



# XVII EURALEX

International Congress

6-10 September, 2016

Tbilisi

## BOOK OF ABSTRACTS

Lexicography and Linguistic Diversity



Lexicographic Centre  
Ivane Javakhishvili  
Tbilisi State University





---

XVII EURALEX International Congress

# Book of Abstracts

6 – 10 September, 2016

Tbilisi

Lexicographic Centre  
Ivane Javakhishvili Tbilisi State University



## Contents

<b>Abstracts of Plenary Lectures.....</b>	<b>11</b>
Complex Morphology and its Impact on Lexicology: the Kartvelian Case.....	13
<i>Jost Gippert</i>	
Three Kinds of Semantic Resonance .....	13
<i>Patrick Hanks</i>	
Why Dictionaries are no Better Than They are – and no Worse .....	14
<i>Robert Ilson</i>	
Bilingual Dictionaries and Theories of Word Meaning.....	15
<i>Pius ten Hacken</i>	
In Praise of Lexicography, and Lexicographers.....	15
<i>Geoffrey Williams</i>	
<b>Abstracts of Papers .....</b>	<b>17</b>
<b>The Dictionary-Making Process .....</b>	<b>19</b>
Descriptive and Prescriptive Lexicography in the Norwegian Context.....	21
<i>Sturla Berg-Olsen</i>	
Pictorial Illustration in Dictionaries The State of Theoretical Art .....	21
<i>Monika Biesaga</i>	
The Role of E-lexicography in the Confirmation of Lexicography as an Independent and Multidisciplinary Field .....	22
<i>Theo J.D. Bothma, Rufus H. Gouws, Danie J. Prinsloo</i>	
Using DPS to Compose Digital Dictionary Pages from Diverse Sources .....	22
<i>Holger Hvelplund, Philippe Climent, Vincent Lannoy</i>	
Editing and Presenting Complex Source Material in an Online Dictionary: The Case of ONP .....	23
<i>Ellert Thor Johannsson, Simonetta Battista</i>	
Dictionary for a Minority Language: the Case of Ket .....	24
<i>Elizaveta Kotorova</i>	
What, When and How? – the Art of Updating an Online Dictionary .....	25
<i>Henrik Lorentzen, Lars Trap-Jensen</i>	
Definizioni illustrate. La selezione di un repertorio di immagini per migliorare la trasparenza semantica delle voci del dizionario di apprendimento <i>Tedesco Junior.</i> .....	25
<i>Silvia Verdiani</i>	

<b>Research On Dictionary Use .....</b>	<b>27</b>
Analysing Translators' Language Problems (and Solutions) Through User-generated Content.....	29
<i>Jaka Čibej, Vojko Gorjanc, Damjan Popič</i>	
Construction of a Tool for the Identification of Electronic Dictionary Users' Skills: Test Specification and Content Validity.....	29
<i>Zoe Gavrilidou, Stavroula Mavromatidou</i>	
Dictionary User Typology: The Slovenian Case.....	30
<i>Špela Arhar Holdt, Iztok Kosem, Polona Gantar</i>	
Pictorial Illustration Versus Text in a Digital English Monolingual Learner's Dictionary: An Eye-tracking Study .....	31
<i>Rafał Kaźmierczak, Robert Lew</i>	
Lexicographical Resources Assisting Vocabulary-building: an Irish-language Experience .....	31
<i>Liam Mac Amhlaigh</i>	
A Modest Proposal for Preventing E-dictionaries from Being a Burden to Teachers and for Making Them Beneficial to the Public.....	32
<i>Carla Marello</i>	
Examining the Effectiveness of Dictionary Instruction on Meaning Determination of Polysemous Words .....	32
<i>Osamu Matsumoto</i>	
Integrating a LSP Dictionary via Mobile Assisted Language Learning in a Multilingual University Setting.....	33
<i>Michele F. van der Merwe</i>	
<b>Lexicography and Language Technologies .....</b>	<b>35</b>
Dictionary as an Instrument of Linguistic Research .....	37
<i>Valentina Apresjan, Nikolai Mikulin</i>	
API-Powered Dictionary Websites .....	37
<i>Sandro Cirulli</i>	
Displaying Language Diversity on the European Dictionary Portal - COST ENeL - Case Study on Colours and Emotions and their Cultural References .....	38
<i>Amelie Dorn, Alina Villalva, Voula Giouli, Wiebke Blanck,         Kira Kovalenko, Eveline Wandl-Vogt</i>	
Designing and Building a Web-based Bilingual Dictionary of South African Sign Language and Afrikaans for Foundation Phase Learners at a School for the Deaf .....	39
<i>Hanelle Fourie Blair, Hanno Schreiber</i>	

The Taming of the Polysemy: Automated Word Sense Frequency Estimation for Lexicographic Purposes .....	40
<i>Anastasiya Lopukhina, Konstantin Lopukhin, Boris Iomdin, Grigory Nosyrev</i>	
GESL Vocabulary and Innovation Technologies .....	40
<i>Tamar Makharoblidze, George Mirianashvili</i>	
The Network Integrator of Croatian Lexicographical Resources .....	41
<i>Marko Orešković, Juraj Benić, Mario Essert</i>	
The Development of a Network Thesaurus with Morpho-semantic Word Markups.....	42
<i>Marko Orešković, Mirko Čubrilo, Mario Essert</i>	
Croatian Linguistic System Modules Overview .....	43
<i>Marko Orešković, Jakov Topić, Mario Essert</i>	
The GEROM Project Search Engine .....	43
<i>Marco Stefano Tomatis</i>	
Furthering the Exploration of Language Diversity and Pan-European Culture:	
The DARIAH-CC Science Gateway for Lexicographers .....	44
<i>Eveline Wandl-Vogt, Roberto Barbera, Giuseppe La Rocca, Antonio Calanducci, Carla Carrubba, Giuseppina Inserra, Tibor Kalman, Gergely Sipos, Zoltan Farkas, Davor Davidović</i>	
<b>Lexicography and Corpus Linguistics .....</b>	<b>45</b>
RuSkELL: Online Language Learning Tool for Russian Language.....	47
<i>Valentina Apresjan, Vít Baisa, Olga Buivolova, Olga Kultepina, Anna Maloletniaia</i>	
Lexicographic Potential of the Georgian Dialect Corpus .....	47
<i>Marina Beridze, David Nadaraia</i>	
What Do Graded Decisions Tell Us about Verb Uses .....	48
<i>Silvie Cinková<sup>1</sup>, Ema Krejčová<sup>1</sup>, Anna Vernerová<sup>1</sup>, Vít Baisa</i>	
Enriching Georgian Dictionary Entries with Frequency Information.....	49
<i>Sophiko Daraselia, Serge Sharoff</i>	
Corpus of the Georgian Language .....	49
<i>Nino Doborjginidze, Irina Lobzhanidze</i>	
From Diachronic Treebank to Dictionary Resource: the Varangian Rus Project ....	50
<i>Hanne Martine Eckhoff, Aleksandrs Berdičevskis</i>	
Semantic Annotation of Verbs for the Tatar Corpus.....	50
<i>Alfiya Galieva, Olga Nevzorova</i>	

A Corpus-based Approach to Lexicography: Towards a Thesaurus of English Idioms .....	51
<i>Guzel Gizatova</i>	
French Cross-disciplinary Scientific Lexicon: Extraction and Linguistic Analysis.....	52
<i>Sylvain Hatier, Magdalena Augustyn, Thi Thu Hoai Tran, Rui Yan, Agnès Tuttin, Marie-Paule Jacques</i>	
Corpus-based Extraction of Collocations for Near-Synonym Discrimination .....	52
<i>Mariusz Kamiński</i>	
The Division into Parts of Speech in the <i>Corpus-based Dictionary of Polish Sign Language</i> .....	53
<i>Jadwiga Linde-Usiekiewicz, Paweł Rutkowski</i>	
The Life and Death of Neologisms: On What Basis Shall We Include Neologisms in the Dictionary? .....	54
<i>Kilim Nam, Soojin Lee, Hae-Yun Jung, Jun Choi</i>	
Parallel Corpora as a Source of Defining Language-specific Lexical Items.....	54
<i>Dmitri Sitchinava</i>	
<i>Nach Belieben kombinieren? Korpusbasierte Beschreibung präpositionaler Mehrworteinheiten im Sprachvergleich</i> .....	55
<i>Kathrin Steyer, Katrin Hein</i>	
<b>Bi- and Multilingual Lexicography .....</b>	<b>57</b>
Problems of Marine Terminology in Georgia .....	59
<i>Inga Adamia</i>	
Multilingual CPA: Linking Verb Patterns across Languages .....	59
<i>Vit Baisa, Sara Može, Irene Renau</i>	
Идиомы в грузино-осетинском и осетино-грузинском словаре .....	60
<i>Наура Белиева, Нино Попиашвили</i>	
L'Accentuation des Emprunts Français dans les Dictionnaires Anglais de Prononciation.....	60
<i>Pierre Fournier</i>	
Problems with the Description of Words Relating to Gender and Sexuality in English-Japanese Learner's Dictionaries .....	61
<i>Yuri Komuro, Saeko Nagashima</i>	
Theoretical Aspects of Rendering Kartvelian Verb Stems in the Georgian-Megrelian-Laz-Svan-English Dictionary .....	62
<i>Ramaz Kurdadze, Maia Lomia, Ketevan Margiani</i>	

Reflecting on Grammar Information in Some Tshivenda Bilingual Dictionaries: A Challenge to Vhavenda Lexicographers .....	62
<i>Munzhedzi James Mafela</i>	
The <i>wdlpOst</i> Toolset for Creating Historical Loanword Dictionaries .....	63
<i>Peter Meyer</i>	
Spezifik der Darstellung von Verben in georgisch-deutschen Wörterbüchern .....	64
<i>Russudan Sekalashvili</i>	
Bilingual Lexicography in the Republic of Tatarstan in 1990-2010.....	65
<i>Gulshat Safullina</i>	
<b>Lexicography for Specialised Languages, Terminology and Terminography....</b>	<b>67</b>
Using Argument Structure to Disambiguate Verb Meaning .....	69
<i>Miriam Buendía Castro, Beatriz Sánchez Cárdenas</i>	
Encoding Context in Bilingual Specialized Dictionaries .....	69
<i>Pamela Faber, Arianne Reimerink, Pilar León-Araúz</i>	
Bridging the Gap between Description and Standardization: a Frame-based Version of NATO Glossaries .....	70
<i>Pamela Faber, Pilar León-Araúz</i>	
Aspects Lexicaux d'une Langue Inventée: la Langue des Schtroumpfs.....	71
<i>Thierry Fontenelle</i>	
Contemporary Legal Terminology – Some Problems of Translation .....	71
<i>Irina Gvelesiani</i>	
French Specialised Medical Constructions: Lexicographic Treatment and Corpus Coverage in General and Specialised Dictionaries .....	72
<i>Ornella Wandji Tchami, Ulrich Heid, Natalia Grabar</i>	
A Descriptive Approach to Medical English Vocabulary.....	72
<i>Renáta Panocová</i>	
Социопрагматический аспект лексикографирования специальной лексики ....	73
<i>Людмила Рычкова</i>	
Collocational Information for Terminological Purposes .....	74
<i>Elsabé Taljard</i>	
<b>Historical and Scholarly Lexicography and Etymology.....</b>	<b>75</b>
Wanderworten Etymological Format: a 'Hemp' Case Study, Starting from Votic and Baltofinnic. ....	77
<i>Manuel Barbera</i>	
The Russian Contributions to the English Language: a Case of OED Record ....	77
<i>Kseniya Egorova</i>	

Lexicography of Language Contact: An Internet Dictionary of Words of German Origin in Tok Pisin .....	78
<i>Stefan Engelberg, Christine Möhrs</i>	
The Word Entry Structure of the Russian Manuscript Lexicons: Evolution through the Centuries .....	78
<i>Kira Kovalenko</i>	
Diachronic Analysis of Word Formation as Basic Studies for Etymologisation ....	79
<i>Iris Metsmägi, Vilja Oja</i>	
Reconsidering the Initial Development of the Dictionaries of Americanisms: Pickering's Legacy in Bartlett's <i>Dictionary</i> (1849) .....	80
<i>Kusujiro Miyoshi</i>	
Etymological Studies Based on a Conceptual Framework .....	80
<i>Vilja Oja</i>	
The Da Vinci Codices. Between Historical Lexicography and Knowledges History .....	81
<i>Margherita Quaglino</i>	
Moth's Danish Dictionary Publishing a Dictionary with a 300-year Delay .....	81
<i>Thomas Troelsgård, Marita Akhøj Nielsen</i>	
Enabling Selective Queries and Adapting Data Display in the Electronic Version of a Historical Dictionary .....	82
<i>Tim van Niekerk, Heike Stadler, Ulrich Heid</i>	
Digitizing the Second Edition of Furetière's <i>Dictionnaire Universel</i> : Challenges of Representing Complex Historical Dictionary Data Using the TEI .....	83
<i>Geoffrey Williams, Ioana Galleron</i>	
<b>Lexicography and Semantic Theory .....</b>	<b>85</b>
Adding Polarity Information to Entries of the Database of Bavarian Dialects in Austria .....	87
<i>Thierry Declerck, Amelie Dorn, Eveline Wandl-Vogt</i>	
Towards a Sense-based Access to Related Online Lexical Resources .....	87
<i>Thierry Declerck, Karlheinz Mörth</i>	
Discursive Constructions in the Russian-Swedish Dictionary Database: A Case Study of <i>v tom-to i N</i> .....	88
<i>Dmitrij Dobrovolskij, Ludmila Pöppel</i>	
Tatar Co-compounds as a Special Type of Classifiers .....	89
<i>Alfiya Galieva, Dzhavdet Suleymanov</i>	
Semantic Approach to Phraseological Patterns in Karstology .....	89
<i>Larisa Grčić Simeunović, Paula de Santiago Gonzalez</i>	

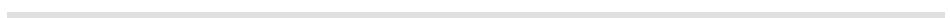
Displacement of Senses, Cluster Equivalence and Bilingual Dictionaries .....	90
<i>Barbara Lewandowska-Tomaszczyk</i>	
Systematic Polysemy of Estonian Colour Adjectives .....	91
<i>Maria Tuulik, Margit Langemets</i>	
<b>Phraseology and Collocation.....</b>	<b>93</b>
Online Dictionary of Idioms .....	95
<i>Irina Lobzhanidze</i>	
On Distinction between Phraseology Conceptions in a General Monolingual Dictionary and in a Special Phraseological Dictionary.....	95
<i>Renáta Neprašová</i>	
Starting from Scratch – The Creation of an Irish-language Idiom Database .....	96
<i>Katie Ní Loingsigh, Dr Brian Ó Raghallaigh</i>	
The Compilation of a Printed and Online Corpus-based Bilingual Collocations Dictionary.....	97
<i>Adriane Orenha-Ottaiano</i>	
<b>Lexicological Issues of Lexicographical Relevance.....</b>	<b>99</b>
On <i>bullying</i> , <i>mobbing</i> (and <i>harassment</i> ) in English and Polish: Foreign-language-based Lexical Innovation in a Bilingual Dictionary .....	101
<i>Arleta Adamska-Sałaciak</i>	
Representing Usage in Dictionaries .....	101
<i>Janet DeCesaris</i>	
Lexical Change in Times of Upheaval and War – And the Dictionary .....	102
<i>Donna M.T. Cr. Farina</i>	
Lexicalized Prosody and the Polysemy of some Discourse Particles in ADR ....	102
<i>Irina Levontina</i>	
Importance of the Issue of Partial Equivalence for Bilingual Lexicography and Language Teaching .....	103
<i>Tinatin Margalitadze, George Meladze</i>	
Towards the Meaning of a Word .....	103
<i>Innes Merabishvili</i>	
<i>Von vielrednerischen Weibern und alles erforschenden Männern –</i> Genderasymmetrien in der Datenbank der bairischen Mundarten in Österreich (DBÖ) im Vergleich mit Großwörterbüchern der Gegenwart .....	104
<i>Maria Pöber, Ingeborg Geyer, Eveline Wandl-Vogt, Barbara Piringer</i>	
Automatic Extraction of Lexical Patterns from Corpora .....	105
<i>Irene Renau, Rogelio Nazar</i>	

---

---

Basic Russian Dictionary: a Corpora-Based Learner's Dictionary of Russian ..	105
<i>Maria Shvedova, Dmitri Sitchinava</i>	
Automatically Identifying Selectional Restrictions of Predicates: A Computational Methodology for Practical Lexicography that Challenges Linguistic Theory .....	106
<i>Alexandra Anna Spalek, Marco Del Tredici</i>	
Paysage: Lexical Variation in the Lexis of Landscape in Romance Dictionaries .....	106
<i>Geoffrey Williams, Araceli Alonso, Janet DeCesaris, Ioana Galleron,     Chris Mulhall</i>	
<b>Reports on Lexicographical and Lexicological Projects .....</b>	<b>109</b>
Un Projet Accompli: le <i>Lexicon de Buda (1825)</i> en Édition Électronique .....	111
<i>Maria Aldea</i>	
Von einem <i>Orthographischen Wörterbuch</i> georgischer Äquivalente deutscher Toponyme bis zum <i>Kulturgeographischen Lexikon Deutschlands</i> .....	111
<i>Marina Andrazashvili</i>	
On Compiling a Norwegian Academic Vocabulary List .....	112
<i>Ruth Vatvedt Fjeld, Arash Saidi</i>	
Lexicographic Description of Czech Complex Predicates: Between Lexicon and Grammar.....	112
<i>Václava Kettnerová, Petra Barančíková, Markéta Lopatková</i>	
An Interactive Solution: a Grammar Handbook Linked to a Prescriptive Dictionary.....	113
<i>Tiina Leemets, Maire Raadik</i>	
Bilingual Dictionary Drafting: Connecting Basque Word Senses to Multilingual Equivalents .....	114
<i>David Lindemann, Iñaki San Vicente</i>	
Online Bibliography of Electronic Lexicography. The Project OBELEX <sup>meta</sup> ....	114
<i>Christine Möhrs</i>	
Straddling the Boundaries of Traditional and Corpus-Based Lexicography: A Latvian-Czech Dictionary .....	115
<i>Michal Škrabal</i>	
The SWEGOT Lexical Resources Framework .....	115
<i>Monica von Martens, Louise Holmer</i>	
<b>List of Authors.....</b>	<b>117</b>

# **Abstracts of Plenary Lectures**



---

## Complex Morphology and its Impact on Lexicology: the Kartvelian Case

**Jost Gippert**

Goethe University Frankfurt am Main  
Honorary Doctor of Ivane Javakhishvili Tbilisi State University  
e-mail: gippert@em.uni-frankfurt.de

The keynote deals with the impact of complex morphological structures on essential aspects of lexicology. On the basis of data from the so-called Kartvelian (South-Caucasian) language family (consisting of Georgian and its sister-languages), it discusses questions of how to lemmatize nominal and verbal word forms so that they can be arranged in a consistent and appropriate way in dictionaries and databases. The most prominent approaches that have been developed since the first attempt to provide a dictionary of Georgian (Paolini & Irbachi 1629) are analyzed with a view to their applicability, usability, and conclusiveness.

## Three Kinds of Semantic Resonance

**Patrick Hanks**

Research Institute for Information and Language Processing  
University of Wolverhampton  
Honorary Member  
European Association for Lexicography  
e-mail: patrick.w.hanks@gmail.com

In this presentation, I explore creativity in association with the use of words, and I ask, what should dictionaries do about this aspect of language? Creativity can be associated with cognitive resonance. Three kinds of cognitive resonance are identified. Examples are given, taken from corpus data and from other literature. They are:

- **Verbal resonance** (including metaphor and other kinds of figurative language)
- **Experiential resonance** (words and phrases such as summer, which activate sensory memory)
- **Intertextual resonance** (words and phrases taken from previous users, ranging from Shakespeare and the Bible and other great writers, to the causal or imaginative coinages made by friends, family, and colleagues)

The full presentation will give as many examples of each kind as time allows. It will suggest a variety of possible actions that could be taken by dictionaries, especially on-line dictionaries, to illustrate the rich cognitive potential of ordinary words in these respects.

## Why Dictionaries are no Better Than They are – and no Worse

*Dice Dios: Toma lo que quieras – y págalos*

**Robert Ilson**

University College London

Honorary Member

European Association for Lexicography

e-mail: roberti@dsl.pipex.com

### 1) Why Dictionaries are no better than they are:-

Lexicography is beset with problems at three levels:-

- 1.1) Metalexicography (ie lexicographic theory), which includes linguistic theory;
- 1.2) Lexicography Proper (ie the confection of dictionaries);
- 1.3) Dictionnaire (ie the organisation of lexicographic projects and the selling of their products).

Or, in plainer words, problems about what to say, how to say it, and how to present it.

Many such problems stem from the failure to generalise best practice across genres and languages, and from the failure to exploit fully the Explanatory Technique of Exemplification.

### 2) Why Dictionaries are no worse than they are:-

- 2.1) Dictionary-users, as native speakers of at least one human language, already know a great deal about language in general and about the world, which enables them to supplement and interpret correctly the information offered by dictionaries;
- 2.2) Many dictionary-users come from dictionnaire communities, in which dictionaries and their conventions are widespread and well known.

## Bilingual Dictionaries and Theories of Word Meaning

**Pius ten Hacken**

Leopold-Franzens-Universität Innsbruck  
e-mail: pius.ten-hacken@uibk.ac.at

When theory is applied to dictionaries, it is first of all monolingual dictionaries that are addressed. However, there is no reason to consider bilingual dictionaries as somehow less theory-prone than their monolingual counterparts. The domain of linguistics that is most relevant for lexicography is word meaning. Therefore, I present four theoretical perspectives of word meaning and discuss how they can be applied to bilingual dictionaries. Word meaning has been conceived as the set of objects the word can refer to, the lexical relations the word has with other words, the way the word has been used, or the semantic information stored with the word in a speaker's brain. As I show, the reference-based view of meaning and the one based on lexical relations would require radically different types of bilingual dictionaries than the ones we currently have. The perspective based on the use of a word gives rise to corpus linguistics and frame semantics. It is particularly useful when it comes to compiling a bilingual dictionary. The concept-based, mental view of meaning offers a good basis for the evaluation and interpretation of dictionaries. These two perspectives can be seen as supplementing each other and need not be in competition.

**Keywords:** reference; lexical fields; corpora; frame semantics; Conceptual Structure; prototypes; dictionary compilation; dictionary evaluation; dictionary interpretation

## In Praise of Lexicography, and Lexicographers

**Geoffrey Williams**

LiCoRN-LIDILEM, Université Grenoble Alpes  
e-mail: geoffrey.williams@orange.fr

The title of this paper draws deliberately on that of John Sinclair's keynote in Lorient in 2004, *In praise of the dictionary*. Whereas John Sinclair concentrated on the product, the dictionary, I shall concentrate on the field, and the harmless drudges who undertake the activity, and more particularly in the area of lexicographical research. The starting point will be Georgian lexicography and its importance to the national language and culture. This leads naturally to what I consider two major challenges facing lexicographical research today and the science policy issues it faces: funding policy and notably our place in EC Framework Programmes and the effects of research evaluation. In a second part, I want to give praise to two particular "harmless drudges" – Antoine Furetière and Henri Basnage de Beauval – who illustrate the profile of lexicography in their time. This will allow me to discuss the importance of historical lexicography within the context of digital humanities.

**Keywords:** lexicography; research funding; research evaluation; Digital Humanities

---

# Abstracts of Papers

---

---

# **The Dictionary-Making Process**



---

## Descriptive and Prescriptive Lexicography in the Norwegian Context

**Sturla Berg-Olsen**

The Language Council of Norway  
e-mail: sturla.berg.olsen@sprakradet.no

The paper examines some selected Norwegian dictionaries from the perspective of descriptivity and prescriptivity. The main conclusion is that the element of prescriptivity in the dictionaries is larger than what is openly signalled to the users. In the case of the one-volume standard dictionaries, prescriptivity also extends outside the domains of spelling and inflection, although these are the only domains that are officially regulated. It is argued that a revision of these standard dictionaries should encompass a reevaluation of their prescriptive profile. The fact that dictionaries contain elements of prescriptivity is neither unexpected nor a problem in principle. However, prescriptive statements as a principle should not remain hidden or implicit. To the extent that they do so, this can easily be interpreted as misleading the users rather than guiding or enlightening them.

**Keywords:** Norwegian lexicography; prescriptivity; descriptivity; language norms; language usage; language regulation

## Pictorial Illustration in Dictionaries The State of Theoretical Art

**Monika Biesaga**

The Institute of the Polish Language  
Polish Academy of Sciences (PAS)  
e-mail: monika.biesaga@ijp.pan.krakow.pl

The aim of this paper is to introduce the theory of pictorial illustration in lexicography. This kind of summary could be useful both for lexicographers with previous experience in incorporating visual materials as well as scientist that would like to add this kind of content to their projects. Concerning the list of arguments to be presented, firstly, a difference between encyclopedias and dictionaries, crucial for including pictorial illustration, will be shown. Secondly, I will characterize the results of recent user studies related to this subject. The third part of the text is a report on verbal and visual mode in lexicography (typology of facilities, significance of color, problem of caption, mutual relation between the two modes). Then I will indicate which groups of meanings and which parts of speech are or should be illustrated. Subsequently, characteristics of different illustrating strategies will be presented. Finally, I will formulate some general rules concerning pictorial illustration in dictionaries, related to the typology of dictionaries. In conclusion some propositions of further research will be pointed out.

**Keywords:** pictorial illustration; multimodal lexicography; multimedia in lexicography; theory of lexicography

## **The Role of E-lexicography in the Confirmation of Lexicography as an Independent and Multidisciplinary Field**

**Theo J.D. Bothma, Rufus H. Gouws, Danie J. Prinsloo**

University of Pretoria, University of Stellenbosch, University of Pretoria

e-mail: theo.bothma@up.ac.za, rhg@sun.ac.za, danie.prinsloo@up.ac.za

This paper argues in favour of an adopted, extended theory of lexicography to cater for the fast-growing e-lexicography practice. We believe that such a modern general lexicographic theory should be rooted in three fundamental phases of the development of lexicographic theory, i.e. the focus on linguistic contents – Zgusta, Wiegand's focus on dictionary structures and Bergenholz and Tarp's emphasis on lexicographic functions with the user in focus. Although the linguistic contents and the functions of printed and e-dictionaries can be similar, the presentation differs and that demands innovative formulations of dictionary structures. Lexicographic theory is expanded by describing and taking cognisance of all innovative technological features enabled by the computer era, including data collection and processing and data representation and interface design, in order to respond to the different needs of the envisaged target users in an optimal way. Insights from the field of information science facilitate the compilation of a much more sophisticated specification of user needs and the way in which users interact with the information. The lexicographer has an obligation to provide users with guidance and required information whilst guarding them against information overload.

**Keywords:** lexicographic theory; interdisciplinary; multidisciplinary; information science; user support tools

## **Using DPS to Compose Digital Dictionary Pages from Diverse Sources**

**Holger Hvelplund, Philippe Climent, Vincent Lannoy**

e-mail: hvelplund@idm.fr, climent@idm.fr, lannoy@idm.fr

When dictionaries are published in digital format users typically from a single interface: (1) have access to many dictionaries; and (2) on dictionary entry pages have access to content from several dictionary datasets and from other dictionary resources. A data processing chain composes the content of the pages by combining content from several dictionary datasets with other relevant dictionary resources.

This presentation proposes a refined approach based on our experience with publishing dictionaries in digital format and focuses on:

- empowering editors using the dictionary editing system to preview, proof and edit the composed dictionary pages;
- making each dictionary entry page a ‘home page’ with content that can be individually crafted;
- instant update of the dictionary entry pages;
- minimize risk of publishing thin or duplicate content.

The approach is based on associating ‘super-entries’ to dictionary entry pages. The ‘super entries’ contain content references to the ‘documents’ that are contained in dictionary entry pages.

Publishing the dictionary content in digital format gives human users endless new ways of consuming the content. In addition, websites can give apps API access to the dictionary content and apps can use the annotated links in the dictionaries for producing new applications of the dictionary content.

**Keywords:** refined digital format; digitally published content; dictionary production process

## Editing and Presenting Complex Source Material in an Online Dictionary: The Case of ONP

**Ellert Thor Johannsson, Simonetta Battista**

Dept. of Nordic Research, University of Copenhagen  
e-mail: nkv950@hum.ku.dk, sb@hum.ku.dk

*Ordbog over det norrøne prosasprog/A Dictionary of Old Norse Prose* (ONP) is a dictionary project that aims at bringing to light the vocabulary of medieval West Scandinavian prose texts, with a large number of examples of word use and various kinds of supplementary secondary information. In this paper, we give a brief overview of the project, discuss its background and approach to the source material and some of the related problematic aspects. We account for the way collected data has been treated and presented in dictionary entries and discuss the editing system that ONP has developed to suit its specialized needs. The dictionary has gone through the transformation from a printed publication to a digital online version, in the middle of a multi-volume publication process. We discuss how this has affected the editing work and how editorial procedures have evolved to better accommodate the lexicographic information through an online medium. We give some examples of how complex information is presented to the user in ONP Online and share our ideas about further development and improvement of the online version.

**Keywords:** dictionary making; editing principles; online dictionary; editing system; user interface

## Dictionary for a Minority Language: the Case of Ket

**Elizaveta Kotorova**

Zielona Góra, Poland / Tomsk

e-mail: e.kotorova@gmail.com

Making a dictionary of a minority language, especially, an endangered one requires solving several problematic issues specific to this kind of lexicographical work. The primary issues to be dealt with in this case are as follows: (1) Identification of the target audience. There are three possible variants: language community, academic community, or both communities. The choice depends on what lexicographical work has already been done for a given language and what are the needs and expectations of the communities. (2) Collecting data and compiling the basic vocabulary. There are several approaches to this issue including a) translation of the list of the most frequent words from a European language, b) extraction of the wordlist from a corpus of texts, and c) thematic elicitation from native speakers in order to obtain basic lexical units belonging to a particular semantic field. (3) Creating dictionary entries. In general, every dictionary entry consists of two main components – a lemma and a commentary. Both components usually have certain peculiarities in dictionaries of minority languages. These may include, for example, orthographic representation of the lemma, hierarchic arrangement of word meanings, presentation of encyclopedic information in the dictionary, and so on.

The present paper provides an account of how these problems were dealt with in the Comprehensive dictionary of Ket, a highly endangered indigenous language spoken in Central Siberia.

**Keywords:** lexicography; dictionary making; minority languages; Siberian languages; Ket

## What, When and How? – the Art of Updating an Online Dictionary

**Henrik Lorentzen, Lars Trap-Jensen**

Society for Danish Language and Literature

e-mail: hl@dsl.dk, ltj@dsl.dk

Online dictionaries are not published in numbered editions and impressions which makes it difficult for users to know how old a particular entry is. As e-dictionaries grow older, the question of updating steadily becomes more pressing. Is everything up-to-date? If not, when was a particular entry composed, and when was it last revised? These are questions the users might ask themselves, and in this paper we deal with some of the aspects involved: what is most likely to change, when should revisions take place, and how can the staff keep track of entries in need of updating. The paper is based on experience from an online monolingual dictionary of Danish, but it is assumed that the problems and their solutions are independent of the particular project and language.

**Keywords:** updating; revising; time aspect; diachronic lexicography; logistics

## Definizioni illustrate. La selezione di un repertorio di immagini per migliorare la trasparenza semantica delle voci del dizionario di apprendimento *Tedesco Junior*.

**Silvia Verdiani**

Università di Torino/Universität Potsdam

e-mail: silvia.verdiani@unito.it

Il ruolo delle illustrazioni può essere determinate per disambiguare la semantica delle voci verbali in un dizionario. Il contributo intende suggerire un percorso di ricerca e selezione del corredo di immagini a partire dal motore di ricerca *Google Images*.

**Keywords:** *Lerner's Dictionary*; dictionary illustration; semantics; polysemy; multimodality

---

# Research On Dictionary Use

---

---

## Analysing Translators' Language Problems (and Solutions) Through User-generated Content

**Jaka Čibej, Vojko Gorjanc, Damjan Popić**

Department of Translation, Faculty of Arts, University of Ljubljana

e-mail: [jaka.cibej@ff.uni-lj.si](mailto:jaka.cibej@ff.uni-lj.si), [vojko.gorjanc@ff.uni-lj.si](mailto:vojko.gorjanc@ff.uni-lj.si), [damjan.popic@ff.uni-lj.si](mailto:damjan.popic@ff.uni-lj.si)

This paper focuses on dictionary use among translators and presents the results of a pilot study into translators' use of language resources when solving language problems. The paper first provides an overview of related work in this field and continues by presenting the results of a pilot study into translators' use of language resources when solving language problems. By analysing a number of discussions in *Prevajalci, na pomoč!*, a dedicated, self-managed Facebook group for Slovene translators aimed at solving translation problems, a taxonomy of typical language problem scenarios is developed. Through an analysis of the problems encountered by translators and their suggested solutions, two aspects are established: the areas in which problems occur and the ways the solutions are reached. This analysis is followed by the final segment, which discusses the suitability of the proposed method and the degree to which this approach yields results that are of use for the compilation of monolingual dictionaries, as well as for lexicographical user research.

**Keywords:** translators; social media; language resources; monolingual dictionary; dictionary users

## Construction of a Tool for the Identification of Electronic Dictionary Users' Skills: Test Specification and Content Validity

**Zoe Gavriilidou, Stavroula Mavrommatidou**

Democritus University of Thrace

e-mail: [zoegab@otenet.gr](mailto:zoegab@otenet.gr), [stavrjav@hotmail.com](mailto:stavrjav@hotmail.com)

The present paper reports results regarding the development and validation of the Strategy Inventory for Electronic Dictionary Use (S. I. E. D. U.) which is a reliable 5-Likert scale self-report instrument consisting of 59 Questions for assessing users' skills and strategies in online electronic dictionary searches. Following Gavriilidou's (2013) Strategy Inventory for Dictionary Use (S.I.D.U.) structure, for assessing dictionary use strategies, the S.I.E.D.U includes five subscales: a. Dictionary use awareness skills, b. Dictionary selection strategies, c. Strategies for lemmatization and acquaintance with dictionary conventions, d. Navigation skills and e. Look-up strategies in the new electronic environments. The method of multiple judges was adopted for the control of content validity of the pilot version of S.I.E.D.U. The Cronbach Alpha coefficient was

employed in order to provide a measure of the reliability of S.I.E.D.U. Internal consistency of the five subscales and the overall scale of the S.I.E.D.U. ranged from good to excellent.

**Keywords:** Online dictionary; electronic lexicography; dictionary strategies; dictionary reference skills; instrument validation

## Dictionary User Typology: The Slovenian Case

Špela Arhar Hold<sup>1,2</sup>, Iztok Kosem<sup>1,2</sup>, Polona Gantar<sup>1</sup>

<sup>1</sup>Centre for Applied Linguistics, Trojina Institute

<sup>2</sup>Faculty of Arts, University of Ljubljana

e-mail: spela.arhar@trojina.si, iztok.kosem@trojina.si, apolonija.gantar@ff.uni-lj.si

Understanding the preferences, needs and habits of dictionary users is an important prerequisite for the compilation of a dictionary. When researching these topics, an appropriate balance has to be struck between over-generalising or over-individualising user profiles. Following this goal, the paper presents a typology of dictionary users, developed as part of the conceptualization of the Dictionary of Contemporary Slovene Language (a general monolingual dictionary of the Slovene language). The typology consists of two levels. On the first one, the situations of dictionary use are divided into the ones arising from: (i) the process of (formal) education; (ii) professional context; and (iii) spare time activities. On the second level, a number of user groups for each of these situations are formed, based on their specific language needs that have been identified. The presented typology will enable quality user studies and consequently improve the status of dictionary use research in Slovenian lexicography. Internationally, the typology can be integrated into the existing methodology of dictionary user research to facilitate more representative sampling of user groups, and to provide a better understanding of the obtained results.

**Keywords:** dictionary user research; user typology; Dictionary of Contemporary Slovene Language

## Pictorial Illustration Versus Text in a Digital English Monolingual Learner's Dictionary: An Eye-tracking Study

**Rafal Kaźmierczak, Robert Lew**

Faculty of English, Adam Mickiewicz University in Poznań  
e-mail: rlew@amu.edu.pl

The paper reports on an eye-tracking study of ten Polish advanced learners of English consulting twenty on-screen monolingual entries of low-frequency but high-illustrability items, adapted from the PC version of the *Longman Dictionary of Contemporary English*. Half of the items featured pictorial illustrations depicting the designate in a context (e.g. *banister* shown next to a staircase), while the other half only showed the designate (no context). We asked participants to supply (verbally) Polish equivalents of the 20 words defined, based on the dictionary entries presented on the screen, and their eye-gaze as well as their responses were recorded. Participants were also given a lexical recall test. The data reveal patterns of relative use of the verbal and pictorial elements of the entries (to what extent do pictorial illustrations and definitions compete for the users' attention?), vocabulary retention, and how these outcome measures are moderated by the contextuality of the pictorial illustration.

**Keywords:** learner's dictionary; monolingual dictionary; illustration; definition; eye-tracking

## Lexicographical Resources Assisting Vocabulary-building: an Irish-language Experience

**Liam Mac Amhlaigh**

Maynooth University  
e-mail: liam.macamhlaigh@nuim.ie

Language-learning adults encounter a dictionary of some form as they develop a familiarity with their chosen language of learning. For children, the process is similar albeit that they may not have 'chosen' to learn the language in question in the same manner – their innate geographical location and the choices of their parents and guardians will play very important roles. All learners use dictionaries principally as an aid to build, and re-confirm, vocabulary in a language. Dictionaries facilitate a sense of ownership and enfranchising that has always had a presence in language learning, even though the concept of self-learning, and/or self-help is something that has only been a powerful force from the 20<sup>th</sup> century onwards. Building on the theoretical overview of learning vocabulary by utilising dictionaries, this paper aims at garnering an insight into how dictionaries can currently assist vocabulary-learning in an Irish-language immersion setting.

**Keywords:** vocabulary-building; dictionaries; Irish language; immersion setting

## **A Modest Proposal for Preventing E-dictionaries from Being a Burden to Teachers and for Making Them Beneficial to the Public**

**Carla Marello**

Università di Torino

e-mail: carla.marello@unito.it

Some simple suggestions, already tested in Italian secondary school classes, aimed at making e-dictionaries an everyday reference tool, as paper dictionaries were (and still are, in some part of the world). By e-dictionaries we mean fully fledged digital versions of paper dictionaries, structured databases which users can access via the Internet by paying a subscription and which are still sold on a DVD or USB key support. We will briefly state why accessing free e-dictionaries is good for practice but not as rewarding as it might be to access paid web site displaying all the range of possible query modes. We conducted a survey about what Italian secondary school students turn to when they do not understand a word and we maintain that e-dictionaries can enhance metalinguistic knowledge and enlarge students' vocabulary, but such goals are better met by task-based activities which do not have a straightforward language lesson appearance. A task based language teaching approach mixed with CLIL (Content and Language Integrated Learning) is an ideal arena where e-dictionaries might win their battle in order to become truly accepted and widespread in the classroom.

**Keywords:** dictionary users; reference skill training; e-dictionary use

## **Examining the Effectiveness of Dictionary Instruction on Meaning Determination of Polysemous Words**

**Osamu Matsumoto**

Waseda University

e-mail: matsumoto036@icloud.com

This study investigates the effects of dictionary use instruction on meaning determination of polysemous words among beginner-level Japanese learners of English, or more specifically, its effects on selecting an appropriate L1 equivalent for target L2 polysemous words from the multiple senses provided by a bilingual dictionary. Twenty-five Japanese high school students participated and were divided into two groups based on having received a dictionary instruction class or a normal reading comprehension class. The dictionary instruction was conducted so as to increase linguistic and dictionary-specific knowledge and raise awareness of surrounding contexts of a word in the text. The quantitative analysis of a pre- and post-test revealed that the group with dictionary instruction considerably improved in locating an appropriate L1 equivalent within the

entry for a polysemous word. In addition, after dictionary instruction, participants were more likely to pay careful attention to familiar words when they appeared in a syntactically and/or semantically unfamiliar way. These findings indicate that, even though dictionary instruction was conducted in a relatively short period, it largely contributed to the learners' development of ability for meaning determination of polysemous words. This effect occurred while reading with and without a dictionary.

**Keywords:** dictionary instruction; polysemous word; meaning determination; familiar word

## Integrating a LSP Dictionary via Mobile Assisted Language Learning in a Multilingual University Setting

**Michele F. van der Merwe**

Stellenbosch University

e-mail: michelevdm@sun.ac.za

Mobile technologies have become worldwide phenomena, opening up new opportunities for language teaching and learning. A literature study on Mobile Assisted Language Learning as well as user research on digital dictionaries is presented. A LSP dictionary on a mobile phone is described, with reference to the aims, structure and functions, as well as the target group of the dictionary. The integration of the LSP dictionary in a multilingual lecture hall is explored, with regard to the structure and functions of the dictionary. A LSP dictionary, called MobiLex was compiled by academic specialists at the faculty of Education for pedagogical purposes at the University of Stellenbosch. The aim of the article is to demonstrate how the dictionary could be integrated in lecture halls at university. MobiLex is trilingual and fits within the university's framework of multilingualism. Afrikaans and English are used as languages of learning and teaching (LOTL) within the university, with the focus on the development of isiXhosa as academic language. A task-based activity for Social Sciences was designed with regard to the dictionary functions of MobiLex.

**Keywords:** user research; mobile dictionaries; LSP dictionary; multilingualism

---

# Lexicography and Language Technologies





## Dictionary as an Instrument of Linguistic Research

**Valentina Apresjan, Nikolai Mikulin**

National Research University “Higher School of Economics”,  
School of Linguistics  
Vinogradov Russian Language Institute  
e-mail: [vapresyan@hse.ru](mailto:vapresyan@hse.ru), [nickmikulin@gmail.com](mailto:nickmikulin@gmail.com)

While linguistic corpora have long been established as an efficient instrument of linguistic research, dictionaries are underrated in this capacity. Their utilitarian function is mostly defined by their usefulness to a general user. Accordingly, electronic dictionaries are primarily designed to satisfy the needs of a language learner, and not of a linguist. But dictionaries of an active type, with their wealth of structured multi-level information about words, could feed linguistic theory in a variety of ways. Active dictionary of the Russian language (ADR) is an advanced active dictionary aimed at recording all linguistic information necessary to use a word correctly. ADR includes information on the morphological, stylistic, prosodic, syntactic, and combinatorial properties of lexical items, as well as on their major semantic relations (synonyms, antonyms, converse synonyms, derivatives). The current project involves turning ADR into an electronic database with a variety of search functions, that would both advance linguistic theory and serve the needs of a general user. The paper is structured as follows: 1) General principles of ADR; 2) ADR Online as an electronic database; 3) Sample uses of ADR Online as an instrument of linguistic research.

**Keywords:** active dictionary; lexicographic type; semantic class; database; integrated description of language; negative polarity

## API-Powered Dictionary Websites (Software Demonstration)

**Sandro Cirulli**

Oxford University Press  
e-mail: [sandro.cirulli@oup.com](mailto:sandro.cirulli@oup.com)

Oxford University Press (OUP) recently started the Oxford Global Languages (OGL) initiative whose focus is to provide language resources for digitally under-represented languages. In August 2015 OUP launched language websites for isiZulu and Northern Sotho and in March 2016 added websites for Malay and Urdu. The backend of these websites is based on an API retrieving data originally modelled in RDF and delivering data to the frontend in JSON.

The software presentation focuses on the API (Application Programming Interface) developed to power these websites. We show API calls to search dictionary entries, add

new content on the website in real-time and delete it if need be. We discuss the advantages of API-powered websites, how the API allowed OUP to crowdsource linguistic data from online communities, and how APIs facilitate the integration of data with external systems and developers. Finally, we outline future work for the next phase of development of the API and OGL websites.

**Keywords:** API; dictionary; isiZulu; Northern Sotho; Malay; Urdu

## Displaying Language Diversity on the European Dictionary Portal - COST ENeL - Case Study on Colours and Emotions and their Cultural References

**Amelie Dorn<sup>1</sup>, Alina Villalva<sup>2</sup>, Voula Giouli<sup>3</sup>, Wiebke Blanck<sup>4</sup>,  
Kira Kovalenko<sup>5</sup>, Eveline Wandl-Vogt<sup>1</sup>**

Austrian Academy of Sciences, Austrian Centre for Digital Humanities<sup>1</sup>;  
Universidade de Lisboa<sup>2</sup>; Institute for Language and Speech Processing, Athena RIC<sup>3</sup>;  
Göttingen Academy of Sciences,  
Deutsches Wörterbuch von J. Grimm und W. Grimm<sup>4</sup>;  
Russian Academy of Sciences, Institute for Linguistic Studies<sup>5</sup>  
e-mail: Amelie.Dorn@oeaw.ac.at, AlinaVillalva@letras.ulisboa.pt, voula@ilsp.gr,  
wiebke.blanck@gmx.net, kira.kovalenko@gmail.com, Eveline.Wandl-Vogt@oeaw.ac.at

In this paper we present a case study on colour and emotion terms and their cultural references in the framework of the COST European Network of e-Lexicography (ENeL), working towards Pan-European lexicography. We take an initial use case of *red* in connection with emotions (*anger*) and look at its roots across different European languages, including Russian. Our data model offers the possibility of connecting these fields in the context of digital lexicography using markup for etymological information with description standards like ONTOLEX or TEI. This is particularly relevant for using and displaying such data on the European Dictionary Portal, potentially offering access to detailed diachronic and synchronic lexicographic knowledge across a variety of languages.

**Keywords:** Linguistic diversity; digital lexicography; European roots

## Designing and Building a Web-based Bilingual Dictionary of South African Sign Language and Afrikaans for Foundation Phase Learners at a School for the Deaf

(Software Demonstration)

**Hanelle Fourie Blair, Hanno Schreiber**

Bureau of the WAT; Stellenbosch University

e-mail: hfourie@sun.ac.za, hanno.schreiber@gmail.com

This software demonstration forms part of a more comprehensive theoretical framework in which a concept model for an electronic bilingual foundation phase dictionary of South African Sign Language and Afrikaans is proposed for a school for the Deaf<sup>1</sup> in South Africa (Fourie 2013). The purpose of the proposed dictionary is to give learners of the school electronic access to written school word lists, with signs and examples in sign language displayed in video format – therefore in the form of an electronic dictionary. A written word list does not meet the needs of school learners in this case and a dictionary that is designed for the Deaf target users at the school will be of far greater value in the learners' communication in both sign language and the written language. Certain components of the model can also be implemented by users other than the original target users, so that it may also be of use to the broader Deaf community, both in South Africa and internationally. The software that will be shown is a prototype used to demonstrate the practical implementation of the model mentioned above. It is the work of an honours degree student in computer science at Stellenbosch University.

**Keywords:** database; sign language; American sign language; South African sign Language; electronic dictionary; notation system; foundation phase learner; Australian sign language; sign parameter

---

<sup>1</sup> It is common for authors to use Deaf with a capital "D" when discussing individuals who are members of the Deaf community and consider themselves to be culturally Deaf and are users of a sign language, and to use deaf with a lower case of "d" to describe audiological hearing status.

## The Taming of the Polysemy: Automated Word Sense Frequency Estimation for Lexicographic Purposes

Anastasiya Lopukhina<sup>a</sup>, Konstantin Lopukhin<sup>c</sup>,  
Boris Iomdin<sup>a,b</sup>, Grigory Nosyrev<sup>d</sup>

<sup>a</sup>Vinogradov Russian Language Institute, Russian Academy of Sciences,

<sup>b</sup>National Research University Higher School of Economics,

<sup>c</sup>Scrapinghub, <sup>d</sup>Yandex

e-mail: nastyा.lopukhina@gmail.com, kostia.lopuhin@gmail.com,  
iomdin@ruslang.ru, grigorij-nosyrev@yandex.ru

Although word sense frequency information is important for theoretical study of polysemy and practical purposes of lexicography, the problem of sense frequency distribution is a neglected area in linguistics. It is probably because sense frequency is not easy to estimate. In this paper we deal with the problem of automated word sense frequency estimation for Russian nouns. We developed and tested an automated system based on semantic context vectors, supplied with contexts and collocations from the Active Dictionary of Russian – a full-fledged production dictionary that reflects contemporary Russian. The study was performed on RuTenTen11 web-corpus. This allows us to reach a frequency estimation error of 11% without any additional labelled data. We compared sense frequencies obtained automatically with sense ordering in different dictionaries for several words. The method presented in this paper can be applied to any language with a sufficiently large corpus and a good dictionary that provides examples for each sense. The results may enrich language learning resources and help lexicographers order senses within a word according to frequency if needed.

**Keywords:** semantics; lexicography; word sense frequency; web corpora; polysemy; frequency; semantic vectors; word sense disambiguation; WSD

## GESL Vocabulary and Innovation Technologies

Tamar Makharoblidze, George Mirianashvili

Ilia State University

e-mail: tamar.makharoblidze@iliauni.edu.ge, gmirian@gmail.com

Georgian Sign Language (GESL) is a language of Deaf and Hard of Hearing people (DHH) in Georgia. The current researches on GESL are connected with computational linguistics and lexicology. Our group at Ilia State University works on the sign language universal translator, which will be able to translate the texts from any sign language into spoken and vice versa. This issue is concerning the communication problems of DHH worldwide.

We elaborated the theory of neutral signs (NS). Currently we are building the mini corpora for GESL to test the elaborated software prototype. We should collect data according to the proper methodology structuring the information, analyzing and comparing the signs from different sign languages (SL), creating the special API for integration with any other SL corpora, revealing and then testing the elaborated algorithms for the universal SL translator, elaborating the software and hardware design concepts for the final product.

**Keywords:** Georgian sign language; sign recognizing engine; sign language machinery interpreter; SL soft

## The Network Integrator of Croatian Lexicographical Resources

**Marko Orešković, Juraj Benić, Mario Essert**

National and University Library in Zagreb

Faculty of Mechanical Engineering and Naval Architecture

e-mail: moreskovic@nsk.hr, juraj.benic@gmail.com, messert@fsb.hr

In this paper we describe an online application which connects various Croatian lexicographical sources (the standard dictionary, professional terminology, and encyclopedias) into a new network framework, which solves the problems of the currently available network solutions and offers new features. Although developed for the formation and processing of the professional terminology of technical sciences (primarily engineering), this framework is easily extensible to all other science areas (humanities, medicine, architecture, etc.). From a word formation standpoint, this framework provides an automatic generation of a temporary dictionary of a given text (entered or taken from the web), comparison of and updating a permanent dictionary with a temporary one (detection of neologisms), an automatic display (standard, professional and encyclopedic) of definitions of the words from the main dictionary or external (network) sources of a given text. From the point of view of processing, the framework enables a classic text analysis (frequency, concordances, statistics), converting the collected data (e.g. technical words) to the linked open data (LOD), and storing it into the Virtuoso triplestore repository. On this data, a variety of professional ontologies for viewing and searching can be developed.

**Keywords:** network framework; automatic dictionary generation; integration of lexicographic web information; linked open data

## The Development of a Network Thesaurus with Morpho-semantic Word Markups

**Marko Orešković, Mirko Čubrilo, Mario Essert**

National and University Library in Zagreb

Faculty of Mechanical Engineering and Naval Architecture

e-mail: [moreskovic@nsk.hr](mailto:moreskovic@nsk.hr), [mcubrilo@foi.hr](mailto:mcubrilo@foi.hr), [messert@fsb.hr](mailto:messert@fsb.hr)

This paper presents a part of the network frame of Croatian linguistics which focuses on a new kind of thesaurus, based on morpho-semantic features of words. Instead of the classic (e.g. MULTTEXT-EAST) POS tagging of words for grammatical and some semantic categories (e.g. animate), in this paper every word has its hierarchical T-structure which can hold various data types in its branches (string, integer, link, word list, ordered word list etc.), and in that way words and their various occurrence possibilities in a text can be described even better. Moreover, the known WordNet or other semantic structures (e.g. the Croatian Language Portal, terminology repository or a network encyclopedia) can be presented as T-structure nodes in the same way. During this process each word in the definition of an entry is linked to a lexicon, which results in increasing the semantic connectivity of words by at least one order of magnitude (about ten times more semantic relations). Searching through and browsing such a network dictionary brings a new dimension, and words in the dictionary, beside the paradigmatic, also possess all the syntagmatic properties, because the computer processes their appearance in any utterance or sentence as a series of connected nodes (LOD objects). This provides the possibility of storing all data in triplestore (e.g. on the Virtuoso server).

**Keywords:** lexical markup framework; morpho-semantic tree-structures; semantic networks

## Croatian Linguistic System Modules Overview (Software Demonstration)

**Marko Orešković, Jakov Topić, Mario Essert**

National and University Library in Zagreb

Faculty of Mechanical Engineering and Naval Architecture

e-mail: moreskovic@nsk.hr, jakov.topic131@gmail.com, messert@fsb.hr

In this paper we show several segments of program solutions which are a part of the Croatian linguistic system (CLS) that is being developed in several ways and aims at achieving the final integration of all modules. Although the system aims at programmatically connecting all areas of linguistics (from phonetics to discourse), in the demonstration we will show only the segments that are related to general lexicon building (which includes a standard and terminological dictionary of the Croatian language) and will be connected with online repositories and encyclopedias. These program segments are searching for neologisms in documents (i.e. words that have not been marked in the general lexicon), generating grammatical forms of such words if they are changeable and saving them into a lexicon, adding semantic markups (like morpho-syntactic) characteristics, and, finally, monitoring Croatian words in space and time.

**Keywords:** lexicon; Croatian linguistics system; semantic markup; word evolution

## The GEROM Project Search Engine (Software Demonstration)

**Marco Stefano Tomatis**

Università degli Studi Internazionali di Roma

e-mail: marco.tomatis@unint.eu

This paper aims at describing the technical engine underlying the project “GEROM”, an online bilingual (German-Italian) terminological tool which has been developed by the Johannes Gutenberg - Universität of Mainz/Germersheim, Germany and Unint - University of International Studies of Rome, Italy. GEROM has been designed to facilitate the comprehension, analysis and translation of Italian and German texts dealing with different social and cultural aspects of the two Countries which are taken into account. Therefore, in order to provide users with a terminological support which focuses on examples of real language, a corpus-based approach has been adopted. Furthermore, since the core of the system is founded on a web-oriented service, a robust search interface based on CGI (Common Gateway Interface) technology has been implemented.

**Keywords:** terminology; CGI; search engine

---

## **Furthering the Exploration of Language Diversity and Pan-European Culture: The Dariah-CC Science Gateway for Lexicographers**

**Eveline Wandl-Vogt<sup>1</sup>, Roberto Barbera<sup>2</sup>, Giuseppe La Rocca<sup>2</sup>, Antonio Calanducci<sup>2</sup>, Carla Carrubba<sup>2</sup>, Giuseppina Inserra<sup>2</sup>, Tibor Kalman<sup>3</sup>, Gergely Sipos<sup>4</sup>, Zoltan Farkas<sup>5</sup>, Davor Davidović<sup>6</sup>**

1 Austrian Academy of Sciences, Austrian Centre for Digital Humanities;

2 Italian National Institute of Nuclear Physics (INFN);

3 Gesellschaft für wissenschaftliche Datenverarbeitung mbh Göttingen;

4 EGI Foundation; 5 MTA SZTAKI; 6 Ruđer Bošković Institute

e-mail: eveline.wandl-vogt@oeaw.ac.at, roberto.barbara@ct.infn.it,

giuseppe.larocca@egi.eu, antonio.calanducci@ct.infn.it, carla.carrubba@ct.infn.it,

giuseppina.inserra@ct.infn.it, tibor.kalman@gwdg.de, gergely.sipos@egi.eu,

zoltan.farkas@sztaki.mta.hu, davor.davidovic@irb.hr

The paper introduces into a new Science Gateway, developed in the framework of the European Horizon 2020 project EGI Engage - Dariah Competence Centre, which started in March 2015 co-funded by the European Union, with the participation of about 70 (research) units in over 30 countries.

In this paper the authors focus on transdisciplinary collaboration in the framework of explorative lexicography in cultural context. On the one hand, they give a short overview on the architecture of the Science Gateway, used techniques, and specific applications and services developed during the Dariah Competence Centre. On the other hand, they mainly focus on possible added value and changes concerning work flow for Lexicographers and researchers on Lexical resources.

This is exemplified on the European network of COST action IS 1305 “European Network of electronic lexicography (ENeL)”.

**Keywords:** science gateway; research infrastructure; cloud infrastructure; social infrastructure; open science commons; cultural diversity; cultural lexicography; linguistic diversity; pan-european heritage; explorative scholarship

# **Lexicography and Corpus Linguistics**



---

## RuSkELL: Online Language Learning Tool for Russian Language

**Valentina Apresjan, Vít Baisa, Olga Buivolova,**

**Olga Kultepina, Anna Maloletnjaja**

National Research University “Higher School of Economics”,

School of Linguistics

Vinogradov Russian Language Institute

NLP Centre, Masaryk University

e-mail: [vapresyan@hse.ru](mailto:vapresyan@hse.ru), [xbaisa@fi.muni.cz](mailto:xbaisa@fi.muni.cz), [ovbuyvolova@edu.hse.ru](mailto:ovbuyvolova@edu.hse.ru),

[okultepina@edu.hse.ru](mailto:okultepina@edu.hse.ru), [maloletnyaya@gmail.com](mailto:maloletnyaya@gmail.com)

RuSkELL (“Russian + Sketch Engine for Language Learning”) is a new online resource intended for researchers and learners of Russian. It incorporates a specially pre-processed corpus and the interface allows users to search for phrases in sentences, extract salient collocates and show similar words. The tool builds upon its English counterpart SkELL (Baisa & Suchomel 2014). The aim of the project is to adapt the existing SkELL tool to Russian, improve its performance and make it user-friendly to Russian users. The existing problems include errors in query output and insufficiently transparent interface. The project aspires to solve them by 1) modifying Sketch grammar rules to exclude irrelevant output and to add informative collocations unaccounted for in the existing Sketch grammar; 2) providing collocation groups with easy-to-understand labels in Russian. We describe the process of building the language data and problems we need to address to accommodate the tool for the specificities of the Russian language.

**Keywords:** online language tool; Sketch Engine for language learning; sketches; collocations

## Lexicographic Potential of the Georgian Dialect Corpus

**Marina Beridze, David Nadaraia**

Tbilisi Ivane Javakhishvili State University,

Arn. Chikobava institute of Linguistics

e-mail: [marine.beridze@gmail.com](mailto:marine.beridze@gmail.com)

The project *Linguistic Portrait of Georgia* envisages various aspects of documentation of Georgian linguistic reality by means of corpus methodologies. This title has been an umbrella for three large-scale projects within the framework of which *The Georgian Dialect Corpus – GDC* (<http://corpora.co>) was developed.

Presently, the architecture and text base of the corpus have been designed, being permanently developed and updated. Besides, the lexicographic base of the corpus has been organized, agglomerating data from printed dialect dictionaries.

The lexical stock of the corpus has been presented based on text, lexicographic and encyclopedic data.

The total quantity of tokens in the corpus is estimated to be up to 2 000 000, while the lexicographic base has 60 000 items (lemmas with entries) by now; this quantity is considerably increased owing to phonetic and grammatical variations, frequently associated with a single lexical item.

**Keywords:** Georgian Dialect Corpus; lexicographic base; encyclopedic data

## What Do Graded Decisions Tell Us about Verb Uses

**Silvie Cinková<sup>1</sup>, Ema Krejčová<sup>1</sup>, Anna Vernerová<sup>1</sup>, Vít Baisa<sup>2</sup>**

Charles University in Prague<sup>1</sup>, Masaryk University Brno<sup>2</sup>

e-mail: {cinkova, krejcová, vernerová}@ufal.mff.cuni.cz,  
xbaisa@fi.muni.cz

We work with 1450 concordances of 29 English verbs (50 concordances per lemma) and their corresponding entries in the Pattern Dictionary of English Verbs (PDEV). Three human annotators working independently but in parallel judged how well each lexical unit of the corresponding PDEV entry illustrates the given concordance. Thereafter they selected one best-fitting lexical unit for each concordance – while the former setup allowed for ties (equally good matches), the latter did not. We measure the interannotator agreement/correlation in both setups and show that our results are not worse (in fact, slightly better) than in an already published graded-decision annotation performed on a traditional dictionary. We also manually examine the cases where several PDEV lexical units were classified as good matches and how this fact affected the interannotator agreement in the best-fit setup. The main causes of overlap between lexical units include semantic coercion and regular polysemy, as well as occasionally insufficient abstraction from regular syntactic alternations, and eventually also arguments defined as optional and scattered across different lexical units despite not being mutually exclusive.

**Keywords:** Word Sense Disambiguation; usage patterns; computational lexicography; graded decisions; Likert scales; Corpus Pattern Analysis; Pattern Dictionary of English Verbs; regular polysemy; coercion

## Enriching Georgian Dictionary Entries with Frequency Information

**Sophiko Daraselia, Serge Sharoff**

University of Leeds

e-mail: mlsd@leeds.ac.uk, s.sharoff@leeds.ac.uk

In this paper we will discuss the integration of corpus analysis into the dictionary making process for the Georgian language. In general, corpus-based lexicography is not a common practice in lexicography in Georgia. This paper presents the first attempt to introduce the corpus-based dictionary wordlist, entries and examples from the Georgian web-corpus – KaWaC. This is a large web corpus of modern Georgian language covering recent 10-15 years of the language development. It contains a wide range of text types, topics and regions from the Internet excluding translations and poetry on the assumption that the language of translation and poetry deviate from the naturally produced language, as the corpus aims to represent naturally occurring modern Georgian language. Within this research we have defined the dictionary wordlist - 10,000 lemmas from the corpus that is a core vocabulary for the Georgian language, compiled the dictionary entries and extracted dictionary examples from the corpus.

**Keywords:** corpus linguistics; web-corpus; corpus-based lexicography; learner's dictionaries

## Corpus of the Georgian Language

**Nino Doborjginidze, Irina Lobzhanidze**

Ilia State University

e-mail: nino\_doborjginidze@iliauni.edu.ge,  
irina\_lobzhanidze@iliauni.edu.ge

The paper is the public presentation of the research program Corpus of Modern, Middle and Old Georgian texts including the Knight in the Panther's Skin and Georgian Chronicles. The Corpus was compiled with the support of the Ilia State University and the Shota Rustaveli Science Foundation aiming at building a new, extensive and representative corpus of the Georgian language.

This paper is structured as follows: Section 1 includes background and research questions; Section 2 presents a methodological approach and briefly summarizes its theoretical prerequisites; Section 3 includes the findings and hypotheses, which refer to the compilation of a corpus including Morphological Analyzer, Corpus Access and Meta-Data Information; and Section 4 provides the answers to the research questions and corpus itself.

**Keywords:** corpus of the Georgian language; morphological analyzer of the Georgian language

## From Diachronic Treebank to Dictionary Resource: the Varangian Rus Project

**Hanne Martine Eckhoff, Aleksandrs Berdičevskis**

UiT The Arctic University of Norway

e-mail: [hanne.m.eckhoff@uit.no](mailto:hanne.m.eckhoff@uit.no), [aleksandrs.berdicevskis@uit.no](mailto:aleksandrs.berdicevskis@uit.no)

In this paper we present the Varangian Rus' dictionary resource, which is based on the Tromsø Old Russian and OCS<sup>1</sup> Treebank (TOROT),<sup>2</sup> a diachronic treebank of Russian containing a balanced selection of 11th–17th century Old East Slavic and Middle Russian texts. The treebank is lemmatised and has rich morphological and syntactic annotation. With simple glossing of the word meanings found in the treebank, we are able to generate a dictionary resource with rich grammatical information. The dictionary work also enables us to improve the lemmatisation of the treebank considerably, as well as annotation at other levels. The dictionary resource will be published as a searchable online tool towards the end of 2016 and is intended for students and scholars alike.

**Keywords:** treebank; historical dictionary; Russian

## Semantic Annotation of Verbs for the Tatar Corpus

**Alfiya Galieva, Olga Nevzorova**

Research Institute of Applied Semiotics of Tatarstan Academy of Sciences, Kazan Federal University

e-mail: [amgalieva@gmail.com](mailto:amgalieva@gmail.com), [onevzoro@gmail.com](mailto:onevzoro@gmail.com)

This paper discusses the problem of developing the metalanguage for linguistic applications and introduces a tag set for the semantic annotation of verbs for the Tatar National Corpus. At present, there are no generally accepted standards for the development of corpus semantic annotation. In many cases it is made by individual researchers or teams for one or another research project, and characteristics of tag sets used in thesauri and electronic corpora differ in many respects.

Using available semantic classifications of vocabulary for different languages and relying upon data from Tatar lexicons, we created a model of the semantic system of Tatar verbs and divided them into semantic classes (3,200 words). We distinguished semantic tags of two types: constructional (categorial) tags, independent of semantic classes of verbs, and semantic (thematic) tags, determining semantic classes of verbs.

1 Old Church Slavonic is abbreviated OCS throughout.

2 The TOROT annotation and browsing interface is found at <https://nestor.uit.no>. Versioned releases of annotated texts in XML format can be accessed via <http://torottreebank.github.io/>

For separating these classes we used the hierarchical and the overlapping classifications, so the same verb may belong to more than one class. The approach is based on the data from explanatory dictionaries of the Tatar language, bilingual Russian-Tatar dictionaries and the system of semantic annotation of the Russian National Corpus. In the current version of our semantic annotation we use 3 categorial and 59 thematic tags.

**Keywords:** Tatar verb; semantics; corpus; semantic annotation

## A Corpus-based Approach to Lexicography: Towards a Thesaurus of English Idioms

**Guzel Gizatova**

Kazan State Agrarian University

e-mail: guzelgizatova@hotmail.com

This paper deals with the principles of constructing a new ideographic dictionary of English idioms (DEI) based on corpus data. Idioms are arranged by their figurative meaning rather than alphabetically. The purpose of this paper is to explore English idioms along two revealing lines of inquiry. The first line of inquiry suggests constructing an ideographic dictionary (thesaurus) of English idioms; the second line illustrates advantages of organizing such a dictionary on an analysis of corpus data.

The need for a new dictionary of English idioms is motivated by the fact that at present there is no corpus-based dictionary of English idioms built on the thesaural principle. Ideographic description of idioms enables the reader to find the biggest possible amount of idiomatic word combinations of the language that expresses the given concept.

Due to corpus data, the dictionary presents a range of syntactic patterns, polysemous idioms and unexpected variants which cannot be retrieved from the existing idiomatic and monolingual dictionaries of the English language. Many dictionaries fail to register all meanings of idioms. The corpora help reveal the majority of meanings of polysemous idioms. Apart from its theoretic relevance as an instrument of description of the mental lexicon, a new ideographic dictionary can be used for the purposes of translation and language acquisition.

**Keywords:** thesaurus; corpus-based; English idioms

## French Cross-disciplinary Scientific Lexicon: Extraction and Linguistic Analysis

**Sylvain Hatier, Magdalena Augustyn, Thi Thu Hoai Tran,**

**Rui Yan, Agnès Tutin, Marie-Paule Jacques**

LIDILEM – EA 609 – Université Grenoble Alpes

e-mail: {sylvain.hatier, magdalena.augustyn, thi-thu-hoai.tran, rui.yan, agnes.tutin, marie-paule.jacques}@univ-grenoble-alpes.fr

This paper presents the work we carried out to extract and structure a specialized lexicon based on a corpus of French scientific articles in the fields of humanities and social sciences. The characteristics of the targeted lexicon may be summarized as follows: it is not domain-related inasmuch as it is shared by various disciplines; it serves to express the specific operations, naming the objects and exposing the results of research processes. In this view, the targeted lexicon studies here is genre-related. We designed this cross-disciplinary scientific lexicon (CSL) as a resource for several purposes: it may serve natural language processing, e.g. as a stoplist for automatic terms identification, as well as foreign language teaching. Indeed, students and scholars in the sciences need to become familiar with the rhetoric of the research article, thus needing to master these words. We present here the two-stage creation of this lexicon: first, it was semi-automatically extracted from a corpus of 500 research articles spanning ten disciplines. Second, it was manually structured to reflect the semantics and rhetoric of science. This structure takes into account the lexico-syntactic properties of CSL nouns, adjectives, verbs and adverbs. The resource will be freely available for academic purposes.

**Keywords:** corpus linguistics; academic writing; open lexical resources; natural language processing

## Corpus-based Extraction of Collocations for Near-Synonym Discrimination

**Mariusz Kamiński**

University of Applied Sciences, Nysa

e-mail: mariusz.kaminski@pwsz.nysa.pl

This paper shows a method of extracting collocations from a corpus for the purpose of discrimination of near-synonyms. Information on synonym discrimination, which is often part of usage notes appended to entries in general dictionaries, specifies points of differences between words of similar meaning. Using computer technology, one can automate the process of collocation retrieval and reduce the amount of time and effort needed to prepare such information. The paper emphasises the importance of

examining collocational behaviour when establishing meaning differences, but it also underlines the necessity of a critical analysis of existing records of word meaning, in particular dictionaries. An evidence from a parsed corpus is used in order to find collocations for a set of synonymous adjectives (*false, counterfeit, artificial, bogus, forged*) standing either in attributive or predicative positions. The corpus findings are compared against definitions in *MacMillan Dictionary* online.

**Keywords:** collocation; synonymy; discrimination; extraction; corpus; parser

## The Division into Parts of Speech in the *Corpus-based Dictionary of Polish Sign Language*

**Jadwiga Linde-Usiekiewicz, Paweł Rutkowski**

University of Warsaw

e-mail: jlinde@uw.edu.pl, p.rutkowski@uw.edu.pl

Recent decades have seen a spectacular development in sign language lexicography, both in technological and theoretical terms. This new subfield of lexicography has encountered numerous challenges, related to the structural differences between spoken and sign languages. One of these has been how the issue of dividing words into different parts of speech (POS) should be handled for sign languages.

For both theoretical and sociolinguistic reasons (Linde-Usiekiewicz & Olko 2006, Linde-Usiekiewicz et. al 2014, 2016), the *Corpus-based Dictionary of Polish Sign Language* compiled at the University of Warsaw provides Polish descriptive sense definitions instead of sets of equivalents. Since the morpho-syntax of spoken Polish makes it impossible to adequately define a sense without imposing a POS interpretation on it, we decided to introduce a functional division of senses in the entry structure. Thus, four usage types have been distinguished for PJM (*polski język migowy*, Polish Sign Language), defined by their resemblance to the traditional POS pattern: a noun-like, verb-like, adjective-like, and adverb-like usage. Each of these types of usage was established partly on semantic grounds (Wierzbicka 2000), and partly on syntactic grounds (Meir 2013).

Interestingly, this procedure has also led us to identify usage types with no obvious counterparts in the traditional POS system, namely signs used autonomously, i.e. in a separate utterance.

**Keywords:** Polish sign language (PJM); corpus linguistics; parts of speech; homonymy; polysemy; autonomous use

## **The Life and Death of Neologisms: On What Basis Shall We Include Neologisms in the Dictionary?**

**Kilim Nam, Soojin Lee, Hae-Yun Jung, Jun Choi**

Kyungpook National University

e-mail: nki@knu.ac.kr, sjmano27@naver.com,  
haeyun.jung.22@gmail.com, c-juni@hanmail.net

This research analyzes web crawling corpora and examines how many of the neologisms that are coined every year are dying out and how many endure. It seeks to grasp what implications the results of the analysis have for the inclusion of these neologisms in the dictionary. The Korean government initiated the investigation into neologisms in 1992 and has been supervising this research project ever since. Some 400 to 500 coinages that meet definite criteria are being extracted every year, compiled and printed out in the form of a glossary. This paper focuses on the years 2005 and 2006, for which 408 and 530 respectively, that is, 938 new words in total, were recorded. The study turns then to the analysis of the usage changes in the Korean mass media which these neologisms have been undergoing for the past decade. On a quantitative level, the investigation shows that 27% of those neologisms have been in consistent usage for the last ten years.

**Keywords:** neologisms; usage changes; web crawling corpus; frequency; news articles

## **Parallel Corpora as a Source of Defining Language-specific Lexical Items**

**Dmitri Sitchinava**

Institute of Russian language, Russian Academy of Sciences

e-mail: mitrius@gmail.com

The paper presents an attempt to propose an exact method for identifying the so-called “language-specific” lexicon, a controversial notion often reasonably questioned. An aligned bilingual parallel corpus is chosen as an instrument for finding “specificity”, and statistical entropy and other indices are used as markers of the dispersion of translation patterns (viz. stimuli). For example, a word can be deemed (maximally) language-specific if it occurs multiple times in a given bilingual corpus and is translated each time in a different way. A word is minimally (or simply not) language-specific if it is translated each time identically. Some problems relative to the application of this method are discussed. These data can be explicitly used in bilingual dictionaries.

**Keywords:** parallel corpora; language-specific lexicon; translation patterns; statistics

## ***Nach Belieben kombinieren? Korpusbasierte Beschreibung präpositionaler Mehrworteinheiten im Sprachvergleich***

**Kathrin Steyer, Katrin Hein**

Institut für Deutsche Sprache (IDS), Mannheim

e-mail: steyer@ids-mannheim.de, hein@ids-mannheim.de

Präposition-Substantiv-Verbindungen mit rekurrentem Nullartikel in adverbialer Verwendung – z.B. *nach Belieben, auf Knopfdruck, ohne Ende* oder *bei Nacht* – sind ein in der Mehrwortforschung bisher eher vernachlässigter Typ. Sie sind Untersuchungsgegenstand des laufenden Forschungsprojekts „Präpositionale Wortverbindungen kontrastiv“ (beteiligte Institutionen: IDS Mannheim, Universität Santiago de Compostela, Universität Trnava), in das wir in unserem Vortrag einen Einblick vermitteln.

Es wird skizziert, wie sich solche Wortverbindungen sowie abstraktere präpositionale Wortverbindungsmuster vom Typ *[in + SUB<sub>X-Zeit(en)</sub>* (z.B. *in Echtzeit, in Krisenzeiten*) aus kontrastiver Sicht (Deutsch – Spanisch – Slowakisch) korpusbasiert untersuchen und lexikografisch beschreiben lassen. Von großem Interesse – gerade auch für Fremdsprachenlerner – sind dabei insbesondere die semantisch-funktionalen Restriktionen, denen solche Entitäten unterliegen. Basierend auf den theoretischen und empirischen Grundannahmen des am IDS entwickelten Modells „Usuelle Wortverbindungen“ (vgl. Steyer 2013) werden im Projekt zunächst Kollokations- und Kotextmuster für die binären deutschen Mehrworteinheiten induktiv in sehr großen Korpora ermittelt; im Anschluss werden sie einem systematischen Vergleich mit dem Spanischen und Slowakischen unterzogen. Methodisch greifen wir – in allen drei Sprachen – u.a. auf Kookkurrenzprofile zu den Wortverbindungen sowie auf Slotanalysen zu definierten Suchmustern zurück. Ziel des Projekts ist u.a. die Entwicklung eines neuartigen Prototyps für eine multilinguale Aufbereitung des Untersuchungsgegenstands (speziell für Fremdsprachenlerner).

**Keywords:** korpusbasierte Phraseologie; Kollokationsforschung; Corpus Pattern Analysis; Gebrauchsbasiertheit; Schnittstelle Konstruktionsgrammatik – Phraseologie; Äquivalenztheorien

---

# Bi- and Multilingual Lexicography



---

## Problems of Marine Terminology in Georgia

**Inga Adamia**

Ivane Javakhishvili Tbilisi State University

e-mail: [inga\\_adamia@mail.ru](mailto:inga_adamia@mail.ru)

The article opens with an overview of the history of formation and development of the Georgian technical (i.e. specialized) terminology. Discussed in further detail are the issues associated with the development of Georgian marine terminology and marine dictionaries. Their importance for ever intensifying scientific and economic relations is also highlighted.

**Keywords:** marine terminology; Georgian special terminology

## Multilingual CPA: Linking Verb Patterns across Languages

**Vít Baisa<sup>1</sup>, Sara Može<sup>2</sup>, Irene Renau<sup>3</sup>**

<sup>1</sup>Masaryk University, Czech Republic; <sup>2</sup>University of Wolverhampton;

<sup>3</sup>Pontificia Universidad Católica de Valparaíso

e-mail: [xbaisa@fi.muni.cz](mailto:xbaisa@fi.muni.cz), [S.Moze@wlv.ac.uk](mailto:S.Moze@wlv.ac.uk), [irene.renau@pucv.cl](mailto:irene.renau@pucv.cl)

This paper presents the results of a pilot study in linking corresponding English and Spanish verb patterns using both automatic and manual procedures. Our work is rooted in Corpus Pattern Analysis (CPA) (Hanks 2004, 2013), a corpus-driven technique that was used in the creation of existing monolingual pattern dictionaries of English and Spanish verbs, which were used in our experiment to design a gold standard of manually annotated verb pattern pairs. Research in CPA has inspired parallel projects in English, Spanish, Italian and German. Our study represents the first attempt to build a multilingual lexical resource by linking verb patterns in these languages. Verb has special difficulties related to grammar and argument structure that we do not find in other parts-of-speech, and for that reason we think that it is necessary to create a specific resource for them. After applying the automatic matching to a set of 87 Spanish verbs linked to 176 English verbs, an evaluation of a random selection of 50 of these pairs show 80% precision.

**Keywords:** Corpus Pattern Analysis; corpus lexicography; multilingual resources; verb patterns

## Идиомы в грузино-осетинском и осетино-грузинском словаре

**Найра Бепиева, Нино Попиашвили**

Тбилисский государственный университет

e-mail: bepieva.naira@gmail.com, nino.popiashvili@tsu.ge

Идиома появляется в языке сотворившего его народа и закрепляется как механизм художественного, фразеологического восприятия и выражения мира. Посредством идиомы люди обогащают выразительную форму языка, его содержание. Идиомы являются носителями культуры, истории, материальных и духовных ценностей народа.

Идиома – это один из видов фразеологических единиц. Ею часто пользуются в речи с целью более художественной передачи мысли. Идиома состоит, по крайней мере, из двух слов, однако ее значение не определяется значением входящих в него компонентов, взятых по отдельности, хотя у некоторых идиом общее значение совпадает с прямым значением входящих в нее слов. Вместе с тем, идиома имеет художественно оформленное значение, т. е. она одновременно и свободное словосочетание, и устойчивое выражение с фразеологической семантикой.

Как в грузинском, так и в осетинском языках часто встречаются структурно и семантически идентичные идиомы, однако мы специально не останавливаемся на вопросе происхождения и заимствования того или иного оборота или крылатой фразы; мы хотели лишь отметить, что в результате длительного исторического взаимодействия грузин и осетин, в обоих языках отобразились параллельные художественные формы, идиомы, схожие и лексически, и семантически, что указывает на длительные культурные и общественные взаимосвязи носителей этих языков.

**Keywords:** словарь; грузинский; осетинский, идиома

## L'Accentuation des Emprunts Français dans les Dictionnaires Anglais de Prononciation

**Pierre Fournier**

Université Paris 13 - Pléiade (EA7338)

e-mail: pierre.fournier@univ-paris13.fr

La langue anglaise a beaucoup emprunté à la langue française. Ainsi, les emprunts français sont massivement attestés dans les dictionnaires anglais de prononciation. A partir de deux sources dictionnaires anglaises de référence et d'un corpus représentatif d'emprunts français en anglais, cet article étudie l'accentuation des emprunts français

en anglais britannique et américain, tout en introduisant la distinction nom commun / nom propre, et ce, dans une perspective synchronique mais également diachronique, avec une comparaison intra-dictionnaire et inter-dictionnaire. Les deux sources sont le *Longman Pronunciation Dictionary* (LPD) et le *Cambridge English Pronouncing Dictionary* (EPD). Les résultats démontrent l'impact du paramètre variétal sur l'accentuation des emprunts français en anglais mais aussi la présence de morphologies déterminantes avec influence distincte selon la variété d'anglais considérée, ces deux phénomènes étant interdépendants. L'anglais américain reproduit plus fidèlement l'accentuation démarcative finale du français que l'anglais britannique. Toutefois, l'étude explore également les limites d'un corpus d'analyse élaboré à partir de terminaisons.

**Keywords:** emprunts français en anglais; accentuation lexicale; dictionnaires de prononciation; *loan phonology*

## Problems with the Description of Words Relating to Gender and Sexuality in English-Japanese Learner's Dictionaries

**Yuri Komuro, Saeko Nagashima**

Chuo University

e-mail: yurik@tamacc.chuo-u.ac.jp, s-naga@ tamacc.chuo-u.ac.jp

While English-Japanese learner's dictionaries have developed relatively sophisticated ways of describing sexist terms, they are not fully successful in the treatment of words related to gender and sexuality as a whole. Our survey reveals that (1) some English-Japanese learner's dictionaries translate non-discriminatory English words into derogatory Japanese words; (2) there seems to be no clear policy about providing translation equivalents with discriminatory English words; and that (3) non-derogatory words are given carelessly as synonyms of derogatory headwords. This paper will report mismatches and inconsistency in the selections and arrangements of translation equivalents for words related to gender and sexuality in current English-Japanese learner's dictionaries, and argue that low awareness of the issues relating to sexuality in Japanese society underlies such deficiencies. It will also propose how those lexical items should be treated, suggesting immediate improvements to be made in order not to reproduce discrimination against sexual and gender minorities.

**Keywords:** gender and sexuality; English-Japanese learner's dictionaries; translation equivalents

## Theoretical Aspects of Rendering Kartvelian Verb Stems in the Georgian-Megrelian-Laz-Svan-English Dictionary

**Ramaz Kurdadze, Maia Lomia, Ketevan Margiani**

Ivane Javakhishvili Tbilisi State University

e-mail: ramaz.kurdadze@tsu.ge, maia.lomia@tsu.ge,  
keteven\_margiani@mail.ru

The paper addresses principles of classification of Kartvelian verb forms featured in the Georgian-Megrelian-Laz-Svan-English dictionary. Voice being the main morphological category of verbs in the Kartvelian languages, the distinction based on active and passive forms was deemed the most relevant criterion for the description of verb forms. Such description also enabled to determine the possibility of derivation of active and passive forms from one and the same stem and to establish the presence of common or different voice-related morphological patterns and markers in the Kartvelian languages. The verbs presented in the dictionary can be classified into *native* and *borrowed* stock of words. Most of the loanwords in Megrelian and Svan have origins in the Georgian language, while Laz appears to have borrowed largely from the Turkish language. In the Kartvelian languages verb forms may derive either from verb stems proper or nominal stems.

Verb forms in the Kartvelian languages are modelled according to essentially the same principle and are characterised by common morphological features. Sound correspondences are evident in the case of vocalic prefixes (Geo/Svan a-; Megr./Laz o-). Diachronic and synchronic linguistic evidence reveals asymmetry of passive voice forms between literary and non-written languages (Georgian -d: Megr./Laz/Svan i-).

**Keywords:** Kartvelian; lexicography; verb

## Reflecting on Grammar Information in Some Tshivenda Bilingual Dictionaries: A Challenge to Vhavenda Lexicographers

**Munzhedzi James Mafela**

University of South Africa

e-mail: mafelmj@unisa.ac.za

Dictionaries are used by people of all types and all skills. The users consult them mainly for the meaning of words. In the case of bilingual and trilingual dictionaries, users search for both the meaning and equivalents in the target languages because these serve as learning tools. Therefore the entry definition in a bilingual dictionary should provide the equivalent, accompanied by grammatical information to provide credible meaning.

Grammatical information is essential to learners of a foreign language. All dictionaries in Tshivenda are bilingual, except one monolingual dictionary and one trilingual dictionary. There are concerns from users that bilingual and trilingual dictionaries in Tshivenda do not provide enough information for the users. As a result, learners of Tshivenda struggle to learn the language due to lack of the necessary information. This prompted the researcher to investigate the existence of grammatical information in some Tshivenda bilingual dictionaries. Three most used Tshivenda bilingual dictionaries have been selected for the purpose of investigation, i.e. *Venda Dictionary: Tshivenda – English, Improved Trilingual Dictionary: Venda – Afrikaans –English, and Tshivenda/English: ɏhalusamaipfi/Dictionary*.

**Keywords:** Tshivenda; bilingual dictionary; trilingual dictionary; grammatical information; lexicographers; foreign language; learners

## The *wdlpOst* Toolset for Creating Historical Loanword Dictionaries (Software Demonstration)

**Peter Meyer**

Institut für Deutsche Sprache, Mannheim  
e-mail: meyer@ids-mannheim.de

The *wdlpOst* dictionary writing system to be presented in this paper has been developed for the specific purposes of a lexicographical project on German loanwords in the East Slavic languages Russian, Belarusian, and Ukrainian. The project's main objectives are (i) to document those loanwords for which a cognate lexical borrowing from German is known in Polish and (ii) to establish possible borrowing pathways for these lexical items. In a first phase of the project, the collaborative client/server architecture of the *wdlpOst* system has been used for excerpting detailed lexicographical information from a large range of historical and contemporary East Slavic dictionaries, taking the entries in a large dictionary of German loanwords in Polish as a common frame of reference. For the project's second phase, the *wdlpOst* system provides innovative tooling for compiling entries for the East Slavic loanwords. Most importantly, the numerous word sense definitions for a set of cognate loanwords, as excerpted from different lexicographical sources, are mapped onto a system of newly defined cross-language word senses; in a similar vein, the phonemic and graphemic variation in the loanwords and their derivatives is captured through a tool that abstracts from dictionary-specific idiosyncrasies.

**Keywords:** loanword lexicography; dictionary writing system; Slavic languages

## Spezifik der Darstellung von Verben in georgisch-deutschen Wörterbüchern

**Russudan Sekalashvili**

Iwane-Dschawachischwili-Universität Tbilissi

e-mail: [rusudan.zekalashvili@tsu.ge](mailto:rusudan.zekalashvili@tsu.ge), [rusikoz@yahoo.com](mailto:rusikoz@yahoo.com)

Die vorliegende Arbeit gibt einen Überblick über georgisch-deutsche Wörterbücher, die im Ausland und in Georgien herausgegeben worden sind. Der Schwerpunkt liegt dabei auf der Betrachtung des Darstellungsprinzips des georgischen Verbs als Stichwort, welches sehr problematisch in der georgischen Lexikografie ist. Das Verb kann in Wörterbüchern hauptsächlich zweierlei dargestellt werden: 1. in der konjugierten Form (Präsens oder Futur, die 1. oder 3. Person) und 2. in der Form des Verbalnomens, s.g. *"Sacqisi"* oder *"Masdar"*. Das Hauptziel des Vortrags ist es, die Spezifik der georgischen Verbalnomina zu erklären und die Nach- und Vorteile ihres Gebrauchs als Stichwort in Übersetzungswörterbüchern aufzuzeigen.

Zu diesem Zweck werden georgisch-deutsche Wörterbücher analysiert, besonders ausführlich das Wörterbuch von Kita Tschenkeli (1960-1965), wo die Verben nach den Wurzeln alphabetisch angeordnet und mit dem Index versehen sind. Mit einer Wurzel sind alle möglichen konjugierten Verbformen aufgeführt. Außerdem ist das Verbdarstellungsprinzip im Verbenwörterbuch von H. Fähnrich (2013) und im georgisch-deutschen Wörterbuch von R. Sekalashvili (2009) besprochen, in dem das Verb in der Form des Verbalnomens dargestellt ist, aber mit der Angabe der konjugierten Formen. Außerdem werden die Verbklassen, Konjugationstypen und einige grammatische Merkmale der Verben beschrieben, die in genannten Wörterbüchern angegeben sind und für die Spracherlernung sehr nützlich scheinen.

**Keywords:** georgisch-deutsche Wörterbücher; Verb; Verbalnomen; konjugierte Form des Verbs

## Bilingual Lexicography in the Republic of Tatarstan in 1990-2010

**Gulshat Safiullina**

Kazan Federal University

e-mail: safiullina\_gulshat@mail.ru

This paper overviews the contemporary development of bilingual Tatar lexicography which started with the Declaration of Sovereignty of the Republic of Tatarstan in 1992 and its future perspective. The Law on Languages in 1992 proclaimed the Tatar language to be the second state language in the Republic of Tatarstan along with the Russian to safeguard and promote the language of Tatars. The renaissance of national awareness resulted in multitude of lexicographic studies. First, Tatar-Russian terminological dictionaries, containing study material in spheres of oil refinery and management were compiled, then the dictionaries for learners, primary school pictorial dictionaries, college students' dictionaries, finally, full academic dictionaries. The obvious progress has been made in compiling the entries that meet the requirements of theoretical paradigms and practical applicability and present the proper information on the level of morphology, phonetics, semantics and stylistics. The steps to be taken are to include the example sentences from the corpus for implicit information about syntactic relationships at the level of phrases and sentences; to produce online versions of dictionaries relevant to digital reality; to make academic dictionaries adjusting them to the needs of the Tatar-speaking audience in mastering the English language.

**Keywords:** dictionary; Tatar; lexicography

---

# Lexicography for Specialised Languages, Terminology and Terminography



---

## Using Argument Structure to Disambiguate Verb Meaning

**Miriam Buendía Castro, Beatriz Sánchez Cárdenas**

University of Castilla-La Mancha, University of Granada

e-mail: miriam.buendia@uclm.es, bsc@ugr.es

This study proposes a methodology to disambiguate verb meaning in terminographic resources. To this end, the underlying approach used for verb entries description in the environmental knowledge base EcoLexicon has been applied (Buendía, Montero and Faber 2014; Faber and Buendía 2014). The description is based on three parameters: (i) the nuclear meaning of the verb (i.e. its lexical domain, as proposed by the Lexical Grammar Model (Faber & Mairal 1999); (ii) its meaning dimension (i.e. the lexical subdomain); (iii) its predicate-argument structure highlighting the semantic categories of the arguments. Our study proves that a verb can activate different meaning dimensions in different lexical domains, depending on the semantic categories, or different roles (Van Valin 2005) of its arguments. This way of describing verb meaning according to lexical domains and subdomains, semantic roles and semantic categories helps disambiguate verbs and represent their meaning in specialized resources.

**Keywords:** verbs; disambiguation; specialized resources

## Encoding Context in Bilingual Specialized Dictionaries

**Pamela Faber, Arianne Reimerink, Pilar León-Araúz**

Department of Translation and Interpreting, University of Granada

e-mail: pfaber@ugr.es, arianne@ugr.es, pleon@ugr.es

Specialized dictionaries and knowledge resources encode the meaning of specialized knowledge units but they rarely include contextual information. Context is crucial for language comprehension and production, since they both largely depend on users ability to activate the right frame. Moreover, context can be defined in many different ways. Therefore, it is timely and necessary to parametrize context with a view to more effectively facilitating knowledge representation in term entries. In this paper we propose a three-level parametrization of context based on scope (i.e. local vs. global) and according to the type of information conveyed (i.e. syntactic, semantic and pragmatic). The term 'absorption' is used to exemplify what kind of contextual information is normally included in bilingual specialized dictionaries and how this information could be expanded to provide more useful information for text comprehension and production.

**Keywords:** context parameters; bilingual dictionaries; specialized knowledge

## Bridging the Gap between Description and Standardization: a Frame-based Version of NATO Glossaries

**Pamela Faber, Pilar León-Araúz**

Department of Translation and Interpreting, University of Granada

e-mail: pfaber@ugr.es, pleon@ugr.es

Promoting successful communication in multilingual military scenarios entails more than facilitating a standardized list of alphabetically arranged concepts. Given that misinterpreted messages can have dramatic consequences, text senders and receivers should also possess the same type of shared domain knowledge to facilitate mutual understanding, which we believe can be acquired if terminological resources are context-oriented or frame-based. Knowledge of terminological units and their meanings also signifies the awareness of how these units combine with others and in what scenarios these combinations may occur. Consequently, it is also necessary to understand the typical contexts activated within the specialized domain, and to have a grasp of the concepts and categories participating in them. In NATO, the need for terminology management has long been recognized, however, their glossaries do not provide a meaningful access to knowledge. This paper describes how the AMedP-13 (A) NATO glossary of medical terms could be adapted to a frame-based model in order to enhance knowledge acquisition in the medical military domain. The glossary was first converted into a pre-network structure derived from the glossary's definitions and corpus data. After that, different interrelated categories were rearranged in the form of semantic frames, such as the evacuation process frame, which activates different object categories (medical officer, patient/casualty, medical facility, vehicle), that are better acquired in association with the process in which they participate.

**Keywords:** terminology management; frame-based terminology; NATO term; NATO glossaries; context

## Aspects Lexicaux d'une Langue Inventée: la Langue des Schtroumpfs

**Thierry Fontenelle**

Centre de traduction des organes de l'Union européenne

e-mail: Thierry.Fontenelle@cdt.europa.eu

Le phénomène des langues inventées n'est pas rare et la liste des langues artificielles (ou construites) s'allonge avec les nouvelles formes d'art que sont le cinéma ou la bande dessinée. Certaines langues, comme le klingon ou les langues elfiques de Tolkien, ont attiré l'attention des linguistes qui se sont penchés sur leurs structures, leur morphologie, leur vocabulaire, allant parfois même jusqu'à créer des ouvrages de références (dictionnaires, grammaires...) ou en traduisant certaines œuvres littéraires dans ces langues. Nous nous penchons ici sur la langue des Schtroumpfs, les petits lutins bleus issus de l'imagination du dessinateur belge Peyo. Nous démontrons que la structure de cette langue artificielle est loin d'être arbitraire et que son vocabulaire répond à une logique qui permet d'expliquer pourquoi le langage schtroumpf reste compréhensible pour le lecteur francophone, même si les règles lexicales semblent totalement libres.

**Keywords:** langues inventées; lexicographie des langues inventées; collocations et combinatoire lexicale; langue des Schtroumpfs

## Contemporary Legal Terminology – Some Problems of Translation

**Irina Gvelesiani**

Ivane Javakhishvili Tbilisi State University

e-mail: irina.gvelesiani@tsu.ge

The juridical systems of different countries of the world have not escaped the “overwhelming” process of globalization. It is vividly felt in every branch of the legal sphere. The rapid developments have been detected and predicted in the comparative legal studies. The increasing role of the comparative law stipulates the rise of interest in terminological studies. The greatest attention is paid to the researches based on the comparative analysis of the contemporary and historical legal terms, because the language of a particular country as well as the law “is not a static, but a dynamic phenomenon” (The Cambridge companion to comparative law 2012: 91). It is the result of “a historical evolution which did not happen on its own, as if in a vacuum, but in constant interchange with other legal languages and legal systems” (The Cambridge companion to comparative law 2012: 91). The present paper involves the study of concepts, conceptual systems and their labels (*terms*) according to the comparative analysis of the data given in different European languages. The major emphasis is put on the concept of *trust* and its European “counterparts”.

**Keywords:** fiducie; globalization; trust

## French Specialised Medical Constructions: Lexicographic Treatment and Corpus Coverage in General and Specialised Dictionaries

Ornella Wandji Tchami<sup>12</sup>, Ulrich Heid<sup>2</sup>, Natalia Grabar<sup>1</sup>

STL<sup>1</sup> UMR 8163 CNRS, University of Lille3; IWIST<sup>2</sup>,

University of Hildesheim

e-mail: ornwandji@yahoo.fr, heidul@uni-hildesheim.de,

natalia.grabar@univ-lille3.fr

The aim of this paper is to examine four French dictionaries: the *on-line Larousse*, the *Petit Robert (2009)*, the *Larousse Médical*, and the *TLFi*, with regard to the way specialised verbal constructions (extracted from two medical subcorpora which are different according to the level of expertise of their author and readership) are described across the resources. According to our results, specialised verbal constructions are not given an important place in the queried resources, even in the medical dictionaries. Very few of the studied constructions are found in the analysed dictionaries, and those which are presented are not given a homogeneous and proper description.

**Keywords:** verb argument structure; specialised verb usage; lexicography; terminography; corpus analysis; phraseology; preferred co-occurrence and collocation

## A Descriptive Approach to Medical English Vocabulary

Renáta Panocová

Pavol Jozef Šafárik University in Košice

e-mail: renata.panocova@upjs.sk

This paper presents research into the characterization of medical vocabulary in English. It aims at developing an optimal methodological approach to the characterization of medical vocabulary in English. It is based on the analysis of data from the medical subcorpus of the Corpus of Contemporary American English (COCA). Earlier corpus-based research into medical vocabulary was carried out mainly from a pedagogical perspective and resulted in medical word lists. In those approaches, all criteria are based on absolute frequencies. It would not be sufficient to replace absolute frequency with relative frequency, because a minimal degree of absolute frequency is also necessary. What I show is that the threshold to be set for the absolute frequency interacts with the relative frequency. Therefore a measure based on the interaction of absolute frequency and relative frequency is shown to provide a better tool for identifying medical vocabulary than previously used measures.

**Keywords:** relative frequency; absolute frequency; corpus; language for specific purposes (LSP)

## Социопрагматический аспект лексикографирования специальной лексики

Людмила Рычкова

Гродненский государственный университет имени Янки Купалы  
e-mail: rychkova@grsu.by

Специальная лексика, как правило, составляет основу энциклопедических и отраслевых терминологических словарей. Общественные словари также традиционно отражают определенные разряды специальной лексики, хотя до сих пор теория лексикографии не выработала единых критериев отбора лексики профессионального и терминологического характера для реестров таких словарей и к принципам ее представления в словарных статьях с учетом консубстанциональности. Интеллектуализация языков, обусловленная влиянием информационного общества, актуализировала задачу лексикографирования различных разрядов специальной лексики для непрофессиональных пользователей, то есть за пределами собственно терминографии. Особую актуальность в условиях новых независимых государств, к каковым относится и Республика Беларусь, получили языки для социокультурных целей, отражающие и призванные формировать у граждан новую систему ценностей, включая понимание самой государственности как абсолютной ценности общества. На примере создания pilotной версии нового типа словаря белорусского языка – Словаря социально значимой лексики специального и профессионального характера, ориентированного на широкий круг пользователей, – обоснована значимость учета социопрагматического аспекта при отборе лексем для словарного реестра; показана важность использования аутентичного языкового материала, необходимого для построения словарных макростатей; описана структура макростатей и особенности включения в нее составных номинаций. Подчеркнута роль Белорусского республиканского фонда фундаментальных исследований в поддержке научных проектов, направленных на создание и широкое применение электронных языковых ресурсов, без которых создание Словаря социально значимой лексики специального и профессионального характера белорусского языка было бы невозможным.

**Keywords:** белорусский язык; специальная лексика; языки для специальных целей; социопрагматический аспект; словарь социально значимой специальной лексики; реестр словаря; структура словарной макростатей

---

## Collocational Information for Terminological Purposes

**Elsabé Taljard**

Department of African Languages, University of Pretoria  
e-mail: elsabe.taljard@up.ac.za

Traditionally, collocational information was sparsely, if at all, provided in terminological tools. This is particularly true for LSP dictionaries in paper format. One possible reason for this state of affairs is that LSP dictionaries had in the past often been compiled by subject field experts, who did not deem this kind of information as useful from a terminological point of view. This paper critically considers the importance of collocational information in terminological tools such as terminological databases, term banks and LSP dictionaries. It is argued that collocational information can assist the user on both conceptual and usage-related level and that such information is therefore of critical importance. Secondly, two methodologies for the identification of collocates, viz. introspection and corpus-based methods are compared, and it is concluded that access to huge amounts of real-life language data, also for LSP purposes, leads to insights which surpass those arrived at through introspection.

**Keywords:** collocational information; corpus-based terminology; terminological database

# **Historical and Scholarly Lexicography and Etymology**



---

## Wanderworten Etymological Format: a ‘Hemp’ Case Study, Starting from Votic and Baltofinnic.

**Manuel Barbera**

University of Turin

e-mail: b.manuel@inrete.it

The main theme of this short paper will be the format of the etymological section of dictionaries for the *Wanderworten*. The case study will be provided by the etymology of the word for ‘hemp’ in the Baltofinnic languages, starting with the Votic *kańevo*, an entry that I have already dealt with in my VoEG.

In the VoEG, I attempted to forge an intermediate structure (expounded in Barbera 2002; for a general introduction to VoEG cf. Barbera 2000) between the ample and circumstantiate formats, most exemplarily exhibited by Wartburg in the FEW, and the (often too much) shorter formats. I will argue that even if more models can exist between the shorter and the longer ones (and the VoEG is one of those), even the shorter for the *Wanderworten* should be more detailed than usual for more “normal” etymologies.

**Keywords:** etymology; *Wanderworten*; Baltofinnic languages

## The Russian Contributions to the English Language: a Case of OED Record

**Kseniya Egorova**

Authentic Russian Project

e-mail: kseniya.a.egorova@gmail.com

The paper looks at Russian contributions to the English language, which are in evidence of the Third Edition of the *Oxford English Dictionary* (*OED3*). With the number of 402 loanwords the Russian language figures among the twenty-five most prolific language sources of word-borrowing in the vocabulary of modern English (Durkin, 2014). This survey is an attempt to produce a far fuller language profile of Russian investigating *OED*’s data further. In addition to the loanwords, it also looks at those lexical items in *OED3*, which appeared in English under the lexical influence of the Russian language – namely, loan translations or calques, semantic borrowing, proper names of Russian origin and their derivatives. By performing a number of advanced searches on *OED Online*, additional subsets of ‘Russianisms’ were generated and carefully examined. The paper concludes that the total number of the Russian contributions to the English language exceeds 1000 lexical items and the input of the Russian language to the lexicon of modern English should be assessed as more prolific.

**Keywords:** OED; etymology; lexical borrowing; loan translation; semantic borrowing; Russian

## Lexicography of Language Contact: An Internet Dictionary of Words of German Origin in Tok Pisin

**Stefan Engelberg, Christine Möhrs**

Institute for the German Language in Mannheim

e-mail: engelberg@ids-mannheim.de, moehrs@ids-mannheim.de

The paper presents an ongoing project in the domain of lexicography of language contact, namely, the “Internet Dictionary of Words of German Origin in Tok Pisin”. The German influence onto the lexicon of the main pidgin language of Papua New Guinea has its roots in the German colonial empire, where Tok Pisin played an important role as a lingua franca in the colony of German New Guinea. Tok Pisin also served as an intermediate language for many borrowing processes; that is, German loans entered many languages in the South Pacific via Tok Pisin.

The Internet Dictionary of Words of German Origin in Tok Pisin is based on all available lexicographical sources from the early 20th century up to now. These sources are systematically evaluated within our project; the results will be documented in the dictionary. The microstructure of the dictionary will be presented with respect to its major features: documentation of sources, examples for word usage, audio files, and lexicographic comments.

**Keywords:** language contact; borrowing; dictionary of language contact; electronic lexicography

## The Word Entry Structure of the Russian Manuscript Lexicons: Evolution through the Centuries

**Kira Kovalenko**

Institute for Linguistic Studies, Russian Academy of Sciences

e-mail: kira.kovalenko@gmail.com

Early Russian lexicography is represented by diverse types of manuscript glossaries and lexicons, compiled in Kievan and then in Moscow Rus'. At first they were intended to help in reading and understanding texts translated from Hebrew and Greek and were based on glosses to obscure words. That determined the word entry structure; entries in such glossaries consisted of the head word and explanation. Later, glossaries were developed into complicated lexicographical works, which were orientated towards scribes and editors of Russian literature. The compilers made the word entry structure more complicated, and the last representatives of the lexicographical genre to appear had such information categories as headword, explanation, language mark, references to the literary sources, examples of use and collocations and a reference to words with close semantics or the same topic group. A detailed description of how the word entry structure evolved in manuscript glossaries and lexicons from the 13<sup>th</sup> to 17<sup>th</sup> centuries will be presented in this article.

**Keywords:** glossary; lexicon; azbukovnik; Russian manuscripts; word entry; loan words

## Diachronic Analysis of Word Formation as Basic Studies for Etymologisation

**Iris Metsmägi, Vilja Oja**

Institute of the Estonian Language

e-mail: iris.metsmagi@eki.ee, vilja.oja@eki.ee

The paper is focused on some problems of word formation in Estonian and Finnic, which are etymologically relevant. Diachronic derivational relations may be not ascertained by the synchronic approach, e.g. either due to sound changes in the stem (cf. Estonian *kõõl-us* ‘tendon, string’ and *keel* ‘tongue; language; string’), or because a root does not occur separately or is rare. A suffix may have been dropped from active use or is subject to extensive variation, e.g., many Estonian verb stems containing the component *-ka*, *-ki*, *-ku* etc. (*kilka-ma* ‘to scream’) could be interpreted as derivatives with respective suffixes (*kil-ka-ma*, cf. *kil-ise-ma* ‘to clink’). The interpretation of the morphological structure of loanwords is often complicated. The loanwords may contain derivational suffixes of the donor language, sometimes even resembling genuine ones, e.g. Estonian *lusikas*, Finnish *lusikka* ‘spoon’ does not contain historically the Finnic suffix *-k(k)a* but is borrowed from Old Russian *luzhka*. Sometimes foreign-sounding loanwords are subjected to folk-etymological adaption and the results are similar to a derivative or a compound, e.g. Estonian *puuslik* ‘idol’, looking like a derivative of the stem *puu* ‘tree; timber’, is rather a folk-etymological version of Russian *apostol* ‘apostle’.

**Keywords:** etymology; derivation; loanwords; Estonian; Finnic languages

## Reconsidering the Initial Development of the Dictionaries of Americanisms: Pickering's Legacy in Bartlett's *Dictionary* (1849)

**Kusuiro Miyoshi**

Soka Women's College

e-mail: KusuiroMiyoshi@gmail.com

It has widely been acknowledged among authorities that the application of scientific lexicographical principles in American English dictionaries is pioneered by John Bartlett's first edition of *Dictionary of Americanisms* (1849), the second American English dictionary. The general recognition about the dictionary is that it divorced from John Pickering's *Vocabulary, or Collection of Words and Phrases* [...] (1816), which is usually regarded as a normative dictionary compiled for the purification of the language in America. Actually, however, when we collate the *Dictionary* with the *Vocabulary*, it becomes clear that more than 10% of entries in the *Dictionary* is heavily dependent on the *Vocabulary*, in terms of explanations of words and citations, as well as verbal examples which Pickering uniquely made, signifying the fact that the *Vocabulary* was an indispensable basis for Bartlett to compile his historic *Dictionary*. Here, the credibility of Reinhard Hartmann's (1986:vii) saying is strengthened, which goes "Most dictionaries have forerunners, and all have imitators", in the development of American English dictionaries.

**Keywords:** Bartlett; Pickering; American; history; dictionary

## Etymological Studies Based on a Conceptual Framework

**Vilja Oja**

Institute of the Estonian Language

e-mail: vilja.oja@eki.ee

The purpose of the poster is to demonstrate that vocabulary analyses by concept groups can contribute to etymological research. The paper focuses on some groups of the Indo-European influences on Estonian and cognate languages. Many nautical and fishing terms were borrowed from the Swedish in the Middle Ages, when the coastal areas of Estonia and Finland were populated by Swedes. The Estonian colour terms system is quite similar to the German system, and many words for the specific tones are German loanwords in Estonian (Oja 2001: 32–33). Studying the names of insects we found very few German loanwords, but there are many similarities in comparison with the corresponding Baltic/Slavic thematic groups (Vaba 2015: 370–383). The systematic analysis helps get more reliable results in etymological research. Single concepts will be approached as elements of the semantic system in order to explain the initial motivation for naming the concept.

**Keywords:** etymology; onomasiology; thematic groups; Indo-European loanwords; Estonian

## **The Da Vinci Codices. Between Historical Lexicography and Knowledges History**

**Margherita Quaglino**

Università di Torino

e-mail: margherita.quaglino@unito.it

The paper discusses the principal issues connected with the study of lexica of ancient specialist languages in the early centuries of vernacular in Italy, focusing on the formal typologies of definitions in Leonardo da Vinci's works. The insertion of definitions means, above all, an attempt to establish fixed points in the confused linguistic composition of annotations; in the second place, it indicates a growing awareness of a new cognitive fact, preceded or followed by a new linguistic elaboration; the coincidence between model (drawn, at times only sketched) and a word; the temporary suspension of the transferability of terms between fields of knowledge and different linguistic registers; the possibility to verify one or a short list of meanings, at the synchronic level and in a diachronic perspective; the identification of syntactic structures that are becoming conventional for one sole type of communication.

**Keywords:** historical lexicography; Leonardo da Vinci; definitions

## **Moth's Danish Dictionary Publishing a Dictionary with a 300-year Delay**

**Thomas Troelsgård, Marita Akhøj Nielsen**

Society for Danish Language and Literature

e-mail: tt@dsl.dk, man@dsl.dk

The first comprehensive Danish dictionary was compiled by the high official Matthias Moth (1649-1719). To get material for the dictionary he organized a collection of words from all over the Danish king's realm, asking especially for the daily language of ordinary people, including dialect material. This was quite unusual for his time – and for a long period after. The collection was finished shortly after 1700, and Moth immediately started editing the dictionary. He compiled several editions and changed the overall layout before starting the final edition that comprises more than 20 volumes in folio, all written by himself. This huge work was never printed, but for 200 years the manuscripts were utilized as an important source for comprehensive Danish dictionaries because of the valuable collection of words and sayings. In 2013-2015 an online edition was published, and to demonstrate the development of this edition an account is given of Moth's dictionary through three centuries focusing on the digitization of the handwritten material. Finally, the online edition is presented.

**Keywords:** historical dictionary; daily language; Elder Danish; digitization; mark-up; online edition

## Enabling Selective Queries and Adapting Data Display in the Electronic Version of a Historical Dictionary

**Tim van Niekerk, Heike Stadler, Ulrich Heid**

Dictionary Unit for South African English, Grahamstown;

Department of Information Science and Natural Language Processing, University of Hildesheim

e-mail: t.vanniekerk@ru.ac.za, stadle@uni-hildesheim.de, heid@uni-hildesheim.de

We report on an ongoing project aimed at making the scholarly reference work *A Dictionary of South African English on Historical Principles (DSAE, Silva et al. 1996)* accessible for selective online querying<sup>1</sup>. On the basis of the XML version of the current online edition of the dictionary, we perform two major steps towards this goal: (i) the design of interactive presentation devices for the dictionary, and (ii) the enhancement of the available lexicographic data categories. We take inspiration from the Function Theory of Lexicography (Tarp 2008), even though its main focus is on communicatively oriented dictionaries, while the *DSAE* primarily serves cognitive purposes. Since we are reporting on an ongoing project, we only present examples of the work.

**Keywords:** historical lexicography; electronic lexicography; print adaptation; user interfaces; XML

---

<sup>1</sup> We gratefully acknowledge partial financial contributions to some of the work presented in this paper, from the Niedersächsische Staatskanzlei and MWK Niedersachsen (2015) and from the German Academic Exchange Service, DAAD (2015).

## Digitizing the Second Edition of Furetière's *Dictionnaire Universel*: Challenges of Representing Complex Historical Dictionary Data Using the TEI

Geoffrey Williams, Ioana Galleron

LiCoRN-LIDILEM, Université Grenoble Alpes

e-mail: Geoffrey.williams@orange.fr, ioana.galleron@gmail.com

This poster will present preliminary conclusions about the pitfalls and the best practices in marking up entries from the *Dictionnaire universel* by Antoine Furetière, in the revised and expanded version published in 1701 by Henri Basnage de Beauval (Furetière 1701). As the universality of the *Dictionnaire* is its main characteristic, the project decided to concentrate on terms for a start. Four domains have been spotted (architecture, maritime affairs, law and theatre), which are of interest for all 17<sup>th</sup> and 18<sup>th</sup> century French studies, and give a good idea about Furetière's lexicographical principles and techniques.

After an initial analysis conducted on the PDFs by using a CAQDAS system (Atlas.ti), encoding experimentations have been realised, aiming at:

- 1° Modeling the structure, while allowing to observe the experimentations conducted by Furetière in this regard.
- 2° Proposing a consensual conceptualisation of scientific domains covered by the *Dictionnaire*.
- 3° Linking the quotations to their sources.

**Keywords:** Furetière; Basnage de Beauval; CAQDAS; TEI; maritime terms; architecture terms; law terms; theatrical terms; Digital Humanities

---

# Lexicography and Semantic Theory



---

## Adding Polarity Information to Entries of the Database of Bavarian Dialects in Austria

**Thierry Declerck<sup>1</sup>, Amelie Dorn<sup>2</sup>, Eveline Wandl-Vogt<sup>2</sup>**

<sup>1</sup>Saarland University, <sup>2</sup>Austrian Center for Digital Humanities

e-mail: declerck@dfki.de, amelie.dorn@oeaw.ac.at,  
eveline.wandl-vogt@oeaw.ac.at

In this short paper we describe on-going work, which consists in adding polarity information to entries that are part of the Database of Bavarian Dialects in Austria (DBÖ). By “polarity information” we mean the positive or negative interpretation a word can carry. The starting point of our study is given by SentiMerge, a lexical resource that encodes polarity information for standard German words on the basis of integration processes performed on four pre-existing polarity lexicons. The lexical information of the entries of SentiMerge is encoded using the Ontolex model, which has been developed in the context of the W3C Ontology-Lexica Community Group. The polarity information, on the other hand, is encoded using the MARL ontological model, which has been developed at the Universidad Politécnica de Madrid. The use of such formal representation frameworks is aimed at supporting the publication of the resulting data in the Linked Open Data cloud. In the first phase of our work we focus on headwords describing colour terms, taking also compound words into consideration.

**Keywords:** Polarity; DBÖ; colour terms; Ontolex; MARL

## Towards a Sense-based Access to Related Online Lexical Resources

**Thierry Declerck, Karlheinz Mörth**

DFKI GmbH, Language Technology Lab,

Austrian Centre for Digital Humanities

e-mail: declerck@dfki.de, Karlheinz.Moerth@oeaw.ac.at

We present an approach aiming at a method to support sense-based cross-dialectal – and cross-lexicon – access to dictionaries of spoken Arabic varieties. The original lexical data consists of three TEI P5 encoded dictionaries describing varieties spoken in Cairo, Damascus and Tunis. This data is included in the Vienna Corpus of Arabic Varieties (VICAV). We briefly present this data, before summarizing the TEI approach for encoding senses in lexical resources. We discuss certain issues related to the TEI representation when it comes to the possibility to provide for a sense-based access to the lexical data. We investigate the use of the recent W3C Ontology-Lexicon Community Group (OntoLex) modeling work and present first results of the mapping of the TEI encoded data onto its *ontolex* model and show how this new representation format can efficiently support cross-dictionary sense-based access.

**Keywords:** senses; ontolex; TEI; SKOS; Arabic dialects

## Discursive Constructions in the Russian-Swedish Dictionary Database: A Case Study of *v tom-to i N*

**Dmitrij Dobrovolskij, Ludmila Pöppel**

Russian Academy of Sciences, Russian Language Institute; Stockholm University, Department of Slavic and Baltic languages, Finnish, Dutch and German  
e-mail: dm-dbrv@yandex.ru, ludmila.poppel@slav.su.se

The primary goal of the present study is to develop a new way for the description of discursive units in a Russian-Swedish dictionary database which is currently under construction. Discursive units are important elements of communication, but have not been fully described as yet. Using large text corpora for the empirical data, we are going to analyze the type of construction *v tom-to i N* / в том-то и N (that's [just] the N) which is realized in a wide variety of tokens and demonstrate that some tokens of the construction can be so frequent that they can be considered to be cognitively entrenched units and are preserved in memory as separate units of the language, i.e. phrasemes. Such units should be described as separate items of the lexicon. The task of the investigation is to refine our notions of the structural properties of *v tom-to i N* / в том-то и N (that's [just] the N) and identify additional distinctive features of the construction which should be included in its lexicographic description.

**Keywords:** bilingual lexicography; phrasemes; parallel corpus; constructions; discursive units; Sketch Engine

## Tatar Co-compounds as a Special Type of Classifiers

**Alfiya Galieva, Dzhavdet Suleymanov**

Research Institute of Applied Semiotics of Tatarstan Academy of Sciences,  
Kazan Federal University  
e-mail: amgalieva@gmail.com, dvdt.slt@gmail.com

The so called co-compounds, or dvandva constructions, exist in many languages as specific type of lexical items, nevertheless they are poorly described in lexicology. The paper analyses the semantics of Tatar co-compounds formed on the basis of names of natural phenomena. The main hypothesis of the study is that co-compounds are a special type of nomination used to designate provisional, non-rigid classes of objects in different situations.

We study co-compounds proceeding from the assumption that there is no strict dichotomy between linguistic and encyclopedic knowledge and that there are no stable word meanings.

Co-compounds as signs expressing concepts imply comparing properties of objects and phenomena by means of segregating some common features of these objects. Simultaneously the objects are differentiated by their properties and classified according to their similarities and differences, so co-compounds serve as a special type of classifiers. The analysis of the Tatar language data allows us to conclude that generalization of meanings, distinctive for co-compounds, can help overcome the discreteness of linguistic units and may lead to the overlap of different classes of phenomena with correlated or adjacent concepts.

All the examples are taken from the Tatar National Corpus.

**Keywords:** co-compounds; the Tatar language; semantics; lexical items

## Semantic Approach to Phraseological Patterns in Karstology

**Larisa Grčić Simeunović, Paula de Santiago Gonzalez**

University of Zadar, King's College London

e-mail: larisa.grcic@gmail.com, pauladesantiago@gmail.com

As the previous research about collocational patterning (Stubbs 1996, 2001, Flowerdew 2009) has demonstrated, words are restricted in their collocational behaviour. This article aims at illustrating how specialised phraseology reveals restrictional meaning of terminological units as well as patterned syntagmatic relations between terms or terms and other lexical units as a result of a semantic analysis. Two kinds of paradigmatic sets were established: conceptual classes formed by term bases, and semantic sets formed by adjective collocates. Their distributional behaviour is described in order to identify the most recurrent phraseological patterns. Authors point out the importance of combinatorial information for determining specialized meaning and examine the possibility of offering additional information to the terminological definition.

**Keywords:** collocation; specialized corpus; semantics; selectional preference; terminography

## Displacement of Senses, Cluster Equivalence and Bilingual Dictionaries

**Barbara Lewandowska-Tomaszczyk**

State University of Applied Sciences in Konin  
e-mail: blt@uni.lodz.pl

Basing on the data from the BNC and National Corpus of Polish, and relevant parallel corpora, the concepts of *lexical displacement of senses* and *cluster-for-cluster equivalence* are discussed and exemplified on the materials from English and Polish. Direct relevance to bilingual lexicography is further presented and inter-lingual correspondences involving two main types are analysed - lexical-conceptual *L1-cluster-for-L2 cluster mapping* and a *higher syntactic schematization* of the Target as contrasted with the more fine-grained construed Source language version. In order to receive a more complete picture of the *cross-language cluster mapping*, the two-directional translational corpora are consulted and the clusters referring to the mental space of *possibility* generated in both languages, analysed and juxtaposed for the qualitative and quantitative properties. The present project aims at extending the depth of cluster presentation in both languages and the refinement of a broad sense of equivalence between them. A special emphasis in bilingual lexicography project presented here is put on *Part-of-Speech cross-linguistic re-categorization* to recover more complex (asymmetric, incommensurable) inter-linguistic relations. The Principle of *Cluster Equivalence* is shown to be observed in a variety of styles and registers both in the written and in spoken mode.

**Keywords:** bilingual lexicography; cluster equivalence; collocational profile; construal; dictionary; displacement of senses; meaning approximation; parallel corpora; re-conceptualization; translation

## Systematic Polysemy of Estonian Colour Adjectives

**Maria Tuulik<sup>1,2</sup>, Margit Langemets<sup>1</sup>**

Institute of the Estonian Language<sup>1</sup>, University of Tartu<sup>2</sup>  
e-mail: Maria.Tuulik@eki.ee, Margit.Langemets@eki.ee

This article introduces patterns of systematic adjectival polysemy in Estonian focusing on colour terms. The aim of our work is to create a theoretical framework for presenting adjectives in the Estonian Collocations Dictionary (to be published in 2018). To explore systematic polysemy patterns, semantic types were attached to the word senses and similarities in the semantic shifts of the words were searched for. Semantic types provided for describing Estonian (polysemous) adjectives (Tuulik 2014) were used in the study. In the case of senses denoting nouns, semantic types developed for Estonian nouns (Langemets 2010) were used.

The colour adjectives are examined in terms of both polysemy and morphological structure. A polysemy index, i.e. the average number of meanings per word, is ascertained on the basis of a single-volume explanatory dictionary of Estonian (to be published in 2018) for root adjectives, derivatives and compound words in order to explore the relations between polysemy and morphology based on the example of Estonian colour adjectives (463 colour terms altogether).

**Keywords:** semantics; lexicology; lexicography; morphology; adjectives; Estonian



# Phraseology and Collocation

---



## Online Dictionary of Idioms

**Irina Lobzhanidze**

Ilia State University

e-mail: irina\_lobzhanidze@iliauni.edu.ge

The first edition of this dictionary was prepared with the financial support of the Shota Rustaveli Science Foundation (Project No Y-04-10). The Dictionary was bilingual in Modern Georgian and Modern Greek, included approximately 12000 entries and reflected how the idioms function. It combined features of translational and learner's dictionaries. The dictionary was not published. Thus, it was decided to compile an online dictionary of idioms both monolingual and bilingual, equipped with a user-friendly environment to find idioms, their meanings and contexts of use. The project was prolonged with the support of the Shota Rustaveli Science Foundation (Project No LE/17/1-30/13). Thus, at this moment, the beta version of dictionary is available at <http://idioms.iliauni.edu.ge/>.

The present paper is a part of the research project dedicated to the compilation of an online dictionary of Idioms, bidirectional in two languages: Modern Georgian and Modern Greek. The paper is structured as follows: Section 1 provides brief descriptions of Dictionaries and briefly summarizes their prerequisites; Section 2 represents the methodological approach; Section 3 includes findings and hypotheses, which refer to the distinguishing features, attributes of dictionaries and characteristics of idioms and Section 4 summarizes the work implemented.

**Keywords:** online dictionary; idioms; Georgian; Modern Greek

## On Distinction between Phraseology Conceptions in a General Monolingual Dictionary and in a Special Phraseological Dictionary

**Renáta Neprašová**

Institute of the Czech Language of the Academy of Sciences  
of the CR, v. v. i.

e-mail: neprasova@ujc.cas.cz

This contribution focuses on the comparison of approaches to the treatment of phraseology in the Academic Dictionary of Contemporary Czech (*Akademický slovník současné češtiny*) and in the special Czech Dictionary of Idioms (*Slovník české frázeologie a idiomatiky*). In the introduction the paper briefly presents the approach to the treatment of phraseology in previous Czech monolingual dictionaries, advantages of this approach and possible difficulties which the upcoming dictionary attempts to avoid. The contri-

bution also deals with the definitions of phrasemes in the monolingual and in the special dictionary and with the selection of included units, which differs in both dictionaries. The fundamental lexicographical basis in the monolingual dictionary is one-word lexical unit, whereas for the special dictionary it is the whole multi-word lexical unit. It also operates with different extent possibilities. The paper therefore presents differences between basic forms of a lemma, structures of a phraseme entry in both dictionaries and briefly compares main differences between approaches to the meaning explanation and exemplification.

**Keywords:** phraseology; lexicography; monolingual dictionary; dictionary-making principles; entry structure

## Starting from Scratch – The Creation of an Irish-language Idiom Database

**Katie Ní Loingsigh, Brian Ó Raghdhaigh**

Fiontar, Dublin City University

e-mail: katie.niloingsigh@dcu.ie, brian.oraghallaigh@dcu.ie

This paper presents a database of Irish-language idioms along with a description of its internal schema. This is a dialect-specific synchronic study of idioms in the Irish of West Munster from the published work of Canon Peadar Ó Laoghaire (1839-1920). Firstly, a brief overview is given of previous research undertaken on Irish-language idioms in the area of phraseology. The database schema is outlined in the latter section of the paper. Each of the individual categories in which idioms are recorded, categorized, explained and cross-referenced in the database are described in detail. A selection of sample entries is provided in the Appendix to this paper. This research provides an innovative approach to the collection and categorization of Irish-language idioms and also a foundation for further analysis of idioms in other dialects and languages and for future studies in this area.

**Keywords:** lexicography; language technology; phraseology; collocation; database schema

## The Compilation of a Printed and Online Corpus-based Bilingual Collocations Dictionary

**Adriane Orenha-Ottaiano**

Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP)

e-mail: [adriane@ibilce.unesp.br](mailto:adriane@ibilce.unesp.br)

Collocations are conventionalized, recurrent and arbitrary lexical combinations. Due to the fact that they are highly specific for a particular language and may be contextually restricted, collocations pose a problem to foreign language learners and translators with regard to the production or encoding. Taking that into account, the compilation of dictionaries of monolingual and bilingual collocations for the learners of both English and Brazilian Portuguese as a foreign language and professional translators is crucial and highly significant. The aim of this paper is to give an account of the compilation of a *printed and online Corpus-based Bilingual Collocations Dictionary*, in the English-Portuguese and Portuguese-English directions. The dictionary will focus on all types of collocations (verbal, nominal, adjectival and adverbial collocations) in order to help the referred audience use them more accurately and productively. The idea of having the proposed dictionary in online format will allow to incorporate more qualitatively and quantitatively collocational information. Furthermore, more examples can be included, different from conventional printed collocations dictionaries. Being the first bilingual collocations dictionary in the aforementioned directions, it is hoped to achieve the challenge of meeting learners' collocational needs as the collocations will be selected according to Brazilian learners' difficulties.

**Keywords:** corpus-based collocations dictionary; collocations; bilingual collocations dictionary; collocational patterns

---

# **Lexicological Issues of Lexicographical Relevance**



---

## **On *bullying*, *mobbing* (and *harassment*) in English and Polish: Foreign-language-based Lexical Innovation in a Bilingual Dictionary**

**Arleta Adamska-Sałaciak**

Faculty of English, Adam Mickiewicz University

e-mail: arleta@wa.amu.edu.pl

The present paper deals with the treatment of incipient borrowing in a bilingual dictionary. While not completely new, the topic has so far received little attention in the lexicographic literature, Adamska-Sałaciak (2004) and (2007) being among the few exceptions. Yet it is worthy of note that, in a period when dictionary making as a whole is being overhauled, the lines previously drawn regarding the lexicographic treatment of borrowing seem to be shifting as well. In particular, the requirement – not always explicitly voiced – that unless a loanword has been recorded in a monolingual dictionary of the recipient language, it is not eligible for the inclusion in a bilingual dictionary (whether as an equivalent or as a headword), seems to have lost much of its force, especially with regard to the digital medium. This gives bilingual lexicographers more freedom, but at the same time puts more responsibility on their shoulders. The issue is illustrated by a discussion of a specimen case encountered in the process of updating a major English-Polish, Polish-English dictionary.

**Keywords:** lexical innovation; borrowing; bilingual dictionary; English; Polish

## **Representing Usage in Dictionaries**

**Janet DeCesaris**

IULA-Institut de Lingüística Aplicada, Universitat Pompeu Fabra

e-mail: janet.decesaris@upf.edu

This paper examines the nature of “usage information” in monolingual dictionaries of English. English, unlike many other European languages like French, Italian, or Spanish, has never had a language academy to dictate ‘good’ English as opposed to ‘bad’ English, and as a result dictionaries have a privileged role in determining what constitutes the standard language. We first review the historical development of usage information in dictionaries of English, and particularly in dictionaries of American English, in which the representation of usage has been both noteworthy and controversial. ‘Usage’ in the past has often referred to dialectal forms and differences in pronunciation and spelling, as well as to commentary about what the editors considered acceptable (or unacceptable) changes in grammar and word meaning. We then address what constitutes ‘usage’ in the context of corpus-based dictionaries, which presumably provide a more descriptive – and less prescriptive – view of the language. We compare the information about usage in several current monolingual dictionaries aimed at both native speakers and at advanced learners by analyzing usage notes and comments in headwords belonging to the letter ‘D’.

**Keywords:** English monolingual lexicography; usage; prescriptivism in dictionaries

## Lexical Change in Times of Upheaval and War – And the Dictionary

**Donna M.T. Cr. Farina**

New Jersey City University

e-mail: [dfarina@njcu.edu](mailto:dfarina@njcu.edu)

We consider how to address lexicographically words, expressions, and usages that are for the most part quite recent in the Russian language – when their newness may mean that known dictionaries or lexicographic corpora do not cover them fully. We refer to “usages” because the prevailing linguistic situation in the Russian language is rife not only with new words, but also with new word combinations, new connotations, new meanings, broadened or extended meanings as well as other linguistic phenomena. Our focus is the examination of a variety of real contexts for these neologisms, with comment on the lexicographic strengths and weaknesses of each context. The essential lexicographic questions are: What are the characteristics of any successful lexicographic context? When a usage is new and existing corpora may not reflect it, how should lexicographers evaluate and use other types of contexts in their work? How should connotation, given its propensity to change quickly, be treated in lexicographic definition and in illustrative examples, whether in online or in print dictionaries?

**Keywords:** connotation; denotation; explanation of meaning

## Lexicalized Prosody and the Polysemy of some Discourse Particles in ADR<sup>1</sup>

**Irina Levontina**

Russian Language (Vinogradov) Institute

e-mail: [irina.levontina@mail.ru](mailto:irina.levontina@mail.ru)

The paper is dedicated to some polysemous Russian particles as they are represented in the Active dictionary of Russian (ADR), which is currently being compiled under the guidance of Yury Apresjan at the Russian Language Institute (RAS), and to the problem of lexicalized prosody. General principles of ADR presuppose that all important properties of a lexeme are described in the lexical entry including lexicalized prosody. The phenomenon of lexicalized prosody in Russian drew the linguists' attention about 30 years ago. The investigation is usually confined to phrasal stress as the most frequently lexicalized and therefore the most lexicographically interesting prosodic pattern. However, as far as discourse particles are concerned, not only phrasal stress but also intonation pattern is of greatest interest. Two Russian discourse particles will be discussed. One of them is the particle *-to*. Its different usages imply very different prosodic patterns. The other one is *vot* which can be used not only as a demonstrative particle, but also as a xenomarker (quotation marker), which requires a specific prosody.

**Keywords:** lexicalized prosody; polysemy; discourse particles; Russian

<sup>1</sup> The research was supported by RSF (project No 16-18-02054).

## Importance of the Issue of Partial Equivalence for Bilingual Lexicography and Language Teaching

**Tinatin Margalitadze, George Meladze**

Lexicographic Centre, Ivane Javakhishvili Tbilisi State University  
e-mail: tinatin.margalitadze@tsu.ge, meladzegeorge@yahoo.com

The paper addresses the paramount importance of the issue of semantic equivalence of interlingual synonyms for various aspects of general linguistic, lexicological, and lexicographic research and related activities.

Drawing on the studies conducted by prominent lexicologists and lexicographers, the authors highlight several aspects making the successful solution of the problem of semantic equivalence especially important: (a) the production of adequate, comprehensive bilingual dictionaries reflecting whole spectrum of semantic fields of words included in their wordlists and thus bridging the divide between semantic systems of various languages; (b) the role of such dictionaries for successful acquisition of foreign languages; and (c) the role of dictionaries with maximum number of native equivalents of foreign headwords in their word-entries, for the preservation of lexical richness of languages and linguistic diversity in general.

In the article the authors also describe the methods they have been using for several decades in order to interconnect the semantic systems of Georgian and English. The semantic asymmetry between these languages and different methods of its treatment are shown on the examples extracted from the Comprehensive English-Georgian Dictionary (CEGD).

**Keywords:** partial equivalence; bilingual lexicography; Comprehensive English-Georgian Dictionary; linguistic diversity

## Towards the Meaning of a Word

Innes Merabishvili

Ivane Javakhishvili Tbilisi State University

e-mail: innesmerabishvili@hotmail.com

In gnosiology, the image is a fundamental category and is used in connection with sensuous reflection and abstract thinking. The presence of a **conceptual image** in a word is beyond doubt. As for the **sensuous image** in a word, it is only intuitively noted by scholars. The paper argues that the extralinguistic reality is reflected in the form of a sensuous imprint but developed into a sensuous image in the process of formation of a word as a second signal system (I. Pavlov's theory of two signal systems).

The author regards the sensuous image existing in a word as a **linguistic image**. A linguistic image is enclosed in every word but is called forth when figurativeness of the word is created. The acknowledgment of a linguistic image in a word permits us to present the meaning of a word as a correlation of three main components: the sound form of a word as its nomination, the conceptual image as the significatum and the sensuous image as a linguistic image. The afore-presented study of the meaning of a word appears to be of great importance in translated matters, especially when translating poetry that abounds with occasional word-combinations.

**Keywords:** sensuous imprint; linguistic image; three components of the meaning of a word

## *Von vielrednerischen Weibern und alles erforschenden Männern – Genderasymmetrien in der Datenbank der bairischen Mundarten in Österreich (DBÖ) im Vergleich mit Großwörterbüchern der Gegenwart*

**Maria Pober<sup>1</sup>, Ingeborg Geyer<sup>2</sup>, Eveline Wandl-Vogt<sup>3</sup>,  
Barbara Piringer<sup>3</sup>**

<sup>1</sup>Universität Wien, Institut für Germanistik; <sup>2</sup>bis 31.12.2015:  
Österreichische Akademie der Wissenschaften, Austrian Centre for Digital Humanities;

<sup>3</sup>Österreichische Akademie der Wissenschaften,  
Austrian Centre for Digital Humanities

e-mail: maria.pober@univie.ac.at, ingeb.geyer@gmail.com,  
eveline.wandl-vogt@oeaw.ac.at, barbara.piringer@oeaw.ac.at

Unterscheidet sich die lexikographische Artikel- bzw. Belegdarstellung älterer Wörterbücher in Bezug auf Gendersymmetrie und ihrer -repräsentanzen tatsächlich grundlegend von derjenigen neuerer gegenwartssprachlicher? Den Fokus der Arbeit bildet der genderanalytische Vergleich von historischen (dialektalen) Belegbeschreibungen (DBÖ 1913-33) mit solchen gegenwartssprachlicher Großwörterbücher im Deutschen (Duden, Wahrig, Küpper 1980-2012).

Hierzu werden bewusst keine Basislexeme unter den Personenreferenzen wie Frau und Mann herangezogen, sondern die Nomina Agentis mit dem Erweiterungssuffix -ler und -lein, da erstere lexeminhärent die Merkmale + (weiblich) bzw. + (männlich) und damit die Bedeutungskategorie [SEXUS] aufweisen, letztere hingegen nur auf Suffixebene und so mögliche implizite Genderimplikationen aufgezeigt werden können. Anhand von folgenden zwei Bedeutungskategorien [REDEFÄHIGKEIT] und [SEXUALITÄT] wurden diese Personenreferenzen (-ler und lein) ausgewählt und unter Einbeziehung von Zusatzkategorien wie z. B. [HANDEL] oder [ALTER] bzw. [HETEROSEXUELLE AKTIVITÄT] oder [HOMOSEXUELLE AKTIVITÄT] symmetrisch gegenübergestellt. Die Ergebnisse dieser Genderanalyse der Beleg- und Lemmatabeschreibung finden in den „fiktiven Lexikoneinträgen 2016“ ihre praktische gendersymmetrische und wertneutrale Umsetzung. Die vorliegende Studie führt in ein qualitatives System ein, das am Lexicography Laboratory der ÖAW zur Modellierung und Basis einer quantitativen Analyse ausgearbeitet wird (Linked Open Data, TEI).

**Keywords:** Metalexikographie; Genderkonzeptionalisierungen in der Lexikologie; Lexikographische Gendersymmetrien; Metasprachliche Paradigmen der Geschlechterordnung; Pragmatik; Diachronie; Varietät; Pejorisierung; Standards; Gendersymmetrie

## Automatic Extraction of Lexical Patterns from Corpora

Irene Renau, Rogelio Nazar

Pontificia Universidad Católica de Valparaíso

e-mail: irene.renau@pucv.cl, rogelio.nazar@pucv.cl

We present our first attempt to extract lexical patterns using corpus statistics. A pattern is a structure that combines syntactic and semantic features and is linked to a conventional meaning of a word. This means, for example, that the verb *to die* does not have intrinsic meanings, but potential meanings which are activated by the context: in ‘His mother died when he was five’, the meaning of the verb differs from ‘His mother is dying to meet you’, due to collocational restrictions and syntactic differences. With the automatic analysis of thousands of concordances per verb, we can make a first approach to the problem of detecting these structures in corpora, a very time-consuming task for lexicographers. The average precision is around 50%. The next step to increase precision is adding a dependency parser to the system and make adjustments to the automatic taxonomy we have created for semantic labeling.

**Keywords:** computational lexicography; lexical patterns; Spanish verbs; taxonomy

## Basic Russian Dictionary: a Corpora-Based Learner's Dictionary of Russian

**Maria Shvedova, Dmitri Sitchinava**

Kyiv National Linguistic University

e-mail: masha.shvedova@gmail.com

The paper presents a project of a learner's dictionary of Russian based on the most frequent Russian lexicon, and the most common words are also used for definitions. Both frequencies and examples are defined using the Russian National Corpus. The corpora methods make possible to include into the dictionary some meanings, constructions absent from the existing dictionaries of Russian let alone learner's dictionaries. Some elements of thesaurus are used in the dictionary. The structure of metalanguage, the study of collocations, and the use of corpora are discussed and exemplified in the paper.

**Keywords:** learner's dictionary; frequency lists; Russian National Corpus; metalanguage

## Automatically Identifying Selectional Restrictions of Predicates: A Computational Methodology for Practical Lexicography that Challenges Linguistic Theory

**Alexandra Anna Spalek, Marco Del Tredici**

Universitetet i Oslo, Universitat Pompeu Fabra

e-mail: a.a.spalek@ilos.uio.no, marco.deltredici@upf.edu

In this article we explore a methodology based on unsupervised learning that aims at automatically classifying the arguments of a verb, a classification crucially important in the process of sense isolation in lexicographic projects. More specifically, we strive to exploit the lexical semantic information obtained by combining a clustering algorithm with Word Embeddings and argue that the combination of both represents a fertile methodological ground for automatically classifying the selectional behaviour of a predicate. From the perspective of practical lexicography, our methodology should help to significantly shorten the time invested in isolating context-related senses of a predicate, in our case a verb. In particular, we illustrate, based on a case study of two Spanish verbs, how our methodology can provide a lexicographer with a general picture of a verb's behaviour in a corpus and argue that the clusters we provide automatically encode insightful generalizations that facilitate the process of decision-making for a lexicographer as to which senses are most common for a particular predicate, according to the different types of contexts it is most frequently used in.

**Keywords:** practical lexicography; selectional restrictions of verbs; automatic clustering of arguments; Word Embeddings; *k*-means

## Paysage: Lexical Variation in the Lexis of Landscape in Romance Dictionaries

**Geoffrey Williams, Araceli Alonso, Janet DeCesaris,**

**Ioana Galleron, Chris Mulhall**

LiCoRN-LIDILEM, Université Grenoble Alpes, Universitat Pompeu Fabra,  
Waterford Institute of Technology

e-mail: geoffrey.williams@orange.fr, aalonsocampo@gmail.com,  
ioana.galleron@gmail.com, janet.decesaris@upf.edu, cmulhall@wit.ie

This paper centres on lexical variations across languages and across time using the theories of collocational networks and collocational resonance. The word ‘paysage’ serves as a starting point to observe the evolution of landscape words in four Romance languages – French, Spanish, Italian, Romanian – from both a diachronic and a synchronic perspective. Data from several dictionaries, beginning with 16<sup>th</sup> century French dictionaries, indicate that earliest usage of ‘paysage’ is related to painting. From French, the word spread to Spanish (*paisaje*), in which it has been recorded in dictionaries since the early 18<sup>th</sup> century and is again related fundamentally to painting. The corresponding word in Italian, *paesaggio*, on the other hand, is much more sparsely represented in Italian lexicography, and in Romanian, *peisaj* does not appear until the 19<sup>th</sup> century under a combined French and Italian influence. In our paper, we use the collocations for ‘paysage’ to show how collocational networks is a proper framework for analysing and comparing lexical development across languages and for identifying – and representing – nuances in meaning and usage over time.

**Keywords:** landscape; lexical variation; collocational networks; collocational resonance

---

# Reports on Lexicographical and Lexicological Projects



---

## Un Projet Accompli: le *Lexicon de Buda (1825)* en Édition Électronique

(Software Demonstration)

**Maria Aldea**

Université Babeş -Bolyai

e-mail: aldea\_maria@yahoo.com

Le but de notre ouvrage est de présenter le résultat du projet lexicographique accompli au sein de l'Université Babeş-Bolyai de Cluj-Napoca, intitulé *Le Lexicon de Buda (1825). Édition informatisée en vue d'une consultation en ligne, id est l'édition électronique de ce dictionnaire quadrilingue ancien ayant pour langue de base le roumain* (<http://dspace.bcucluj.ro/handle/123456789/48723>).

**Keywords:** le Lexicon de Buda; lexicographie roumaine numérisée; e-dictionnaires

## Von einem *Orthographischen Wörterbuch* georgischer Äquivalente deutscher Toponyme bis zum *Kulturgeographischen Lexikon Deutschlands*

**Marina Andrazashvili**

Ivane Javakhishvili Tbilisi State University

e-mail: andarezi@yahoo.de

Der vorliegende Beitrag widmet sich den Ausgangs- und Gestaltungsprinzipien eines zweisprachigen Wörterbuchs, das ursprünglich in einem linguistischen Rahmen erschlossen und als ein *Orthographisches Wörterbuch georgischer Äquivalente deutscher Toponyme* aufgebracht werden sollte. Im Verlauf der Arbeit hat sich das Format aber wesentlich erweitert und sich durch die Ergänzung der Kurzkommentare geodätisch-administrativen, landeskundlichen, kulturellen, geschichtlichen Charakters in ein *Kulturgeographisches Lexikon Deutschlands* umgewandelt.

Im Beitrag wird über die Ergebnisse berichtet, die wir durch mehrere Arbeitsvorgänge erzielt haben. Zunächst haben wir deutsche Toponyme und deren Äquivalente in georgischen Quellen einer kontrastiv-translatorischen Analyse unterzogen, wobei wir die als unkorrekt signierten Formen exzerpiert und versucht haben, den Ursprung ihrer Fehlübertragung zu qualifizieren. Als nächstes haben wir alternative Übertragungskriterien auf verschiedenen Sprachebenen erarbeitet und eine möglichst objektive Gesetzmäßigkeit eines adäquateren Transfers der Eigennamen mit Rücksicht auf die Individualitäten des georgischen Sprachsystems abgeleitet. Danach haben wir die Korpora für jedes Bundesland erstellt, und zwar entsprechend den vorher festgelegten linguistischen Anforderungen sowie inhaltlichen Gestaltungsprinzipien für jede semantische Subklasse der Toponyme (Oikonyme, Choronyme, Oronyme, Hydronyme, Dryronyme etc.).

Die noch anstehenden Abschlussarbeiten beinhalten die Revision der bearbeiteten Lexikonartikel und ihre alphabetische Einbettung in das Buch sowie die Ergänzung des Buches durch das Illustrationsmaterial (Landkarten, CD-Aufnahmen jeweils auf Deutsch und auf Georgisch etc.) und das wissenschaftliche Corpus.

**Keywords:** Toponyme Deutschlands; georgische Äquivalente

## On Compiling a Norwegian Academic Vocabulary List

**Ruth Vatvedt Fjeld, Arash Saidi**

Department of Linguistics and Scandinavian Studies, University of Oslo  
e-mail: r.e.v.fjeld@iln.uio.no, arashsa@student.ifi.uio.no

In the last few years, several dictionaries of academic vocabulary have been developed, based on automatic analysis of academic corpora. Our presentation documents the first attempt to create a Norwegian Academic Vocabulary list (NAV) for the Norwegian Bokmål variety. This work is part of the Nordic project LUNAS (Language Use in Nordic Academic Settings) where the goal is to develop academic vocabulary lists and dictionaries for the Scandinavian languages. Part of the goal is to support Nordic language use in academic work instead of English, and to maintain each and every Nordic language also as a full-fledged academic language.

**Keywords:** academic vocabulary; academic words; academic corpus; statistical lemma selection; multiword expressions (MWEs); phraseology

## Lexicographic Description of Czech Complex Predicates: Between Lexicon and Grammar

**Václava Kettnerová, Petra Barančíková, Markéta Lopatková**

Charles University in Prague, Faculty of Mathematics and Physics  
e-mail: kettnerova,barancikova, lopatkova@ufal.mff.cuni.cz

In this paper, we propose a representation of Czech complex predicates with light verbs in a valency lexicon. We demonstrate that if such a representation is to be theoretically adequate and at the same time economical, the information on complex predicates should be divided between two components of the lexicon, a data component and a grammar component. The data component stores all the information necessary for the generation of well-formed deep and surface syntactic structures of complex predicates, namely valency frames of both light verbs and predicative nouns, links between these frames, mapping of verbal and nominal valency complementations, and mapping of the

semantic participant Instigator onto a verbal complementation. The grammar component of the lexicon, representing a part of the overall grammar of the language, contains formal rules. These rules, which are instantiated by the information stored in the data component, allow users to obtain deep and surface structures of complex predicates. Finally, the proposed model is applied in the annotation of 1.215 Czech complex predicates selected from the Czech National Corpus on the basis of frequency and saliency. The resulting data forms a solid foundation for further survey into various semantic and syntactic aspects of Czech complex predicates.

**Keywords:** valency lexicon; complex predicates with light verbs; predicative nouns; syntactic structure formation

## An Interactive Solution: a Grammar Handbook Linked to a Prescriptive Dictionary

**Tiina Leemets, Maire Raadik**

Institute of the Estonian Language

e-mail: [tiina.leemets@eki.ee](mailto:tiina.leemets@eki.ee), [maire.raadik@eki.ee](mailto:maire.raadik@eki.ee)

The paper introduces a new project at the Institute of the Estonian Language. 2018 will mark the 100th anniversary of Estonian dictionaries of correct usage. This prescriptive dictionary type was founded in 1918 as a concise spelling dictionary with basic morphological information. Since the 1999 edition it has many new layers and levels, including examples of words in sentences (word order, capitalization, case government etc.) and showing the difference between sublanguages. The last two editions (2006, 2013) were published both on paper and online.

The concise form of a dictionary does not enable explanations about the reasons for certain preferences or the history of codification. The aim of the current project is to create an interactive combination of a dictionary and an online handbook of standard Estonian, using customized MediaWiki software. The handbook will provide 1) spelling and grammar rules, 2) comments and advice on frequent questions about correct usage, 3) written records that show the development of language planning ideas and theory in Estonia. Dictionary entries can be linked to handbook chapters and vice versa.

**Keywords:** electronic lexicography; prescriptive lexicography; dictionary appendices

## Bilingual Dictionary Drafting: Connecting Basque Word Senses to Multilingual Equivalents

**David Lindemann, Iñaki San Vicente**

UPV/EHU University of the Basque Country, Elhuyar Foundation  
david.lindemann@ehu.eus, i.sanvicente@elhuyar.com

This paper presents a simple method for drafting bilingual dictionary content using existing lexical and NLP resources for Basque. The method consists of five steps, three belonging to a semi-automatic drafting, and another two to semi-automatic and manual post-editing: (1), the building of a corpus-based frequency lemma list; (2) the drafting of syntactical entities belonging to a lemma-sign; (3) the drafting of word senses belonging to syntactical entities; (4) a semi-automatic detection of gaps regarding syntactical entities, and (5) manual detection of word sense gaps. The described method relies on the exploitation of existing resources for Basque, and the multilingual cross-references present in WordNet. The application of the described method follows two goals: (1) drafting of a series of bilingual dictionaries with Basque, and (2) contribution to the updating and enrichment of two Basque NLP resources used for the drafting, EDBL and EusWN.

**Keywords:** bilingual lexicography; dictionary drafting; Basque language; WordNet

## Online Bibliography of Electronic Lexicography. The Project OBELEX<sup>meta</sup>

**Christine Möhrs**

Institute for the German Language in Mannheim  
e-mail: moehrs@ids-mannheim.de

The Online Bibliography of Electronic Lexicography (OBELEX<sup>meta</sup>) is a bibliographic database which is developed for researchers working in the field of dictionary research. The platform is hosted at the Institute for the German Language (Institut für Deutsche Sprache, IDS) in Mannheim. The poster presentation aims at presenting the current status of the ongoing project.

**Keywords:** electronic lexicography; bibliographic database; research literature

## Straddling the Boundaries of Traditional and Corpus-Based Lexicography: A Latvian-Czech Dictionary

Michal Škrabal

Institute of the Czech National Corpus, Charles University in Prague  
e-mail: michal.skrabal@ff.cuni.cz

The aim of the presentation is to introduce an ongoing project of a Latvian-Czech dictionary, the first bilingual lexicographical description for this combination of languages. My quite logical tendency to employ corpus methods faces, however, a significant problem – a limited data base. Thus, I necessarily find myself straddling the boundaries between the traditional, “card-based”, and “new”, corpus-based, way of lexicographical work. However, this should not be seen as a disadvantage – the dictionary is being built on the solid foundations provided by traditional, time-proven lexicographical practice. Nevertheless, whenever possible, I try to employ modern-age tools and methods (corpora, internet, software, etc.) which save a substantial amount of work and time. Several sample entries on the poster demonstrate the sought-after priorities of the dictionary. These are – at least in the field of Czech-Baltic lexicography – innovative features: a modern, user-friendly approach; a rich description of collocations; descriptiveness rather than prescriptiveness; the “emancipation” of certain lemmas (often excluded from the main list of entries, taboo-words, or words traditionally nested – e.g. feminine derivates, etc.); usage notes. If the existing, albeit prevailing, lexicographical description is inconsistent with the language reality, it should be – contrary to the tradition – discarded and replaced with another description.

**Keywords:** Latvian-Czech dictionary; traditional vs. corpus lexicography; bilingual lexicography; limited database

## The SWEGOT Lexical Resources Framework (Software Demonstration)

Monica von Martens, Louise Holmer

Dept. of Swedish, University of Gothenburg

e-mail: <mailto:monica.martens@gu.se>, [louise.holmer@svenska.gu.se](mailto:louise.holmer@svenska.gu.se)

The SWEGOT lexical resources framework consists of a database model, sample SQL scripts for database update, configuration text files, and program files written in the PHP programming language providing an easy-to-use web interface for searching, comparing and accessing collections of lexical data, such as glossaries. The database model is flexible and additional resources can easily be added with limited configuration needs. The user interface provides several methods for limiting the search results and the presentation grid as suitable for the user. An advanced query interface facilitates comparisons between different lexical resources such as ‘entries found in BookA and BookC, but not in BookB’. Actual access to each lexical resource is provided via image files or links to external resources, as defined separately for each ‘book’.

**Keywords:** lexical resources; glossary; database

---

**List of Authors**

<b>Adamia Inga</b> .....	<b>59</b>
<b>Adamska-Salaciak Arleta</b> .....	<b>101</b>
<b>Akhøj Nielsen Marita</b> .....	<b>81</b>
<b>Aldea Maria</b> .....	<b>111</b>
<b>Alonso Araceli</b> .....	<b>106</b>
<b>Andrazashvili Marina</b> .....	<b>111</b>
<b>Apresjan Valentina</b> .....	<b>37, 47</b>
<b>Arhar Holdt Špela</b> .....	<b>30</b>
<b>Augustyn Magdalena</b> .....	<b>52</b>
<b>Baisa Vít</b> .....	<b>47, 48, 59</b>
<b>Barančíková Petra</b> .....	<b>112</b>
<b>Barbera Manuel</b> .....	<b>77</b>
<b>Barbera Roberto</b> .....	<b>44</b>
<b>Battista Simonetta</b> .....	<b>23</b>
<b>Benić Juraj</b> .....	<b>41</b>
<b>Bepieva Naira</b> .....	<b>60</b>
<b>Berdičevskis Aleksandrs</b> .....	<b>50</b>
<b>Berg-Olsen Sturla</b> .....	<b>21</b>
<b>Beridze Marina</b> .....	<b>47</b>
<b>Biesaga Monika</b> .....	<b>21</b>
<b>Blanck Wiebke</b> .....	<b>38</b>
<b>Bothma Theo J.D</b> .....	<b>22</b>
<b>Buendía Castro Miriam</b> .....	<b>69</b>
<b>Buivolova Olga</b> .....	<b>47</b>
<b>Calanducci Antonio</b> .....	<b>44</b>

<b>Carrubba Carla</b> .....	<b>44</b>
<b>Choi Jun</b> .....	<b>54</b>
<b>Čibej Jaka</b> .....	<b>29</b>
<b>Cinková Silvie</b> .....	<b>48</b>
<b>Cirulli Sandro</b> .....	<b>37</b>
<b>Climent Philippe</b> .....	<b>22</b>
<b>Čubrilo Mirko</b> .....	<b>42</b>
<b>Daraselia Sophiko</b> .....	<b>49</b>
<b>Davidović Davor</b> .....	<b>44</b>
<b>DeCesaris Janet</b> .....	<b>101, 106</b>
<b>Declerck Thierry</b> .....	<b>87</b>
<b>Del Tredici Marco</b> .....	<b>106</b>
<b>de Santiago González Paula</b> .....	<b>89</b>
<b>Doborjginidze Nino</b> .....	<b>49</b>
<b>Dobrovolskij Dmitrij</b> .....	<b>88</b>
<b>Dorn Amelie</b> .....	<b>38, 87</b>
<b>Eckhoff Hanne Martine</b> .....	<b>50</b>
<b>Egorova Kseniya</b> .....	<b>77</b>
<b>Engelberg Stefan</b> .....	<b>78</b>
<b>Essert Mario</b> .....	<b>41, 42, 43</b>
<b>Faber Pamela</b> .....	<b>69, 70</b>
<b>Farina Donna M.T. Cr.</b> .....	<b>102</b>
<b>Farkas Zoltan</b> .....	<b>44</b>
<b>Fjeld Ruth Vatvedt</b> .....	<b>112</b>
<b>Fontenelle Thierry</b> .....	<b>71</b>
<b>Fourie Blair Hanelle</b> .....	<b>39</b>
<b>Fournier Pierre</b> .....	<b>60</b>
<b>Galieva Alfiya</b> .....	<b>50, 89</b>

<b>Galleron Ioana</b> .....	<b>83, 106</b>
<b>Gantar Polona</b> .....	<b>30</b>
<b>Gavriilidou Zoe</b> .....	<b>29</b>
<b>Geyer Ingeborg</b> .....	<b>104</b>
<b>Giouli Voula</b> .....	<b>38</b>
<b>Gippert Jost</b> .....	<b>13</b>
<b>Gizatova Guzel</b> .....	<b>51</b>
<b>Gorjanc Vojko</b> .....	<b>29</b>
<b>Gouws Rufus H.</b> .....	<b>22</b>
<b>Grabar Natalia</b> .....	<b>72</b>
<b>Grčić Simeunović Larisa</b> .....	<b>89</b>
<b>Gvelesiani Irina</b> .....	<b>71</b>
<b>Hanks Patrick</b> .....	<b>13</b>
<b>Hatier Sylvain</b> .....	<b>52</b>
<b>Heid Ulrich</b> .....	<b>72, 82</b>
<b>Hein Katrin</b> .....	<b>55</b>
<b>Holmer Louise</b> .....	<b>115</b>
<b>Hvelplund Holger</b> .....	<b>22</b>
<b>Ilson Robert</b> .....	<b>14</b>
<b>Inserra Giuseppina</b> .....	<b>44</b>
<b>Iomdin Boris</b> .....	<b>40</b>
<b>Jacques Marie-Paule</b> .....	<b>52</b>
<b>Johannsson Ellert Thor</b> .....	<b>23</b>
<b>Jung Hae-Yun</b> .....	<b>54</b>
<b>Kalman Tibor</b> .....	<b>44</b>
<b>Kamiński Mariusz</b> .....	<b>52</b>
<b>Kaźmierczak Rafał</b> .....	<b>31</b>
<b>Kettnerová Václava</b> .....	<b>112</b>

<b>Komuro Yuri</b> .....	<b>61</b>
<b>Kosem Iztok</b> .....	<b>30</b>
<b>Kotorova Elizaveta</b> .....	<b>24</b>
<b>Kovalenko Kira</b> .....	<b>38, 78</b>
<b>Krejčová Ema</b> .....	<b>48</b>
<b>Kultepina Olga</b> .....	<b>47</b>
<b>Kurdadze Ramaz</b> .....	<b>62</b>
<b>Langemets Margit</b> .....	<b>91</b>
<b>Lannoy Vincent</b> .....	<b>22</b>
<b>La Rocca Giuseppe</b> .....	<b>44</b>
<b>Lee Soojin</b> .....	<b>54</b>
<b>Leemets Tiina</b> .....	<b>113</b>
<b>León-Araúz Pilar</b> .....	<b>69, 70</b>
<b>Levontina Irina</b> .....	<b>102</b>
<b>Lew Robert</b> .....	<b>31</b>
<b>Lewandowska-Tomaszczyk Barbara</b> .....	<b>90</b>
<b>Lindemann David</b> .....	<b>114</b>
<b>Linde-Usiekiewicz Jadwiga</b> .....	<b>53</b>
<b>Lobzhanidze Irina</b> .....	<b>49, 95</b>
<b>Lomia Maia</b> .....	<b>62</b>
<b>Lopatková Markéta</b> .....	<b>112</b>
<b>Lopukhin Konstantin</b> .....	<b>40</b>
<b>Lopukhina Anastasiya</b> .....	<b>40</b>
<b>Lorentzen Henrik</b> .....	<b>25</b>
<b>Mac Amhlaigh Liam</b> .....	<b>31</b>
<b>Mafela Munzhedzi James</b> .....	<b>62</b>
<b>Makharoblidze Tamar</b> .....	<b>40</b>
<b>Maloletnjaja Anna</b> .....	<b>47</b>

<b>Marello Carla</b> .....	<b>32</b>
<b>Margalitadze Tinatin</b> .....	<b>103</b>
<b>Margiani Ketevan</b> .....	<b>62</b>
<b>Matsumoto Osamu</b> .....	<b>32</b>
<b>Mavrommatidou Stavroula</b> .....	<b>29</b>
<b>Meladze George</b> .....	<b>103</b>
<b>Merabishvili Innes</b> .....	<b>103</b>
<b>Metsmägi Iris</b> .....	<b>79</b>
<b>Meyer Peter</b> .....	<b>63</b>
<b>Mikulin Nikolai</b> .....	<b>37</b>
<b>Mirianashvili George</b> .....	<b>40</b>
<b>Miyoshi Kusujiro</b> .....	<b>80</b>
<b>Möhrs Christine</b> .....	<b>78, 114</b>
<b>Mörth Karlheinz</b> .....	<b>87</b>
<b>Može Sara</b> .....	<b>59</b>
<b>Mulhall Chris</b> .....	<b>106</b>
<b>Nadaraia David</b> .....	<b>47</b>
<b>Nagashima Saeko</b> .....	<b>61</b>
<b>Nam Kilim</b> .....	<b>54</b>
<b>Nazar Rogelio</b> .....	<b>105</b>
<b>Neprašová Renáta</b> .....	<b>95</b>
<b>Nevzorova Olga</b> .....	<b>50</b>
<b>Ní Loingsigh Katie</b> .....	<b>96</b>
<b>Nosyrev Grigory</b> .....	<b>40</b>
<b>Oja Vilja</b> .....	<b>79, 80</b>
<b>Ó Raghallaigh Brian</b> .....	<b>96</b>
<b>Orenha-Ottaiano Adriane</b> .....	<b>97</b>
<b>Orešković Marko</b> .....	<b>41, 42, 43</b>

<b>Panocová Renáta</b> .....	<b>72</b>
<b>Piringer Barbara</b> .....	<b>104</b>
<b>Pober Maria</b> .....	<b>104</b>
<b>Popiashvili Nino</b> .....	<b>60</b>
<b>Popič Damjan</b> .....	<b>29</b>
<b>Pöppel Ludmila</b> .....	<b>88</b>
<b>Prinsloo Danie J</b> .....	<b>22</b>
<b>Quaglino Margherita</b> .....	<b>81</b>
<b>Raadik Maire</b> .....	<b>113</b>
<b>Reimerink Arianne</b> .....	<b>69</b>
<b>Renau Irene</b> .....	<b>59, 105</b>
<b>Rutkowski Paweł</b> .....	<b>53</b>
<b>Rychkowa Liudmila</b> .....	<b>73</b>
<b>Safiullina Gulshat</b> .....	<b>65</b>
<b>Saidi Arash</b> .....	<b>112</b>
<b>Sánchez Cárdenas Beatriz</b> .....	<b>69</b>
<b>San Vicente Iñaki</b> .....	<b>113</b>
<b>Schreiber Hanno</b> .....	<b>39</b>
<b>Sekalashvili Russudan</b> .....	<b>64</b>
<b>Sharoff Serge</b> .....	<b>49</b>
<b>Shvedova Maria</b> .....	<b>105</b>
<b>Sipos Gergely</b> .....	<b>44</b>
<b>Sitchinava Dmitri</b> .....	<b>54, 105</b>
<b>Škrabal Michal</b> .....	<b>115</b>
<b>Spalek Alexandra Anna</b> .....	<b>106</b>
<b>Stadler Heike</b> .....	<b>82</b>
<b>Steyer Kathrin</b> .....	<b>55</b>
<b>Suleymanov Dzhavdet</b> .....	<b>89</b>

---

<b>Taljard Elsabé</b> .....	<b>74</b>
<b>ten Hacken Pius</b> .....	<b>15</b>
<b>Tomatis Marco Stefano</b> .....	<b>43</b>
<b>Topić Jakov</b> .....	<b>43</b>
<b>Tran Thi Thu Hoai</b> .....	<b>52</b>
<b>Trap-Jensen Lars</b> .....	<b>25</b>
<b>Troelsgård Thomas</b> .....	<b>81</b>
<b>Tutin Agnès</b> .....	<b>52</b>
<b>Tuulik Maria</b> .....	<b>91</b>
<b>van der Merwe Michele F.</b> .....	<b>33</b>
<b>van Niekerk Tim</b> .....	<b>82</b>
<b>Verdiani Silvia</b> .....	<b>25</b>
<b>Vernerová Anna</b> .....	<b>48</b>
<b>Villalva Alina</b> .....	<b>38</b>
<b>von Martens Monica</b> .....	<b>115</b>
<b>Wandji-Tchami Ornella</b> .....	<b>72</b>
<b>Wandl-Vogt Eveline</b> .....	<b>38, 44, 87, 104</b>
<b>Williams Geoffrey</b> .....	<b>15, 83, 106</b>
<b>Yan Rui</b> .....	<b>52</b>



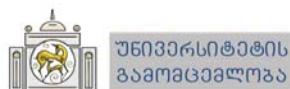
**Under the patronage  
of Tbilisi City  
Assembly**

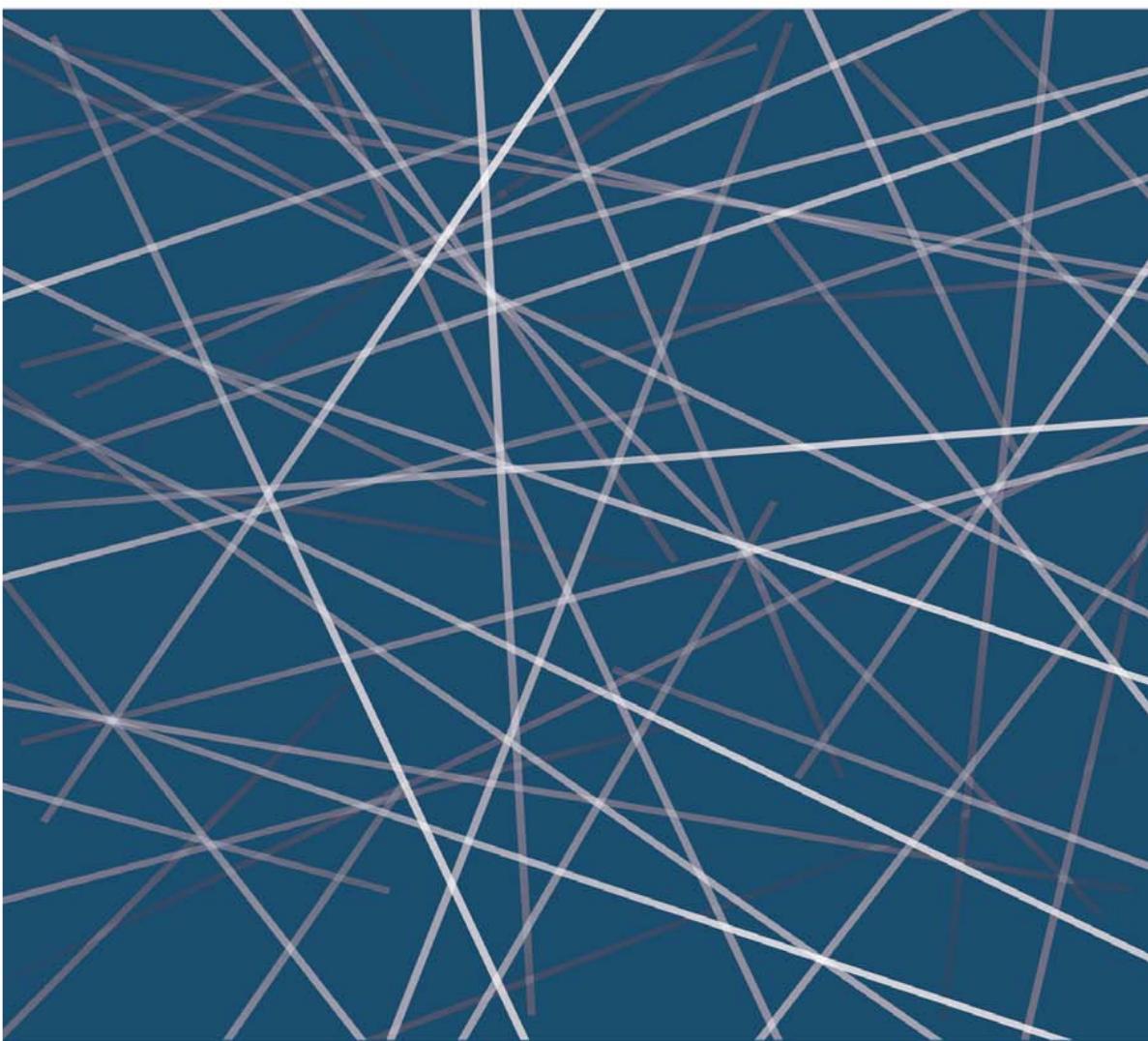


**Under the scientific patronage  
of Georgian National Academy  
of Sciences**



**And with the support of:**





უნივერსიტეტის  
გამომცემება

