connecting creative minds

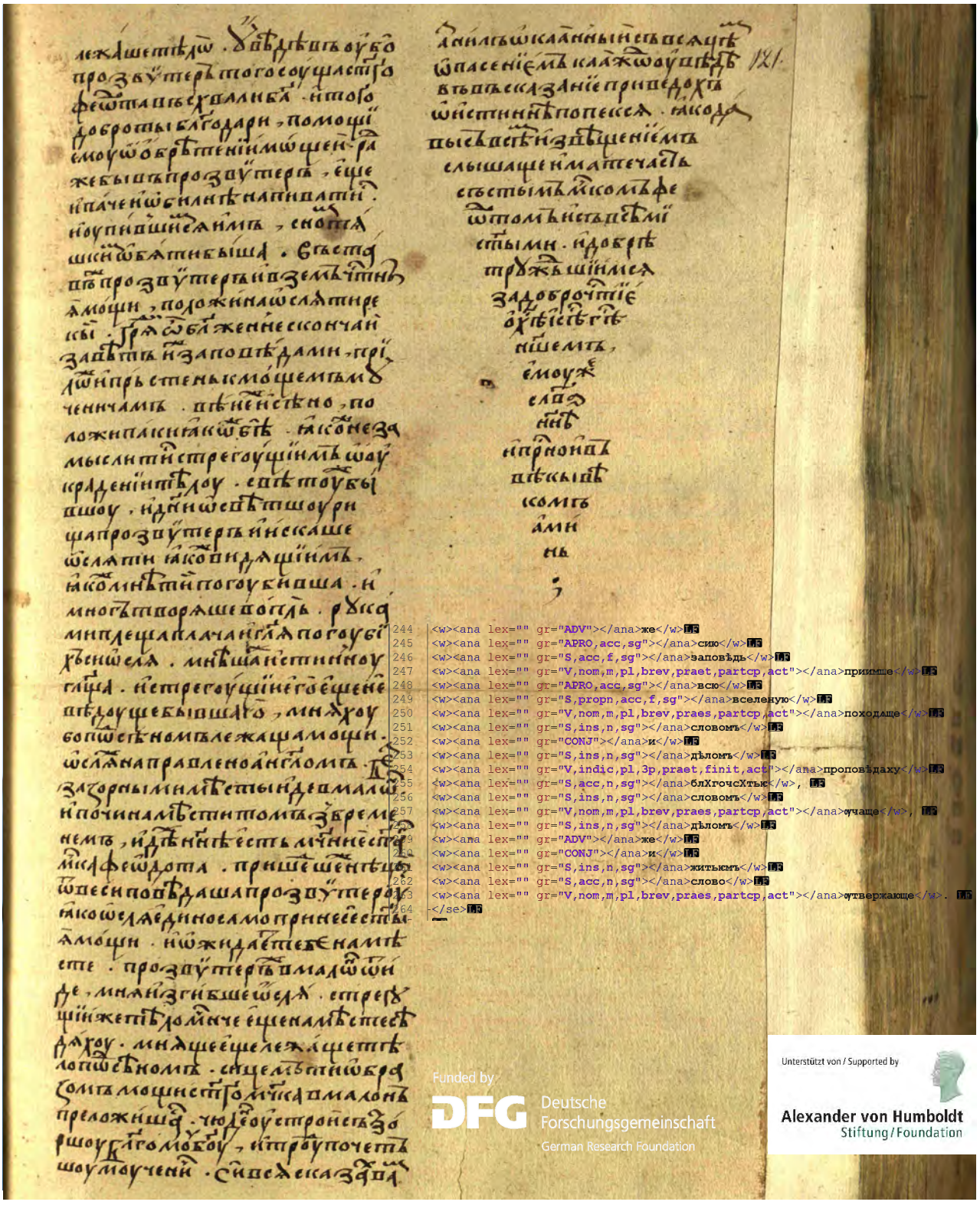
El'Manuscript 2021

8th International Conference on Textual Heritage and Information Technologies

12.04.2021 – 15.04.2021

Albert-Ludwigs-Universität Freiburg

FREIBURG



```
<w><ana lex="" gr="ADV"></ana>же</w>
<w><ana lex="" gr="APRO,acc,sg"></ana>сию</w>
<w><ana lex="" gr="S,acc,f,sg"></ana>заповѣдь</w>
<w><ana lex="" gr="V,nom,m,pl,brev,praet,partcp,act"></ana>приимше</w>
<w><ana lex="" gr="APRO,acc,sg"></ana>всю</w>
<w><ana lex="" gr="S,propn,acc,f,sg"></ana>вселенную</w>
<w><ana lex="" gr="V,nom,m,pl,brev,praes,partcp,act"></ana>походаще</w>
<w><ana lex="" gr="S,ins,n,sg"></ana>словомъ</w>
<w><ana lex="" gr="CONJ"></ana>и</w>
<w><ana lex="" gr="S,ins,n,sg"></ana>дѣломъ</w>
<w><ana lex="" gr="V,indic,pl,3p,praet,finit,act"></ana>проповѣдаху</w>
<w><ana lex="" gr="S,acc,n,sg"></ana>блХгочсХтыѥ</w>,
<w><ana lex="" gr="S,ins,n,sg"></ana>словомъ</w>
<w><ana lex="" gr="V,nom,m,pl,brev,praes,partcp,act"></ana>учаще</w>,
<w><ana lex="" gr="S,ins,n,sg"></ana>дѣломъ</w>
<w><ana lex="" gr="ADV"></ana>же</w>
<w><ana lex="" gr="CONJ"></ana>и</w>
<w><ana lex="" gr="S,ins,n,sg"></ana>житьмъ</w>
<w><ana lex="" gr="S,acc,n,sg"></ana>слово</w>
<w><ana lex="" gr="V,nom,m,pl,brev,praes,partcp,act"></ana>утвержающе</w>.
</se>
```

Funded by DFG Deutsche Forschungsgemeinschaft German Research Foundation

Unterstützt von / Supported by Alexander von Humboldt Stiftung/Foundation

El'Manuscript 2021

Textual Heritage and Information Technologies

8th International Conference

Freiburg im Breisgau, Germany, 12.04.2021 – 15.04.2021

Booklet of Abstracts

V.i.S.d.P.: Achim Rabus

Organizing Committee: Achim Rabus, Victor Baranov, Heinz Miklas, Aleksandr Moldovan

Editors: Juliane Besters-Dilger, Achim Rabus

Sponsors: Deutsche Forschungsgemeinschaft, Alexander von Humboldt-Stiftung, Albert-Ludwigs-Universität Freiburg

Technical redaction: Kathrin Eisel, Melanie Miller

© 2021 Albert-Ludwigs-Universität Freiburg

Contents

Ilia Afanasev
The Old Church Slavonic Corpus Development: Preprocessing Stage 9

Marina Aksenova
Literary Analysis of Texts in the Context of Changed Orthography and Grammar 9

Artem Andreev
How Can Textual Corpora Benefit from Distributed Computing Methods? 9

Alexandre Arkhipov
Using HTR on Bilingual Evenki-Russian Manuscripts of 1910s 10

Irina Azarova | Elena Alekseeva | Alexey Lavrentev | Elena Rogozina | Konstantin Sipunin
Content Structuring in St Petersburg Corpus of Hagiographic Texts (SCAT) 10

Victor A. Baranov | Roman M. Gnutikov
An Experiment in Eliminating Variation of Data of the Slavonic Historical Corpus to Facilitate Search, Demonstration and Statistical Analysis 11

Victor A. Baranov | Oksana V. Zuga
Correlation and Cluster Analysis of Fragments of the Earliest Slavonic Gospels 11

Juliane Besters-Dilger | Achim Rabus
Strengths and Weaknesses of Neural Morphological Tagging for Slavic 12

Elias Bounatirou
When *telefon* Became *brzoglas*: The Making of Fascist Croatian (1941–1945) 12

Simon Brenner | Fabian Hollaus | Patricia Engel | Heinz Miklas | Manfred Schreiner | Wilfried Vetter
Interdisciplinary Analyses of the Codex Marianus, Vienna Part (Cod. Vind. slav. 146) 13

George Aaron Broadwell | Brook Danielle Lillehaugen | Xóchitl Flores-Marcial | Eloise Kadlecek | Felipe H. Lopez | Mike Zarafonetis
Digital Editions of Endangered Language Texts as a Form of Pedagogy, Scholarship, and Resistance 14

Gijsjan Brouwer
Docere: A Generic And Customisable Framework For Publishing and Archiving Digital (Scholarly) Editions 14

Daniel Bunčić
OpenType and the Diversity of Slavic Texts 15

Thomas Daiber
Orthographic doubts: grammatika and grammatikija 15

Chiara De Bastiani | Thomas Krause | Martin Klotz | Marco Coniglio | Heike Sahm | Anabel Recker | Jan Christian Schaffert | Svenja Walkenhorst
Building and Using a Parallel Corpus for Analyzing Translations of High and Low German Incunabula 15

Tsvetana Dimitrova
On Sentence Segmentation in (Historical) Corpora 16

Dmitriy Dobrowolski
Style of Old Russian Sermons in the 11–12th Centuries: A Quantitative Approach 16

Quinn Dombrowski
From Annotation to Modeling: Computational Horizons for Medieval Slavic Studies 16

Nataliia Drozhashchikh | Elena Efimova
Lemmatization for Old English: Building Text Analysis Application 17

Hanne Eckhoff
Aligning the Psalterium Sinaiticum with the Greek Psalter 17

Kazuyuki Enami | Yoshihiro Okada | Taketoshi Hibiya | Satoru Sato | Takashi Yokota | Shigeru Sawayama
Scientific Study of Paper Used for Ukiyoe Pictures Published in the Edo-Era by High-Resolution Digital Microscope 18

Jürgen Fuchsbauer | Fabio Maion | Lara Sels
Towards a Database of Older Slavonic Literature: First Steps Towards the Digitisation and Publication of Francis J. Thomson's Card Index 18

Oksana Gorban | Marina Kosova | Elena Sheptukhina | Andrey Svetlov
Don Cossack Army Official Documents of the 18-19th Centuries: Compiling a Linguistic Corpus 18

Lena Henningsen | Duncan Paterson
Manuscript Culture during the Cultural Revolution 19

Kameliya Hristova-Yordanova
Electronic Descriptions of Slavonic Manuscripts at the Bulgarian National Library "St. St. Cyril and Methodius" – Problems and Perspectives 19

Angelina Kalashnikova
Infrared Digital Scanning of the 15th – the First Half of the 16th Centuries Russian Judicial Documents' Watermarks 20

Sebastian Kempgen
Unicode and Slavic Philology - Status quo and Open Questions 20

Viacheslav Kozak | Anastasia Makarova | Diana Balashevich | Anastasiia Kharlamova | Andrey Sobolev
Linguistic Description of the Glagolitic Fragment from the 14th c. (NLR, Berčić II 36) 20

Kristijan Kuhar
In the Quest for Theological and Liturgical Texts in Glagolitic Miscellanies (15th c.) 21

Natalia Kuzemko
Binding Stamping Research Methodology (Trasological Analysis) 21

Gerhard Lauer
Do Manuscripts Have Style? Digital Stylometry for Manuscript Studies 21

Alexei Lavrentiev | Liubov Kurysheva
A Bilingual Digital Edition of La Belle et la Bête and its Russian Translation by Kh. Demidova 22

Alexei Lavrentiev | Elena Markova
Medea Project: Digital Editing and Analysis of Medieval and Renaissance Medical Texts in France 22

Marija Lazar | Michael Hoffert
Constructing a Time Machine to Surf the Past: Compiling and Individualizing a Historical Legal Glossary 23

Timm Lehmberg | Anne Ferger | Daniel Jettka
Cross-Resource Exploitation of Manuscript and Spoken Language Data 23

Yuan Li | Guanwei Liu | Shoju Ikeda
Issues around Glyph Creation of Undefined Chinese Characters in Kanchi'inbon Ruijumyogisho 25

Olga Lyashevskaya | Dmitri Sitchinava | Maria Skachedubova
Lemmatization of the Middle Russian Corpus within the RNC: Choice of Solutions 25

Olga Lyashevskaya
Universal Dependencies for Pre-Modern Russian: Morphology 26

Ludmila Mazur | Oleg Gorbachev
Soviet All-Union Census of 1959 as a Source of Data for Family Reconstruction Studies: Challenges of Developing and Normalization of a Network Data Model 26

Elena Mikhailova
Manuscript Monuments of the Secular Musical Culture of Peter the Great Times (Late 17th – First Quarter of the 18th Century): A Complex Study and a Digital Presentation Project 27

Alexandra Milanova
Digital History in the Making: Case Study from Bulgaria 27

Anissava Miltenova | Ivan I. Iliev
Computer Corpora and Public Challenges 27

Ekaterina Mishina
The Annotation of Verbal Aspect in Diachrony: Parameters, Algorithms and Problems 28

Alexandr Moldovan
Textual Information in the Linguistic Corpus 28

Klaus Müller | Aleksej Tikhonov | Roland Meyer
Scribe vs. Authorship Clustering in Historic Czech Manuscripts with LiViTo: A Case Study with Visual and Linguistic Features 29

Kirill Nazarenko
Digital Visualization Based on the Analysis of Handwritten Texts: Reconstruction of the Russian Naval Costume of the Early 18th Century 29

Vladimir Neumann
"Deep Mining" of the Collection of Old Prints "Kirchenslavica Digital" 29

Ekaterina Nosevich | Ivan Poliakov | Denis Tsypkin
Modern Methods of Manuscripts Marks Studying: Wax Drops 30

Ekaterina Nosova | Dmitriy Weber
Study of Applied Black Seals from the Collection of Nikolay Likhachev: Preliminary Observations 30

Mariia Novak
Cyril of Jerusalem Catechetical Lectures in the 13th-Century Old-Russian Tolstovskii Sbornik: A Textological Study 31

Matija Ogrin
The Repertorium of Slovenian Early Modern Manuscripts. A Place of Cultural Memory and Textual Research 31

Nilo Pedrazzini
Tackling Lack of Linguistic Data with HTR: A Specialized Model for the Transcription of Serbian Church Slavonic Manuscripts 31

Yana Penkova | Achim Rabus
Indefinite Pronouns in Old East Slavic: A Corpus Approach 32

Tatiana Pentkovskaya
О текстологии печатного перевода Корана 1716 г. 32

Štefan Pilát
GORAZD: The Old Church Slavonic Digital Hub – New Developments 33

Anna Ptentsova
Доумати (dumati), гадати (gadati) and their Aspectual Pairs in the Historical Subcorpus of the Russian National Corpus 33

Alla Polianina
The Problem of Age Classification when Publishing Texts of Historical Sources and Documentary Records 33

Vladimir Polomac
Compiling a Diachronic Corpus of Serbian: Theoretical and Methodological Challenges 34

Olga Semenova | Egor Salnikov | Lidiia Ostyakova
Morphological Tagging for 17th Century Russian 34

Anna Senina
Perm Zemstvo and the Right to Public Opinion in the Newspaper "Permskaya Zemskaya Nedelya" 35

Dmitri Sitchinava | Anton Dyshkant

Integration of the Old East Slavic Epigraphical Databases, Corpora and Indices 35

Daniil Skorinkin

Russian Digital Humanities – a View from Inside 35

Maria Smirnova | Ivan Poliakov

Corpus of Autobiographical Notes in Russian Manuscripts in XVII– XVIII Centuries: Methods of Search and Study 36

Andrey Svetlov | Anatoly Komendantov | Alexander Matveev

On Software Development for a Corpus of Archival Fund Documents 36

Walker R. Thompson

Using Text Recognition Tools to Transcribe Multilingual Lexica 37

Cynthia M. Vakareliyska

Tweaking the Digital Menology Collation 37

Allison Vanouse

Transport Protocols and the "Attacker" in Digital Preservation: the Case of the Très Riches Heures du Duc de Berry 38

Liudmila Varlamova

Standardization of the Long Term Preservation of Digital Documents and their Formats 38

Regina A. Vernyaeva

Collocations with the Component *-ьn(o)* in Russian Chronicles: Quantitative and Statistical Analysis (on the Basis of the Corpus of Russian Chronicles of the Historical Corpus "Manuscript") 38

Cristina Vertan | Walther v. Hahn

Annotation of Vague and Uncertain Places and Events in Historical Texts 39

Xiaojie Xu | Kazuyuki Enami

Analysis of Incunabula Paper Quality: Beginning from "The Travels of Marco Polo" 39

Valentina Yakunina | Elizaveta Popova

Economic, Political, and Sociocultural Communications between Russia and Baltic Region in the 15–17th Centuries based on the Archival Collections of Tallinn, Lübeck, Berlin, Stockholm and St. Petersburg 40

Svetlana Zemicheva | Maxim Gromov

Tomsk Dialect Corpus as a Universal Information Search System 40

Ekaterina Zhdanova

Texts of the Corpus of Russian Dialects of Udmurtia as a Source of Linguistic and Culturological Information 41

Larysa Zherebtsova

The Markup of the Act Documents of the 16th Century from the Lithuanian Metrica Concerning the Customs Systems' History with CEI 41

Oleg Zholobov

Online Publication, Fragmentation and Linguistic-Statistical Study of a Megatext from the 13th Century 42

Svetlana D. Zlivko | Liudmila S. Shatunova

The English-Language Versions of Multicomponent Terms in the Electronic Linguistic Dictionary of M. V. Lomonosov 42

Victor A. Baranov | Oksana V. Zuga

(Kalashnikov Izhevsk State Technical University | Udmurtia State University)

Correlation and Cluster Analysis of Fragments of the Earliest Slavonic Gospels

Keywords: Slavonic manuscripts of Gospels, linguistic textual criticism, correlation, cluster analysis

One of the tasks in the field of textual criticism of the Slavonic Gospel is finding efficient methods of classification of the many manuscripts. The majority of scholarship on this topic uses textologically significant variants (variant readings) of contexts, available both in short and complete aprakoses and the four Gospels, as material for studying the topic.

We present the results of a correlation analysis, followed by a clustering of the fragments of several gospel copies corresponding to each other from the historical corpus "Manuscript" (http://manuscripts.ru/mns/por-tal.main?p1=9&p_lid=1&p_sid=1). Lists of textual forms (bag of words) of the codices parts which match each other textologically (relating to the same cycle of readings written by the same Evangelist) were juxtaposed, and their corre-lation in pairs was evaluated. The results are used to find the degree of closeness (distance) of the fragments of this type between the manuscripts.

The experiments are based on the following theses: that the text of the compiled manuscripts cannot be studied without consideration of their textual composition, that the degree of closeness of the copies should be revealed, among other things, on the basis of linguistic analysis of textually compared texts (fragments) of different codices, and

that the statistical methods can be applied for comparison of the quantitative characteristics of the lists of words extracted from the fragments of different manuscripts matching each other.

The experimental part of the work was supported by the Russian Foundation for Basic Research (RFBR) in the framework of the project "A linguistic statistical analysis of one- component and multicomponent lexical units of the historical corpus Manuscript" (grant No. 18-012-00463).