

The background of the entire page is a complex, abstract geometric pattern in shades of blue. It consists of numerous interconnected points (nodes) and lines (edges), forming a network-like structure. Some areas are more densely connected, while others are more sparse. The lines and nodes vary in brightness, creating a sense of depth and movement. The overall effect is a futuristic, digital, or network-themed aesthetic.

AZ OKTATÁS, A KUTATÁS ÉS A KÖZGYŰJTEMÉNYEK DIGITÁLIS TRANSZFORMÁCIÓJA FELSŐFOKON

NETWORKSHOP 2024
33. Országos Informatikai Konferencia

2024. április 3–5.
Eszterházy Károly Katolikus Egyetem, Eger

AZ OKTATÁS, A KUTATÁS ÉS A KÖZGYŰJTEMÉNYEK DIGITÁLIS TRANSZFORMÁCIÓJA FELSŐFOKON

NETWORKSHOP 2024
33. Országos Informatikai Konferencia

2024. április 3–5.
Eszterházy Károly Katolikus Egyetem, Eger

Szerkesztette: Tick József, Kokas Károly, Holl András

HUNGARNET Egyesület
Budapest, 2024



HUN-REN
Magyar Kutatási Hálózat

NETWORKSHOP

Szerkesztette: Tick József, Kokas Károly, Holl András

Tipográfia és tördelés: Vas Viktória

Korrektúra: Danyi Melinda

Angol nyelvi lektor: Cseresnyés Dorottya

Networkshop 2024 konferencia előadásainak közleményei

Eszterházy Károly Egyetem, Eger

2024. április 3–5.

ISBN 978-615-82243-2-1

DOI: <https://doi.org/10.31915/NWS.2024>

Kiadja a HUNGARNET Egyesület
az MTA Könyvtár és Információs Központ közreműködésével

Budapest

2024

Borítókép: [freepik.com](https://www.freepik.com)

TARTALOMJEGYZÉK

Előszó	5
Ungváry Rudolf	
A MARC21 formátum kettős szerkezete és a formátum felhasználói szintjének fordításai	7
Holl András, Andódy Katalin	
Adatbányászati gyakorlatok repozitóriumra és MTMT-re.....	16
Simon András	
Mesterséges intelligenciával támogatott adatgazdagítás a Nemzeti Levéltárban.....	22
Soós Gábor, Rövid András, Ormos Pál	
V2X – A járművek közötti kommunikáció kihívásai	29
Csernai Zoltán	
Egy online tanulást támogató portál kurzusának vizsgálata Big Data adatelemző módszerekkel.....	36
T. Nagy László, Németh Áron	
A mesterséges intelligencia (MI) teológiai kompetenciái	45
Mészáros Erika	
Kodaktól a jövőig – Egy könyvtári digitalizálás szinterei.....	55
Frankó Máté, Sándor Ákos	
Adatvizualizáció a könyvtári menedzsmentben: fejlesztések az SZTE Klebelsberg Könyvtár döntéstámogató rendszerében.....	63
Hernek István	
A felhasználóképzés szintjei az SZTE Klebelsberg Könyvtárban: az elsőévesektől a kutatókig	72
Némethi-Takács Margit, Borbély Mária	
Bibliográfiai kapcsolatok az általános megjegyzés adatmezőben.....	78
Dobás Kata, Tüskés Anna	
A magyar irodalomtörténet bibliográfiájának migrációja az ITIdata szemantikus adatbázisba	87
Horváth Péter	
A kanonikus magyar költészet versformakeresője.....	96
Sebestyén Ádám, Sárközi-Lindner Zsófia	
Történeti források szemantikus feldolgozása – Az ELTEdata adatbázis új gyűjteményei	105

Bolya Mátyás

Lyukkártya és népdalrendezés – Egy mechanikus népzenei adatbázis digitális rekonstrukciójának lehetőségei.....112

Kovácsházy Tamás

Az idő, mint alapvető infrastruktúra, az idő szerepe az adatközpontban.....121

Albert Ágota Katalin

A mesterséges intelligencia használatának követelményei az oktatási szektorban, különös tekintettel a mesterséges intelligencia használatáról szóló rendelethez.....129

Varga Emese

Digitális szövegszerkesztés a dHUpa keretrendszerében135

Nemoda Zsuzsanna, Héjja Balázs, Nagy Andor, Tóth Máté

A Pest Megyei Digitális Könyvtár fejlesztése 141

Nagy Dóra, Sándor Ákos

Voice2text: a hanganyagátírás lehetőségei MI segítségével.....149

Kalcsó Gyula

Képek és metaadataik gyűjteményezése scrapingtechnológiával közösségi képmegosztó oldalról157

Péter Róbert, Szántó Zsolt, Biacsi Zoltán, Kocsis Zoltán, Berend Gábor, Bilicki Vilmos

Az AVOBMAT (Analysis and Visualization of Bibliographic Metadata and Texts) többnyelvű kutatási eszköz munkafolyamata és új funkciói163

Máray Tamás

Kvantum-számítástechnika: ez már a „jövő”?171

Fellegi Zsófia

Digitális kiadások migrációja: gépi és emberi intelligencia együttműködése.....177

Palkó Gábor

Posztmodern intertextualitás és digitális szövegkiadás184

Antal Dániel

A szlovák adatkicserélési tér magyarországi föderációjának lehetőségei.....192

Vass Johanna

Kutatási adatok megosztása a gyakorlatban – Adatrepozitóriumok használata az Ökológiai Kutatóközpont publikációiban199

Mihály Eszter, Micsik András, Nagy Kadosa

Irodalmi levélváltások nyomában TEI-vel és térképpel.....208

Posztmodern intertextualitás és digitális szövegkiadás

Postmodern intertextuality and digital text publishing

Palkó Gábor

ELTE Digitális Bölcsészeti Tanszék, Digitális Örökség Nemzeti Laboratórium

palko.gabor@btk.elte.hu

ORCID: [0000-0002-4394-8577](https://orcid.org/0000-0002-4394-8577)

Abstract

The radical intertextuality in postmodern prose can also be interpreted as an attack on the traditional concept of the modernist authorial identity. How can the identity of the implied author (Wayne C. Booth) be constructed when a work of literary art is a collage of both marked and unmarked quotations from heterogeneous sources and genres? Péter Esterházy, arguably the most influential Hungarian prose writer of the second half of the 20th century, questioned the unity of authority and identity of the writer in both his literary works and essays. But how does this subversive rhetoric impact the practice of critical text editing? This question is particularly timely, as we are currently in the second year of a three-year NKFIH project aimed at producing a genetically annotated edition of Esterházy's first and perhaps most important novel, in both print and digital formats. In this brief study, I will address the context and status of the edition, with a particular focus on how to manage the dense network of intertextual connections within and between texts, a network that cannot be uncovered through traditional "close reading" methodologies. Practices of „distant reading” (Franco Moretti), or more precisely, computational literary analysis, may present a viable alternative, yet integrating these methods into a philological workflow – especially in a born-digital format—is far from straightforward.

Keywords: digital scholarly editing, genetic criticism, intertextuality, text similarity, Text Encoding Initiative

A radikális intertextualitás a posztmodern prózában a modernista írói identitás hagyományos fogalma elleni támadásként is értelmezhető.¹ Hogyan építhető fel az implicit szerző (Wayne C. Booth) identitása, amikor a szövegművészet egy alkotása heterogén forrásokból és műfajokból származó jelölt és jelöletlen idézetek kollázsából áll? Esterházy Péter, a 20. század második felének talán legnagyobb hatású magyar prózaírója, irodalmi munkáiban és esszéiben egyaránt az írói autoritás és identitás egységét kérdőjelezte meg. De hogyan hat ez a szubverzív retorika a kritikai szövegszerkesztés gyakorlatára? A kérdés annál is időszerűbb, mivel jelenleg egy hároméves NKFIH-projekt második évében tartunk,² amelynek célja Esterházy első, és talán

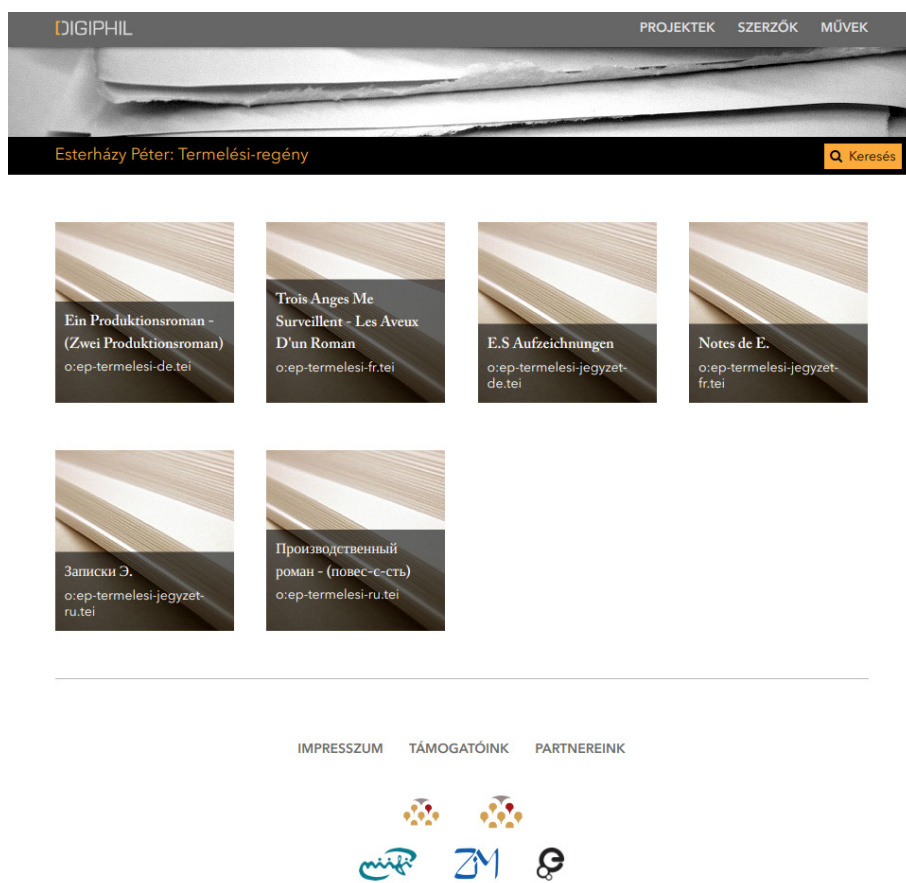
1 Az intertextualitás fogalmához l. általában: Mary Orr, „Intertextuality”, in *The Encyclopedia of Literary and Cultural Theory, Volume II. Literary Theory from 1966 to the Present: A-Z*, Wiley, 2010, <https://doi.org/10.1002/9781444337839>. A témáról magyarul l.: Kulcsár-Szabó Zoltán, „Intertextualitás: létmód és/vagy funkció?”, *Irodalomtörténet*, 1995/4, 495–541. Esterházy Péter és a posztmodern intertextualitás forgalmáról l.: Wernitzer Julianna: *Idézetvilág avagy Esterházy Péter, a Don Quijote szerzője*, Jelenkor-Szépirodalmi, Pécs-Budapest, 1994.

2 A *Filológia, szövegköziség, értelmezés (Esterházy Péter Termelési-regényének kritikai kiadása és újraolvasása)* című projekt az NKFIH K 143536 számú támogatás keretében valósul meg. A projektben résztvevő kutatók: Bónus Tibor (vezetőkutató), Fellegi Zsófia, Kulcsár, Szabó Ernő, Lőrincz Csongor, Molnár Gábor

legfontosabb regényének genetikus annotált kiadása nyomtatott és digitális formában. E rövid tanulmányban kitérek a kiadás kontextusára és státuszára, kiemelt hangsúly kerül arra a kérdésre, hogy hogyan lehet kezelni a szövegen belüli és a szövegek közötti intertextuális kapcsolatok sűrű hálózatát, amelyet a hagyományos, 'közeli olvasás' módszertanával még feltárni sem lehetséges. A 'távoli olvasás' (Franco Moretti), vagy, pontosabban, a számítógépes irodalomvizsgálat gyakorlatai kézenfekvő alternatívát jelenthetnek, ezek integrálása egy filológiai munkamenetbe, legyen ez akár *born digital*, koránt sem magától értetődő.

A digitális kiadás előzményei

A regény digitális kritikai kiadásának első produktuma 2017-ben jelent meg a DigiPhil szolgáltatás keretei között. Ez a kiadás a regény nyomtatott szövege mellett a német, francia és orosz fordításokat is közzétette, az átiratok a Text Encoding Initiative XML ajánlásának megfelelő jelölőnyelvi kódolás formájában készültek el.



1. kép: A Termelési-regény fordításai a DigiPhil szolgáltatás weboldalán
(<https://digiphil.hu/context:ep-tr>)

Tamás, Palkó Gábor, Szirák Péter. Jelen digitális kiadás technikai előkészítésében részt vesz: Palkó Gábor (szöveg hasonlósági vizsgálatok, TEI-specifikáció), Fellegi Zsófia (TEI-specifikáció), Kolozsi Blanka (Esterházy-bibliográfia, TEI-rögzítés), Kiss Júlia (Esterházy-bibliográfia, TEI-rögzítés), Szarka Szilvia (Esterházy-bibliográfia).

A regény magyar szövegének kódolása során szembesültünk egy nem elhanyagolható filológiai problémával. Esterházy Péter tagja a Digitális Irodalmi Akadémiának, műveit a DIA weboldalán lehet olvasni. Azonban az író életében megjelent nyomtatott változatok összehasonlítása során Balajthy Ágnes irodalomtörténész megállapította, hogy a DIA elektronikus kiadása egyik nyomtatott változatnak sem felel meg, ami azt jelenti, hogy az elektronikus változatot önálló szövegnek kell tekinteni. Bár a nyomtatott szövegek közötti különbségek feltárása, valamint a DIA szövegváltozatának összevetése megtörtént, egy teljes értékű tudományos kiadás nem képzelhető el további munkafázisok nélkül. Az első az annotálás, melynek fontos része a regény értelmezésében kulcsfontosságú intertextuális hálózat legalább részleges feltárása. A másik a szöveg keletkezésének nyomon követése, a genetikus folyamat leképezése. Esterházy munkásságában az intertextualitás szerepét már röviden érintettük a bevezetőben. Most vessünk egy pillantást a *Termelési-regény* szöveg-genezisének problémájára.

A szöveg genezise, a genezis reflexiója

Ahogy a műalkotás létrehozásának folyamatára való reflektálás a regény értelmezésének alapvető eleme, úgy a kritikai szövegkiadás előállításában is alapvető feladat a genetikai folyamat feltérképezése. Bár 2017-ben, amikor a tudományos kiadás előkészítése megkezdődött, sokat tudtunk Esterházy kijelentéseiből és műveiben színre vitt írási folyamatról, a genetikai folyamat feltárását is lehetővé tevő filológiai munka valódi lehetősége csak akkor adódott, amikor Esterházy kéziratok hagyatéka hozzáférhetővé vált. A korábban említett NKFIH-támogatás lehetőséget adott arra, hogy a regényhez tartozó kéziratokat a Berliini Művészeti Akadémiával együttműködve digitalizáljuk. A munkát azonban akadályozza, hogy a hagyatékot még nem dolgozták fel egészében, azaz nem minden részét készítették elő a kutatásra, még nem férünk hozzá sem az analóg levelekhez, sem az e-mailekhez, amelyek kulcsfontosságúak a genetikai folyamat és az intertextuális hálózat szempontjából, és bár a *Termelési-regény*hez rendelt kéziratok digitalizálása megtörtént, több mint 1000 oldal terjedelemben, a későbbi prózai művek egy részének dossziéi még nem érhetőek el. Ez utóbbi különös nehézség, mert a regényben emlegetett, és az írói módszer szempontjából oly fontos cetlik, amelyek a regény szerkezetének vázlatait és más szövegfragmentumokat tartalmazhatnak, csak nagyon kis részben találhatók meg a *Termelési-regény* kéziratdossziéiban. Ahogy egyre több anyaghoz férünk hozzá, az a reményünk, hogy a cetlik más regények anyagaiban bukkannak elő, szertefoszlani látszik.

A csekély számú cetlin és egyéb szálas anyagon túl a mappák Esterházy kézzel írott füzetait tartalmazzák. Ezen kéziratokat nagyjából három csoportra lehet osztani. Időben a legközelebb a nyomtatott kiadáshoz a tisztázatok állnak, ezek csak kis mértékben térnek el a végleges megjelent szövegtől. A szöveg könnyű olvashatósága, az egyértelmű javítások arra utalnak, hogy Esterházy egy gépíró vagy szedő számára készítette őket. (Sajnos a hagyatékban nem található meg a gépelt szöveg vagy a korrektúra valamely példánya.) A digitális kiadásban ezek a szövegváltozatok a legkönnyebben kódolhatók, mind technikailag, mind filológiailag kevés problémát vetnek fel. Azonban izgalmas a jövőbeli szövegértelmezés szempontjából, hogy néhány intertextuális hivatkozás csak ebben a majdnem végleges előkészítési szakaszban kerül be a regénybe. A következő csoportba tartozó füzetek korai vagy köztes vázlatokat tartalmaznak. Néhány fejezet alig változott az első vázlatok óta, másokat a szerző többször átdolgozott. Ezek a szövegek a regény értelmezése szempontjából nagy

jelentőségű változásokat tárnak fel. Néhány cselekményszál végül kikerült a regényből, és több (főként női) szereplő kimaradt végül a végleges szövegből. Izgalmas jelenség, hogy a regény első felének néhány fejezete az első fogalmazványokban egyes szám első személyű elbeszélés, ezt, ismételt újrairások után végül harmadik személyű elbeszélés váltotta fel (a hős, vagy ahogy a regényben íródik, a „hoes” történetéről van itt szó). Ez a réteg több száz oldal kidolgozott prózaszöveget tartalmaz, amely végül nem került be a *Termelési-regény*be, és más, későbbi művekbe sem. A harmadik réteg a legkorábbi vázlatokat tartalmazza, amelyeknek csak kis része mutat konkrét szöveghozzávetőleges kapcsolatot a végleges szöveggel. Ezek kíséreltek a szöveggenézis első fázisából származnak, annak nyomai, ahogy a regény díszlet- és cselekvéstere, elbeszélőtechnikája lassan kialakul. Meglepő módon ebben a szakaszban Esterházy még párhuzamosan használta a füzeteket munkahelyi jegyzetek készítésére és a regény tervezésére és vázlatok, fogalmazványok lejegyzésére. Bár nem világos, hogy ezek a korai jegyzetek végül bekerülnek-e majd a korlátozott terjedelmű nyomtatott kiadásba, itt érdemes megjegyezni, hogy ezek az anyagok, amelyek Esterházy munkahelyi feladataival kapcsolatosak, korántsem mellékesek, mivel a munka és a művészet szorosan összefonódik, nemcsak ezeknek a füzeteknek az oldalain, hanem a regényben is. (Mind a fiatal Esterházy, mind a regény hőse, Tomcsányi, a hoes: termelési folyamatokat szervező szoftver fejlesztéséért felelős.)

A kéziratok jelölőnyelvi kódolása

A genetikus folyamat szövegszintű feltárásakor egy egyszerre analóg és digitális kiadást készítő, ugyanakkor *born digital* alapokra épülő vállalkozásánál különös problémák adódhatnak. A kutatócsoport két, Esterházy-értelmezésben jártas PhD. Hallgató (Kiss Júlia, Kolozsi Blanka) közreműködésével, a projektvezető, Bónus Tibor irányításával a DigiPhil specifikációját követve, Oxygen XML Editort szoftver felhasználásával elkészítette a több mint 1000 kéziratoldal XML-átíratát, jelölve a kéziratban az összes releváns szövegműveletet és szövegjellemzőt: a törléseket, javításokat, margináliákat stb. Az XML önmagában azonban kevésbé alkalmas egy analóg szövegműveletekre épülő szövegválogató és az irodalomtudós értelmezési praxisára épülő kapcsolati mintázatkeresés kiszolgálására. Mivel a rendelkezésre álló XML-ből nyomtatható dokumentumot előállító kész eszközök³ nem voltak képesek kezelni a kéziratokat leíró XML-ek komplexitását, végül a DigiPhil megjelenítő *Nyomtatási nézet* funkciójával sikerült könnyen olvasható és áttekinthető, szétvágható és színes jelölőkkel ellátható materiális szövegtestet előállítanunk, közvetlenül az XML-ekből.

3 TEI XML-ből különféle dokumentumformátumokat az Oxygen XML Editor beépített XSLT stíluslapjai segítségével is készíthetünk. Az XML-formátumot konvertáló legismertebb eszköz a Pandoc (<https://pandoc.org/>), ez azonban nem használható TEI XML-fájlok átalakítására. A TEIGarage eszköz ugyan képes TEI XML-t különféle dokumentumformátumokba konvertálni, de egy csak a TEI-ajánlásban szereplő tag-ek és tag-kombinációk kisebb részét támogatja (<https://github.com/TEIC/TEIGarage>).

Metadata

Nyomtatás dátuma: 2024. 06. 20.

e. péter

1975 aug →

Esterházy

[lr]

{76 {[]² {[[[hiány]]⁴ }³ {márc}⁵ . 18.) április 14 → többmint ötlet május 10 → {irodalmaró}² }¹

¹ Beszűrés : margó

² [bizonytalan olvasat]

³ [törölt]

⁴ **Egyéb:** karakter | **Ok:** olvashatatlanság

⁵ Beszűrés

Termelési=regény (kisregény) {((temesi Ferenc))¹¹ }⁶ "A minap magamhoz kérttem a Főnököt." (IHIM),
a {}⁷ {[(~~egy~~)]⁸ gyönyörűszip nyomozó {~~hagyatékát~~]⁸ }⁸ (Ha, akkor az véletlen.)

⁶ [törölt]

⁷ **Korrekciók:**teiComponents.popOver.metamarkProperties.insert

⁸ Beszűrés : alá

Most, Tisztelt Olvasóm, hogy kifogad hatlovas hintomból a lovakat {Ekkor jelent meg E.P. első könyv;
kifogták a lovakat}⁹ (hat darabot, leltári számnak a megfelelő, de elveszett mellékletben), s magad
áltál a rúd mellé, visszaállítottad {vén ekképp}¹⁰ az író múlt századi hatalmát, {→ s ezzel elszabadult
a pokol, (egyszersmind a mennyország {is}¹⁰ – de ez nem könnyebbség; túl nagy a rizikó!) {[
angyalok könykölnék az ördögök mellett, söröznek, ki a túró ismeri ki magát szeszszesrűvületben (mert
kikapott a Fradi)]¹⁰ }¹⁰ kezemben pattog az ostar, gyia azanyád (satöbbi), s úgy érzed neked nagyon
nehéz, az instráng vágja a hátat, más – imi=olvasni, nem tudok – rajtad röhögnek, s ujjal mutogatnak
rád, és még avval is megrágalmaznak, hogy a lószar is téled származik, pedig nem és nem, pedig ne
hidd, hogy itt fõnn sokkal jobb, a szél fúj, és a {~~kez~~]¹¹ karom szinte leszakad {~~az engem üdvözöl~~
~~tömeg~~]¹¹ az üdvözlõ intésektõl {}⁹

⁹ Beszűrés : teiComponents.popOver.other-page-0v

¹⁰ Beszűrés : fölé

¹¹ [törölt]

mindenesetre most, hogy eligazodom már a világ dolgaiban, bízz bennem, {~~fin~~]¹² eligazítalak téged
is, nem középiskolás fokon, {már}¹³ ami a majd megnevezhető papírokat illeti.

¹² [törölt]

¹³ Beszűrés : fölé

Fiam, hajtson az Intézet felé. Örökök, mely egyre pontosabb, mutat

2. kép: A regény kéziratának részlete a DigiPhil TEI XML megjelenítő Nyomtatási nézet funkció segítségével vizualizálva

Intertextualitás

Amikor a kutatócsoport megalakult, már sejtettük, hogy a kiadások elkészítésének fő problémája az intertextuális hálózatok felfedezése és beépítése lesz a kiadásba, valamint a kéziratfragmentumok értelmes elrendezése és a regény végleges szövegével való összekötése.

Esterházy számára az intertextualitás, ahogy erre már utaltunk, az írói identitás problematizálása. Mint szövegválogató,⁴ az író vidám szertelenséggel idéz kanonizált és elfeledett irodalmi műveket, használati utasítást, vicceket, kommunista munkadalokat, politikai beszédek és archaikus imákat. Elméleti szempontból ez a kavalkád az irodalom nyelvi meghatározottságának és általában a nyelvi megelőzöttség tudatának humoros színre állítása. A filológus számára mindez egy rendkívül nehezen megoldható probléma (bár Esterházy számos szövegforrást megnevez a kéziratfüzetek lapjain). A probléma megoldá-

4 Balajthy Ágnes, Bódi Katalin, Szirák Péter: *A kortárs magyar irodalom. VII/4. Idézettechnika, „szövegválogatás”,* Debreceni Egyetemi Kiadó, 2021, <https://doi.org/10.5484/9789636150761>

sára tett kísérletet Jankovics József, aki közzétette a *Bevezetés a szépirodalomba* egy fejezetét annotált kiadás formájában, amelyben az intertextusokat és szövegmagyarázatokat a nyomtatott médiumban megszokott módon a margóra helyezte.⁵

A 21. században azonban a filológus rendelkezésére állnak a távoli olvasás olyan módszerei, amelyekről korábban csak álmodni lehetett. Ahogy a digitális korpuszok mérete és elérhetősége nő, úgy nő a számítógép által feltárt idézetkapcsolatok azonosításának képessége is. Az irodalmi intertextusok tanulmányozásában régóta használt szöveg hasonlóság-mérési technológia alkalmazható az Esterházy-kutatásban is.⁶ Az ELTE BTK Történeti Intézetében működő Digitális Bölcsészeti Tanszék munkatársai, a Digitális Örökség Nemzeti Laboratórium támogatásával, azzal a nem titkolt céllal folytatnak kísérleteket kvantitatív szöveg hasonlósági méréseken alapuló intertextuális kapcsolati hálózatok megalkotásával, hogy a Tanszék irodalmi korpuszainak keresőfelületeit⁷ intertextuális kapcsolatok kereshetővé tételével fejlesszék tovább. A három korpusz közül az egyik, a dráma korpusz anyaga már most is vizsgálható kvantitatív hálózatkutató módszerrel segítségével, itt azonban a kapcsolati hálózatot a drámaszereplők kapcsolati hálózata adja, nem a szövegek közötti intertextuális kapcsolatok.



3. kép: A Bánk bán szereplőinek kapcsolati hálózata (<https://dramakorpusz.elte-dh.hu/view.php?id=56&chapter=2>, <https://dracor.org/hun/katona-bank-ban>)

⁵ Jankovics József (szerk.): *Esterházy Péter: Függő* (Matúra klasszikusok), Ikon, Budapest, 1993.

⁶ A szöveg szimilaritás digitális bölcsészeti, valamint az intertextualitás irodalomtudományi fogalmainak viszonya rendkívül bonyolult, jelen rövid szövegben nincs tér ezek részletes tárgyalására. Kiindulópontként szolgálhat Mike Kestemont és munkatársainak tanulmánya: Manjavacas, Enrique, Folgert, Karsdorp, and Mike, Kestemont: „A Statistical Foray into Contextual Aspects of Intertextuality”, in *Proceedings of the Workshop on Computational Humanities Research (CHR 2020) Amsterdam, the Netherlands, November 18–20, 2020*. ed Folgert, Karsdorp, Barbara, McGillivray, Adina, Nerghees and Melvin, Wevers (<http://ceur-ws.org/Vol-2723/long28.pdf>)

⁷ <https://verskorpusz.elte-dh.hu/>, <https://regenykorpusz.elte-dh.hu/>, <https://dramakorpusz.elte-dh.hu/>

A Tanszék munkatársaival a Yale Digital Humanities Laboratory (DHLab) által fejlesztett Intertext szoftver továbbfejlesztésén dolgozunk. Az eszköz jelenleg a MinHash-algoritmus alapján állít elő szöveghasonlóság-értékeket,⁸ ahhoz azonban, hogy különféle, nem betű szintű egyezésre építő intertextusok, így pl. parafrázisok detektálására is alkalmas legyen az eszköz, azt az újabb szóbeágyazások (pl. fastText, BERT), szemantikus hasonlóságokra is érzékeny kalkulációval kell kiegészíteni.

A módszer gyakorlati alkalmazása érdekes, váratlan eredményeket hozott. A szakirodalomban régóta ismert, hogy Esterházy intertextusainak jelentős része önidézet. Eredményeink szerint a *Termelési-regény* a legtöbb és legkiterjedtebb intertextuális kapcsolata az író első kötetével, a *Fancsikó és Pintával* van, a két mű között 57 idézetkapcsolatot azonosítottunk gépi úton. Szintén sikerrel alkalmaztuk a módszert a szakirodalomban már említett, de szövegszinten nem vizsgált kapcsolatok adatszerű pontosítására. Erdélyi Zsuzsanna *Hegyet hágék, lőtőt lépék, Archaikus népi imádságok* című kötete 1974-ben, majd 76-ban jelent meg, Esterházy 1976-ban zárta le a kéziratot, de csak viszonylag későn, a már szinte kész kéziratba illesztette bele a kötet imádság-részleteit, akkor viszont több esetben betű szerint, még a pontuációkat is változtatlanul hagyva.

The screenshot displays the Intertext software interface. On the left, there's a sidebar with search filters: 'Találatok (27)', 'Összetett Keresés', 'Egyezések száma' (a slider), 'Szimilitás' (a slider), 'Korábbi Szöveg' (with buttons for Szerző, Cím, Fajl Id), and 'Későbbi Szöveg' (with buttons for Szerző, Cím, Fajl Id). The main area shows search results for 'Hegyet hágék' by Erdélyi Zsuzsanna (szerk.). It lists two results, each with a '63%' similarity score. The first result is 'Hegyet hágék' and the second is 'Termelési-regény' by Esterházy Péter, 1979. The interface includes buttons for 'Keresés' and 'Rendezés: Hossz'.

4. kép: Szövegszerű egyezések Erdélyi Zsuzsanna *Hegyet hágék, lőtőt lépék népi imádságai* és a *Termelési-regény* szövege között

Kvantitatív szövegszimilitás digitális szövegkiadás

A gépi szöveghasonlósági-vizsgálatok során azonosított, kvantitatív hasonlóságot mutató szövegszegmentumok beemelése egy kiadásba, illetve egy digitális kiadásokat kezelő szolgáltatásba önmagában is komplex probléma, amelyre itt csak röviden utalunk. A gépi asszisztenciával vagy anélkül, a filológus szoros olvasási praxisa által azonosított

⁸ <https://dhlabs.yale.edu/projects/intertext/>. A MinHash módszerről részletesen: Leskovec, Jure, Anand Rajaraman, and Jeffrey David Ullman. "Finding Similar Items." Chapter in *Mining of Massive Datasets*, 68–122. Cambridge University Press, 2014.

intertextuális kapcsolatok tárolása, jelölése és kereshetővé tétele eltér a nyomtatott, és a *born digital* kiadások esetében. A nyomtatott kiadásokban az intertextuális kapcsolatokat akkor rögzítik, ha jelentőségük és értelmezhetőségük szempontjából a kiadást készítő kutatók jelentősnek minősítik őket. Szükség esetén, különösen, ha módosított idézetről van szó, a forrásszöveget is közlik a kiadások, és értelmezik a vendégsszöveg funkcióját. De el kell-e hagynunk a szövegkapcsolatok jelölését, ha az azonosított hasonlósági kapcsolatok száma, aránya meghaladja az értelmezhetőség szintjét? A nyomtatott kiadásban valószínűleg igen, a digitális kiadásban nem feltétlenül. A jelölőnyelvi átírat (azaz a TEI XML) adott pontjához szemantikus állítások formájában akkor is kapcsolhatunk idézetkapcsolatot, ha a szimilitás tényén és forrásán (és esetleg a hasonlósági arányon) kívül más annotáció nem készül. Ez a DigiPhil projekt keretei között a szemantikus web technológiájának segítségével megoldható, hiszen a szövegkiadás bibliográfiai és egyéb jellemzőit eleve szemantikus adattérben rögzítjük. Ugyanakkor annak módszertanát, hogy a DigiPhil-ben közzétett kiadások filológiai adatait tartalmazó ITIdata adatbázisban hogyan rögzítünk szövegssimilitás-kapcsolatokat, még ki kell dolgozni.