

eHumanities: Nutzen für die historischen Philologien

Freiburg im Breisgau, 8–10 October 2015*

From 8 to 10 October 2015, the Department of Slavonic Studies at the University of Freiburg hosted the international interdisciplinary conference titled ‘eHumanities: Nutzen für die historischen Philologien’ (‘eHumanities: Benefits for Historical Philologies’). The conference launched the final phase of the project ‘SlaVaComp – COMputer-aided research on VARIability in Church SLAvonic’, sponsored by the Federal Ministry of Education and Research (BMBF) and carried out in cooperation between the Department of Slavonic Studies and the University IT Services (<<http://www.slavacomp.uni-freiburg.de>>, last accessed 10 March 2016). The aim of the conference was to discuss in a broad interdisciplinary exchange recent findings and current research approaches in the realm of Digital Humanities, in particular when applied to Slavic studies.

The hosting SlaVaComp project was introduced by the Project Director Juliane Besters-Dilger. The goals of the project are to establish an extensive bilingual (Church Slavonic-Greek resp. Greek-Church Slavonic) glossary and to create a lemmatizer to return the respective lemma of any valid Church Slavonic word regardless of its specific graphic features. The aim is to make analyzable the lexical and graphic variation of the Church Slavonic written heritage in its regional and chronological development until the sixteenth century. Irina Podtergera and Susanne Mocken reported about how a lemmatizer for Church Slavonic can be accomplished. In particular, Irina Podtergera concentrated in her contribution entitled *From Historical Paper-Lexicography to Historical E-Lexicography* on philological and linguistic aspects of the issue. She highlighted the macro- and microstructure of the eighteen glossaries data mined by the project, underlining how differently the same kind of information may be presented. She brought out the formal and substantive advantages of an electronic dictionary. From a philological point of view, the greatest benefit of the Church Slavonic electronic dictionary is that it facilitates signif-

* This publication appears as a part of the project ‘SlaVaComp—COMputer-aided research on VARIability in Church SLAvonic’, sponsored by the German Federal Ministry of Education and Research (BMBF), funding code 01UG1251, as well as of the habilitation project of Dr. Irina Podtergera in the framework of the Margarete von Wrangell Habilitation Programme supported by the Ministry of Science, Research and the Arts (MWK Baden-Württemberg). An extended analytical report of the conference shall appear in *Studi Slavistici*.

icant investigations of history of lexemes and concepts and helps distinguish more exactly the mediaeval schools of translation according to lexical properties of the translated texts. Susanne Mocken subsequently dealt with issues of markup in her paper *How Can Diversity Be Unified*. She gave a description of the XML-structure of all encoded glossaries and showed how the preliminary version of the Church Slavonic-Greek MetaGlossary works. Simon Skilevic reported on the tool for converting non-Unicode files into Unicode format he developed as a student assistant in the SlaVaComp project. Evgenii Filimonov spoke of *Greek-Slavonic Asymmetries in Syntax and Lexis*. Dealing with discrepancies between Church Slavonic translation and the Greek original is an urgent problem. Lexical asymmetries may include free or vague translation of the original term and multiword expressions for one-word equivalents and vice versa in the source and target language.

A number of papers were dedicated to theoretical issues of digital humanities. The keynote speaker, Manfred Thaller (University of Cologne), spoke on the application of computational technologies to philological studies, discussing the modern concepts of ‘Digital Humanities’, ‘Big Data’ and ‘big’ and ‘small’ academic disciplines (‘große’ und ‘kleine’ Fächer) and ‘big’ and ‘small’ philologies respectively. He defined ‘Digital Humanities’ as an intellectual agenda that seeks to achieve substantive results which are unavailable or unverifiable otherwise. The emphasis is placed on the analytical value of the digital methods in the Humanities and not only on the sharing of results. Discussing the concept ‘Big Data’, Manfred Thaller formulated criteria for two paradigms, of the ‘big’ and of the ‘small’ philologies. Ralph Cleminson (Winchester) provided some theoretical and practical reflections on the topic of *Encoding Text and Encoding Texts*. He emphasized that an encoded text as a digital edition must accommodate the cultural function of the text to be encoded—both the ‘ideal’ function and its particular realisations in manuscripts.¹ In his talk on *Internet Lexicography and the Lexicon Dynamics* Stefan Engelberg (University of Mannheim) focused on a gap between everyday language usage and our current state of knowledge in linguistics: because of the strong dynamics of the lexicon, traditional lexicography finds it difficult to record word usage entirely. Corpus-based studies show that only one per cent of the contemporary inventory of lexemes is documented in paper dictionaries.

Advances in computer linguistics were a core topic of the conference. Alexander Mehler (Goethe University Frankfurt) introduced *Wikidition*, a new text technology that allows automatic lexiconization, i. e. lemmatization and grammatical analysis of each syntactical word, and cross-linking of

1 See his paper in this Bulletin issue [red.].

text corpora (cf. <<http://capitwiki.hucompute.org/>>, last accessed 10 March 2016). The faceted search, which enables researchers to browse the information space by playing audio files, surveying the location where it was recorded, differentiation by speakers' sex and age, etc. and results in new geotemporal and interpretive contexts, was the focus of the lecture by Thomas Efer (University of Leipzig), *Use of Graph Databases in the Analysis of historical corpora*. Setting the tone by pointing to the limitation of the text processing with XML as a simple hierarchy of elements, he brought to attention the benefits of graph databases for text technology, which can cover many parallel hierarchies, using the example of the Leipzig historical project 'eXChange'. Stylianos Chronopoulos (University of Freiburg) presented his ongoing research project on *Pollux' WordNet*, concerning a digital edition of a famous Greek thesaurus from the second century AD. The thesaurus consists of ten books and contains ca. 120,000 words which are pooled in hierarchically-structured semantic fields organised according to subject-matter. Lists of words are embedded in a continuous text, so that the microstructure of the semantic field depends on syntax of this text.

Specifically in the field of Slavonic lexicography, Lora Taseva (Bulgarian Academy of Science, Institute of Balkan Studies in Sofia) spoke on *Multiple Translations as a Research Object of Philological Mediaeval Studies and Challenge for Computational Linguistics*. She showed how lexical factors play a key role for the dating and localisation of translated texts as well as for the description of translation techniques. An accurate dating and exact location have to be tackled only by means of statistical analyses of 'big data'. Roland Meyer (Humboldt University Berlin) illustrated the application of specific computational linguistic methods to the study of Slavonic languages. He evaluated the *Data Driving Identification of Registers in the Historical Texts* by a synchronous and diachronic comparison of the relative pronouns in Polish, Czech, and Russian. Aleksandr Moldovan (Russian Academy of Sciences, Moscow) spoke on *Essentials of Language Documentation* by focusing upon the old Cyrillic written heritage. He discussed complications in the encoding of old Cyrillic texts emphasising linguistic relevance of graphic and orthographic distinctive features as well as of regional and historical variations of writing and grammatical systems. Achim Rabus (Friedrich Schiller University, Jena), in his paper on *Multiple Use of Data and Code*, focused on two recent Slavonic dia- and synchronic corpus-linguistics projects, in which he was involved or which he initiated. He took the Freiburg diachronic VMČ corpus as a starting point in order to argue that graphical user interfaces, data, and codes can be recycled and subsequently performed, as applied for the project 'Rusyn Language as a Minority Language across National Boundaries:

Dynamic Processes'. Christine Grillborzer (University of Freiburg), signalled the difficulties faced by linguists by searching for clauses with zero dative subject in the Russian National Corpus (RNC) and comparing them to the clauses with a nominative subject in her paper *Annotation of Zeros*. The main emphasis of the lecture by Toma Tasovac (Belgrad Center of Digital Humanities), *The Devil is in the Detail: From Data Modelling to Data Enrichment in Legacy Dictionaries*, was on how historical dictionaries function nowadays not as reference works for the contemporary language usage but as research objects. The top challenge for eLexicography must be to incorporate the available electronic editions into an efficient research environment for the exploration of historical semantics, as attempted by the 'Plattform for the Transcription and Digital Editions of the Serbian Manuscript' (<<http://prepis.org>>, last accessed 10 March 2016). Tasovac shortly reported on his experience with encoding of Vuk Karadžić's 'Lexicon Serbico-Germanico-Latinum' (1818, 1852), and with digitising some 23,000 lexicographic paper slips compiled by Serbian amateur lexicographer Dimitrije Čemerikić (1882–1960).

Digital philology as text editing was the focus of the report of David J. Birnbaum (University of Pittsburgh) about his collaborative work with Hanne M. Eckhoff (University of Tromsø) on the digital edition of the Codex Suprasliensis. His paper was devoted to the *Machine-Assisted Normalization* of the encoded Old Church Slavonic manuscript text. At present, the electronic edition of the Codex Suprasliensis is supplied with diplomatic transcriptions of all Slavonic texts, parallel Greek correspondences, and high-quality facsimile of the manuscript (cf. <<http://suprasliensis.obdurodon.org/>>, last accessed 10 March 2016), but we still lack a normalized reading view of it. Birnbaum and Eckhoff have developed a machine-assisted method to convert a diplomatic edition of the manuscript into normalized canonic Old Church Slavonic. Anissava Miltenova (Bulgarian Academy of Science in Sofia) presented a talk on *Rethinking Old Church Slavonic Digital Library by Ontologies* giving insights into the project 'Scripta Bulgarica'. This innovative project aims at collecting data concerning mediaeval Bulgarian written heritage and providing models and samples for the presentation of metadata, terminological articles, and articles on Byzantine writers, etc. The integrated thesaurus contains terms and concepts in Palaeoslavistics in eight languages. The text resources and metadata are extracted from already existing databases and corpora, for instance from the 'Repertorium of Old Bulgarian Literature and Letters' (cf. <<http://repertorium.obdurodon.org/>>, last accessed 10 March 2016), the aforementioned digital edition of the Codex Suprasliensis, electronic collection of Bulgarian manuscripts, etc. Jürgen Fuchsbauer (University of Regensburg) dealt with *Paralleling Different Versions of Slavic Texts*.

Using the example of Church Slavonic and Balkan Slavic Lives of Paraskeva of Epibatai (Petka Tarnovska), he raised the question of how several versions of one text should be aligned within one digital edition and what preliminary work would be necessary for this. He gave an overview of the whole corpus, from the original Church Slavonic text situated in the Middle Bulgarian 'Miscellany of German' (Germanov Sbornik, 1358/59), through shortened and extended Church Slavonic redactions of the text, which had been composed by Patriarch Euthymius of Tarnovo between 1376 and 1382, up to Russian Church Slavonic redactions and Bulgarian vernacular versions from seventeenth, eighteenth, and nineteenth century. In order to achieve comparability between all versions, thematic and text units must be linked to each other, possibly including the predication level, a challenging task when dealing with such highly complicated tradition.

This conference stressed how Slavistics and especially Palaeoslavistics benefits from Digital Humanities. Not only are computer-assisted methods of great importance because they offer new perspectives for analysing written heritage, but also the eHumanities per se because they stimulate interdisciplinary networking and the exchange of knowledge between representatives of different disciplines and different scientific cultures. One of the results of this networking is a unification of research instruments and tools which leads to the elimination of the boundaries between 'big' and 'small' philologies. Conference abstracts can be downloaded from <<http://www.slavacomp.uni-freiburg.de/konferenz.html>> (last accessed 10 March 2016).

Irina Podtergera
Department of Slavonic Studies, University of Freiburg