
TANULMÁNYOK

MARÓTHY SZILVIA

A vizuális és az adatközpontú szemlélet

Bevezető gondolatok

A *Helikon* 1969/3–4. száma ugyan kivételesen nem volt tematikus, ám több tanulmány és recenzió kapcsolódott a „számítógép és a humán tudományok” témaköréhez, ezt emelte ki a címlap is. Közel fél évszázada Petőfi S. János a következő kutatási területekre hívta fel a figyelmet rövid bevezetőjében: automatikus szöveg-előállítás, gépi dokumentáció, számítógépes nyelvelemzés, számítógépes stílus-elemzés.¹ A *Helikon* jelen *Számítógépes irodalomtudomány* című száma is bizonyítja, hogy ezek a területek a digitális bölcsészet fogalmának és intézményesülésének megjelenésével mit sem változtak.²

A tágabb kontextust egy Tito Orlandival készült interjú fordítása adja, mely többek között a digitális bölcsészeti, s egyben a számítógépes filológiai kutatások kanonizált és mitizált ősforrását, Roberto Busa kutatásainak tényleges hozadékait vitatja. Miben rejlik a valódi paradigmaváltás, milyen volt és milyen ma a bölcsészettudósok viszonya a számítógépes bölcsészettudományokhoz, hogyan változott a számítógépes bölcsészettudományokkal foglalkozó tudományos közösség az évtizedek során? Az interjú a digitális bölcsészeti kutatások elmúlt közel hét évtizedébe nyújt személyesebb hangvételi betekintést. A számítógépes irodalomtudományról általában, illetve a diszciplínát ért új keletű támadásokról, s azok lehetséges elhárításáról Almási Zsolt ír.

A cikkek zöme a számítógépes filológia elméleti és gyakorlati kérdéseit tárgyalja. A téma elsődlegessége jól mutatja, milyen intenzíven foglalkoztatja a történeti források számítógépes feldolgozása a kutatóközösséget. De rávilágít arra is, hogy ezen közösség legfontosabb intézménye, a Textológiai Munkabizottság 2004-ben közzétett kiadási alapelvei³ óta elhanyagolta az eltelt másfél évtizedben rohamosan fejlődő kutatási ágat. Jelen tematikus számban Golden Dániel az utóbbi évtizedek eredményeit, beváltott és megvalósításra váró ígéreteit tekinti át, s amellet érvel, hogy a kulturális örökség digitalizációjában korunk humanis-

¹ PETŐFI S. János, „A számítógépek és a humán tudományok”, *Helikon* 15, 3–4. sz. (1969): 372–373.

² A *Helikon* ezen számáról és a folyóirat más, a számítógépes bölcsészet témájába vágó kiadványairól Golden Dániel emlékezik meg részletesebben jelen *Helikon*-szám lapjain.

³ „Alapelvek az irodalmi szövegek tudományos kiadásához”, *Irodalomtörténet* 35, 3. sz. (2004): 328–330. A problémáról, illetve az *Alapelvek* alkalmazhatóságáról lásd: MARÓTHY Szilvia, „Tudományos szövegkiadások a hálózaton: Áttekintés”, *Irodalomtörténeti Közlemények* 122, 5. sz. (2018): 617–633.

táinak, filológusainak különösen nagy szerepe és felelőssége van. Kalcsó Gyula a kéziratok források tudományos elektronikus kiadásának elméleti kérdéseivel foglalkozik, követendő kiadási elveket fogalmazva meg. Fellegi Zsófia és Palkó Gábor az Arany János Emlékév során készült digitális szövegkiadások elkészítésének és webes publikálásának komplex munkafolyamataiba nyújt betekintést. Szénási Zoltán szemléje pedig a hazai tudományos elektronikus szövegkiadások három nagy műhelyének (az ELTE BTK volt Bölcsészinformaticai Önálló Programja, illetve az MTA–DE Klasszikus Irodalmi Textológiai Kutatócsoport és a DigiPhil) kiadványait mutatja be.

*

De mihez képest hozott változást a számítógépes rögzítés, feldolgozás és megjelenítés? A '70-es és '80-as években zajló textológiai viták középpontjában az állt, hogy a pozitivizmus örökségét, a kronologikus irodalomtörténeti narratívát a szövegek elgondolásában, értelmezésében, kontextusba ágyazásában és gondozásában felválthatjuk-e, s ha igen, mivel. Az 1976-ban Horváth Iván sajtó alá rendezésében megjelent Balassi-kiadás az eredeti kompozíció, szerzői intenció szerint – illetve azt platonikus értelemben „eszményítve” – közli a verseket. Mint utószavában írja:

Azért nevezhető kísérletinek ez a kiadás, mert – a korábbi kiadói gyakorlatól eltérően – Balassi Bálint verseit nem a soha pontosan meg nem állapított kronológia alapján, hanem elvszerűen a költő saját versgyűjteményeinek (ciklusainak) sorrendjében közli.⁴

A kiadás a szemléletváltással nem csupán azt kezdeményezi, hogy a versek időbeli elhelyezése helyett azok költői dispositiójára irányuljon a figyelem, hanem (Bori Imre szavaival) a költő „művészi öntörvényeinek leírása kerül előtérbe az eddigi életrajz-központú szemlélet ellenében.”⁵ Szigeti Csaba az *Appendix Balassianá*-ban a kronologikus gondolkodás problematikusságát a régiség irodalmát illetően Koháry István és Beniczky Péter költészetén mutatja be. Alább elméleti megfonto-

⁴ HORVÁTH IVÁN, szerk., *Balassi Bálint összes versei: a versek helyreállított eredeti sorrendjében* (Novi Sad: Újvidéki Egyetem, 1976), 143. Az eszményített kompozíció egyúttal a Gerézdi Rabán és Klaniczay Tibor 3×33-as és Varjas Béla 2×33-as szerkezeti hipotézisére adott válasz.

⁵ Uo., 145. Bori Imre a kiadáshoz írt előszóban hozzászól, hogy szándékuk nem magában álló, ez az igény az 1976-ban Egerben megrendezett Balassi-értekezleten is megfogalmazódott. Vö. *Irodalomtörténeti Közlemények* 80, 5–6. sz. (1976). Az *Itk* ezen számában Szörényi László Balassi török bejtjeit mint kompozíciót elemzi. Tanulmányának keletkezéséről ezt írja: „Ez a dolgozat a Balassi Bálint életművét tárgyaló tudományos ülészakon, Egerben, 1976. május 14-én elhangzott előadás kibővített változata. Létrejöttében szerepet játszott Horváth Iván baráti készítése, aki e sorok szerzőjével együtt már évek óta törekszik arra, hogy a klasszikus költők életművének kiadásában jelenleg kizárólagos időrendi elv mellett a kellő súllyal érvényesüljenek a kötetkompozíció követelményei.” SZÖRÉNYI László, „»Valahány török bejtje«”, *Irodalomtörténeti Közlemények* 80, 5–6. sz. (1976): 706–713, 706.

lásaiból ragadok ki szemelvényeket a kronologikus és a kompozíciós értelmezés egymásnak feszülését illusztrálándó.

[... hogy] a kronologikus sor szemléletszerűen mennyire áthatja az irodalomtörténetet, úgy gondolom, bizonyítás nélkül is megálló közhely és igazság.

[...] az adatvirágok elvileg végtelen mennyiségben fűzhetők fel a relatív kronológia madzagjára, ami a filológus számára a gazdagság (a kimeríthetlenség) érzetének alapját alkotja meg. Az adat vagy tény, esetleg szöveg beillesztése az időrendi sorba: a behelyezés. E művelet steril irodalomtörténetész eljárásnak tűnik, holott – könnyen belátható, hogy – ideologikus, manipulatív gesztus, sőt minden későbbi értelmezés bázisa. [...] A kronologikus gondolkodású irodalomtörténetész lényegében egyetlen pillanat érdeklí, egyetlen pillanatot értelmez: a mű keletkezésének pillanatát.⁶

Noha a kronologikus gondolkodást tekinthetjük egyeduralkodónak az ezredforduló környékéig,⁷ fontos látni, hogy a textológiai kritikában korábban is fel-felmerült ennek tarthatatlansága. Szigeti Csaba idézett cikkében Toldy Ferenc egy közleményét hozza fel példaként, melyben felhívja a figyelmet Beniczky addig ismeretlen korai költeményeire, s a *Magyar rithmusok* verseinek feltehetőleg jóval korábbi keletkezésére. Ez alapján módosítja a Beniczkyról, a kötet korabeli költészetébe ágyazottságáról alkotott képet: „Beniczkyt tehát ezen túl az irodalmi történetíró nem Zrínyi után, hanem elibe fogja helyezni, s Rimai mellé, kinek ifjabb kortársa volt.”⁸

Horváth Iván fent hivatkozott, Balassi kötetkompozíciójával foglalkozó cikkében Klaniczay Tibor 1957-es tanulmányát idézi, melyben Eckhardt kiadása apropóján kitér Balassi és Rimay nagykompozíciós törekvéseire is. Mint írja, Balassi és filológiája ebben is úttörő: „A kronologizmusnak Balassi vetett véget. Klaniczay Tibor (1957) javaslatot tett a költő »maga kezével írt könyvének« verssorrendjére.” Klaniczay így foglalja össze hozzászólásának célját: „felülvizsgálom Balassi- és Rimay kéziratos és nyomtatott szöveg hagyományának egész problematikáját, s ezen az úton megpróbálom a kronológia és szövegkritika terén eddig elért ered-

⁶ SZIGETI Csaba, „Appendix Balassiana: Kronológia, tradíció, hagyománytudat a XVII. századi Balassi-követő nemesi költészetben”, *Irodalomtörténeti Közlemények* 89, 6. sz. (1985): 675–687, 675–676.

⁷ A narratológiai-textológiai fordulatról lásd legújabban Horváth Iván cikkének bevezetőjét, melyben nem annyira a pozitívizmus örökségével való leszámolásként, inkább a megszállás időszakájának ideológiai meghatározottságától való elszakadásként exponálja a bekövetkezett változást. HORVÁTH Iván, „Balassi Bálint verseinek fragmentumi”, *Credo* 24, 3–4. sz. (2018): 18–39.

⁸ TOLDY Ferenc, „Beniczky Péter kora, s némely fontos kicsiség”, *Új Magyar Múzeum* 3, 1. sz. (1853): 352–356, 355. Igaz, ahhoz, hogy Rimay kortársának tekintsük, el kellene fogadnunk Toldy azon spekulációját („De ha okoskodni szabad – pedig szabad a historiában is, adatok létében és nem létében! ...”), hogy Beniczky 1632-re elhunyt. Ezt azonban csupán egy possessorbejegyzésre alapozza, s pár évtizeddel később Komáromy közlése nyomán kiderült, Beniczky 1664-ben hunyt el. KOMÁROMY András, „Adalékok Beniczky Péter életéhez”, *Történelmi Tár* 11, 3. sz. (1888): 435–449, 448.

ményeket továbbfejleszteni.”⁹ Mind Toldy, mind Klaniczay esetében láthatjuk, hogy a szándék nem forradalmi, nem törekszik kizárólagosságra: a kronológiai és a kompozíciós értelmezői narratívát kívánja ötvözni.

A számítógépes szövegfeldolgozás lehetőségével nemcsak a tudományos szövegkiadás fogalma kérdőjeleződött meg, de a szövegé is. A médiumváltással összefüggésben jelenik meg a szöveg pluralitásának felismerése,¹⁰ vagy legalábbis láthatóvá tétele, a szöveg határainak megkérdőjelezése és a hipertext jelenség, valamint a szöveg mint hierarchikus felépítmény modellezhetősége is.¹¹ Bernard Cerquiglini szövegvariánsoknak szentelt könyvében már kijelöli azokat a lehetőségeket, melyeket a számítógépes feldolgozás a könyv megjelenésekor (1989) ígért a középkor filológiája számára¹² – például a szövegvariánsok párhuzamos megjelenítését, az annotálás határtalanságát, indexek és konkordanciák generalitását. Karina van Dalen-Oskam mintegy negyedszázaddal később tanulmányában azt vizsgálja, milyen módon és mértékben valósultak meg ezek:¹³ a párhuzamos megjelenítésre ma számos példát találhatunk, a különféle annotációknál pedig legfeljebb az okozhat gondot, hogyan helyezték el azokat, milyen interaktív lehetőségeket biztosítsanak a felhasználó számára, hogy a befogadás gördülékenyen menjen.

⁹ KLANICZAY Tibor, „Hozzászólás Balassi és Rimay verseinek kritikai kiadásához”, *MTA Nyelv- és Irodalomtudományok Osztály Közleményei* 17, 1–4. sz. (1957): 265–338, 266.

¹⁰ A plurális szövegfogalomhoz lásd: Bernard CERQUIGLINI, *In the Praise of the Variant*, trans. Betsy WING, (Baltimore, London: John Hopkins University Press, 1999). Eredeti kiadvány: Bernard CERQUIGLINI, *Éloge de la variante: Histoire critique de la philologie* (Paris: Ed. du Seuil, 1989); HORVÁTH Iván, „Szöveg”, 2000, 11. sz. (1994): 42–53; HORVÁTH Iván, „A hálózati kultúra fenomenológiája: Egy általános textológia vázlata”, *Gépeskönyv*, 2001, <http://magyar-irodalom.elte.hu/vita/thi2.html>. Ebben a szellemben készült továbbá a régi magyar versek repertórium is: *Répertoire de La Poésie Hongroise Ancienne I–II*, dir. Iván HORVÁTH (Paris: Le Nouvel Objet, 1992). Hálózati kiadás: ua., 4.0 v, hozzáférés: 2020.03.21, <http://rpha.oszk.hu/>.

¹¹ Steven J. DE ROSE, David G. DURAND, Elli MYLONAS and Allen H. RENEAR, „What is Text, Really?”, *Journal of Computing in Higher Education* 1, no. 3. (1990): 3–26. Az OHCO-modellhez és verzióihoz lásd még: Allen RENEAR, Elli MYLONAS and David DURAND, „Refining Our Notion of What Text Really Is”, 1993, <http://cds.library.brown.edu/resources/stg/monographs/ohco.html#sec13>. Filológiai történeti áttekintés: LABÁDI Gergely, „A filológiai tudás formái”, in *Textológia, filológia, értelmezés: Klasszikus magyar irodalom*, szerk. CZIFRA Mariann és SZILÁGYI Márton, Csokonai Könyvtár: Bibliotheca Studiorum Litterarium 55, 173–190 (Debrecen: Debreceni Egyetemi Kiadó, 2014).

¹² Illetve részint be is váltott addigra, lásd pl. a Roberto Busa által az 1950-es években kezdeményezett *Corpus Thomisticum*ot, vagy az 1970-es években indult már idézett *Répertoire de La Poésie Hongroise Ancienne*-t. Roberto BUSA, „L’Index Thomisticus e l’informatica filosofica”, *Revue Internationale de Philosophie* 27, no. 1. (1973): 31–36; RPHA MUNKACSOPT, „Szegedi kísérlet a XVI. századi magyar vers gépi feldolgozására”, *Irodalomtörténeti Közlemények* 84, 5–6. sz. (1980): 630–638.

¹³ Karina VAN DALEN-OSKAM, „In Praise of the Variant Analysis Tool: A Computational Approach to Medieval Literature”, in *Texts, Transmissions, Receptions: Modern Approaches to Narratives*, eds. André LARDINOIS, Sophie LEVIE, Hans HOEKEN and Christoph LÜTHY, *Radbound Studies in Humanities* 1, 35–54 (Leiden: Brill, 2015), <https://www.jstor.org/stable/10.1163/j.ctt1w76wgh.7>. A jelen folyóirat számban hasonlóra tesz kísérletet Szenási Zoltán cikke.

A visszakereshetőség az, ami talán a leginkább felülmúlta a korabeli várakozásokat, hiszen nemcsak tetszőleges indexek generálhatók a rögzített adatokból, hanem a rendelkezésre álló lekérdező eszközök és a ráépülő szoftveres támogatás (például adatvizualizációk) is olyan rálátást adnak a szövegekre és azok adataira, melyek új, a korábbi eszköztárral nem hozzáférhető felismeréseket hozhatnak. Jelen pillanatban azonban a filológia még mindig időigényes és csak részben automatizálható műveletek sorából áll, így a nagy méretű (*big data* szintű), filológiai-
lag is megbízható szövegtörzsekért, s az azokon végezhető vizsgálatokért még dolgozni kell.

Az elmúlt közel huszonöt év szövegkiadásából válogatva¹⁴ úgy tűnik, többségben vannak azok, amelyek nem kezelik külön, tehát nem regisztrálják és jele-
nítik meg az eredeti forrásbeli szövegsorrendet. Ehelyett jellemzően tematikus, szerkezeti és formai szempontok szerint indexált, illetve kereshető megjelenítés-
sel találkozunk, különösen a 2010-es évek előtti, HTML-alapú kiadások esetében. A szövegek egymásutánosságára a fájlközpontú (számítógépes adatkezelési egysé-
gekben építkező) szemlélet miatt, mely szerint egy szövegegység egy fájl, ritkán találunk reflexiót. Ez azt jelenti, hogy az egyes szövegforrásokat feldolgozó fájlok nem tartalmazzak információt a szövegforrásban (kéziratban vagy nyomtatvány-
ban) előttük lévő, illetve utánuk következő szövegről, esetleg magáról a szöveg-
forrásról sem. Többnyire ezen adatok is eljutnak más módon (pl. index oldalon, kísérőtanulmányban) a felhasználóhoz, de a szöveg kódolásában ezek az adatok nincsenek explicite jelen.

Az a kérdéskör tehát, hogy milyen módon adjunk közre szépirodalmi szöve-
geket: a kronológia és így közvetetten az életrajzi referencialitás, vagy a szerzői
intenciók elsődleges figyelembevételével, megszűnt problémának látszani. Talán
a szöveg plurális megközelítése vonta el erről a figyelmet. A számítógépes textó-
lógia egy forrásközpontú, pontosabban a források kevésbé hierarchikus felfogású
közreadási tevékenységére váltott át. A számítógépes feldolgozás, illetve a webes
sajtó alá rendezés és megjelenítés eltérő jellege miatt a szövegek feletti kompozí-
ció explicit regisztrálása gyakran csak a kiadásban kódolt szövegeken kívül kap
helyet – a művek tulajdonképpen darabokra esnek, minden drámaiság nélkül.

A helyzet dramatizálására azért sincsen okunk, mert ez a hiányosság nem a
médiium sajátossága. Ha jobban meggondoljuk, explicite a nyomtatott kiadások-
ban sincs általában rögzítve a szövegek egymásutánosságának ténye (tehát az egy-
másutánosság mint a szövegegységek egy tulajdonsága) – egészen addig, míg az
említett fordulatok, a textológia kritika a '70–'80-as években azzá nem tette a

¹⁴ A két legátfogóbb katalógus digitális szövegkiadásokra: Patrick SAHLE, „A catalog of Digital Scholarly Editions, v 3.0” (2008–), <http://www.digitale-edition.de/>; Greta FRANZINI, „Catalogue of Digital Editions” (GitHub, 2012–), https://github.com/gfranzini/digEds_cat, ennek böngészhető változata: <https://dig-ed-cat.acdh.oew.ac.at/>. Cikkem ezen részében a 2019-ben vitára bocsátott disszertációm eredményeire támaszkodom. MARÓTHY Szilvia, *Szerzői verskötetek a 17. századi magyar irodalomban* (ELTE BTK Régi Magyar Irodalom Tanszék, 2019), 139–152, <https://edit.elte.hu/xmlui/handle/10831/44817>.

korábban inkább implicit létmódú állításokat (ti. hogy a költői életmű elemeinek sorrendje kronologikus). A kronologikus irodalomtörténeti gondolkodás hatásal volt a szövegkiadásokra, de fordítva is: a kronologikus szövegkiadások hatással voltak (vannak) az irodalomtörténeti gondolkodásra. A számítógépes filológia adatközpontú szemlélete az, ami felhívta a figyelmet a hagyományos filológia vizualitáshoz, tipográfiai megoldásokhoz való erős kötődésére. A papíralapú kritikai kiadás magától értetődően alkalmaz tipográfiai megoldásokat filológiai megállapítások kifejezésére.¹⁵ Így ezek a megállapítások adatszerűen nincsenek rögzítve a kiadásban, vagy nem a szöveghez kapcsolódnak közvetlenül. Ezt a funkciót is többnyire a kísérőtanulmány látja el.

A kísérőtanulmány számos eleme a számítógépes szövegtárolás esetében annak metaadatai közé vándorol. A TEI Header, mely a kódolt szöveghez kapcsolódó metaadatokat rögzíti,¹⁶ épp annak érdekében jött létre és finomodott az évek során, hogy a tanulmányban összegzett információkat szegmentálja, s ezáltal újrafeldolgozhatóvá, összehasonlíthatóvá és nem utolsósorban a weben könnyebben megtalálhatóvá tegye. Mindez természetesen nem azt jelenti, hogy a kísérőtanulmányt egy metaadathalmaz váltja fel (egyelőre), hanem azt, hogy a számítógépes feldolgozás megköveteli a filológiai eljárásokra való reflektálást strukturált adathalmaz formájában is.

A hagyományos és a számítógépes filológia összevetésére visszatérve, a tudományos kiadások elemzése arra világít rá, hogy a hagyományos filológia vizualitásból kiinduló, vizuális eszközökre alapozó szemlélete a számítógépes eszközökkel létrehozott tudományos kiadásokra is sok esetben jellemző. Bár a kiadások és a filológiai gyakorlatok igen eltérőek, talán mégis ábrázolhatjuk ezt egy folyamatként, mely a vizuálistól az adatszerű reprezentáció felé halad. A folyamat egy jelentősebb állomása a HTML kódolásról az XML kódolásra való áttérés, mely által különvált a tudományos adat és annak megjelenítése. Ezzel párhuzamosan egy jelentős ellenmozgás, visszalépés a HTML-ről a PDF formátumokra való áttérés, mely újra a vizuális kódokat részesíti előnyben. Előrefelé egy következő állomás az XML szintaxis meghaladására törekvő ún. stand-off markup, illetve a Linked Open Data, vagy gráfalapú szemlélet, mely még inkább a szöveg sokféleségének (multiplicitásának és pluralitásának) megragadására koncentrál.¹⁷

Míg a HTML kódolással az alapvető probléma az volt, hogy a tartalom és annak megjelenítése nem válik külön, az XML-lel az, hogy nem kezeli jól az átfedé-

¹⁵ A hagyományos és a digitális filológia szemléleti ütközéseire hoz szemléletes példákat: Magdalena TURSKA, James CUMMINGS and Sebastian RAHTZ, „Challenging the Myth of Presentation in Digital Editions”, *Journal of the Text Encoding Initiative* 9 (2016), <https://doi.org/10.4000/jtei.1453>.

¹⁶ „The TEI Header”, in TEI CONSORTIUM, *P5: Guidelines for Electronic Text Encoding and Interchange*, v 4.0.0, hozzáférés: 2020.03.19, <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html>.

¹⁷ „Stand-off markup”, TEIWiki, hozzáférés: 2020.03.21, https://wiki.tei-c.org/index.php?title=Stand-off_markup.

seket, s a beágyazottság miatt redundánssá, zavarossá válhat.¹⁸ Az XML szintaxis számára megalkotott TEI kódolás pedig úgy látszik, nem egészen váltotta be az olyan vele kapcsolatos reményeket, mint a platformfüggetlenség, az újrafelhasználhatóság, vagy az interoperabilitásra való alkalmasság. Ezért került az utóbbi években a figyelem középpontjába a szövegközi (*inline*) helyett a szövegen kívüli (*stand-off*) kódolás, markup. A módszer lényege, hogy külön rétegekként és külön számítógépes fájlkként is kezeli a nyers szöveget (*plain text*) és a hozzá tartozó szerkezeti, tartalmi, szövegkritikai stb. állításokat. Így lehetővé válik a szöveggel kapcsolatos, egymással a szöveg linearitása miatt átfedésben lévő állítások megfogalmazása. A stand-off maga nem új keletű, az 1960-as évek végéhez és Ted Nelson munkásságához kötik első megjelenését,¹⁹ a humán tudományok területén azonban elsősorban a nyelvészek alkalmazták eddig.²⁰

Az alábbiakban két nem kifejezetten nyelvészeti alkalmazást mutatok be röviden, Desmond Schmidt, valamint Elena Spadini és Magdalena Turska megoldási javaslatát. Schmidt a kódolásnak négy szintjét különíti el: metaadat, szöveg (variánsok), jegyzetelés (értelmezői réteg), szövegtulajdonságok (pl. bekezdések, kiemelések).²¹ A TEI XML kódolás szerint ezek egy fájlban belül találhatóak, az adott szövegforrás(ok)hoz kapcsolódnak. A stand-off módszer szerint azonban szétválnának különálló, más formátumú fájlokra: metaadat, nyers szöveg, különféle annotációk (pl. értelmező jegyzetek, szövegtulajdonságok, szövegkritikai apparátus).

Kérdés azonban, ha a szövegkritikai apparátust leválasztjuk a szövegről (szövegforrásunk átiratáról), mi marad, mi az a nyers szöveg, melyhez az annotációkat kapcsoljuk? Ez ugyanis elvben nem tartalmazhat semmilyen formázást,

¹⁸ Az XML szintaxisához lásd: <https://www.w3.org/XML/>. Példa átfedésre:

Az ifju visszaretent: „Hah! megint te?

Mindig te!” és elfordítá szemét, (*Bolond Istók*, I/44/1–2, AJÖM III/146.)

A két verssorban egy idézet szerepel, mely az egyik versorból a másikba nyúlik át. Ez az átfedés az XML szintaxis szerint nem lehetséges, az egyes elemek egymásba kell hogy ágyazódjanak, itt azonban nem ez a helyzet. Hibás átírás lenne tehát:

<|>Az ifju visszaretent: <q>Hah! megint te?</|><|>Mindig te! </q>és elfordítá szemét,</|>

Ezt át lehet ugyan hidalni többféleképpen, ám egyik sem nyújt természetes, illetve az XML feldolgozás szempontjából hatékony megoldást. Például üres <q> elemek beszúrásával, melyek kezdő (sID) és záró (eID) attribútumot kapnak.

Pl. <|>Az ifju visszaretent: <q sID="idezet1" />Hah! megint te?</|><|>Mindig te!<q eID="idezet1" /> és elfordítá szemét,</|>

¹⁹ Desmond Allan SCHMIDT, „Using standoff properties for marking-up historical documents in the humanities”, *Information Technology* 58, no. 2. (2016): 63–69, 64, <https://doi.org/10.1515/itit-2015-0030>.

²⁰ A megoldás újszerűségét mutatja, hogy a feltehetőleg első nagyobb volumenű TEI XML–stand-off konverzióról egy tavalyi cikk számol be. Giuseppe G. A. CELANO, „Standoff Annotation for the Ancient Greek and Latin Dependency Treebank”, in *Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage – DATeCH2019*, 149–153 (Brussels: ACM Press, 2019), <https://doi.org/10.1145/3322905.3322919>.

²¹ Desmond Allan SCHMIDT, „Standoff properties as an alternative to XML for digital historical editions” (Semantic Scholar, 2016), 1–23, 6, <https://www.semanticscholar.org/paper/Standoff-properties-as-an-alternative-to-XML-for-Schmidt/6056afc3c25fc0e9b3e677c04ea4bc34b8151ab>.

szegmentálást, de szövegváltozatokat sem. Ha egy szövegforrás javításokat, törlést, betoldást tartalmaz, az egyes szövegállapotokat miként rögzítjük stand-off modellben? Schmidt javaslata szerint ilyenkor egy-egy szövegforrásnak több átiratát kell elkészítenünk, melyek megfelelnek az egyes szövegállapotoknak.²² Ez az elképzelés azonban azt feltételezi, hogy az adott szövegegészen esett változtatásokat rétegekké tudjuk szétfésülni. Például a szöveg eleji törlésről és a szöveg végén található betoldásról tudjuk, hogy ugyanazon szövegréteghez tartoznak, tehát közel egy időben keletkeztek – erről ritkán vannak információink. A nyers szöveg Schmidt megoldásában ráadásul – mint azt maga is elismeri – nem teljesen mentes a szövegtulajdonságoktól, metaadatoktól, a paragrafushatár jelölésére például a Markdownhoz hasonlóan a dupla sorközt vagy nagyobb térköz alkalmazását javasolja, illetve az oldalszám nyers szövegben való szerepeltetését bizonyos esetekben indokoltnak tartja.

Bár elméleti szempontból nagyon vonzó a stand-off markup azon elképzelése, hogy a szöveget és a rá vonatkozó állításainkat elkülönítse, ennek textológiai és technológiai gyakorlatba ültetése egyaránt rengeteg kérdést, problémát vet fel. A stand-off ugyanúgy magában rejt a kód sokszor felesleges bonyolításának lehetőségét. Szemléletes példákat hoz a lekérdezés bonyolultságára Spadini és Turska: például míg egy adott versszöveg negyedik sorának XPath lekérdezése TEI XML-ben 13 karakter hosszú, stand-off rendszerben háromsornyi, tehát 2–300 karakternyi is lehet.²³ A szerzők azonban elismerik, hogy a TEI XML is gyakran válik áttekinthetatlenné vagy redundánssá. Elismerve a stand-off előnyeit a két technológia ötvözésére tesznek javaslatot. A TEI XML jelenlegi két fő részét (TEI Header és Text) egy harmadikkal bővítenék ki, melybe a szövegen kívüli annotációk kerülnének, rétegekként rögzítve. A <teiHeader> továbbra is a metaadatokat tartalmazza, a <text> a (mondatra, sorra, szóra stb.) szegmentált szöveget, ahol minden szegmens egyedi azonosítót kap. A harmadik rész, a <standoff> tartalmazná rétegekre bontva az egyes szövegszegmensekre, intervallumokra vonatkozó annotációkat.²⁴ Például az egyik réteg a szöveg szerkezetét írja le (cím, strófa, verssor stb.), a másik a szövegtulajdonságokat (idézet, kiemelés stb.), a harmadik a szövegkritikai megállapításokat (törlés, betoldás, hiány stb.).

Schmidt, illetve Spadini és Turska megoldásait összevetve az egyik szembetűnő különbség, hogy míg előbbi a szövegek megtöbbszörözésével éri el, hogy a szövegalkulás egyes állomásait rögzítse, addig utóbbi – a TEI hagyományait követve – az annotációban rögzíti azokat. Mindkét esetben kérdéses marad, mit tekinthe-

²² Uo., 7–11.

²³ Elena SPADINI and Magdalena TURSKA, „XML-TEI Stand-off Markup: One Step Beyond”, *Digital Philology: A Journal of Medieval Cultures* 8, no. 2. (2019): 225–239, 234, <https://doi.org/10.1353/dph.2019.0025>.

²⁴ Uo., 228–230. Az elemkészletet bevezető cikk: Javier POSE, Patrice LOPEZ and Laurent ROMARY, „A Generic Formalism for Encoding Stand-off Annotations in TEI”, 2014, <https://hal.inria.fr/hal-01061548/>.

tünk alap- vagy nyers szövegnek, illetve hogyan körvonalazzuk az egyes szövegrétegeket. A másik fontos különbség a szövegnek mint adatok halmazának hierarchikus vagy nem-hierarchikus értelmezése. A TEI XML és a mögötte rejlő OHCO modell hierarchikus felépítményként értelmezi a szöveget. A stand-off markup nem, vagy csak részben hierarchikusként, elismerve, hogy a szöveg csak külön értelmezési keretek rétegeiként írható le. Mindkét modell helytálló lehet: a különféle textológiai hagyományok és a szöveg egyéni vonásai között kell egyensúlyt teremteni, s ahhoz érdemes technológiát választani.

Lehet ugyan, hogy az XML-kódolásnál a stand-off technikailag jobban támogatja az együttműködést és a platformfüggetlenséget, de az is előfordulhat, hogy a technikai mellett az emberi tényezőkben kell keresnünk a hibát. A tudományos elektronikus kiadást készítőik ugyanis csak ritkább esetben teszik közzé az eredeti, kódolt szövegállományt és járulékos fájljait (pl. TEI specifikációt),²⁵ amelyre a webes megjelenítés épül. Kutatásuk leginkább innovatív, s a tudományos közösség számára jelentős eredményeik rejtve maradnak, nem vizsgálhatók. Mintha egy nyomtatott kiadásban csak a jegyzetek felét tennék közzé, vagy az alkalmazott szöveggözlési elveknek csak egy részébe avatnák be az olvasót. Az összehasonlítás végett: Franzini jelenleg 304 tételes katalógusának valamivel több mint 20 százaléka (63 darab) biztosítja a TEI XML források letöltését. Ebben a katalógusban természetesen számos HTML alapú és más, nem TEI XML szabványt alkalmazó hálózati kiadás van (nem TEI XML összesen 135). Ha a TEI XML-ben készült kiadásokat nézzük, az arány már biztatóbb, azok több mint 37 százaléka szabadon hozzáférhető (63/169). Sajnos egy magyar sincs köztük.²⁶

A számítógépes feldolgozás, elemzés és webes közzététel egy másik (nem csupán a számítógépes irodalomtudományt érintő) megoldásra váró problémája annak kutatásértékelésben elfoglalt helye. A tudományometriai és kutatásértékelési szempontokat magába olvasztó rendszer, a Magyar Tudományos Művek Tára (MTMT) a számítógépes eszközökkel létrehozott kutatási eredményeket nem, vagy alig díjazza, függetlenül azok tartalmától, minőségétől és publikáltságától. Nemcsak a kutatás viseli tehát magán a vizualitás, a vizuális alapú gondolkodás nyomait, hanem a kutatásértékelés is. A grafikus megjelenített tartalom (főleg, ha nyomtatott könyv): tudományos közlemény. A digitális fájlok, adatok halmaza (ha nem nyomtatott – s ez a jellemző): egyéb. A döntés során a tartalmi jellemzőket a formaiak írják felül. A közlemény és a közleményekre történő hivatkozások elszámolása tekintetében egyaránt hátrányban van még ma is az a magyarországi kutató, aki korszerű technikai eszközök bevonásával végzi és teszi közzé kutatásait.

²⁵ Utóbbira jó példa a *Deutsches Textarchiv* példásan részletes, mindenki számára hozzáférhető dokumentációja. Martin GRÖTSCHEL, Wolfgang KLEIN und Alexander GEYKEN, *Deutsches Textarchiv (DTA)* (Berlin: Brandenburgische Akademie der Wissenschaften, 2007–), <http://www.deutschestextarchiv.de/>.

²⁶ Az adatok Franzini hivatkozott katalógusából származnak, hozzáférés: 2019.05.03.

Egy konkrét példát említve: az *Ómagyar Mária-siralom* TEI XML-kiadását²⁷ többéves kutatói munka előzte meg. A kiadás egy OTKA-projekt (119355) támogatásával valósult meg, megjelenése előtt szakmai fórumokon vitára bocsátottuk eredményeinket,²⁸ a kiadás bírálói, szakmai tanácsadói a téma kiemelkedő szakértői voltak. A kiadás megfelel az MTA Textológiai Munkabizottsága által a kritikai kiadással szemben támasztott követelményeknek (például: bevezető tanulmány, mely ismerteti az előzményeket és a szövegekölési elveket, kritikai apparátus). Végezetül – számos internetes szövegkiadással ellentétben – DOI azonosítóval is rendelkezik. Mindezek ellenére nem számít tudományos közleménynek, az MTMT adminisztráció szerint azért, mert nincs publikálva. Noha ez az egyetlen jelenleg szabadon hozzáférhető, a weben közzétett hazai TEI XML szövegekölés.²⁹

Változást minden bizonnyal valóban a 2020-as évek hoznak, erre utal, hogy a tudomány nyílt hozzáférést támogató Európai Unió ajánlások idén a magyarországi pályázati rendszert, s így a kutatási infrastruktúra felelős intézményeit is elérték. A kutatási adatkezelés (*research data management*), s ennek jelenleg legkidolgozottabb ajánlása, a FAIR alapelvek (Findable, Accessible, Interoperable, Reusable)³⁰ remélhetőleg rövidesen általános gyakorlat részeivé válnak.³¹

A számítógépes filológia megoldatlan kérdéseire a jelen *Számítógépes irodalomtudomány* című szám több cikke is felhívja a figyelmet. Alapvető elméleti problémákra térnek vissza újra és újra, melyek a nemzetközi szakirodalomban is tematizálódtak az elmúlt közel fél évszázadban. A hazai textológiai szakirodalom az utóbbi két évtizedben³² viszonylag kevésbé foglalkozott a számítógépes filológia elméletével ahhoz képest, hogy a webes szövegkiadások mennyisége – s lassan tudományos értéke is – jelentősen növekedni kezdett. Az 1990-es évektől a 2000-es évek közepéig tartó első fellendülést, mely elsősorban a Horváth Iván vezette Bölcsészinformatikai Önálló Program kiadványaihoz és rendezvényeihez kötődik,

²⁷ HORVÁTH Iván és MARÓTHY Szilvia, szerk., *Ómagyar Mária-siralom: kritikai kiadás*, v1.0 (GitHub-Zenodo, 2018), <http://doi.org/10.5281/zenodo.1287583>. Ez a kiadás, s más, újabban keletkezett kiadványok Franzini már hivatkozott katalógusába még nem kerültek fel.

²⁸ Pl. HORVÁTH Iván és MARÓTHY Szilvia, „Felkészülés a HBK kiadására: az ÓMS próbakiadása”, *A Pray-kódex-munkacsoport első műhelykonferenciája*, Budapest, Országos Széchényi Könyvtár, 2017. június 13; MARÓTHY Szilvia, „Az ÓMS elektronikus kritikai kiadása”, *III. Forráskutatás, forráskiadás, tudománytörténet konferencia*, Budapest, Eötvös Loránd Tudományegyetem, 2017. november 16–17.

²⁹ Ti. semelyik másik tudományos szövegkiadást TEI XML-ben kódoló, weben közzétevő személy, illetve intézmény nem teszi hozzáférhetővé ezen kiadások lényegét, a TEI XML forrásfájlokot, lásd fentebb.

³⁰ Mark D. WILKINSON, Michel DUMONTIER, IJsbrand Jan AALBERSBERG, Gabrielle APPLETON, Myles AXTON, Barend MONS et al., „The FAIR Guiding Principles for Scientific Data Management and Stewardship”, *Scientific Data* 3 (2016): 160018, <https://doi.org/10.1038/sdata.2016.18>.

³¹ A kutatási adat-kezelés, nyílt hozzáférés, FAIR alapelvek témaköréhez lásd: MARÓTHY Szilvia, „A nyílt és a zárt tudományról”, in *Kulturális iparágak, kánonok és filterbuborékok*, BÁRÁNY Tibor et al., szerk. (Budapest: BME-ELTE, 2020, megjelenés alatt).

³² A témával foglalkozó cikkeket lásd a jelen kötet válogatott bibliográfiájában. A hazai tudományos elektronikus kiadások bibliográfiája elérhető: <https://www.zotero.org/groups/2199751/> és https://github.com/marothyszilvi/digEds_hun. Vö. MARÓTHY, „Tudományos szövegkiadások a hálózaton...”.

úgy tűnik, csak most, a 2010-es évek végén, 2020-as évek elején követi egy újabb. Akkor az elméleti viták hullámai a Textológiai Munkabizottságig is elértek, ennek ékes bizonyítéka a 2004-ben közreadott, már idézett *Alapelvek*.³³ Mára ezek az alapok is újragondolásra, kiegészítésre várnak. Különszámunk cikkei mind elméleti megfontolásaikkal, mind gyakorlatiasabb javaslataikkal nagyban hozzájárulhatnak ehhez, s egyúttal a számítógépes irodalomtudományról szóló diskurzust is megélénkíthetik.

³³ Szakirodalmi előzményeihez lásd: „Filológia és digitális barbárság”, ELTE BTK Bölcsészettudományi Informatika Önálló Program (BIÖP), 2004. március 4., <http://magyar-irodalom.elte.hu/biop/barbar/>.