Pázmány Péter Catholic University

Faculty of Humanities and Social Sciences

Digital Philology in Hungary: Principles and Practices

Doctoral (PhD) Dissertation: Thesis Booklet

Zsófia Fellegi

Doctoral School of Literary Studies

Thesis Advisor: Dr. Zsolt Almási, Head of Department, Associate Professor

Head of the Doctoral School: Dr. István Dobos, Head of Department, Professor

Budapest

2024

1. Research background and problem statement

The transition to digital media impacts all current cultural practices, including science, humanities, literary studies, and philology. The highest quality philological editions of national cultural heritage, the critical editions, have always shaped what becomes the subject of research by producing controlled, 'stable' text corpora for humanities research. Over recent decades, the development of computing and the internet has brought changes in scientific methodologies. In the fields of literary studies and philology, there has been a demand for developing a research infrastructure that enables research networks to extend across physical boundaries.

This dissertation does not attempt to write the history of Hungarian digital philology, a task that could alone be the subject of a comprehensive monograph. During the research, it became clear that the genesis of Hungarian digital philology workshops and projects is only partially documented, with much key information unavailable, exceeding the scope of this dissertation. The dissertation also does not provide a systematic theoretical overview of digital philology, a task that would be even more challenging to tackle in monograph form due to the rapid changes in technology. However, the dissertation touches on certain chapters of Hungarian digital philology's history and addresses theoretical aspects within the context of the DigiPhil (Portal of the Philology of Hungarian Literary Studies) practice.[1] The primary aim of this dissertation is to aid in the overview of practical and theoretical problems arising in the field of Hungarian philology by highlighting good practices. This is driven by the effort to convey the experiences gathered around DigiPhil over the past decade to other philological workshops and particularly to future digital philologists. Thus, the dissertation endeavours to provide a problem-oriented overview of the genesis and current practice of DigiPhil. To achieve this, it is essential to provide a brief introduction to the Hungarian and international context preceding and surrounding the formation of DigiPhil, as well as to discuss theoretical questions, debates, and current discourses that have directly influenced the micro- and macro-decisions of DigiPhil's research methodology. The dissertation covers both DigiPhil 1.0 and 2.0 practices and the theoretical contexts that played a role in their development.

---

[1] Website of the DigiPhil service: https://digiphil.hu/.

2. Research methodology and structure of the thesis

The introductory chapter of the dissertation is divided into two parts. The first part introduces the basic concepts on which the argument relies:

digital philology (and digital humanities)

- digital and electronic (less often networked) edition
- informatized edition
- born-digital edition
- markup language transcript
- database, data structure

Although the aim of the dissertation was not to write the history of Hungarian digital philology, to place the argument of the subsequent chapters (primarily built around examples from the editions of DigiPhil starting in 2014) in a historical context, a brief and selective overview of the most significant developments in digital philology over the past two decades is provided in the second part of the introductory chapter.

The first chapter of the dissertation (*Transition from Analog to Digital Philology*) deals with the meta-philology of the transition from analogue to digital philology (considering theoretical aspects) but within the context of specific text editions published with the professional support of the DigiPhil group. The presented digital scholarly and critical editions are diverse; the chapter strives to provide an overview of the most important questions of digital philology by listing and analyzing different philological schools and text corpora. First, we see examples of the informatization of printed editions, focusing on the difference between digitization and informatization. This part illustrates the informatization through retroconversion of critical editions of János Arany's works and the *Synoptic Critical Edition of Imre Madách's The Tragedy of Man*, showing how different editorial principles of different edition types can be represented in one system. This is followed by the analysis of born-digital editions. For born-digital editions, the impact of digital philology on the principles of text publication and the changing role of the philologist cannot be ignored. The analysis examines several edition types. In the first case (*Critical Edition of Zsigmond Móricz's Correspondence*), the born-digital edition was not part of a series, so the editorial principles did not have to conform to a series' guidelines. In the second case (*Critical Edition of Miklós Oláh's Correspondence*, second volume), the born-digital edition had a precedent in the first volume,

which was created with dedicated software allowing the editor to complete the entire process, from transcription to incorporating critical apparatus to producing the indexed, formatted print-ready version, within one system. Although the dedicated software used for the first edition allowed for a standard markup language transcript, the necessary settings could not be made without digital philological skills. The second edition was produced in a user-friendly environment imitating office software tailored to the edition's needs, resulting in a markup language transcript without the editor needing to know the TEI standard or XML language. The edition is part of the *Bibliotheca Scriptorum Medii Recentisque Aevorum* series, so the creation of the markup language transcript had to consider both the series' and DigiPhil's text dissemination principles. The last presented enterprise, the *Complete Works of Mihály Babits* series, is unique in Hungary and likely in Europe. The XML representation of the volumes following the principles of genetic criticism is extremely complex due to the nature of the material and would be impossible to produce without the latest technologies that emerged with the spread of the semantic web approach. The foundation of the concept of the semantic web, or Web 3.0, is based on the phenomenon that with the spread of Web 1.0 and 2.0, an unmanageable amount of unorganized data and documents became available on the internet by the early 2000s. According to the concept, the problem can be solved with intelligent machine solutions, but for this, the data must be available in a format that machines can interpret.

The chapter on *Digital Philology and the Semantic Web* examines the impact of these technologies and the semantic web approach on digital philology. This impact is evident in both the theory and practice of digital philology. The dissertation demonstrates the principle and operation of the semantic web through the analysis of international experiments and new theories. From the beginning, the DigiPhil project connected digital editions with a background database containing bibliographic and biographical data, an integrated library system which entailed many compromises. By applying semantic web technology, a graph-shaped knowledge base can be created that not only manages data stored in the library system but also organizes research data into a semantic network and draws further conclusions from the data.

Beyond Digital Editions: Analysis of Markup Language Transcripts examines the tools available for analyzing the XML representations of critical editions. For the successful research of digitally represented literary texts, it is necessary to manage metadata and texts together and to process the text from multiple perspectives. By using international standards and ontologies, meaningful and sophisticated searches can be conducted across physically distant databases. The first part of the chapter presents the process of metadata aggregation considering current

web theories. The second part deals with the possibilities of network visualizations of markup language transcripts. The tools available for analysis are demonstrated through the informatized edition of János Arany's Correspondence 1866–1882, specifically processing the metadata of the letters, and comparing various visualization tools for the metadata. Since the launch of DigiPhil, a corpus of nearly 10,000 XML files has been built, mainly consisting of works by 19th- and 20th-century authors. The size of the corpus and the temporal proximity of the authors provide the opportunity to perform linguistic statistical analyses on the available critical editions. The final part of the chapter demonstrates some of the available methods through the examination of the critical editions of Dezső Kosztolányi's works.

The rise of digital philology in recent decades is undeniable; however, despite decades of experience, the education of digital philology is still underdeveloped in Hungary, with no comprehensive training available for future experts. This is a particularly timely issue, considering the growing number of research groups dealing with digital or digitized editions, but current Hungarian education does not ensure the "succession" of experts. The concluding chapter of the dissertation reflects on this problem, presenting the opportunities available in recent years and today for teaching digital philology and what aspects should be considered when developing a comprehensive training program. The suggested aspects apply to both undergraduate and postgraduate education.

3. Results and conclusion

One of the main results of the dissertation is the delineation of the terminology used in the field of digital philology, defining the concepts used in the discipline. The dissertation aimed to provide a problem-oriented overview of the genesis and current practice of DigiPhil, and in the process, it managed to systematize previously scattered knowledge and experiences in digital philology. The dissertation provides an overview of the theoretical considerations behind DigiPhil's decisions and offers a broad view of the borderline areas of digital philology, including the possibilities of computer analysis of markup languages. The dissertation reflects on the theoretical and methodological questions arising in the field of digital philology. Currently, there is no comprehensive digital philology education available in Hungary; the dissertation, considering previous good practices, makes suggestions for developing short- and long-term training program.

4. Publications related to the thesis

DOBÁS Kata, FELLEGI Zsófia, PALKÓ Gábor. „A kis gömböc meséje – az ITIdata irodalomtudományos adatbázis fejlesztése 2022–2023-ban". In *Új technológiákkal, új tartalmakkal a jövő digitális transzformációja felé: 32. Networkshop: országos konferencia: 2023. április 12-14. Pannon Egyetem, Veszprém*, szerkesztette TICK József, KOKAS Károly és HOLL András, 192–198. Budapest: Hungarnet Egyesület, 2023. https://doi.org/10.31915/NWS.2023.28

FELLEGI Zsófia. „WIKIDATA – ELTEdata – WIKIMEDIA: A szövegkiadások technikai háttere". In *A tudományos tudásáramlás mintázatai a Magyar Királyságban, 1770-1830: DigiTudásáramlás (DITU) /DigiCirculation of Knowledge (DICIKO)*, szerkesztette BALOGH Piroska, BODNÁR-KIRÁLY Tibor, GURKA Dezső, KRÁSZ Lilla, KURUCZ György, Sz. KRISTÓF Ildikó, Budapest: ELTE BTK Digitális Bölcsészet Tanszék, 2022. https://doi.org/10.5281/zenodo.7584606

SZEMES Botond, BAJZÁT Tímea, FELLEGI Zsófia, KUNDRÁTH Péter, HORVÁTH Péter, INDIG Balázs, DIÓSSY Anna, és mtsai. „Az ELTE Drámakorpuszának létrehozása és lehetőségei". In *Valós térben - Az online térért: Networkshop 31: országos konferencia. 2022. április 20–22. Debreceni Egyetem*, szerkesztette TICK József, KOKAS Károly és HOLL András, 170–178. Budapest: Hungarnet Egyesület, 2022. https://doi.org/10.31915/NWS.2022.21.

FELLEGI Zsófia. „A digitális filológia infrastruktúrái. A DigiPhil megújulásáról". In *Valós térben - Az online térért: Networkshop 31: országos konferencia. 2022. április 20–22. Debreceni Egyetem*, szerkesztette TICK József, KOKAS Károly és HOLL András, 338–344. Budapest: Hungarnet Egyesület, 2022. https://doi.org/10.31915/NWS.2022.43.

FELLEGI Zsófia. „Átmenet az analógból a digitális filológiába: médiumváltás? Madách Imre: Az ember tragédiája. Szinoptikus kritikai kiadás". *Helikon* 67, 1. sz. (2021): 84–102.

FELLEGI Zsófia. „A digitális filológia oktatása: Gyakorlati tapasztalatok". In *Digitális eszközök a középiskolai irodalomoktatásban*, szerkesztette MOLNÁR Gábor Tamás és FEJES Richárd, 157–178. Budapest: Prae.hu, 2021.

FELLEGI Zsófia és PALKÓ Gábor. „Arany-kéziratok és kritikai kiadások közzététele az Arany János Emlékévben". *Helikon*, 66. sz. (2020): 82–98.

FELLEGI Zsófia. „Digital Philology on the Semantic Web. Publishing Hungarian Avant-Garde Magazines". In *The (Web)Sites of Memory*, edited by Donald E. MORSE, Zsófia RÉTI és Miklós TAKÁCS, 105–116. Loci Memoriae Hungaricae ; 8. Debrecen: Debrecen University Press, 2018.

FELLEGI Zsófia és VADÁSZ Noémi. „Digitálisbölcsészet-oktatás Magyarországon". In *Digitális (szöveg)kultúrák a bölcsészképzésben*, szerkesztette L. VARGA Péter, MOLNÁR Gábor Tamás és PALKÓ Gábor, 168–180. Budapest: ELTE-BTK, 2018.

CSÉVE Anna, FELLEGI Zsófia és KÓMÁR Éva. „Móricz Zsigmond levelezésének (1892–1913) digitális kritikai kiadása". *Digitális Bölcsészet* 1 (2018. szeptember 19.): 159–174. https://doi.org/10.31400/dh-hun.2018.1.227.