of our lexicographic project, we intend to apply our approach to NLP word sense disambiguation.

References

DiMuccio, P./Giacomini, L. (2017): Designing an Italian learner's dictionary with phraseological disambiguators. In: Mitkov, R. (ed.): Computational and Corpus-Based Phraseology. Second International Conference. Europhras 2017, LNAI 10596. Cham, pp. 290–305.

Hanks, P. (2004): Corpus pattern analysis. In: Proceedings of the XI EURALEX International Congress. Volume 1. Lorient, pp. 87–98.

Jezek, E./Hanks, P. (2010): What lexical sets tell us about conceptual categories. In: Lexis 4, pp. 7–22.

Langacker, R. W. (1987): Foundations of cognitive grammar. Stanford.

Louw, B. (1993): Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In: Baker, M./Francis, G./Tognini-Bonelli, E. (eds.): Text and technology. In Honour of John Sinclair. Amsterdam, pp. 157–176.

Mennes, J./van der Waart van Gulik, S. (2020): A critical analysis and explication of word sense disambiguation as approached by natural language processing. In: Lingua 243, pp. 1–14.

Sinclair, J. M. (1991): Corpus, concordance, collocation. Oxford.

Sinclair, J. M. (1998): The lexical item. In: Weigand, E. (ed.): Contrastive lexical semantics. (= Current Issues in Linguistics Theory 171). Amsterdam, pp. 1–24.

Sinclair, J. M. (2004): New evidence, new priorities, new attitudes. In: Sinclair, J. M. (ed.): How to use corpora in language teaching. Amsterdam, pp. 271–299.

Sinclair, J. M./Jones, S./Daley, R. (2004): English collocation studies: The OSTI Report. London/New York, pp. xvii–xxix.

UK Dictionary (2022): https://www.lexico.com (last access: 06-04-2022).

Contact information

Paolo DiMuccio-Failla
University of Hildesheim
muccio@uni-hildesheim.de

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Daniele Franceschi

LEXICOGRAPHIC REPRESENTATIONS OF ANGLO-SAXON AND LATINATE NEAR-SYNONYMS IN ENGLISH MONOLINGUAL AND ENGLISH-ITALIAN BILINGUAL LEARNERS' DICTIONARIES

Keywords Anglo-Saxon; Latinate; near-synonyms; lexical-semantic change; lexicographic representations

It is a known fact that the lexicon of English consists of a basic indigenous vocabulary of Germanic origin with many foreign borrowings especially from French, Latin and Greek. According to Minkova/Stockwell (2006), only 31,8% of the 10,000 most frequent words in the spoken component of the BNC are of Anglo-Saxon origin, while over 60% of them are loanwords that were imported into English from the classical languages, typically via French. This has produced an etymologically diverse word-stock characterized by distinct features (Baugh/Cable 1993; Hughes 2000; Durkin 2014, 2020): the Anglo-Saxon core is made up of morphologically simple and semantically indispensable words referring to common concepts and situations from everyday life, e. g. body parts (*hand*, *foot*, *arm*), animals (*horse*, *cow*, *sheep*), elements of the natural landscape (*land*, *field*, *hedge*), etc.; on the other hand, borrowed (non-Anglo-Saxon) words tend to be polysyllabic and to have a higher level of phonological complexity (e. g., *abdomen*, *cerebellum*, *halitosis*), in addition to describing more elaborate and abstract notions from various areas of specialization, e. g. politics (*capitalism*, *administration*, *bureaucracy*), economics (*money*, *commerce*, *finance*), law (*jurisdiction*, *constitution*, *justice*), etc. In many cases, the addition of Latinate words has produced a duplication of meanings which now complement those of the pre-existing Anglo-Saxon words.

The aim of this presentation is twofold: it intends to provide an initial analysis and a preliminary classification of the meaning relations holding between Anglo-Saxon and Latinate equivalents in contemporary English (Franceschi 2019), such as *speed*/*velocity*, *sweat*/*perspire*, *lunatic*/*insane*, *before*/*prior*, etc. from the theoretical perspective of cognitive lexical semantics, and then discuss the possible lexicographic representations of these relations in English monolingual and English-Italian bilingual learners' dictionaries.

Previous studies within the field of lexical semantics (Bauer 1998; Burnley 1992; Cruse 1986, 2000; Firth 1951; Geeraerts 2010; Hanks 2013, 2015; Hoey 1991; Leech 1981; Sinclair 1998; Pinnavaia/Brownlees (eds.) 2010; among others) have only marginally addressed the phenomenon of near-synonymy when it involves words of different origin; those works that focus more specifically on synonymy (Murphy 2003, 2010) tend to explain meaning variation in terms of contextual use, i. e. they are pragmatics oriented. After an in-depth analysis of both empirical data from corpora and Google Books as well as of example sentences accompanying lexical entries in dictionaries, it has instead been possible to observe that pairs of apparently equivalent words actually present differences at the level of semantics, too. It thus makes sense to explain variation in terms of truth values before addressing the non-denotational differences between them. The use of a near-synonym may for instance be justified by the need to expand or restrict the semantic "contour" of an already existing word.

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

There are a series of cognitive factors that appear to motivate the use of words borrowed from Latin and French. In addition to causing the narrowing or broadening of the meaning of the pre-existing Anglo-Saxon items through metonymy, meronymy and metaphor, Latinate words also seem to determine semantic shifts of focus, Aktionsart, implicature, etc.

Monolingual and bilingual dictionaries of English, however, tend to distinguish between etymologically unrelated synonyms and only in terms of style, register and connotation: Latin-based words are typically labelled as (more) formal, technical or as belonging to a specialized domain, e. g. medicine, biology, engineering, and so on. Examples sentences do not always clarify the limits of substitutability and interchangeability between near-synonyms, nor do they motivate restrictions on semantic grounds. It would instead be appropriate to provide information regarding the connections and the processes of meaning differentiation between Anglo-Saxon and Latinate words through codes, labels and/or usage notes. This metadata would be particularly useful for learners of English whose L1 is a Romance language. Italian EFL learners, for instance, often struggle with words such as *velocity*, *embrace*, *courageous*, etc., in that they are similar in form to their Italian counterparts, i. e., *velocità*, *abbracciare*, *coraggioso/a*, but which cannot be used in the same way, because they are either false friends (Chamizo-Domínguez 2008; Ferguson 1994) or partial cognates and thus different in their scope of reference. A lexicographic improvement with respect to the representation of what appear as recurrent patterns in near-synonymous relations is thus called for. Some examples of possible, finer-grained metadata will be provided.

References

Bauer, L. (1998): Vocabulary. London/New York.

Baugh A. C./Cable, T. (1993): A history of the English language. Englewood Cliffs.

Burnley, D. (1992): Lexis and semantics. In: Blake, N. (ed.): The Cambridge history of the English language. Volume II: 1066–1476. Cambridge.

Chamizo-Domínguez, P. J. (2008): Semantics and pragmatics of false friends. New York/Abingdon.

Cruse, D. A. (1986): Lexical semantics. Cambridge.

Cruse, D. A. (2000): Meaning in language. Cambridge.

Durkin, P. (2014): Borrowed words: a history of loanwords in English. Oxford.

Durkin, P. (2020): The relationship of borrowing from French and Latin in the Middle English period with the development of the lexicon of Standard English: some observations and a lot of questions. In: Wright, L. (ed.): The multilingual origins of Standard English. Berlin/Boston.

Ferguson, R. (1994): Italian false friends. Toronto.

Firth, J. R. (1951): Modes of meaning. In: Firth, J. R. (ed.) (1957): Papers in linguistics 1934–51. London. [Reprint]

Franceschi, D. (2019): Anglo-Saxon and Latinate synonyms: the case of speed vs. velocity. In: International Journal of English Linguistics 9 (6), pp. 356–364.

Geeraerts, D. (2010): Theories of lexical semantics. Oxford.

Hanks, P. (2013): Lexical analysis. Cambridge, MA.

Hanks, P. (2015): Cognitive semantics and the lexicon. In: International Journal of Lexicography 28 (1), pp. 86–106.

Hoey, M. (1991): Patterns of lexis in texts. Oxford.

Hughes, G. (2000): A history of English words. Oxford.

Leech, G. (1981): Semantics: the study of meaning. 2nd edition. Harmondsworth.

Minkova, D./Stockwell, R. (2006): English words. In: Bas, A./McMahon, A. (eds.): The handbook of English linguistics. Malden, pp. 461–482.

Murphy, L. (2003): Semantic relations and the lexicon: antonymy, synonymy, and other paradigms. Cambridge.

Murphy, L. (2010): Lexical meaning. Cambridge.

Pinnavaia, L./Brownlees, N. (eds.) (2010): Insights into English and Germanic lexicology and lexicography: past and present perspectives. Monza.

Sinclair, J. (1998): The lexical item. In: Weigand, E. (ed.): Contrastive lexical semantics. Amsterdam, pp. 1–24.

Contact information

Daniele Franceschi
Dipartimento di Lingue, Letterature e Culture Straniere
Università degli Studi “Roma Tre”
daniele.franceschi@uniroma3.it

Robert Krovetz

AN INVESTIGATION OF SENSE ORDERING ACROSS DICTIONARIES WITH RESPECT TO LEXICAL SEMANTIC RELATIONSHIPS

Abstract This paper discusses an investigation of how senses are ordered across eight dictionaries. A dataset of 75 words was used for this purpose, and two senses were examined for each word. The words are divided into three groups of 25 words each according to the relationship between the senses: Homonymy, Metaphor, and Systematic Polysemy. The primary finding is that WordNet differs from the other dictionaries in terms of Metaphor. The order of the senses was more often incorrectly figurative/literal, and it had the highest percentage of figurative senses that were not found. We discuss leveraging another dictionary, COBUILD, to re-order the senses according to frequency.

Keywords Lexical semantics; word senses; corpus analysis

Contact information

Robert Krovetz
Lexical Research
rkrovetz@lexicalresearch.com

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Index of Authors

INDEX OF AUTHORS

A

Abel, Andrea 136
Alcina, Amparo 238
Aldea, Maria 182
Alves, Ieda Maria 223
An, Jinsan 224
Arapopoulou, Maria 200
Arnau Garcia, Mariona 36

B

Baek, Mikyung 224
Bajčetić, Lenka 112
Balbiani, Laura 63
Ballestracci, Sabrina 137
Bartels, Hauke 159
Biber, Hanno 160
Bichlmeier, Harald 183
Bielinska, Monika 192
Brač, Ivana 100

C

Caiazza, Angela 242
Calabrese, Rita 210
Cartier, Emmanuel 229
Caruso, Valeria 138, 242
Chadjipapa, Elina 213
Chambat, Anaïs 201
Chervino, Alessandra 138
Choi, Jun 228
Costa, Rute 69
Cotgrove, Louis 82

D

Dang-Anh, Mark 38
Daniele, Giulia 138
Declerck, Thierry 86
De Martin Pinter, Patrizio 50
de Oliveira Silva, Maria Eugênia Olímpio 118
de Schryver, Gilles-Maurice 70, 233
Despotidou, Isidora 162
Deutsch, Andreas 194
Diewald, Nils 76
Dimou, Athanasia-Lida 108
DiMuccio-Failla, Paolo 50, 250
Dobrovoljc, Helena 234
Domínguez Vázquez, Maria Josè 192
Dorn, Nico 110
Doupas, Athanasios 177
Dringó-Horváth, Ida 140
Ducassé, Mireille 111
Dzhuranyuk, Vova 116

E

Ecker, Alexandre 41
Efthimiou, Eleni 108
Elizbarashvili, Archil 111
Engelberg, Stefan 46
Ermakova, Maria 245

F

Filipović Petrović, Ivana 67
Fliatouras, Asimakis 213
Flinz, Carolina 137, 164
Flores Flores, Walter Amaru 184
Flouda, Christina 177
Fotinea, Stavroula-Evita 108
Franceschi, Daniele 253
Füreder, Birgit 87

G

Ganslmayer, Christine 47
Gantar, Polona 79, 168
Garoufos, Apostolos 169
Gärtig-Bressan, Anne-Kathrin 63
Gavriilidou, Zoe 143, 162, 169, 175, 213
Geyken, Alexander 245
Giacomini, Laura 50, 164
Gînsac, Ana-Maria 77
Giouli, Voula 177
Girshin, Rina Zviel 128
Gloning, Thomas 28
González Ribao, Vanessa 80, 170
Goulas, Theodoros 108
Gouws, Rufus H. 29, 192
Go, Yelin 224
Grønvik, Oddrun 99

H

Harm, Volker 197
Heid, Ulrich 145
Herbers, Birgit 216
Holdt, Špela Arhar 128
Hollós, Zita 134

I

Ilić, Velibor 112

J

Jarosch, Julian 113
Johannsson, Ellert Thor 215
Jung, Hae-Yun 228

K

Kalafikis, Georgios 200
Kallas, Jelena 156
Karamitsou, Dimitra 200
Kernerman, Ilan 116
Klaes, Christiane 93
Klosa-Kückelhaus, Annette 51
Koeva, Svetla 156
Konstandinidou, Evi 143
Koplenig, Alexander 82
Koppel, Kristina 128
Kosem, Iztok 78, 79, 128, 156, 178
Kouassi, Konan 66
Kovářiková, Dominika 186
Kovbasyuk, Larysa 246
Krek, Simon 79, 168
Kroiß, Daniel 184
Krovetz, Robert 256
Kruse, Theresa 145
Kupietz, Marc 76
Kupriianov, Yevhen 88, 172

L

Labropoulou, Penny 93
Langemets, Margit 156
Langer, Gabriele 180
Lauer, Claudia 216
Le Du, Bridgitte 132
Lemnitzer, Lothar 245
Lew, Robert 146
Lindemann, David 93
Lišková, Michaela 71
Lohk, Ahti 218
Lonke, Dorielle 116
Lorente Casafont, Mercè 36
Lorentzen, Henrik 232
Luís, Ana R. 128
Lüngen, Harald 76

M

Makino, Takahiro 117
Mambelli, Gloria 52
Mantegna, Sarah 202
Marello, Carla 202
Márkus, Katalin P. 140
Maroneze, Bruno 223
Márquez, Manuel 94
McLelland, Nicola 30
Měchura, Michal 97
Meliss, Meike 80
Meyer, Peter 171
Michaelis, Frank 82, 101
Minde, Trond 99
Miyata, Rei 117
Morris, Lorna 189
Moruz, Mihai-Alex 77, 206
Moshövel, Andrea 198
Müller, Anke 180
Müller-Spitzer, Carolin 54, 82

N

Nabirye, Minah 233
Nandi, Enakshi 55
Nazar, Rogelio 81
Nied Curcio, Martina 32, 63, 192

O

Ochs, Samira 82
Orenha Ottaiano, Adriane 118
Ostapova, Iryna 88, 172
Ostroški Anić, Ana 100
Otte, Felicitas 180

P

Panocová, Renáta 221
Paulsen, Geda 218
Pavlova, Anna 173
Petrović, Snežana 112
Pinnavaia, Laura 57
Plate, Ralf 174
Pozzi, María 191
Proost, Kristel 101

Q

Quarado, João Pedro 118

R

Raaf, Manuel 149
Rees, Geraint Paul 152
Renau, Irene 81
Roll, Bernhard 245
Rüdiger, Jan Oliver 54, 82, 101
Russo, Katherine E. 210

S

Salgado, Ana 69, 125
Salveridou, Kyriaki 175
Sarischoulis, Efstratios 200
Sato, Satoshi 117
Schierholz, Stefan 63, 192
Schmitz, Annabella 122
Scholl, Stefan 38
Šemelík, Martin 71
Shyrokov, Volodymyr 88, 172
Sidiropoulos, Nikos 177
Simões, Alberto 125
Škrabal, Michal 71, 186
Smith, Chris A. 126
Smith Ore, Christian-Emil 99
Španović, Ana 112
Stainhaouer, Gregory 177
Starko, Vasyl 102

Stincone, Clarissa 207
Stöckle, Philipp 105
Storjohann, Petra 58
Sungwon, Seo 117
Sviķe, Silga 155
Szemińska, Weronika 164

T

Tasovac, Toma 69
ten Hacken, Pius 221
Tiberius, Carole 128, 156
Todorović, Branislava Šandrih 128
Trap-Jensen, Lars 232
Tselikas, Sotiris 200
Tuulik, Maria 218

U

Ungureanu, Mădălina 77, 206

V

Vacalopoulou, Anna 108, 177
Vainik, Ene 218
Valêncio, Carlos Roberto 118
Vasilaki, Kiki 108
Vranjek Ošlak, Urška 234

W

Wahl, Sabine 105
Wähl, Sabrina 180
Wigestrand Hoftun, Agnes 157
Wolfer, Sascha 82, 146

X

Y

Yablochkov, Mykyta 88, 172

Z

Zabawa, Marcin 59
Žarković, Marija 208
Zeschel, Arne 101
Zimmer, Ben 33
Zingano Kuhn, Tanara 128

This page is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.