

PÉTER RÓBERT

Rejtett hatások feltárása az eszmetörténetben a *Reception Reader* kutatási eszközzel¹

ORCID: 0000-0002-7972-4751

robert.peter@ieas-szeged.hu

HELIKON

Uncovering Hidden Influences: The *Reception Reader* as a Tool for Intellectual Historians

Abstract

This article critically examines the *Reception Reader*, a digital tool developed by the Helsinki Computational History Group to explore text reuse in early modern print culture. Positioned at the intersection of intellectual history and digital humanities, the tool leverages large-scale corpora such as EEBO-TCP and ECCO to trace borrowing, reprinting, and repurposing of texts between 1475 and 1800. Using a modified BLAST algorithm to address noisy OCR data, the *Reception Reader* enables scholars to combine distant and close reading in the study of reception, influence, and intertextuality. Three case studies – on Bernard Mandeville, David Hume, and eighteenth-century utopian literature – demonstrate its capacity to reassess authorship attribution, map the circulation of ideas, and reveal hidden intellectual influences. While the tool's visualisations and databases present methodological limitations, its integration of computational analysis with traditional scholarship underscores its value for intellectual historians. Looking ahead, the COMHIS group plans to integrate large language models to detect semantic similarity and to develop an edition comparison tool that will help scholars chart changes across multiple editions and works of the same author. By democratising access to vast digitised archives and advancing text reuse detection, the *Reception Reader* exemplifies how interdisciplinary digital methods can transform the study of intellectual history.

Keywords: text-reuse, evolution of ideas, intellectual history, Reception Reader, early modernity

¹ Ez a cikk az alábbi, szabadon felhasználható mű alapján készült, de több tekintetben eltér az eredeti tanulmánytól: PÉTER Róbert, „Uncovering Hidden Influences: The Reception Reader as a Tool for Intellectual Historians (Reception Reader, Helsinki Computational History Group, <https://receptionreader.com/>)”, *Global Intellectual History*, 1–13, hozzáférés: 2025.11.02, <https://doi.org/10.1080/23801883.2025.2474486>. Eredetileg megjelent az Informa UK Limited (Taylor & Francis Group) kiadásában.

1. HÍD AZ ESZMETÖRTÉNET ÉS A DIGITÁLIS BÖLCSÉSZET KÖZÖTT

A nyílt hozzáférésű, illetve előfizetéshez kötött digitális archívumokban elérhető, hatalmas és egyre bővülő adatmennyiség (*big data*) olyan kereshető korpuszokat biztosít a kutatók számára, amelyek gyakran ritka és sérülékeny könyveket, kéziratokat, újságokat, leveleket és efemerákat is tartalmaznak. A digitális bölcsészet megjelenésének köszönhetően az irodalomtudományban az elmúlt húsz év során viszonylag elterjedtté vált az adatalapú módszerek alkalmazása, amelyek új felismerések, mintázatok és összefüggések feltárásában segítik a kutatókat. Ugyanakkor, ahogyan Mark J. Hill 2016-ban megjegyezte a *Global Intellectual History* folyóiratban, „a digitális bölcsészetről bár sok szó esik, mégis alig feltárt terület az eszmetörténet szempontjából”. Tanulmányának záró gondolataiban arra biztatta a digitális bölcsészeket és az eszmetörténészeket, hogy működjenek együtt, és hozzanak létre olyan új eszközöket és módszereket, mint például az e tanulmányban bemutatott *Reception Reader*;² amelyek gazdagítják és elmélyítik kutatások minőségét, és a gondolatok terjedésének nyomon követését nagy idő- és térbeli távolságokban is lehetővé teszik.³ Hill javaslata óta jelentős digitális bölcsészeti projektek indultak el, fejlődtek tovább és értek el komoly eredményeket a történettudomány területén, ezek közül néhányat röviden ismertetek az alábbiakban.

A 2007-ben indított és azóta bővített *French Book Trade in Enlightenment Europe* (FBTEE) projekt a régi rend (*ancien régime*) korszakának könyvkereskedelmét térképezi fel, bibliometrikus módszereket alkalmazva az eszmék terjedésének elemzésére.⁴ A projekt korábban ismeretlen összefüggésekre világított rá: például olyan vallásos szövegek – mint a *L'Ange conducteur* – széles körű elterjedésére, amelyek megkérdőjelezik a felvilágosodás szekularizációjáról alkotott narratívákat.⁵ A *Mapping the Republic of Letters* projekt vizualizációs technikákat alkalmaz a felvilágosodáskori gondolkodók levelezési hálózatainak azonosítására.⁶ Többek között az *Electronic Enlightenment* adatbázis

² ROSSON, „Reception Reader”, *Journal of Open Humanities Data* 9, 1. sz. (2023), <https://doi.org/10.5334/johd.101>, <https://receptionreader.com/>.

³ MARK J. HILL és SIMON HENGCHEN, „Invisible Interpretations”, *Digital Scholarship in the Humanities* 34, 4. sz. (2019): 130–50.

⁴ *The French Book Trade in Enlightenment Europe*, hozzáférés: 2025.02.06, <http://fbtee.uws.edu.au/main/>.

⁵ SIMON BURROWS, „The FBTEE Revolution,” in *Digitizing Enlightenment: Digital Humanities and the Transformation of Eighteenth-Century Studies*, szerk. SIMON BURROWS és GLENN ROSE, 94 (Liverpool: Liverpool University Press, 2020).

⁶ *Mapping the Republic of Letters*, hozzáférés: 2025.02.06, <http://republicofletters.stanford.edu/>.

metaadatait felhasználva elemzi a levelezéseket, kapcsolatokat teremtve olyan kulcsfigurák között, mint Locke, D'Alembert és Franklin, hogy ezek révén szélesebb intellektuális közösségeket tárjon fel. Tér- és időbeli elemző eszközök új generációját alkalmazva a kutatás jelentős felismerésekre jutott – például Voltaire korlátozott brit kapcsolatait illetően –, és megkérdőjelezett általános, a felvilágosodás kapcsolatrendszereire vonatkozó feltételezéseket.⁷ A *Reassembling the Republic of Letters (1500–1800)* projekt szintén digitális ke-retrendszerben tette kutathatóvá a kora újkori levelezési hálózatokat.⁸ Ezek a kezdeményezések összekapcsolt nyílt adattechnológiák (*linked open data*) segítségével interoperábilis metaadat-forrásokat hoztak létre, ezáltal lehetővé téve összekapcsolásukat és a transznacionális intellektuális kapcsolatok átfogó vizsgálatát.

A *Measuring Enlightenment: Disseminating Ideas, Authors, and Texts in Europe* (MEDIATE) projekt magánkönyvtárak aukciós katalógusait digitalizálja annak érdekében, hogy a könyvek és – a felvilágosodás fogalmához hagyományosan társított – eszmék áramlását kövesse nyomon 1665 és 1830 között bibliometrikus módszerekkel és az adatmodellezés eszköztárát mozgósítva.⁹ A *Comédie-Française Registers Project* (CFRP) a Comédie-Française színházi társaság 1680 és 1793 közötti előadásainak és pénzügyi nyilvántartásainak digitalizálását végzi.¹⁰ Az adatok vizualizációját és a metaadatok modellezését ötvözve a projekt a jegyeladási adatokat és a repertoár alakulását vizsgálja.¹¹ A tárgyaltak közül a legújabb, *Modelling Enlightenment: reassembling inter-textual networks through data-driven research* (ModERN) projekt célja pedig egy átfogó francia szövegkorpusz létrehozása, valamint új kutatási módszerek

⁷ Dan EDELSTEIN, „Mapping the Republic of Letters”, in BURROWS és ROSE, szerk., *Digitizing Enlightenment: Digital Humanities and the Transformation of Eighteenth-Century Studies*, 73–88.

⁸ TUOMINEN, „Reassembling the Republic of Letters”, in *Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference (DHN 2018)*, szerk. Eetu MÄKELÄ, Mikko TOLO-NEN, Jouni TUOMINEN (CEUR Workshop Proceedings, 2018), 76–88, <https://doi.org/10.5617/dhnpub.11011>.

⁹ Measuring Enlightenment: Disseminating Ideas, Authors, and Texts in Europe (MEDIATE), hozzáférés: 2025.02.06, <https://mediate18.nl/>; Alicia C. MONTROYA, „Enlightenment? What Enlightenment? Reflections on Half a Million Books,” *Eighteenth-Century Studies* 54, 4. sz. (2021): 909–934, <https://doi.org/10.1353/ecs.2021.0097>.

¹⁰ Jeffrey S. RAVEL, „The Comédie-Française Registers Project”, in *Digitizing Enlightenment: Digital Humanities and the Transformation of Eighteenth-Century Studies*: 133–150.

¹¹ Comédie-Française Registers Project, hozzáférés: 2025.02.06, <https://cfregisters.org/>.

kidolgozása, hogy mélyebb betekintést nyerhessünk a kora újkori francia nyomtatott kultúrába és annak intertextuális dinamikájába.¹²

Az említett projektekben alkalmazott adatalapú eszközök és algoritmusok jelentősen előmozdították az eszmetörténeti kutatásokat, elősegítve például a fogalmak megjelenésének és átalakulásának elemzését, a hosszú távú ideológiai változások, továbbá az intellektuális hálózatok rekonstrukcióját. Az eszközök ugyanakkor nemcsak a fogalmi ideológiai változások időbeli modellezésében segítik a kutatókat, hanem abban is, hogy megadják a fogalmak és gondolatok jellemző kontextusait (például műfajok szerint), ezzel megkönnyítve a hagyományos, szoros olvasásra épülő szövegelemzést is. A mesterséges intelligencia legújabb fejlesztései ugyancsak gazdagították az eszmetörténeti kutatásokat. Például az MI-alapú eszközök segítik a többnyelvű korpuszok gyorsabb és hatékonyabb elemzését, ezáltal a kutatók az eredeti nyelveken tanulmányozhatják a szövegeket, miközben azonosíthatják a kulturális hatásokat, és transznacionális összehasonlításokat végezhetnek.

2. INTERDISZCIPLINÁRIS EGYÜTTMŰKÖDÉS: A *RECEPTION READER* MÖGÖTT ÁLLÓ CSAPAT

Az elmúlt évtizedben a Helsinki Computational History Group (COMHIS), amelynek Hill is tagja volt, jelentős mértékben hozzájárult a bölcsészettudományi kutatások digitális átalakulásához, különösen az eszmetörténet területén.¹³ A Mikko Tolonen által vezetett kutatócsoport az eszmetörténetet a modern adattudomány, a gépi tanulás módszereit pedig a történettudomány és a nyelvészet szemléletével ötvözi. Számukra a „»számítógépes történelem« egy kevert módszertant jelenti, amelyben a nagy adatok elemzését az eszmetörténet és könyvtörténet szakértői tudásával kombinálják”.¹⁴ A COMHIS csoport számos jelentős nemzeti és nemzetközi kutatási projektben vett részt ezen a területen, amelyek alapvető és gyakran hivatkozott publikációkhoz vezettek. A jelen tanulmányban tárgyalt *Reception Reader* – egy hiánypótló, webes alkal-

¹² Dario Maria NICOLSI és Glenn ROE, „Modelling Enlightenment: Reassembling Intertextual Networks through Data-Driven Research (ModERN)”, *Digital Enlightenment Studies* 2, 1. sz. (2024): 62–84, <https://doi.org/10.0.238.219/des.22>.

¹³ Helsinki Computational History Group, hozzáférés: 2025.02.06, <https://researchportal.helsinki.fi/en/projects/helsinki-computational-history-group/publications>.

¹⁴ Helsinki Computational History Group (COMHIS), hozzáférés: 2025.02.06, <https://www.helsinki.fi/en/researchgroups/computational-history>.

mazás – szintén e csapat munkája, és 2023-ban vált elérhetővé a nagyközönség számára.

A *Reception Reader* olyan tudományos eszköz, amely bölcsészettudományi kutatóknak nyújt segítséget szövegek időbeli újrafelhasználásának (*text reuse*) vizsgálatában. Az *Early English Books Online-Text Creation Partnership* (EEBO-TCP)¹⁵ és az *Eighteenth Century Collections Online* (ECCO)¹⁶ adatbázisok – amelyek együttesen jelentős mennyiségű, 1475 és 1800 között megjelent szöveget tartalmaznak – gazdag forrásanyagát használva választhatják ki a felhasználók azt a művet, amelynek recepcióját vizsgálni kívánják. Az eszköz megmutatja szövegek újrafelhasználásait, a találatokat kronológiai előfordulások szerint rendszerezi, és dokumentumrészek alapján csoportosítja. Tervezése során elsődleges szempont volt a távoli és a szoros olvasás módszertanainak együttes alkalmazhatósága, ami művek recepciójának és hatásának árnyalt vizsgálatát biztosítja időbeli távlatban.

3. A SZÖVEG-ÚJRAFELHASZNÁLÁS FELTÉRKÉPEZÉSE AZ ESZMETÖRTÉNETBEN ÉS A RECEPCIÓKUTATÁSBAN

A szöveg-újrafelhasználás tanulmányozása a bölcsészettudományok egyik központi eljárása, amely betekintést nyújt abba, miként terjednek és alakulnak az eszmék és szövegek az idők során. Az ismétlődő vagy közös szövegrészek azonosításával a kutatók nyomon követhetik az idézetek, kölcsönzések, újrayomatások, parafrázisok vagy utalások mintázatait. Mindezek hozzájárulnak a hatástörténeti összefüggések és azok tágabb intellektuális kontextusainak mélyebb megértéséhez.

A kora újkori nyomtatott kultúrában a szövegek újrakiadása megszokott gyakorlat volt: az újságok előszeretettel közöltek teljes bekezdéseket forrás-megjelölés nélkül, a kiadók pedig frissített vagy kibővített kiadásokkal próbálták új vásárlók figyelmét felkelteni, míg az antológiák különféle prózai és lírai műveket gyűjtöttek egybe szélesebb olvasóközönség számára. Az olvasók maguk is részt vettek ezekben a gyakorlatokban – például közhelygyűjteményeket (*commonplace books*) vezettek, amelyekben témák szerint gyűjtötték az idé-

¹⁵ Early English Books Online – Text Creation Partnership (EEBO-TCP), hozzáférés: 2025.02.06, <https://textcreationpartnership.org/tcp-texts/eebo-tcp-early-english-books-online/>.

¹⁶ Eighteenth Century Collections Online – Text Creation Partnership (ECCO-TCP), hozzáférés: 2025.02.06, <https://textcreationpartnership.org/tcp-texts/ecco-tcp-eighteenth-century-collections-online/>.

zeteket, hogy azokat később újra felhasználják. Az efféle átvételek egy olyan kultúra normáit tükrözték, amely nagyra értékelte a klasszikus és kortárs források ismeretét. Ugyanakkor olyan viták, mint a *Régiek és Modernnek vitája* (*Quarrel of the Ancients and Moderns*), arra utalnak, hogy a gyakran idézett klasszikus szerzők tekintélyéhez való hozzáállás változóban volt: noha műveiket gyakran utánozták, a modern szerzők bírálták is ezeket. A kora újkorban a plágium fogalma inkább az engedély nélküli újradíjazásra vonatkozott, semmint az ötletek eltulajdonítására. Ebben a kontextusban az eredetiség jól megférhetett a kiterjedt szövegátemelésekkel.

A szöveg-újrafelhasználás szorosan kapcsolódik a recepciótörténethez is, hiszen a későbbi művekben ismétlődő szövegrészekből következtethetünk arra, hogy miként értelmezték és hasznosították újra a korábbi írásokat. Mindazonáltal fontos megjegyezni, hogy ugyanannak a részletnek az idézése nem feltétlenül jelent közvetlen hatást vagy befolyást: a szövegek gyakran merítnek egy közös, ismert idézetekből vagy motívumokból álló repertoárból is. Ahogyan a *Reception Reader* alkotói is hangsúlyozzák: „a szöveg-újrafelhasználás nem szükségszerűen jelent befolyást, mégis értékes információkat nyújt a recepció nagyléptékű vizsgálatához.”¹⁷ Valóban, egy adott szövegrész – például egy bibliai idézet – pusztán ismétlése nem feltétlenül utal közvetlen hatásra, inkább egy hasonló kulturális háttérrel jelölhet, azonban tanulságos lehet az is, ha különböző szerzők eltérően értelmezhetik vagy más célokra használhatják fel ugyanazokat a sorokat.

Neves kutatók rávilágítottak arra, hogy a szöveg-újrafelhasználás mögött összetett dinamikák rejlenek. Harold Bloom például azt vizsgálta, miként birkóznak meg a későbbi szerzők elődeikkel annak érdekében, hogy saját hangjukat érvényesítsék. Julia Kristeva intertextualitás-fogalma (amely Mihail Bahtyin munkásságára épül) azt hangsúlyozza, hogy a jelentés mindig is szövegek közötti kölcsönhatás révén jön létre. Gérard Genette transztextualitás-elmélete tovább finomítja ezeket a kapcsolattípusokat: az intertextualitás megjelenhet például idézetek, plágium és allúziók formájában is.¹⁸ Ezek az elméleti keretek megvilágítják, miért foglal el kiemelkedő helyet az irodalomtudományi kutatásokban a szöveg-újrafelhasználás, amely az intertextualitás egy specifikus formájaként tekinthető. Ezekből kiindulva egy kutatás irányulhat arra, hogyan és miért maradtak fenn bizonyos eszmék; hogyan viszonyultak a szerzők a meglévő művekhez; illetve miként alakították a kulturális

¹⁷ ROSSON, „Reception Reader”.

¹⁸ Glenn ROE, „Text Reuse as Cultural Practice,” *Digital Enlightenment Studies* 2, 1. sz. (2024): 1–30, <https://doi.org/10.61147/des.23>.

normák ezeket az adaptációs és kölcsönzési folyamatokat. Megfigyelhető továbbá az is, hogyan emeltek át szövegeket egyik korszakból a másikba, néha megerősítve egy korábbi szerző hírnevét, máskor pedig újraértelmezve az eredeti anyagot új célok érdekében. Annak vizsgálata, hogy az írók hogyan választottak ki, módosítottak vagy hasznosítottak újra bizonyos szövegrészeket, betekintést nyújthat a retorikai stratégiákba és az olvasói elvárásokba is. Ily módon a szöveg-újrafelhasználás feltárja az irodalom és a kulturális termelés közötti kapcsolatokat, és szélesebb történeti összefüggéseket vázolhat fel anélkül, hogy szükségszerűen közvetlen hatást vagy lineáris eszmei láncolatot feltételezne.

Ahogy a *Reception Reader* fejlesztői is hangsúlyozzák, „a kontextus és a recepció vizsgálata alapvető fontosságú a kora újkori intellektuális, irodalmi és kulturális történelem megértéséhez.”¹⁹ A digitális fordulat előtt a szövegátvétel azonosítása kézi módszerekkel történt, aminek következtében például a kritikai kiadásokon dolgozó kutatók kisebb léptékben tudták feltárni a szövegszintű kapcsolatokat. Az elmúlt húsz év során azonban számos számítógépes eszköz és módszer született e célra, például a *passim*, a *TRACER* és a *TextPair*, amelyek már nagy számú szövegre vonatkozóan hoznak megbízható eredményeket.²⁰ Ezeket a lehetőségeket hasznosító projekt például az irodalmi szövegekre összpontosító *Versioning Machine*, a latin költészetet vizsgáló *Tesserae*, valamint a Vergilius *Aeneis*-ének tudományos fogadtatását tanulmányozó kutatások.²¹ Számos projekt, mint például az *impresso Media Monitoring of the Past*, az *Oceanic Exchanges* vagy a *Viral Texts*, egyebek mellett a történeti újságokban található szövegátvételek felismerését is lehetővé teszi, így a hírek áramlását is azonosítani képes kiterjedt idő- és térbeli távolságokban.²²

¹⁹ ROSSON, „Reception Reader”.

²⁰ *Passim*, hozzáférés: 2025.02.06, <https://github.com/dasmiq/passim/>; Matteo ROMANELLO és Simon HENGCHEN, „Detecting Text Reuse with Passim”, *Programming Historian*, 10. sz. (2021), <https://doi.org/10.46430/phen0092>; Pairwise Alignment for Intertextual Relations, hozzáférés: 2025.02.06, <https://github.com/ARTFL-Project/text-pair/>. *TRACER*, hozzáférés: 2025.02.06, <https://www.etrapp.eu/research/tracer/>; BÜCHLER és mások, „Towards a Historical Text Re-Use Detection”, in *Text Mining*, szerk. Chris BIEMANN, Alexander MEHLER (Berlin: Springer, 2014): 221–238, 238, https://doi.org/10.1007/978-3-319-12655-5_11.

²¹ SCHREIBMAN, „The Versioning Machine”, 101–107; LEE, „A Computational Model of Text Reuse in Ancient Literary Texts”, 472–79. Neil COFFEE, „The Tesserae Project: Intertextual Analysis of Latin Poetry” *Literary and Linguistic Computing* 28, 2. sz. (2013): 221–228, <https://doi.org/10.1093/llic/fqs033>.

²² DÜRING és mások., „Impresso Text Reuse at Scale,” *Frontiers in Big Data* 6 (2023), <https://doi.org/10.3389/fdata.2023.1249469>; CORDELL, „Oceanic Exchanges,” *Oceanic Exchanges: Tracing Global Information Networks In Historical Newspaper Repositories, 1840–1914*; David

Számos szöveg-újraterjesztéssel foglalkozó projekt, köztük a *Reception Reader* is, kifejezetten a 18. századra fókuszál. Ilyen az *Intertextual Hub*, egy innovatív digitális bölcsészeti kutatási környezet is – az ARTFL projekt²³ részeként –, amely a dokumentumokat szélesebb intertextuális kapcsolathálózatokban helyezik el, legyen szó akár közvetlen vagy közvetett átvételekről, más művekkel vagy szövegrészekkel közös témákról, illetve lexikai átfedésekről.²⁴ Ez segíti a felhasználókat abban, hogy ne csak az egyes szövegeket, hanem a közös eszmék, témák és szövegrészek révén összekapcsolódó szöveghalmazokat is feltárják. Például az ARTFL Encyclopédie projektjének keretében digitalizálták Diderot és D’Alembert *Enciklopédiáját*, amelyet a felvilágosodás kori intellektuális hálózatok viszonyában tettek kutathatóvá: szekvenciaillesztési algoritmusokat²⁵ alkalmazva rejtett, korábban nem ismert idézési mintázatokra mutatnak rá. Kimutatták többek között, hogy például 13 szócikk az *Enciklopédiából* jelentős részben Émilie Du Châtelet *Institutions de physique* című művéből „kölszönzött” részeket, a *Contradiction* szócikk esetében a szöveg 95%-át.²⁶

4. A RECEPTION READER ALAPJÁUL SZOLGÁLÓ ADATBÁZISOK ÉRTÉKELÉSE: KORLÁTOK, KIHÍVÁSOK ÉS MEGOLDÁSOK

Mielőtt bemutatnánk, hogyan működik a *Reception Reader*, illetve összehasonlítanánk más hasonló szoftverekkel, fontos röviden áttekinteni azokat a digitális gyűjteményeket, amelyekre ezek az eszközök épülnek. Az első az *Eighteenth-Century Collections Online* (ECCO), amely az angolszász világban a 18. században nyomtatott anyagok egyik legnagyobb digitális gyűjteménye. A *Reception Reader* fejlesztői arról tájékoztattak, hogy az ezt üzemeltető Gale cég 2016-ban a gyűjtemény első két verzióját, összesen 207 613 művet bocsá-

A. SMITH, Ryan CORDELL, és Elizabeth Maddock DILLON, „Infectious Texts”, in *2013 IEEE International Conference on Big Data*: 86–94, <https://doi.org/10.1109/BigData.2013.6691675>.

²³ The Project for American and French Research on the Treasury of the French Language (ARTFL), hozzáférés: 2025.02.06, <https://artfl-project.uchicago.edu/>.

²⁴ Clovis GLADSTONE, „The Intertextual Hub”, *Digital Enlightenment Studies* 1, 1. sz. (2023), hozzáférés: 2025.02.06, <https://intertextual-hub.uchicago.edu/>.

²⁵ A módszer lényege, hogy a nem pontos egyezéseket is képes azonosítani; a részletekhez lásd például a lapszám *Metronom: versmértékek változatainak azonosítása a lokális szekvenciaillesztés módszerével* című tanulmányát – a szerk.

²⁶ ROE, „Text Reuse as Cultural Practice”.

tott a kutatócsoport rendelkezésére.²⁷ 2020-ban a Gale hozzáfogott további mintegy 90 000 könyv digitalizálásához, amelyek a jövőben az ECCO III részeként kerülnek kiadásra.²⁸ A *Reception Reader* másik jelentős forrása az *Early English Books Online-Text Creation Partnership* (EEBO TCP) 60 327 szövege. Az EEBO gyűjteményben található művek képfájllai azonban gyenge minőségűek, és tudomásom szerint még nem alakították át őket géppel olvasható szöveggé.

Összességében tehát a *Reception Reader* mintegy 56%-át fedi le az *English Short Title Catalogue*-nak (ESTC), amely körülbelül 480 000 rekordot tartalmaz olyan nyomtatványokról, amelyeket 1473 és 1800 között adtak ki a Brit-szigeteken, a brit gyarmati fennhatóság alatt álló tengerentúli területeken, valamint az Egyesült Államokban.²⁹ Valójában azonban ennél jóval nagyobb a lefedettség, ha figyelembe vesszük, hogy az ECCO projekt célja az összes egyedi mű és nem feltétlenül azok minden újrakiadásának feldolgozása volt.³⁰ Fontos hangsúlyozni azt is, hogy az ESTC-ben rögzített művek közül több ezernyi – különböző okok miatt – még nem került digitalizálásra, így a COMHIS csapat ezeket nem is tudta felhasználni. Emellett maga az ESTC – mint minden katalógus – sem teljes. Például jelen sorok szerzője 63 olyan művet azonosított szabadkőműves archívumokban, amelyek nem szerepelnek az ESTC-ben; ezek közül néhány a *British Freemasonry, 1717–1813* című, ötkötetes forráskiadásban jelent meg.³¹ Mivel az ECCO TCP gyűjtemény mindössze 3100 címet tartalmaz, a *Reception Reader* fejlesztői kénytelenek voltak az OCR hibákkal terhelt ECCO verzióra is támaszkodni, ami nemcsak az adatbázis kereshetőségére, hanem a szövegelemzés minőségére is hatással van.

A COMHIS kutatócsoport emiatt igyekezett meghatározni az OCR-hibák arányát, és felmérte ezek hatását az adatalapú szövegelemzésekre. Kimutat-

²⁷ Hálás vagyok Mikko Tolonennek és Ville P. Vaara-nak a *Reception Reader* eszközzel kapcsolatos konkrét kérdéseimre adott válaszáért, amelyek jelentősen gazdagították ezt az áttekintő irodalmat.

²⁸ Stephen H. GREGG, *Old Books and Digital Publishing: Eighteenth-Century Collections Online*. Elements in Publishing and Book Culture (Cambridge: Cambridge University Press, 2021), <https://doi.org/10.1017/9781108767415>.

²⁹ English Short Title Catalogue (ESTC), hozzáférés: 2025.02.06, <https://web.archive.org/web/20230812200936/ESTC>, hozzáférés: 2025.02.06, <https://www.bl.uk/projects/english-short-title-catalogue>.

³⁰ Mikko TOLONEN, Eetu MÄKELÄ és Leo LAHTI, „The Anatomy of Eighteenth Century Collections Online (ECCO),” *Eighteenth-Century Studies* 56, 1. sz. (2022): 95–123, <https://doi.org/10.1353/ecs.2022.0060>.

³¹ Róbert PÉTER, Cécile RÉVAUGER és Jan SNOEK, *British Freemasonry, 1717–1813* (London: Routledge, 2016), <https://doi.org/10.4324/9781315639901>.

ták, hogy az olyan hibák következtében, mint például a hosszú „s” betű vagy a ligatúrák helytelen felismerése, a teljes korpusz tokenjeinek 26,9%-a torzulhatott az OCR-folyamat során.³² Ez azt jelenti, hogy a hagyományos kulcsszavas keresések a szavak legalább 26,9%-át nem képesek megtalálni. Csak ebből a nézőpontból érthetjük meg igazán, hogy a szoftverfejlesztőknek milyen jelentős módszertani kihívásokkal kellett szembenézniük, amikor a rendkívül „zajos” adatbázisokra – mint például az ECCO – épülő szövegátvételt elemző programokat készítették.

Ez a kihívás ösztönözte Clovis Gladstone és Charles Cooney kutatókat, amikor a *Reception Reader*höz hasonlóan a szöveg-újrafelhasználást tanulmányozták az ECCO és az EEBO TCP adatbázisokban, valamint egy klasszikus latin szövegekből álló korpuszban.³³ Céljuk az intellektuális hatások és a szövegátvétel feltérképezése volt ezekben a gyűjteményekben. Ehhez kifejlesztettek egy szekvenciaillesztési algoritmust, a *PhiloLine*-t, amely képes azonosítani a szövegátvétel méreteket és idézeteket nagy méretű, „rendezetlen” korpuszokban is, hiszen kisebb és bizonyos fajtájú (például különböző karakterek ugyanabban a szóban) eltérések ellenére is felismerhetővé teszi az egyezéseket. A megbízhatóság növelése érdekében figyelmen kívül hagyták ugyanazon szerzőhöz tartozó műveket, kiszűrték az újrakiadásokat (a korpusz 43%-át), egységesítették a helyesírást, és szótrigramokká alakították a szövegeket a hasonlóságok elemzése érdekében. Így több mint 60 millió közös szövegrészletet sikerült azonosítaniuk. Az eszköz lehetővé teszi továbbá a felhasználók számára, hogy metaadat-szűrők (úgy, mint szerző, cím, dátum, egyezés mérete), idősoros és kontextusalapú elemzések révén kutassák az újrafelhasznált szövegeket. A projekt többek között betekintést nyújt abba, milyen volt az egyes szerzők, mint például Lucretius recepciója a kora újkori Angliában, illetve rávilágít a Biblia széles körű használatára is (a szövegegyezések több mint 58,5%-a; a felhasználóknak ugyanakkor lehetőségük van kizárni a bibliai forrásokat a keresésekből, ha más típusú összefüggésekre kíváncsiak). Számos olyan átvett szövegrészlet esetében nem lehetett megfelelő forrásazonosítást adni, amelyek az ECCO adatbázis állományát megelőző időszakból (1700 előtt) származtak. Ennek következtében Shakespeare, Milton, Locke és számos más szerző szövegrészletei sem voltak pontosan azo-

³² Mark J. HILL és Simon HENGCHEN, „Quantifying the Impact of Dirty OCR on Historical Text Analysis”, *Digital Scholarship in the Humanities* 34, 4. sz. (2019): 825–843, <https://doi.org/10.1093/lilc/fqz024>. A ligatúra több betűjel összekapcsolása és összesűritése úgy, hogy egyetlen betűtestben jelenjen meg.

³³ Clovis GLADSTONE és Charlse COONEY, „Opening New Paths for Scholarship”, in *Digitizing Enlightenment: Digital Humanities and the Transformation of Eighteenth-Century Studies*: 353–384. Lásd még: <https://commonplacecultures.uchicago.edu/>, hozzáférés: 2025.02.06.

nosíthatók. E szerzők műveinek 18. századi kiadásai és kommentárjai szolgáltak az újr felhasznált szövegrészek legkorábbi forrásaként.

A következőkben azt kívánom bemutatni, miként oldotta meg a COMHIS csapat ugyanezeket a kihívásokat a *Reception Reader* alkalmazás keretein belül. Mindenekelőtt eltérő módszert alkalmaztak a szövegátfedések azonosítására. A *Basic Local Alignment Search Tool* (BLAST) helyi szekvenciaillesztő algoritmus módosított változatát használták, amelyet eredetileg biomolekuláris szekvenciák – például DNS-láncok – összehasonlítására fejlesztettek ki. Ezzel nagyrészt sikerült kiküszöbölni a gyenge OCR minőségből fakadó hibákat, ugyanis a kisebb eltérések ellenére is felismerhetővé váltak a hasonló szekvenciák (például, ha ugyanabban a sorban találhatók különbözőképp átírt szavak). A Turku Egyetem TurkuNLP kutatócsoportja fejlesztette ezt az algoritmust, amelyet korábban már alkalmaztak más digitális bölcsészeti projektekben is, például finn újságcikkek szöveg-újr felhasználásának felismerésére.³⁴ A *Reception Reader* összevetve a *PhiloLine*-alapú eszközzel úgy tűnik, hogy az előbbi több szövegátfedést képes azonosítani. A két eszköz pontos összehasonlítása azonban jelenleg több okból sem lehetséges. Egyrészt a szövegegyezések töredezettsége mindkét algoritmus számára kihívást jelent. Ez a töredezettség több tényezőtől fakad, többek között a gyenge OCR-minőségből, valamint abból, hogy a hosszú fejléc- és lábléc-szakaszok megszakítják a szövegfolyamot, még akkor is, ha a fő szövegtörzs folyamatosnak tűnik. Másrészt ennél is fontosabb, hogy – ellentétben sok természetesnyelv-feldolgozási (NLP) feladattal – ebben az esetben nem létezik arany standard egyértelmű mérőszámokkal, amely alapján meg lehetne mérni és összehasonlítani a szöveg-újr hasznosítást azonosító algoritmusok teljesítményét és sikerességét, illetve az általuk detektált szövegátfedések mértékét.

5. A RECEPTION READER HASZNÁLATA: FUNKCIÓK ÉS LEHETŐSÉGEK

Ebben a fejezetben azt mutatom be, hogyan működik a *Reception Reader* a gyakorlatban.³⁵ A rendszer egy metaadat-alapú keresőfelülettel indul, ahol a felhasználók szerzőnév és könyvcím (egy vagy akár több tag szava) alapján, illetve ezek kombinálva adhatnak meg lekérdezéseket. A keresőmotor támogat-

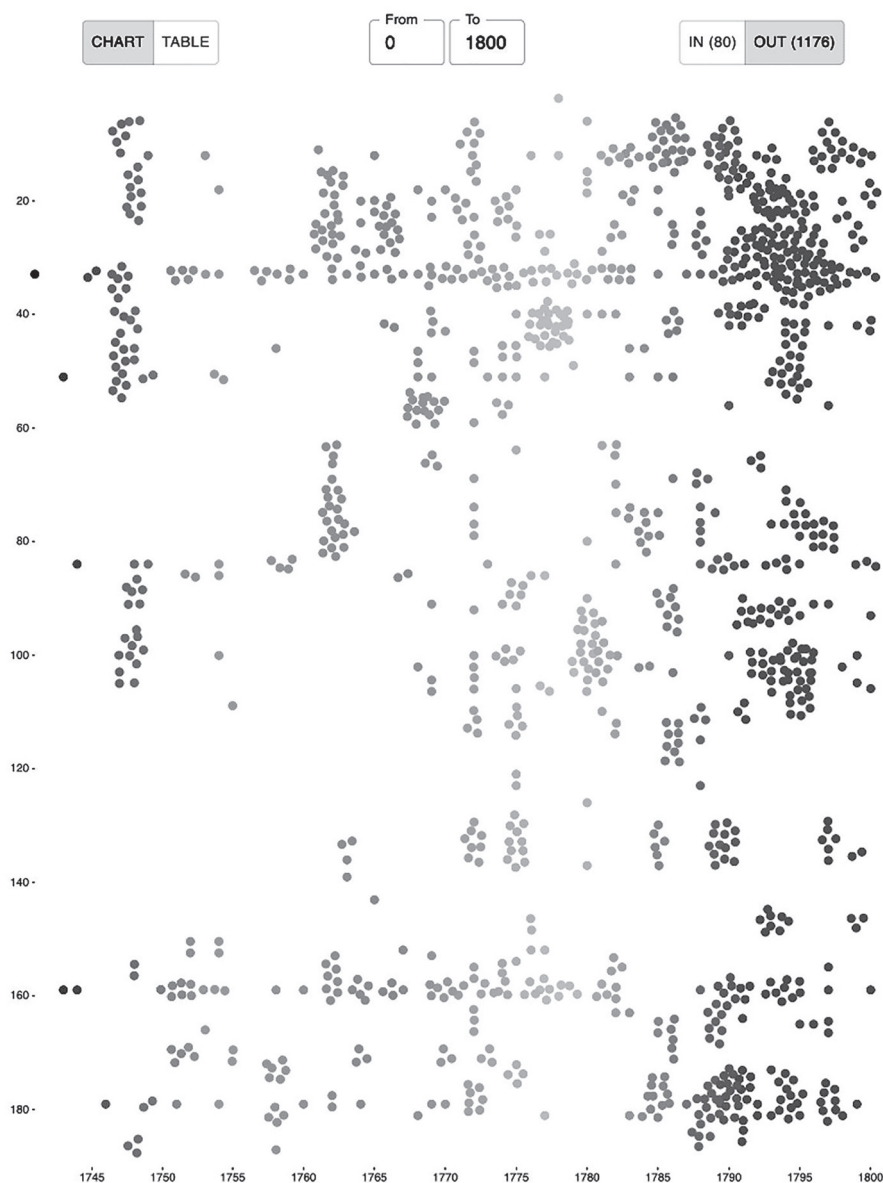
³⁴ Aleks VESANTO, Asko NIVALA, Heli RANTALA, Tapio SALAKOSKI, Hannu SALMI és Filip GINTER, „Applying BLAST to Text Reuse Detection in Finnish Newspapers and Journals 1771–1910”, in *Proceedings of the NoDaLiDa 2017 Workshop on Processing Historical Language*, szerk. Gerlof BOUMA és Yvonne ADESAM (Gothenburg, 2017), 54–58.

³⁵ ROSSON, „Reception Reader”.

ja az előtag-alapú (*prefix*) és az úgy nevezett „hibatűrő” (*fuzzy*) keresést is, így a csak megközelítőleg azonos esetek is megjeleníthetők. A keresés indítása után a rendszer az első száz találatot mutatja, amelyek relevancia szerint, a találati feltételek alapján rangsorolva jelennek meg. Minden találatnál feltűntetésre kerül a dokumentumazonosító, a kiadás éve, a szerző neve és a cím rövidített alakja. Az oszlopokban található adatok tetszés szerint, az oszlopok fejlécéreire kattintva, növekvő vagy csökkenő sorrendbe rendezhetők. A dokumentumazonosítóra kattintva a felhasználó kiválaszthatja azt a művet, amelyet az elsődleges elemzéshez szeretne használni.

Miután kiválasztottuk az elsődleges dokumentumot, megjelenik a fő kezelőfelület, amely bal oldalon egy „méhraj” diagramot (*bee swarm chart*), jobb oldalon pedig egy oldalkép-megjelenítőt tartalmaz. A méhraj diagram minden egyes pontja egy szöveg-újrafelhasználási esetet jelöl. Alapértelmezés szerint az ábra a „kimenő” kapcsolatokat mutatja, vagyis azokat az újrafelhasznált szövegrészeket, amelyek a vizsgált dokumentummal egy időben vagy annál később megjelent művekben fordulnak elő. Amennyiben a felhasználó arra kíváncsi, hogy korábbi művek szövegei hol bukkannak fel az elsődleges műben, akkor átválthat a „bejövő” nézetre. Az x tengely a kapcsolódó művek kiadási évét ábrázolja, míg az y tengely az azonosított részletnek az elsődleges dokumentumon belüli helyzetét mutatja (legtöbbször oldalszám alapján). Például ha egy 100 oldalas mű 40. oldaláról idéztek később, akkor az y tengelyen a 40-es értéknél jelennek meg az idéző művek adatpontjai. Az egyes pontok színe a kiadások közötti időbeli távolságot tükrözi: a „kimenő” nézetben a kék szín a vizsgált dokumentumhoz közeli kiadási dátumokat, míg a piros pontok fél évszázaddal vagy még később megjelent műveket jelölnek – lásd az 1. ábrán, ahol a szürke minél világosabb árnyalata felel meg a minél közelebbi időbeli távolságnak.

A felhasználó megadhat konkrét évszámtartományokat, így csak az adott időszakon belül megjelent könyvek pontjai maradnak láthatók. A diagram egy szimuláció segítségével kissé eltávolítja egymástól a pontokat, hogy ne fedjék le egymást – ezáltal kirajzolódnak a szövegátvétel időbeli mintázatai. Azok a szövegrészek, amelyeket az adott időszak során többször is idéztek, vízszintesen kirajzolódó pontsorokként jelennek meg egy meghatározott oldaltartományban (y-érték). A függőleges sávok segítségével pedig azonosítható a szövegrészlet különböző időszakokban előforduló újrafelhasználása. Egy oldalirányban kiszélesedő „kúpszerű” alakzat fokozatosan növekvő érdeklődést jelezhet a vizsgált szöveg egy adott szakasza iránt az idő előrehaladtával. Ha a felhasználó egy adatpontra viszi a kurzort, megjelenik egy információs buborék, amely



1. ábra: A *Reception Reader* megjelenítője. Az x tengely a kiadási évek szerint rendezi a szövegeket, míg az y tengely az elsődleges dokumentum oldalszámai alapján az azonosított részleteket. Minden adatpont egy dokumentumot jelöl, amelyben szövegátvétel azonosítható.

tartalmazza a kiadás évét, a szerző nevét, a címet és az elsődleges dokumentum oldalszámát. Ha pedig rákattint egy pontra, a jobb oldali nézetben – amennyiben elérhető – megjelennek a szkennelt oldalképek is, amelyeken kiemelve látható az újrafelhasznált szövegrész mind az elsődleges, mind a kapcsolódó dokumentumban. Ez lehetővé teszi annak vizsgálatát, hogy az idézett szakaszt hogyan illesztették be vagy kommentálták későbbi művekben.

A méhrajdiagram alternatívájaként választható táblázatos nézet is, amely soronként tünteti fel ugyanezeket a kapcsolatokat: minden sor megfelel egy diagramon szereplő pontnak. A sorok tartalmazzák az újrafelhasznált szövegrészek adatait (a kiadás évét, szerző nevét és a mű címét), valamint az eredeti dokumentum idézett oldalszámát. Itt szintén rendezhetők a sorok az oszlopfejlécekre kattintva, illetve kereshetők a böngésző beépített szövegkeresőjével. A táblázat egyik sorát kiválasztva ugyanaz a nézet jelenik meg jobb oldalon, mint a diagram nézet esetében: a kiemelt, újrafelhasznált szövegrészek oldalképei. Összességében ezek a funkciók ötvözik a szöveg-újrafelhasználási mintázatok közvetlen vizuális elemzését a beszkennt forrásanyagok részletes megtekintésének lehetőségével. Így a történések átfogó képet kapnak arról, hogy egy művet miként idézték, dolgozták át vagy bírálták különböző időszakokban. A *Reception Reader* visszajelzési lehetőséget is kínál a felhasználóknak, amely hozzájárul az eszköz további fejlesztéséhez és finomhangolásához.

6. A *RECEPTION READER* KORLÁTAI

A *Reception Reader* innovatív vizualizációs megoldása különösen hasznos, mivel kiemeli az ismétlődő szövegátvételeket, és jól szemlélteti, hogyan idéztek vagy alakítottak át részeket az idők során, ami végső soron a gondolatok és eszmék evolúciójába adhat betekintést. Mindezt pedig a hagyományos lineáris narratívákhoz képest összetettebben képes láttatni: az összefüggések időbeli és tematikus hálózata ugyanis bonyolult kapcsolatrendszerekre világít rá, kiemelve a hivatkozások sűrűségét és időrendi eloszlását. Ugyanakkor fontos figyelembe venni, hogy a digitális bölcsészeti eszközök vizualizációi sohasem objektívek vagy semlegesek. Többek között a szoftverfejlesztés során hozott *design* döntések – például a színskála, a pontméret, az adatelrendezés vagy a szűrés alapbeállításai – óhatatlanul befolyásolják az értelmezést, és elhomályosíthatnak más-különb fontos jellemzőket. Például a felhasználóknak tudniuk kell, hogy a *Reception Reader* alapértelmezetten a legjobb egyezési pontszám alapján rende-

zett első száz találatot jeleníti meg. A méhrajdiagram sűrűsége vagy a színátmenetek hatására a kutatók esetleg túlhangsúlyozhatják egyes szövegrészek jelentőségét, vagy éppen figyelmen kívül hagyhatják az adatokban lévő következetlenségeket és hiányokat. A méhrajdiagram sűrűsége és dokumentumcsoportosításai azt hivatottak jelezni, hogy „mely könyvrészeket tárgyalták a legtöbbet, és melyeket nem”, valamint hogy időbeli trendeket világosítsanak meg. A diagram célja a recepcióban megjelenő újrachasznosítás „volumenének” vizualizálása. Ugyanakkor ez a „valódi volumen” torzulhat, mivel járulékos elemek – mint például kiadói impresszumok vagy reklámok – is gyakran újrachasznált szövegekként jelennek meg, és előfordulhat, hogy egyetlen dokumentum a forrásszöveg több részét is kommentálja, melyet később újra kiadnak; ez felnagyítja az újrachasználás mértékét és torzíttja a vizuális reprezentációt. A közeljövőben a *Reception Reader* alkotói terveik szerint összevonják a szomszédos újrachasználási eseteket, kiszűrik a zavaró járulékos szövegelemeket, és finomítják a diagramkészítési mechanizmust, hogy az adatok alapján pontosabb vizuális narratívát nyújtsanak.

A *Reception Reader* jelenlegi verziója dokumentumszintű áttekintést nyújt a szövegátfedésekről, ezért a kutatók nem tudják például egy könyvfejezet utóéletét külön vizsgálni. A szöveghiányokat illetően fontos megemlíteni, hogy az eszköz szándékosan kizárja ugyanazon szerző más műveit, valamint a teljes egészében újranyomott szövegeket. Emellett nem célja a programnak az allúziók vagy szemantikai hasonlóságok azonosítása sem. Mindezek miatt elengedhetetlen az ábrákon alapuló felismeréseket kiegyensúlyozni a szoros olvasás eredményeivel, amelyhez maga az eszköz is támogatást nyújt.

A BLAST-alapú módszer lenyűgözően átfogó, mivel szinte minden releváns szövegátfedést képes azonosítani. A készítőik azonban hangsúlyozzák, hogy a jelentős OCR-zaj miatt az eredmények nem hibátlanok. A *Reception Reader* nem feltétlenül azonosítja a két jelenleg használt digitális gyűjtemény összes újrachasznált szövegrészletét; nem is beszélve arról, hogy a gyűjtemények lefedettsége is hiányos az *English Short Title Catalogue*-hoz viszonyítva. A lábjegyzetek gyakran kimaradnak, és a sokszor ismétlődő szövegrészek pedig felnagyíthatják az újrachasználási esetek számát. Egyes hosszabb átvett szövegrészek töredezett formában jelennek meg, mivel lábjegyzetek, fejlécek, valamint az OCR hibái szét darabolják őket. Ráadásul a páronkénti hasonlóságvizsgálat nem veszi figyelembe a szövegkölcsonzés történeti irányát, és számos „duplikált” eredményt hoz létre, amikor egy szövegrész sok dokumentumban is újrachasználásra kerül. Noha ezek a tényezők megnehezítik az elemzést az adatok mennyiségének és töredezettségének növekedésével,

együttal rámutatnak arra is, mennyire fontos a kritikus értelmezés: az eszmétörténészeknek a detektált példákat továbbra is manuálisan kell ellenőrizniük és kontextusba helyezniük, nem támaszkodhatnak kizárólag a módszer automatikus kimenetére.

7. HÁROM ESETTANULMÁNY

Ebben a fejezetben három esettanulmányt mutatok be, amelyek szemléltetik a *Reception Reader* használatát, és kiemelik a legfontosabb kutatási eredményeket. Az első tanulmány címe *A Comparative Text Similarity Analysis of the Works of Bernard Mandeville*, amely Bernard Mandeville műveinek és recepciójának feltárására törekszik nagyszabású szöveg hasonlósági vizsgálat segítségével.³⁶ Az elemzés empirikus alapokon járul hozzá az eszmétörténeti diskurzushoz azáltal, hogy újraértékeli Mandeville hatását a 18. századi brit és európai gondolkodásra. Pontosabban Mandeville szövegátemelési szokásait, hatásmintázatait és a szerzőség kérdéseit vizsgálja, különös tekintettel az *A Modest Defence of Publick Stews* című művére. Ez utóbbi kapcsán a kutatás felismerése, hogy Mandeville előszeretettel kölcsönöz saját korábbi műveiből is, miközben a *Publick Stews* esetében ilyen átvételek nem találhatók, ezáltal – a negatív bizonyíték elvét használva – megkérdőjelezhető a szöveg Mandeville-nek tulajdonított szerzősége. Ez a megközelítés árnyaltabb szerzőségvizsgálatot eredményezett a pusztán stilisztikai elemzésre építő hagyományos módszereknél.

A módszer továbbá feltárta Mandeville-recepció mintázatait is: írásai nemcsak politikai és filozófiai művekre hatottak, hanem irodalmi és vallási szövegekben is visszhangoztak. Az intertextuális hálózatok segítségével a tanulmány bemutatja, hogyan terjedtek el és alakultak át Mandeville gondolatai az idők során. Mandeville íásaiból 2557 mű (3989 kiadásban) vett át szövegrészeket: például Frederick Eden (akinek társadalmi kérdésekről, például a szegénységről szóló írásai természetes módon kapcsolódnak Mandeville-hez), a régiségtudós John Smith (1747–1807), valamint a whig történész Catharine Macaulay. Az elemzés meggyőzően bizonyítja, hogy a szövegújrafelhasználás-felismerés kulcsfontosságú az eszmétörténet számára, mivel skálázható módszereket kínál az eszmék terjedésének nyomon követésére, a hatás mérésére és a kritikai kiadások hatékonyabb elkészítésére.

³⁶ Yann Ciarán RYAN, Ananth MAHADEVAN, és Mikko TOLONEN, „A Comparative Text Similarity Analysis of the Works of Bernard Mandeville,” *Digital Enlightenment Studies* 1, 1. sz. (2023): 28–58, <https://doi.org/10.61147/des.6>.

A második esettanulmány Mark G. Spencer és Mikko Tolonen kutatása, amely David Hume *Essays* című művének 18. századi brit fogadtatását tárgyalja.³⁷ Megközelítésük ötvözi a hagyományos eszmetörténeti módszereket az automatikus szövegújrafelhasználás-elemzéssel. A kutatás számos, korábban figyelmen kívül hagyott esetet azonosított, amikor Hume esszéit átvették, parafrázták vagy újranyomtatták, ami befolyásolta a közéleti diskurzust a kereskedelem, kormányzás és szabadság kérdéseiről. A szerzők bizonyítják, hogy Hume művei sokkal szélesebb körben terjedtek, és jóval nagyobb hatást gyakoroltak a korabeli értelmiségi kultúrára, mint azt korábban gondolták. Esszéit gyakran név nélkül terjesztették, és váratlan helyeken – például antológiákban, fejezetmottókban vagy folyóiratokban – is felbukkantak. Ezek a szövegátvételek, amelyeket szerkesztők gyakran új kontextusba helyeztek, alakították a kortárs viták kereteit. A számítógépes és a kézi elemzés ötvözésével a tanulmány ezúttal is új utakat nyit a kora újkori szövegek recepciójának és átalakulásának nyomon követésére, és árnyaltabb képet ad az eszmék nyilvános térben való terjedéséről és evolúciójáról.

A harmadik esettanulmány Kira Sophie Hinderks mesterszakos diplomamunkája, amely a *Reception Reader* használta a 18. századi utópisztikus művek kapcsolatainak elemzésére.³⁸ Kutatásának célja, hogy e szövegátvételi gyakorlatokat összekapcsolja a tágabb nyomtatott kultúrával, az eszmetörténettel és a társadalmi-politikai diskurzussal. A tanulmány azt vizsgálja, miként használták az utópisztikus szerzők a szövegátvételt az értelmiségi hagyományok megerősítésére vagy megkérdőjelezésére. Például Thomas Northmore 1795-ben megjelent *Memoirs of Planetes* című művében szándékosan idézett olyan klasszikus szerzőket, mint Cicero és Tacitus, akiket hagyományosan Nagy-Britannia kevert államformájának és a parlament legfőbb hatalmának igazolására alkalmaztak. Northmore azonban e bevett érvek kiforgatására törekedett: ugyanazon szövegeket arra használta, hogy bírálja Nagy-Britannia politikai berendezkedését, és a korabeli radikális eszmének számító népszuverenitás mellett érveljen. A *Reception Reader* kulcsszerepet játszott a kutatásban, bár korlátai is megmutatkoztak: például az OCR-zaj, illetve a kiadói impreszumok és hirdetések ismétlődéséből fakadó álpozitív találatok rontották a pontosságot. Az első két esettanulmányhoz hasonlóan, ezeket a hiányosságokat kézi ellenőrzéssel és műfajalapú kategorizálással sikerült enyhíteni és így

³⁷ Mark G. SPENCER és Mikko TOLONEN, „The Reception of Hume’s Essays in Eighteenth-Century Britain”, in *Hume’s Essays*, szerk. Max SKJÖNSBERG és Felix WALDMANN (Cambridge: Cambridge University Press, 2025): 15–35, <https://doi.org/10.1017/9781009047227.003>.

³⁸ Kira Sophie HINDERKS, *Cromwell on the Moon; Or, Printing, Popularity, Persuasion*, MA tézis, Uppsala University, 2023.

biztosítani az elemzések megbízhatóságát. Hinderks a hagyományos szoros olvasás módszerét alkalmazta az újrafelhasznált szövegrészek tágabb történeti kontextusban való értelmezésére. Kutatása feltárta, hogy az utópisztikus művekben a szöveg-újrafelhasználás nem véletlenszerű, hanem gyakran tudatos stratégia volt az intellektuális hagyományokhoz való kapcsolódás vagy azok kritikája céljából. Az utópisztikus szerzők az átvett szövegeket a tekintély jelzésére, a kortárs politikai vitákba való bekapcsolódásra és társadalom-kritikai nézőpontjaik kiemelésére használták. Hinderks szerint „a szövegújrafelhasználás-felismerő eszközök – mint a *Reception Reader* – legnagyobb erénye abban rejlik, hogy felszínre és előtérbe hozzák ezeket a rejtett szövegeket, így lehetővé teszik a történészek számára, hogy szakterületi tudásukkal értelmezzék a feltárt összefüggéseket.”³⁹

8. ÖSSZEGZÉS ÉS JÖVŐBELI KILÁTÁSOK

Az elmúlt két évtized tömeges szövegdigitalizálási hulláma hatalmas digitális archívumokat hozott létre, amelyek teljes körű feldolgozása új, adatalapú módszereket igényel. Ezek közé tartozik a szövegújrafelhasználás-felismerés is, amely kulcsfontosságú eszköz a recepció és a szövegek utóéletének feltárására: segítségével azonosíthatók az idézetek, átvételek, parafrázisok és szerkesztői módosítások mintázatai. Amint azt a fentiekben bemutattuk, a *Reception Reader* különösen hasznos lehet ebből a szempontból, mivel lehetővé teszi az eszmetörténészek számára a közeli és távoli olvasás együttes alkalmazását. A kutatók nemcsak a visszatérő szövegátvételek időbeli eloszlását vizsgálhatják, hanem azok kontextusát és más művekben való újrakontextualizálását is. Ez a képesség jelentősen gazdagítja az intellektuális hálózatok, az ideológiai átalakulások és az eszmék fejlődésének vizsgálatát. Emellett új lehetőségeket nyit a 17–18. századi hivatkozási stratégiák és a szövegátadási módok feltérképezésére, összekapcsolva kevésbé ismert műveket a kor meghatározó társadalompolitikai vitáival és kulturális gyakorlataival.

A *Reception Reader* módszertani és vizuális kialakítása jelentős technológiai újítást képvisel: a BLAST szekvenciaillesztő algoritmusra épülő robusztus szövegújrafelhasználási eljárás segítségével képes leküzdeni a nagyméretű és „zajos” korpuszokban – például az ECCO-ban és az EEBO-ban – előforduló OCR-hibák okozta kihívásokat. A jövőbeli fejlesztések célja a rendszer korpuszáinak bővítése – például a British Library *Burney* és *Nichols* gyűjteményei-

³⁹ Uo., 36.

nek történeti újságanyagaival –, valamint az ECCO és a brit digitális sajtóarchívumok közötti szövegátfedések azonosítása.⁴⁰ Ez az integráció tovább növeli az eszköz jelentőségét a történeti, irodalmi és kulturális kutatások területén. A *Reception Reader* egyedülálló digitális ökoszisztémája példaértékű interdiszciplináris együttműködés eredménye adattudósok, történészek és nyelvészek között. Ráadásul a platform szabadon elérhető felületének köszönhetően demokratizálja a tudományos kutatáshoz való hozzáférését. A *Reception Reader* segít megkerülni az alapjául szolgáló adatbázisok fizetős falait, így használata az anyagi erőforrásokban szűkölködő intézmények kutatói számára is új lehetőségeket kínál.

A *Reception Reader* jövőbeli fejlesztési irányai kapcsán a COMHIS-csoport nagy nyelvi modellek integrálását tervezi, hogy a szemantikai hasonlóságot is felismerje, így a szó szerinti szövegátvételeken túl a parafrázisok és allúziók azonosítása is lehetővé váljon. Ez az innováció árnyaltabb, „eszmeközpontú” elemzést kínál, amely átformálhatja a meglévő értelmezéseket, és eddig ismeretlen intellektuális hatásokat tárhat fel. Emellett a COMHIS-csoport egy kiadás-összehasonlító eszköz fejlesztésén is dolgozik, amely segít felmérni ugyanazon mű több kiadásán végbement változásokat – ez különösen nagy segítséget nyújt a kritikai kiadások készítőinek. E fejlesztéseket a gyakorlati munka során tesztelik: a kutatócsoport készíti többek között David Hume *The History of England* című munkájának kritikai kiadását az Oxford University Press gondozásában. E fejlesztések rávilágítanak a számítógépes szöveg-elemzés átformáló erejére: új lehetőségeket kínálnak az eszmetörténészek számára az eszmék, szövegek és hagyományok összekapcsolt evolúciójának nyomon követésére, olyan léptékben, amelyre korábban nem volt példa.

⁴⁰ MAHADEVAN, „Optimizing a Data Science System for Text Reuse Analysis,” *arXiv*, 2024.