



# **EURALEX XIX**

**Congress of the  
European Association  
for Lexicography**

**Lexicography for inclusion**

**7-9 September 2021**  
**Virtual**

[www.euralex2020.gr](http://www.euralex2020.gr)

**Proceedings Book  
Volume 2**

Edited by Zoe Gavriilidou, Lydia Mitits, Spyros Kiosses

## **EURALEX Proceedings**

ISSN 2521-7100

ISBN 978-618-85138-2-2

Published by: SynMorPhoSe Lab, Democritus University of Thrace

Komotini, Greece, 69100

e-edition

Publication is free of charge

Edited by: Zoe Gavrilidou, Lydia Mitits, Spyros Kiosses

English Language Proofreading: Lydia Mitits and Spyridon Kiosses

Technical Editor: Kyriakos Zagliveris



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

**2021 Edition**



In Memory of  
**Alain Rey**  
(1928-2020)

## SPONSORS



ΕΤΑΙΡΕΙΑ ΑΞΙΟΠΟΙΗΣΗΣ ΚΑΙ ΔΙΑΧΕΙΡΙΣΗΣ ΠΕΡΙΟΥΣΙΑΣ  
ΔΗΜΟΚΡΙΤΕΙΟΥ ΠΑΝΕΠΙΣΤΗΜΙΟΥ ΘΡΑΚΗΣ

Α.Φ.Μ. 094195820

ΔΠΜΣ Εξειδίκευση στις ΤΠΕ και Ειδική Αγωγή:  
Ψυχοπαιδαγωγική της ένταξης



A. S. Hornby Educational Trust

## Programme Committee

Zoe Gavrilidou  
 Maria Mitsiaki  
 Tinatin Margalitadze  
 Gilles-Maurice de Schryver  
 Simon Krek  
 Annette Klossa-Kückelhaus  
 Tanara Zingano Kuhn  
 George Xydopoulos

## Reviewers

|                            |                      |
|----------------------------|----------------------|
| Andrea Abel                | Dimitra Koukouzika   |
| Arleta Adamska-Salaciak    | Simon Krek           |
| Anna Anastasiadis          | Tita Kyriakopoulou   |
| Battaner Arias             | Lothar Lemnitzer     |
| Xavier Banco               | Robert Lew           |
| Gilles-Maurice de Schryver | Marie-Claude L'Homme |
| Janet DeCesaris            | Phillip Louw         |
| Ioannis Deligiannis        | Carla Marello        |
| Anna Dziemianko            | Tinatin Margalitadze |
| Angeliki Efthymiou         | George Mikros        |
| Asimakis Fliatouras        | Maria Mitsiaki       |
| Thierry Fontenelle         | Rosamund Moon        |
| Angeliki Fotopoulou        | Argyro Moustaki      |
| Zoe Gavrilidou             | Magali Paquot        |
| Alexander Geyken           | Stellios Piperidis   |
| Rufus Gouws                | Natascia Ralli       |
| Sylviane Granger           | Michael Rundell      |
| Oddrun Grønvik             | Max Silbertzein      |
| Patrick Hanks              | Elsabe Taljard       |
| Ulrich Heid                | Carole Tiberius      |
| Anna Iordanidou            | Lars Trap-Jensen     |
| Miloš Jakubiček            | Anna Vacalopoulou    |
| Jelena Kallas              | Geoffrey Williams    |
| Marianna Katsogiannou      | George Xydopoulos    |
| Ilan Kernerman             | Tanara Zingano Kuhn  |
| Annette Klossa             |                      |

## Table of Contents

|   |            |
|---|------------|
| <b>INVITED SPEAKERS</b> .....   | <b>551</b> |
| <b>POUR UN DICTIONNAIRE DE FAMILLES D'UNITÉS (SOUS-)LEXICALES*</b> .....  | 553        |
| <i>Anastasiadis-Symeonidis A.</i>   |            |
| <b>COMBATING LINGUISTIC MYTHS AND STEREOTYPES: THE CONTRIBUTION OF THE PRACTICAL DICTIONARY OF MODERN GREEK OF THE ACADEMY OF ATHENS.</b> .....                     | 567        |
| <i>Charalambakis Ch.</i>  |            |
| <b>DICTIONARIES AND MORPHOLOGY</b> .....  | 577        |
| <i>DeCesaris J.</i>   |            |
| <b>LEXICOGRAPHIC TREATMENT OF SALIENT FEATURES AND CHALLENGES IN THE CREATION OF PAPER AND ELECTRONIC DICTIONARIES.</b> .....                                       | 585        |
| <i>Prinsloo D.</i>  |            |
| <b>PAPERS</b> .....   | <b>593</b> |
| <b>NEOLOGY</b>  |            |
| <b>IT'S A LONG WAY TO A DICTIONARY: TOWARDS A CORPUS-BASED DICTIONARY OF NEOLOGISMS</b> .....   | 597        |
| <i>Afentoulidou V., Christofidou A.</i>   |            |
| <b>BI- AND MULTILINGUAL LEXICOGRAPHY</b>  |            |
| <b>FRISIAN DICTIONARIES, DIGITIZED FROM A TO Z</b> .....  | 609        |
| <i>Drenth E., Sijens H., Van de Velde H.</i>  |            |
| <b>CHARTING A LANDSCAPE OF LOANS. AN E-LEXICOGRAPHICAL PROJECT ON GERMAN LEXICAL BORROWINGS IN POLISH DIALECTS</b> .....  | 615        |
| <i>Meyer P., Hentschel G.</i>   |            |
| <b>HISTORICAL AND SCHOLARLY LEXICOGRAPHY AND ETYMOLOGY</b>  |            |
| <b>DRAWING THE LINE BETWEEN SYNCHRONY AND DIACHRONY IN HISTORICAL AND DIALECTAL LEXICOGRAPHY</b> .....  | 625        |
| <i>Manolessou I., Katsouda G.</i>   |            |
| <b>NEW WORDS IN OLD SOURCES: ADDITIONS TO THE LEMMA LIST OF A HISTORICAL SCHOLARLY DICTIONARY</b> .....   | 635        |
| <i>Johannsson E.T., Battista S.</i>   |            |
| <b>STEREOTYPES AND TABOO WORDS IN DICTIONARIES FROM A DIACHRONIC AND A SYNCHRONIC PERSPECTIVE – THE CASE STUDY OF CROATIAN AND CROATIAN CHURCH SLAVONIC</b> .....   | 643        |
| <i>Lazić D., Mihaljević A.</i>  |            |
| <b>REVISED ENTRIES IN THE MULTI-VOLUME EDITION AND TEI ENCODING: A CASE OF THE HISTORICAL DICTIONARY OF RUSSIAN</b> .....   | 655        |
| <i>Lyashevskaya O., Penkova Y.</i>  |            |
| <b>LEXICOGRAPHY AND CORPUS LINGUISTICS</b>  |            |
| <b>A LEXICOGRAPHIC PLATFORM FOR MIGRATION TERMINOLOGY: PROBLEMS AND METHODS</b> .....   | 665        |
| <i>Chiarì I.</i>  |            |
| <b>USING VERB PATTERNS TO FIND RECURRENT METAPHORS IN CORPUS</b> .....  | 673        |
| <i>Renau I.</i>   |            |
| <b>LES TERMES DES ARTS DANS LES DICTIONNAIRES DE LA TRADITION FRANÇAISE ET DANS LES CORPUS DE DERNIÈRE GÉNÉRATION : UNE RELATION D'INCLUSION RÉCIPROQUE ?</b> ..... | 683        |
| <i>Zotti V.</i>   |            |
| <b>LEXICOGRAPHY AND LANGUAGE TECHNOLOGIES</b>   |            |
| <b>TOWARDS AUTOMATIC DEFINITION EXTRACTION FOR SERBIAN</b> .....  | 695        |
| <i>Stanković R., Krstev C., Stijović R., Gočanin M., Škorić M.</i>  |            |
| <b>LICENSE TO USE: ELEXIS SURVEY ON LICENSING LEXICOGRAPHIC DATA AND SOFTWARE</b> .....   | 705        |
| <i>Kosem I., Nimb S., Tiberius C., Boelhouwer B., Krek S.</i>   |            |
| <b>VERBAL MULTIWORD EXPRESSIONS: A PRELIMINARY STUDY ON THE FIXEDNESS DEGREE, APPLICATION TO MODERN GREEK AND FRENCH</b> .....                                      | 713        |
| <i>Constant M., Fotopoulou A.</i>   |            |

## LEXICOGRAPHY AND SEMANTIC THEORY

|   |     |
|---|-----|
| <b>BUILDING A PARALYMPIC, FRAME-BASED DICTIONARY – TOWARDS AN INCLUSIVE DESIGN FOR <i>DICIONÁRIO PARAOLÍMPICO</i> (UNISINOS/BRAZIL)</b> ..... | 723 |
|---|-----|

*Chishman R., da Silva B., Nardes dos Santos A., Vianna A.L.T., de Oliveira S., Martins M.L., de Schryver G-M;*

## PEDAGOGICAL LEXICOGRAPHY

|  |     |
|--|-----|
| <b>THE DESIGN OF AN EXPLICIT AND INTEGRATED INTERVENTION PROGRAM FOR PUPILS AGED 10-12 WITH THE AIM TO PROMOTE DICTIONARY CULTURE AND STRATEGIES</b> ..... | 735 |
|--|-----|

*Gavrilidou Z., Konstantinidou E.*

## PHRASEOLOGY AND COLLOCATION

|  |     |
|--|-----|
| <b>ΟΙ ΦΡΑΣΕΟΛΟΓΙΣΜΟΙ-ΚΑΤΑΣΚΕΥΕΣ ΤΗΣ ΝΕΑΣ ΕΛΛΗΝΙΚΗΣ ΓΛΩΣΣΑΣ: ΜΙΑ ΛΕΞΙΚΟΓΡΑΦΙΚΗ ΠΡΟΣΕΓΓΙΣΗ</b> ..... | 749 |
|--|-----|

*Onufrieva E.*

## REPORTS ON LEXICOGRAPHICAL AND LEXICOLOGICAL PROJECTS

|   |     |
|---|-----|
| <b>NEW DEVELOPMENTS IN ELEXIFINDER, A DISCOVERY PORTAL FOR LEXICOGRAPHIC LITERATURE</b> ..... | 759 |
|---|-----|

*Kosem I., Lindemann D.*

## POSTERS ..... 767

### LEXICOGRAPHY AND CORPUS LINGUISTICS LEXICOGRAPHY AND LANGUAGE TECH

|   |     |
|---|-----|
| <b>CROWDSOURCING PEDAGOGICAL CORPORA FOR LEXICOGRAPHICAL PURPOSES</b> ..... | 771 |
|---|-----|

*Zingano Kuhn T., Todorović B.Š., Holdt Š.A., Zviel-Girshin R., Koppel K., Luís A.R., Kosem I.*

### PEDAGOGICAL LEXICOGRAPHY

|  |     |
|--|-----|
| <b>ΑΡΧΕΣ ΓΙΑ ΤΗ ΔΗΜΙΟΥΡΓΙΑ ΕΝΟΣ ΕΞΕΙΔΙΚΕΥΜΕΝΟΥ ΛΕΞΙΚΟΥ ΓΙΑ ΠΟΙΗΤΙΚΟΥΣ ΝΕΟΛΟΓΙΣΜΟΥΣ: ΜΕΛΕΤΗ ΠΕΡΙΠΤΩΣΗΣ ΣΤΗΝ ΟΔΥΣΣΕΙΑ ΤΟΥ ΝΙΚΟΥ ΚΑΖΑΝΤΖΑΚΗ</b> ..... | 783 |
|--|-----|

*Μαθιουδάκης Ν.*

## SOFTWARE DEMONSTRATIONS ..... 793

### LEXICOGRAPHY FOR SPECIALISED LANGUAGES, TERMINOLOGY

|   |     |
|---|-----|
| <b>INTRODUCING TERMINOLOGUE: A CLOUD-BASED, OPEN-SOURCE TERMINOLOGY MANAGEMENT TOOL</b> ..... | 797 |
|---|-----|

*Měchura M. B., Ó Raghallaigh B.*

## INDEX OF AUTHORS ..... 801





## Foreword

Given the unprecedented circumstances caused by COVID-19, the XIX Euralex, Congress organized by SynMorPhoSe Laboratory of the Department of Greek of Democritus University of Thrace was initially postponed and finally held online from the 7th till the 9th September 2021.

The motto of Euralex XIX was *Lexicography for Special Needs*. Even though the number of disabled people is rapidly increasing worldwide, modern lexicography has just started to address the needs of disabled or functionally diverse people in order to increase their accessibility and e-inclusion in lexicographic products. This congress aimed to make a strong statement and incite lexicographers to start thinking about how to provide more accessible dictionaries.

The second volume of Euralex XIX Proceedings is dedicated, as a mark of esteem for his scientific career in lexicography, to the memory of Alain Rey, who passed away in October 2020. Alain Rey was the editor-in-chief at French dictionary publisher *Dictionnaires Le Robert* from 1967 until his death. He supervised the publication of many dictionaries under the Le Robert trademark: the *Petit Robert* (1967); the *Micro Robert*, a pocket dictionary; the *Petit Robert des noms propres* (1974), a guide to proper names; the *Dictionnaire des expressions et locutions* (1979), a dictionary of phrases and expressions; the *Grand Robert de la langue française*, a nine-volume work (1985); an updated version, the *Nouveau Petit Robert de la langue française* (1993), and the *Dictionnaire historique de la langue française* (1992). In 2005, he published the *Dictionnaire culturel en langue française*.

This volume includes the papers of four Invited Speakers, 17 papers, 2 poster papers and 1 software demonstration. All submissions have been blind-reviewed by two independent reviewers. In case of doubt, a third independent opinion was asked. As was the case with the previous congresses, contributions were submitted on various topics of lexicography, including, but not limited to, the following fields:

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

The Organizing Committee would like to thank all authors for submitting their papers in order to be included in this volume and all the esteemed colleagues who accepted to review the papers.

As the Chair of the Congress, I would like to acknowledge the precious work done by the members of the Organizing Committee, Elina Chadjipapa, Chryssa Dourou and Stavroula Mavrommatidou, who joined efforts with me to make this second volume of EURALEX 2020 Proceedings possible.

**Zoe Gavriilidou**

Chair of Euralex XIX International Conference,

August 2021

# Revised entries in the multi-volume edition and TEI encoding: a case of the historical dictionary of Russian

Lyashevskaya O.<sup>1,2</sup>, Penkova Y.<sup>2,1</sup>

<sup>1</sup> National Research University Higher School of Economics

<sup>2</sup> V. V. Vinogradov Russian Language Institute RAS  
olesar@yandex.ru, amoenaa@inbox.ru

## Abstract

The Dictionary of Russian Language of the 11th – 17th centuries (DRL11–17), which covers both Old and Middle Russian periods, is an ongoing project of the Russian Academy of Sciences, with volumes 1–31 published in hardcopy in 1975–2019). Up to now, only volumes 28–30 were converted into the database and published free online (<http://web-corpora.net/wsgi/oldrus.wsgi/>). The online edition allows one to search for entries that contain particular grammatical properties, phraseological units, sources of etymology, texts and sources attested in the entry, historical periods they represent, etc. (Aksyonov et al. 2015, Vechkaeva 2016). This paper presents a new initiative aimed at the digitization of earlier volumes, which includes OCR, encoding the dictionary according to a TEI-compatible XML scheme, improving the integrity of entries, and additional data mining and enrichment using external resources. We focus on the issue of how to represent the revised entries, namely, those that were added, deleted, and corrected in subsequent volumes and in a supplementary volume.

The changes in the entries are usually powered by new sources taken into consideration, by new interpretations of the source documents, or by changes in editorial policy. The typology of revisions made by the authors and editors of later volumes includes: adding or deleting entries; adding or deleting certain parts of the entry (senses, examples, etymology, etc.); correcting one or several fields of the entry (definition, example, grammatical properties, bibliographic description of citations, etc.). More complex changes are decomposed into the components listed above.

The TEI-based scheme of the dictionary addresses two ways of presenting the content: (i) an online searchable version and (ii) a retro-digitized version that preserves the layout of the published volumes. In the first case, the revised entry is represented as one merged entry (Target) that incorporates data from Source (entry published in an earlier volume) and Revision (entry published as addendum in later or supplementary volumes). As neither Source nor Revision presents the correct content of the entry in full, the TEI-based representation of the Target should be generated. Besides that, advanced users may have access to the history of changes made by editors and to deleted entries. We use the critical apparatus module of TEI to track the history of changes, in which the lemma contains a “preferred”, corrected reading and another reading corresponds to the content provided in earlier volumes. Taking the perspective of the retro-digitized version, Source and Revision are two separate entries with different metadata. Nevertheless, these two entries are linked to each other using the reference tags. Taken as a whole, the proposed schema outlines the principles for documenting the genetic relationships between different versions of edited lexicographic material.

**Keywords:** historical lexicography; TEI encoding; retro-digitizing; dictionary content revision; Old Russian; Dictionary of Russian Language of the 11th–17th centuries

## 1. The Dictionary of Russian Language of the 11th – 17th centuries and its revision history

The Dictionary of Russian Language of the 11th–17th centuries (DRL11–17) presents different periods of the Old Russian language including Middle Russian (15–17th centuries). The compilation of the DRL11–17 was initially aimed at creating a wide-audience popular-science manual for reading Old Russian texts (cf. the verso of the title page and the preface to (DRL11–17, 1, 5–16)), provided with a minimal number of citations from the Old Russian sources (usually, the earliest and the latest attestations were given). The concept of the DRL11–17 had been changed several times even before the release of the first volume (the editors first planned to publish the so-called “Small Old Russian Dictionary of the 11th – 17th centuries”, see (Krysko 2007, 108)). During the publication process, the concept of the Dictionary underwent significant changes. From a popular-science manual, it gradually turned into an academic dictionary, cf. (Chernysheva 2013).

Changes in the concept caused the revision of many lexicographic principles: the number of source citations highly increased; recently published Old Russian texts are constantly added to the list of the dictionary sources, outdated editions being excluded; unknown originals for Old Russian translations are regularly identified.

One of the consequences of the conceptual change was the need for a revision of the previous volumes, taking into account the recent editions of sources, the emergence of Old Russian corpora, and newly identified originals. Most productively, this work was being carried out when V.B. Krysko became the editor-in-chief of the DRL11–17, i.e., in the 27th–29th volumes, cf. (Krysko 2007).

In 2006, additions and corrections to the first volume of the DRL11–17 were published, both in the appendix to volume 27, and as a separate volume (DRL11–17, vol. 27; Chernysheva 2006). However, the editorial board decided to postpone the systematic publication of additions to the earlier volumes until all the DRL11–17 volumes were completed. The editors decided to provide each current volume only with corrections and additions to the later volumes.

Thus, volume 27 is supplemented with additions for the alphabetic segment “C” and corrections to the previous 26th volume. The 28th volume (DRL11–17, vol. 28) includes only corrections to the 2nd volume (the alphabetic segment

“B”), the additions to the segment “B” being not yet published.

The 29th volume (DRL11–17, vol. 29) contains additions for the alphabetic segment “C”, omitted in previous volumes (including those omitted in additions published in DRL11–17, vol. 27: 214–216). The “Supplements” also indicate words that were mistakenly included in the 28th volume and should be removed. The 30th volume (DRL11–17, vol. 30: 316–318) is supplemented by the additions to the alphabetical segments “C–T” and the corrections to the previous 29th volume. Volume 31 (DRL11–17, vol. 31) also contains some additions to the segments “C–T”; however, it does not include any corrections planned to be published in the upcoming volumes. So, various additions and corrections are scattered across the latest five volumes (27th–31st).

## 2. Variety of revision strategies

The lexicographic principles for submitting corrections and additions were not completely unified in the latest volumes of the DRL11–17. In the 29th volume, references to entries included in the appendix are marked with an asterisk, cf.:

- (1) САМООБЪЩАННО... – Ср. самоотвъщанно\* (DRL11–17, vol. 29: 424)  
 ‘SAMOOBĚŠCANNO... – Cf. samootvěšanno\*’

In the supplementary volume (Chernysheva 2006), this symbol is used wider: it marks all newly discovered lemmas that are missing in the 1st volume. In the 31st volume, the asterisk is used to refer to the unpublished additions that will be included only in the next, 32nd volume.

However, the differences in structuring additions and corrections are not limited to ways of using an asterisk. The crucial difference between the lexicographic strategies used in different volumes is the following. In (Chernysheva 2006), additions and corrections are provided with a special commentary, such as “previous interpretation is removed”; “lemma is corrected”; “previous interpretation is specified”; “misprint in the lemma is corrected”. The previous variant, as a rule, is not indicated. In some cases, it is not easy to understand what changes are being made without consulting the original article from Volume 1, cf.:

- (2) БРЕЩИ. знач.4: цитаты с отрицанием переносятся в статью НЕБРЕЩИ.  
 Б р е щ и с е б е. Исправлены опечатки. (Chernysheva 2006: 60)  
 ‘BREŠČI. sense 4: citations with negation transferred in the entry NEBREŠČI’  
 ‘brešči sebe. Misprints corrected.’

In presenting corrections to the 2nd volume, published in the appendix to volume 28 (DRL11–17, vol. 28: 291–302), another strategy is adopted. The corrections are structured in the form of a table. The left column contains a lemma of the dictionary entry being amended. The lemma is given in the form as presented in the 2nd volume. In the right column, the information that requires correction is placed to the left of the “arrow” symbol (⇒). The Arabic numerals indicate the sense number, and a certain type of font indicates the corresponding field of the dictionary entry: bold capital letters mark lemmas, bold letters – references, italics indicate interpretation, wide font – idiomatic combinations. The correct variant stands to the right of the arrow symbol; a semicolon marks boundaries of the corrections. The initials of the editors who provided the editor-in-chief with the corresponding amendment are put in square brackets. If the correction concerns only the lemma, the arrow symbol is not used, and the corrected lemma variant is put in the right column (see table 1).

|        |   |
|--------|---|
| ВДАНИЕ | 2. Действие по глг. вдатися (в знач. 2) ⇒ Сдача (города) без боя, капитуляция [К.М.]; овоихъ ⇒ ово ихъ (2 раза)                               |
| ВДАНИЕ | ‘2. Action related to the verb vdatisya (sense 2) ⇒ Surrender (of the town) without a fight, capitulation [K.M.]; ovoih” ⇒ ovo ih” (2 times)’ |
| ВДАНИЕ | 3. Вещь, отданная на сбережение; залог ⇒ Вещь или сумма денег, отданная в залог серьезности намерений заключить брак [К.М.].                  |
| ВДАНИЕ | ‘3. A savings item; a pledge ⇒ A thing or amount of money pledged for the seriousness of intent to marry [K.M.]’                              |
| ВДАНИЕ | 2. Действие по глг. вдатися (в знач. 2) ⇒ Сдача (города) без боя, капитуляция [К.М.]; овоихъ ⇒ ово ихъ (2 раза)                               |
| ВДАНИЕ | ‘2. Action related to the verb vdatisya (sense 2) ⇒ Surrender (of the town) without a fight, capitulation [K.M.]; ovoih” ⇒ ovo ih” (2 times)’ |
| ВДАНИЕ | 3. Вещь, отданная на сбережение; залог ⇒ Вещь или сумма денег, отданная в залог серьезности намерений заключить брак [К.М.].                  |
| ВДАНИЕ | ‘3. A savings item; a pledge ⇒ A thing or amount of money pledged for the seriousness of intent to marry [K.M.]’                              |

Table 1: A sample of corrections presented in (DRL11–17, vol. 28: 292).

If the changes are related to removing a segment or transferring it to another dictionary entry, the arrow symbol is not used, and the correction is provided only with a comment: “citation is transferred to article N,” or “entry is removed,” etc.

The types of the changes approve the strategy used in (Chernysheva 2006): the supplementary volume deals mostly with adding new entries. It determined the principles for presenting additions and corrections (see above). However, the

strategy used in the 28th volume seems to be more convenient for the user of the dictionary, on the one hand, and it is easier to integrate the changes into the dictionary database, on the other hand.

The inconsistency with which additions and corrections are presented in different volumes of the DRL11–17 introduces difficulties for integrating the corresponding ones into the dictionary database. However, before discussing particular technical solutions, it is necessary to establish all the types of additions and corrections which we are dealing with.

### 3. Typology of the content revisions

Various reasons cause the changes that have been introduced by the editors. There are many different classifications of lexicographic errors depending on the particular purposes. One can propose a lexicographic error classification based on the reasons behind these errors, cf. (Shapoval 2016). There are various types of such errors:

- errors as a result of an incorrect reading of sources (for example, incorrect word division in a manuscript);
- transmission errors, i.e., errors brought by copying text from a source to dictionary entry;
- errors which go back to the publishers of Old Russian sources cited and uncritically reproduced in the dictionary;
- errors resulting from incorrect interpretation or reconstruction of grammatical forms, etc.

The changes in the entries can also be powered by new interpretations of the source documents, by new sources taken into consideration, or by changes in editorial policy. However, for our purposes, we need a technical classification of changes. The typology of revisions made by the authors and editors of later volumes includes adding missing entries, or parts of the entries, and revision.

#### 3.1. Additions

Types of *additions* include:

- adding a missing dictionary entry;
- adding a missing reference entry;
- adding a missing part in an existing entry (i.e., adding Greek glosses to the citation taken from Old Russian translations, identifying a new sense, etc.);
- expanding already existing fields (i.e., citation extension).

#### 3.2. Simple and complex revisions

We distinguish between simple and complex revisions. The former ones suggest that the correction can be made in one step. Simple corrections include deleting an existing entry or a certain part of it, replacing the contents of one of the fields, moving a part of the entry to another entry, as well as simple corrections in one of the fields (i.e., correction of misprints).

Simple replacements may refer to:

- lemma spelling;
- grammatical information;
- interpretation provided, no changes in the citations are required;
- source citation, unless it causes changes in the chronological order of the citations.

Simple transfers include:

- moving a source citation from one sense to another within the same dictionary entry, provided the latter preserves its structure;
- moving citations to another dictionary entry, provided both entries preserve their structure.

Simple deletions include:

- deleting the dictionary entry, unless it requires deletion of the reference included in another dictionary entry;
- deleting a part of the entry along with all the citations.

Complex revisions are a combination of two or more corrections, i.e., such changes, which, from a technical point of view, should be introduced in several steps. Complex corrections include the following:

- deleting a dictionary entry, along with transferring citations from the deleted entry to other dictionary entries;
- transferring citations along with creating a new dictionary entry;
- transferring citations to another dictionary entry, along with adding to the latter, a new sense field;
- replacing an existing dictionary entry with a reference entry, along with transferring the citations to other dictionary entries.

Table 2 shows an example of a complex change, a replacement of the dictionary entry with the reference entry, while the citation is transferred to another dictionary entry.

|                        |  |
|------------------------|--|
| ВОЖАВСТВО<br>VOŽAVSTVO | ВОЖАВСТВО см. вожевство; цитата переносится в статью ВОЖЕВСТВО<br>'VOŽAVSTVO see voževstvo; citation transferred to the entry VOŽEVSTVO' |
|------------------------|--|

Table 1: A sample entry with a complex change.

### 3.3. Implicit corrections

We discussed the classification of revisions to the earlier volumes of the DRL11–17, proposed by the editors of the latter five volumes. However, there is also another type of correction, which can be called implicit.

Firstly, implicit corrections arise in the case when the source, used for decades, acquires a new dating, due to a separate archaeological research. Secondly, implicit corrections include introduction of new source abbreviations. As a result, different volumes of the dictionary can refer to the same source, using different abbreviations. Thirdly, implicit corrections are powered by the termination of the use of a particular source recognized as unsuitable (i.e., if a manuscript is recognized as not featuring the Russian redaction of Church Slavonic). Fourthly, spelling and interpretation of Greek glosses (and, less commonly, glosses in other languages), is also the subject of editorial revisions in subsequent volumes. Other corrections can also occur.

## 4. The DRL11–17 electronic edition

Starting from 1975, the dictionary was originally published in hardcopy (the 31st volume was released in 2019). The their first attempt to create an online edition, Aksyonov et al. (2015) and Vechkaeva (2016) converted the content of volumes 28–30 into the lexicographic database and made it possible to search for head words, grammatical labels, phraseological units, sources of etymology, texts and sources attested in the entry and historical periods they represent, etc. online (<http://webcorpora.net/wsgi/oldrus.wsgi/>). The new objective is to expand the coverage of the database, so that it eventually includes all printed volumes, and to make access to the content more flexible and user-friendly.

The data is automatically recognized, using the ABBYY FineReader OCR system, manually checked and presented in three different ways:

- TEI-compatible XML scheme;
- SQLite database format;
- screenshots of the printed dictionary pages and individual entries (pdf view).

Figures 1 and 2 illustrate the TEI-based representation and the browser view, based on the SQLite representation of the same entry. By clicking on the “book” pictogram, the user can switch to the pdf view of the printed edition, see figure 3.

To improve the integrity of entries, some implicit information, not present in the printed version, was retrieved and added to the TEI-based representation. For that purpose, we mostly use tags, embedded within the definition and example fields. For example, the definition template elements *Tom, kmo...* ‘The one who (does smth.)’ is marked with the tag `<defTemplate class="agent">`. The corresponding nominalization, denoting the agent’s action (*спасение* ‘salvation’), has a cross-reference to the entry *спасение*. In examples, the mentioned headwords were identified and labeled with `<oRef>`, eg. `<oRef type="pl">сѣсьници</oRef>` for the plural abbreviated form of the headword *спасникъ*. Glosses were classified by language and scope, see the tag `<gloss lang="gr" class="ex">` for the Greek gloss that corresponds to the Greek source of the cited example. Besides that, the editors’ notes, with regard to definitions and examples, were marked and classified. Missing data tags were added, using information from the source titles, see the tag `<date class="hidden" when="1073">` which has the attribute ‘hidden’. In enhanced TEI-based representation, lemmas for all Russian and Old Russian words are provided so that the user could look for particular words in definitions, examples, grammatical data, and other fields throughout the entries.

The TEI scheme of the dictionary is designed to support two ways of presenting the content:

- (i) an online searchable version and
- (ii) a retro-digitized version that preserves the layout of the published volumes.

With such functionality, line break tags (`<lb/>`) are preserved in the TEI-based representation, but used only in (ii). Analogically, information tagged with the attribute ‘hidden’ is used in (i) and not shown in (ii).

The following section addresses the representation of the revised entries, namely, those that were added, deleted, and corrected in the TEI scheme, taking into account (i) and (ii).



```

<entry xml:lang="orv" xml:id="спасникъ" type="mainEntry">
  <form type="lemma">
    <orth>СПАСНИКЪ</orth>
  </form>,
  <gramGrp>
    <gen>м.</gen>
  </gramGrp>
  <sense n="0">
    <def>
      <defTemplate class="agent">Тот, кто</defTemplate>
      несет <ref target="#спасение">спасе-<lb/>ние</ref>, охраняет.
    </def>
    <cit type="example">
      <quote>Да сице убо сѣи добрымъ члкомъ и блговѣрнымъ звѣзды суть и<lb/>
      <oRef type="pl">сѣсьници</oRef>, якоже рече Гѣ къ нимъ: вы есте<lb/>
      свѣтъ мира сего и соль земльная
      <gloss lang="gr" class="ex">(σωτήριοι)</gloss>.
    </quote><lb/>
    <bibl>
      <title level="a" id="izb1073-2">Изб.Св. 1073 г.<hi rend="sup">2</hi></title>,
      <biblScope unit="page">572</biblScope>.
    </bibl>
    <date class="hidden" when="1073"/>
  </cit>
</sense>
</entry>

```

Figure 1: A simplified TEI-based representation of the entry *спасникъ* 'savior'.**СПАСНИКЪ, м.**

Тот, кто несет спасение, охраняет.

1073 Да сице убо сѣи добрымъ члкомъ и блговѣрнымъ звѣзды суть и сѣсьници, якоже рече Гѣ къ нимъ: вы есте свѣтъ мира сего и соль земльная (σωτήριοι).

Изб. Св. 1073 г.<sup>2</sup>, 572

Figure 2: The browser view for the entry *спасникъ* 'savior'.

**СПАСНИКЪ, м.** Тот, кто несет спасение, охраняет. Да сице убо сѣи добрымъ члкомъ и блговѣрнымъ звѣзды суть и сѣсьници, якоже рече Гѣ къ нимъ: вы есте свѣтъ мира сего и соль земльная (σωτήριοι). Изб.Св. 1073 г.<sup>2</sup>, 572.

Figure 3: The pdf view for the entry *спасникъ* 'savior'.

## 5. TEI-based approach to track the revision changes

In the online searchable version (i), the revised entry is represented as one merged entry that incorporates data from both Source (entry published in an earlier volume) and Revision (entry published as addendum in later or supplementary volumes). As none of them presents the correct content of the entry in full, the TEI-based representation of the Target should be generated from both. In the retro-digitized version (ii), Source and Revision are considered as two (or more) separate entries, with different metadata. However, it is useful to link these entries to each other, to facilitate access to the related content. Figure 4 shows the Revision entry of the word *безглавный* 'headless' (see the TEI tag <entry type="supplementaryEntry">), which refers to the Source entry using the TEI tag <ref>.

While compiling the electronic version of the dictionary, we face a challenging task to represent the various kinds of additions and corrections made, over time, by the editors of the DRL11–17. At the same time, advanced users may have access to the revision history of a given entry and to deleted entries. So, the history of changes should also be reflected in the representation.

Taking into account the typology of revisions provided in section 3, we distinguish among:

- adding and deleting the whole entry, reference entry, particular senses;
- transferring data from one entry or subentry (sense) to another, and from one field to another;
- changing the order of certain elements;
- replacing data within a particular entry field.

```

<entry xml:lang="orv" xml:id="безглавный.sup1" type="supplementaryEntry">
  <ref target="#безглавный" type="entry">
    <form type="lemma">
      <orth>БЕЗГЛАВНЫЙ</orth>
    </form>
    <gramGrp>
      <gram type="pos" value="ADJ">прил.</gram>
    </gramGrp>
    <note>...</note></ref>
    <sense n="1">4.
      <def><defTemplate class="equivalent">То же, что</defTemplate><lb/>
        <ref target="#безглавный"><note type="supplementaryEntry">
          <hi rend="sup">*</hi></note>безглавный</ref>.</def>
        </sense>
      </note>
      Сочетание <ref target="безглавная_вѣра"><hi rend="expanded">безглавная вѣра</hi></ref>
      в знач. 3<lb/>
      снимается. Цитата переносится в знач. 4.
    </note>
  </entry>

```

Figure 4: A simplified TEI-based representation of corrections to the entry *безглавный* ‘headless’ in the supplementary volume.

Complex changes are decomposed into the components listed above. Technically, transferring and changing the order of elements is encoded as adding and deleting in the TEI-based encoding.

There is no uniform convention on how to represent the revision history in dictionaries. On the one hand, the tags <add>, <del>, <subst> are good candidates to reflect simple changes, but they are defined to encode changes made in the same primary source (i.e. one tangible medium) (TEI P5 2021). Certainly, this does not hold true in our case, since we deal with multiple, though related, media that carry the changes. On the other hand, tags such as <revisionDesc> and <listChange> are intended to summarize the history of changes in the header, rather than throughout the body of the document. We choose the third option, namely, using the critical apparatus module of TEI that is intended to represent related texts found in different physical witnesses (TEI P5 2021, Section 12).

To illustrate the use of the critical editing tags, let us move back to the previous example. In the Revision entry of the adjective *безглавный* ‘headless’, sense 4 is added provided with a definition only. The editors add a note that the section of the sense 3 that contains a multi-word expression is deprecated; however, the citation from this section should be transferred to sense 4. The reason for restructuring the entry is that the new adjective *безглавный* ‘headless’, of roughly the same morphological structure and with the same meaning as in the multi-word expression, is added to the dictionary. In order to document all the changes, the list of all DRL11–17 volumes is declared as witnesses in the <listWit> element of the TEI header. In the entry, the revised sections are enclosed in the <app> tag (see Figure 5).

Since the entries have been revised only once by now, each <app> section contains a <lem> element (a “preferred”, corrected reading) and one <rdg> element (a reading from an earlier edition identified by the attribute *wit*). In the case of the subentry transfer, there are two <app> sections corresponding to deletion and addition, respectively. The first <app> element contains an empty <lem> element (deletion) and the subentry for the multi-word expression from the Source entry enclosed in the <rdg wit="#V1">...</rdg> tags. The second <app> element contains a lemma with both new sense and its definition explicitly represented in the Revision entry and the transferred citation. The <rdg> element is empty.

Note that in deviation from most common practice, the Source and Revision witnesses are not different versions of the same work, rather, the Revision can be considered as a commentary and supplementary material that substitutes the content of the Source representing the final version of the lexicographic material. Another limitation of the current TEI-based representation is that content provided by the editors and generated content are not explicitly distinguished in the <lem> elements. Yet, in the absence of better solutions, this takes a step towards encoding the “genetic relationships among documents” (Barney 2018).

## 6. Conclusion

Digitizing the Dictionary of Russian Language of the 11th – 17th centuries not only makes the published volumes of this long-lasting project more accessible but also gives researchers a more flexible and powerful resource to work with. Our approach offers a new vision of the critical electronic edition for the multi-volume historical dictionaries. This implies the parallel handling of an online, searchable version, enriched with linguistic and textological information, and a retro-digitized version that preserves the layout of the published volumes.

In this paper, we addressed challenges arising from inconsistencies in editorial practice, strategies to report added, deleted, corrected, and restructured entries. Following the developed typology of simple and complex revisions, we adopted a TEI-compatible XML scheme to represent corrections, content restructuring and transfer, made by editors over many years. However, more efforts to standardize the tracking of longitudinal changes in multi-volume dictionaries are needed.

```

....
<sense n="3">3.
....
  <app>
    <lem/>
    <rdg wit="#V1">
      <entry xml:lang="orv" xml:id="безглавная_вѣра" type="relatedEntry">
        <form type="idiom">Безглавная<lb/>вѣра</form> –
        <def>вероучение, не признающее цер-<lb/>
          ковной иерархии.</def>
        <cit type="example">
          <quote>Маму же нѣкоего... еже<lb/>
            съ Севгиромѣ нѣкоимѣ начатокъ безгла-<lb/>
              вию... поимъ съ собою оѣцъ наша Сава<lb/>
                во Иерѣслмѣ, моляшеся отступити от без-<lb/>
                  главныя вѣры и к соборнѣи общеватися<lb/>
                    цркви.</quote>
          <bibl><title level="a" id="VMCh">ВМЧ, Дек. 1–5</title>,
            <biblScope unit="page">504</biblScope></bibl>.
          <date notAfter="1600" notBefore="1500">XVI в.</date>
        </cit>
      </entry>
    </rdg>
  </app>
</sense>
<app>
  <lem>
    <sense n="1">4.
      <def><defTemplate class="equivalent">То же, что</defTemplate><lb/>
        <ref target="#безглавый"><note type="supplementaryEntry">
          <hi rend="sup">*</hi></note>безглавый</ref>.</def>
        <cit type="example">
          <quote>Маму же нѣкоего... еже<lb/>
            съ Севгиромѣ нѣкоимѣ начатокъ безгла-<lb/>
              вию... поимъ съ собою оѣцъ наша Сава<lb/>
                во Иерѣслмѣ, моляшеся отступити от без-<lb/>
                  главныя вѣры и к соборнѣи общеватися<lb/>
                    цркви.</quote>
          <bibl><title level="a" id="VMCh">ВМЧ, Дек. 1–5</title>,
            <biblScope unit="page">504</biblScope></bibl>.
          <date notAfter="1600" notBefore="1500">XVI в.</date>
        </cit>
      </sense>
    </lem>
  <rdg wit="#V1"/>
</app>

```

Figure 5: A fragment of the generated TEI-based representation of the *безглавный* ‘headless’.

## 7. References

- Aksyonov, K., Bobrik, M., Krivko, R., Orekhov, B., Novosyolova, A., & Vechkaeva A. (2015). Dictionary of Russian Language of the 11th—17th Centuries as a Database: Information Retrieval and Research Perspectives. In *Approaches to the Editing of Slavonic Texts. Tradition and Innovation in Palaeoslavistic Ecdotics* (ATTEST), Regensburg, 11–12 December 2015. Book of Abstracts. Accessed at: [https://www.uni-regensburg.de/sprache-literatur-kultur/slavistik/medien/aktuelles/krivko\\_abstract.pdf](https://www.uni-regensburg.de/sprache-literatur-kultur/slavistik/medien/aktuelles/krivko_abstract.pdf) [20/02/2020]
- Barney, B. (2018). TEI, the Walt Whitman Archive, and the Test of Time. In *Proceedings of JADH 2018*, p. 251.
- Chernysheva, M.I. (2013). Iz istorii izmemeniya koncepcii Slovarja russkogo jazyka XI–XVII vv. In *Acta linguistica petropolitana. Proceedings of the Institute for Linguistic Studies RAS*. Vol. IX (2). *Russkaja istoričeskaja*



- leksikologija i leksikografija XVII–XIX vv.* [Russian historical lexicology and lexicography]. Saint-Petersburg: Nauka, pp. 192–205.
- Chernysheva, M.I. (2006). *Dictionary of Russian language of the 11th–17th centuries. Addenda et corrigenda*. Issue 1 (A–B). Moscow: Nauka.
- Chernysheva, M.I. (ed.) (2020) Resource materials to the dictionary of Russian language of the 11th–17th centuries. Index of sources. Word index (direct). Moscow: Leksrus.
- DRL11–17, vol. 1: *Dictionary of Russian language of the 11th–17th centuries. Volume 1: (A–B)*. Editor-in-chief S.G. Barhudarov. Moscow: Nauka, 1975.
- DRL11–17, vol. 27: *Dictionary of Russian language of the 11th–17th centuries. Volume 27: (Снасъ–Старицынь)*. Editor-in-chief V.B. Krysko. Moscow: Nauka, 2006.
- DRL11–17, vol. 28: *Dictionary of Russian language of the 11th–17th centuries. Volume 28: (Старичекъ–Сулебный)*. Editor-in-chief V.B. Krysko. Moscow: Nauka, 2008.
- DRL11–17, vol. 29: *Dictionary of Russian language of the 11th–17th centuries. Volume 29: (Сулегъ–Тольмиже)*. Editor-in-chief V.B. Krysko. Moscow: Nauka, Azbukovnik, 2011.
- DRL11–17, vol. 30: *Dictionary of Russian language of the 11th–17th centuries. Volume 30: (Томъ–Уберечися)*. Editor-in-chief R.N. Krivko. Moscow–Saint-Petersburg: Nestor-Istorija, 2015.
- DRL11–17, vol. 31: *Dictionary of Russian language of the 11th–17th centuries. Volume 31: (Убивание–Улокъ)*. Editor-in-chief R.N. Krivko. Moscow: Leksrus, 2019.
- Krysko, V.B. (2007). Russkaja istoričeskaja leksikografija (XI–XVII vv.): problemy i perspektivy [Russian historical lexicography of the 11th — 17th]. In *Voprosy Jazykoznanija*, 1, pp. 103–118.
- Shapoval, V. V. (2016). *Teorija i praktika verifikacii slovarnyh dannyh na osnove istočnikov* [Theory and practice of the lexicographic data verification based on sources]. Doct. Sc. dissertation. Moscow.
- Vechkaeva, Anna (2016). *Dictionary of Russian Language of the 11th — 17th Centuries as a Database: Information Retrieval and Research Perspectives*. BA thesis. Moscow, National Research University Higher School of Economics.
- TEI P5 Guidelines, Chapter 11: Representation of Primary Sources. Version 4.0.0. Last updated on 1st March 2021. Accessed at: <https://tei-c.org/release/doc/tei-p5-doc/en/html/PH.html> [20/03/2021].

### Acknowledgements

The work has been supported by the Ministry of Science and Higher Education of the Russian Federation, Agreement No 075-15-2020-793.



**7-9 September 2021**  
Virtual

[www.euralex2020.gr](http://www.euralex2020.gr)

**Index of Authors**



**Roman****A**

|                             |     |
|-----------------------------|-----|
| Afentoulidou V.             | 597 |
| Anastassiadis-Symeonidis A. | 553 |

**B**

|               |     |
|---------------|-----|
| Battista S.   | 635 |
| Boelhouwer B. | 705 |

**C**

|                  |     |
|------------------|-----|
| Charalambakis C. | 567 |
| Chiari I.        | 665 |
| Chishman R.      | 723 |
| Christofidou A.  | 597 |
| Constant M.      | 713 |

**D**

|                  |     |
|------------------|-----|
| da Silva B.      | 723 |
| DeCesaris J.     | 577 |
| de Oliveira S.   | 723 |
| de Schryver G-M. | 723 |
| Drenth E.        | 609 |

**F**

|               |     |
|---------------|-----|
| Fotopoulou A. | 713 |
|---------------|-----|

**G**

|                |     |
|----------------|-----|
| Gavriilidou Z. | 735 |
| Gočanin M.     | 695 |

**H**

|              |     |
|--------------|-----|
| Hentschel G. | 615 |
| Holdt Š.A.   | 771 |

**J**

|                 |     |
|-----------------|-----|
| Johannsson E.T. | 635 |
|-----------------|-----|

**K**

|                   |          |
|-------------------|----------|
| Katsouda G.       | 625      |
| Konstantinidou E. | 735      |
| Kosem I.          | 759, 771 |
| Krek S.           | 705      |
| Krstev C.         | 695      |

**L**

|                 |     |
|-----------------|-----|
| Lazić D.        | 643 |
| Lindemann D.    | 759 |
| Luís A.R.       | 771 |
| Lyashevskaya O. | 655 |

**M**

|                |     |
|----------------|-----|
| Manolessou I.  | 625 |
| Martins M.L.   | 723 |
| Μαθιουδάκης N. | 783 |
| Měchura M. B.  | 797 |
| Meyer P.       | 615 |
| Mihaljević A.  | 643 |

**N**

|                      |     |
|----------------------|-----|
| Nardes dos Santos A. | 723 |
|----------------------|-----|

**O**

|                  |     |
|------------------|-----|
| Onufrieva E.     | 749 |
| Ó Raghallaigh B. | 797 |

**P**

|             |     |
|-------------|-----|
| Penkova Y.  | 655 |
| Prinsloo D. | 585 |

**R**

|          |     |
|----------|-----|
| Renau I. | 673 |
|----------|-----|

**S**

|              |     |
|--------------|-----|
| Sijens H.    | 609 |
| Škorić M.    | 695 |
| Stanković R. | 695 |
| Stijović R.  | 695 |

**T**

|                |     |
|----------------|-----|
| Todorović B.Š. | 771 |
|----------------|-----|

**V**

|                 |     |
|-----------------|-----|
| Van de Velde H. | 609 |
| Vianna A.L.T.   | 723 |

**Z**

|                  |     |
|------------------|-----|
| Zingano Kuhn T.  | 771 |
| Zotti V.         | 683 |
| Zviel-Girshin R. | 771 |







**7-9 September 2021**  
Virtual

[www.euralex2020.gr](http://www.euralex2020.gr)

Index  
Form  
Definition  
Orthography  
Sylla  
Publishing  
Pragmatics  
Spelling  
Ex  
Spoken  
Semantics  
Order  
Context  
Library  
Understand  
Translate  
Graphemics  
Origin  
Reference  
gmentation