

LABÁDI GERGELY

## Matthew Jockers: *Macroanalysis. Digital Methods and Literary History*

A számítógép használata írott szövegek feldolgozására és vizsgálatára a számítógépek modern kori megjelenésével egyidős, itt az 1940-es éveket, Roberto A. Busa *Index Thomisticus*-át szokás elsőként megnevezni.<sup>1</sup> Azonban a számítógépek használatát az irodalomtudományban – mint általában a gépek használatát a nyugati kultúrában a 19. század eleje óta<sup>2</sup> – máig sajátos kettősség uralja: egyrészt úgy tűnik, számos olyan kérdés megoldható használatával, amelyet addig az emberi tényező korlátozott. Egy gép/program nem fárad el és nem unja meg a munkát, nem ugrik a szeme sort, következetesen érvényesíti egy vizsgálat megadott paramétereit. Másrészt a számítógépek használatát meglehetősen ambivalensen értékeli az irodalomtudomány: vagy ünnepli – hiányosságait a technika fejletlenségének tudva be –, vagy lebecsüli.<sup>3</sup> A számítógépeknek a filológiai munkában való használata, amennyiben eszközként tekintünk rájuk, a filológia mindennapjainak részévé vált: könnyebb lett a rögzítés, a módosítás, a kiegészítés, gyorsabbá a szerkesztés, a keresés, és a befoglalható információk szempontjából megszűntek a könyv fizikai korlátai. Sokáig mégis úgy tűnt, a számítógépek elterjedése a humántudományok mindennapjaiban nem írja át azt a textuális kultúrát, amely meghatározza a szövegekkel való foglalatosság gyakorlatait és kijelöli hatalmi viszonyait, valamint nem számolja fel a könyvbeliség sajátosságait sem. A számítógéphasználat a humántudományokban azonban korántsem eszközprobléma, hiszen mediális technikáink meghatározzák magát az észlelést (egy szöveg szöveggé történő azonosítását), és gyökeresen átalakítják a szövegekkel végzett (tudományos) gyakorlatainkat.<sup>4</sup> E kihívásra adott válaszként is értelmezhető Franco

Moretti nagyhatású fogalma, a *distant reading*, illetve a „nem-olvasás” általa művelt gyakorlata,<sup>5</sup> ám a meggyőző eredmények ellenére az elméleti háttér hiánya, vagy inkább kifejtetlensége, illetve a példák, vizsgálatok „esetlegessége” miatt mégis maradhat az olvasóban hiányérzet.

Természetesen Moretti koránt sincs egyedül, amikor az irodalomtudomány módszertani fordulójának szükségessége mellett érvel, mások ugyanakkor talán kevésbé provokatív(an ható) fogalmakat használnak. Stephen Ramsay az algoritmikus irodalomtudomány (*algorithmic criticism*) kifejezést, Matthew L. Jockers pedig a recenzio alapjául szolgáló könyvében a közgazdaságtudományi háttérű makroelemzést (*macroanalysis*) javasolja. A lényeg persze mindhárom fogalom esetében ugyanaz: nagy mennyiségű adat rendszeres, következetes vizsgálata, amely révén olyan trendek, mintázatok azonosíthatók, amelyek a szövegek mikroszintű vizsgálata, azaz a szoros olvasás révén nem elérhetők. Egy olyan szemléletmódról és módszertanról van tehát szó, amely korábban elképzelhetetlen mértékű kontextualizációt tesz lehetővé, amely az egyedi szövegek környezetének, az egyes szövegeket meghatározó irodalmi, társadalmi, kulturális trendeknek az azonosítását, elemzését ígéri: ez a történeti és stilisztikai kérdéseknél egyértelmű előnnyel rendelkezik a szoros olvasással szemben. Ugyanakkor Jockers nem gondolja, hogy el kellene vetni a szoros olvasást, nagy mennyiségű adat számítógépes elemzése egyszerűen csak alternatív módszer adatokat gyűjteni, hipotéziseket felállítani, ám ezeket ugyanúgy kritikával kell kezelni, és értelmezni szükséges.

Ám míg Jockers a negyedik fejezetben eljut a makroelemzés fogalmának bevezetéséig, három rövid, adataiban egyáltalán nem új, ám töménységében és élességében mégis gondolatébresztő fejezetet kell elolvasnunk (*Revolution, Evidence, Tradition*). Az első alig két oldal, legfontosabb és leginkább megfontolandó állítása nem is az, hogy a számítógépes kapacitás mára érte el azt a szintet, hogy az elméleti lehetőségeket meg is valósítsuk, vagy hogy a *digital humanities* a 2000-es évek második felére szervezettel/intézményileg végleg áttört, hanem az, hogy mostanra ha nem is digitalizáltunk mindent, „egy olyan fordulóponthoz, eseményhorizonthoz értünk, ahol már elegendő szöveget és szakirodalmat kódoltunk ahhoz, hogy lehetővé tegye számunkra, sőt hogy kényszerítsen bennünket, új kérdéseket tegyünk fel az irodalomról és az irodalmi adatról”. (4.) Jockers elbeszélésében 2008 a fordulópont, és bár nem reflektál rá, de megállapítása természetesen csak az angol nyelvű szövegekre vonatkozik. A német kultúra kapcsán például Fotis Jannidis és Gerhard Lauer szerint 2011, a TextGrid repozitóriumának megindulása óta áll fenn a kvantitatív vizsgálatok alapfeltétele, a megfelelő mennyiségű és minőségű digitalizált szöveg.<sup>6</sup>

<sup>1</sup> Susan HOCKEY, *The History of Humanities Computing = A Companion to Digital Humanities*, szerk. Susan SCHREIBMAN – Ray SIEMENS – John UNSWORTH, Blackwell, Oxford, 2004. (<http://www.digital-humanities.org/companion/>)

<sup>2</sup> Hermann BAUSINGER, *Népi kultúra a technika korszakában*, ford. SZIJÁRTÓ Zsolt, Osiris, Budapest, 1995.

<sup>3</sup> Két példa az 1970-es évekből: „Úgy tűnik, a számítógépek jelenleg konkordanciák készítésében és a metrikai elemzések szellemtelenebb részének elvégzésében szolgálhatnak bennünket a leginkább.” (Martin L. WEST, *Szövegvizsgálat és szövegvizsgálat*, ford. BOLONYAI Gábor, Typotex, Budapest, 1999 [1973], 90.) „Teljes egészében elemezni tudja a gép a szöveg időmértékes metrikai sajátosságait, a hangsúlyos verselés szerinti elemzést azonban csak mechanikus, külső segítséggel végezheti el (ha a hangsúlyokat, illetve ezek bizonyos fajtáit a gépen kívüli eszközökkel jelöljük a szövegben). Abban az esetben azonban, ha a hangsúlyos verselést külső segítséggel a gép elvégzi, az ilyen vizsgálatot könnyűszerrel összekapcsolja az időmértékes verselés eredményeinek vizsgálatával, és az eredményeket kölcsönösen korrelálja. [...] Konkrét vizsgálatunkban csak az időmértékes verselés szerint elemeztük a szövegeket, a hangsúlyos verselés elemzésére felállított hipotetikus program túl bonyolultnak bizonyult a számítógép számára.” (VOIGT Vilmos, *Számítógépes ritmuselemzési kísérlet*, ItK 1972/2., 203–204.)

<sup>4</sup> A kérdés egyik klasszikus tanulmánya: Steven J. DEROSE – David DURAND – Elli MYLONAS – Allen RENEAR, *What is Text, Really?*, *Journal of Computing in Higher Education* 1990/2., 3–26. Lásd

például a *Text Encoding Initiative* kódolási irányelveit (<http://www.tei-c.org>) és szaklapját (<http://journal.tei-c.org>)

<sup>5</sup> Franco MORETTI, *Graphs, Maps, Trees*, Verso, New York, 2005.; Uő., *Distant Reading*, Verso, New York, 2013. Lásd még a *Stanford Literary Lab* tanulmányait: <https://litlab.stanford.edu/pamphlets/>

<sup>6</sup> Fotis JANNIDIS – Gerhard LAUER, *Burrows's Delta and Its Use in German Literary History = Distant Readings. Topologies of German Culture in the Long Nineteenth Century*, szerk. Matt ERLIN – Lynne TATLOCK, Camden House, Rochester, 2014, 30.

A második fejezet a humántudományokban a szoros olvasás révén elérhető „bizonyíték”, adat régről ismert problematikáját ismerteti. Olyan kérdésekre, amelyek több szöveg egyidejű kezelését igénylik, a szoros olvasás nyilvánvalóan elégtelen, vagy legalábbis kétséges eredményhez vezet, mert már a mintavétel reprezentativitása is kérdéses. Jockers példája Ian Wattnak a regény felemelkedéséről szóló klasszikus munkája, amely mindössze Defoe, Richardson és Fielding munkái alapján készült, jóllehet a szóba jöhető szerzők és szövegek köre nagyságrendekkel nagyobb. Adódik a kérdés, ha ezt a néhány ezer szöveget el tudnánk „olvasni”, más történetet mondanánk? Eddig ezt a kérdést esélyünk sem volt megválaszolni, de most már megvannak az eszközeink.<sup>7</sup>

Az első négy fejezet tehát jobbára a helyzet felmérése (az itt nem tárgyalt harmadik a *digital humanities* expanzióját és fogadtatástörténetét nézi át), ami a magyar irodalomtudomány szempontjából sem tanulságok nélkül való, mindazonáltal a könyv második része, az 5–9. fejezet sokkal izgalmasabb, ahol Jockers konkrét példák nyomán mutatja be a *macroanalysis* módszertanát, használatát és a szoros olvasás nyomán nem elérhető eredményeit. Jockers ugyan nem rögzíti, de áttekintéséből a számítógépes elemzés mindkét alapvető iránya kirajzolódik. Az egyik irány a szövegekhez kapcsolódó metaadatokat vizsgálja. Jockers saját adatbázisának, az ír-amerikai írók alig 758 tételes példáján mutatja be a makroszintű megközelítés előnyeit. Adatbázisa alig két és félszer nagyobb mint a témával foglalkozó alapmonográfia primer irodalmi listája. Charles Fanning az adatok alapján művében arra jutott, hogy az ír-amerikai irodalomban 1900 és 1930 között jelentősen visszaesett az ír-amerikai művek száma, azaz hiányzik egy nemzedék. Jockers adatbázisa az ír-amerikai fogalmát ugyanúgy érti, mint Fanning, ám nemcsak a művek bibliográfiai adatait veszi fel, hanem rögzíti a szerző nemét, hogy keleti vagy nyugati parti szerző-e, városban vagy vidéken él-e, illetve a mű városban vagy vidéken játszódik-e. Az alig nagyobb adatbázisa alapján teljesen más történetek vázolhatók fel: a Fanning észlelte hiányzó nemzedék valójában a keleten élő, a városi világban játszódó történeteket író elveszett férfiak nemzedékét jelenti – tehát az ír-amerikai irodalom kánonjának vakfoltjáról van szó. Az adatbázis alapján koránt sincs jelentős visszaesés, sőt ugyanebben az időszakban a nyugati, vidéki környezetben élő és arról alkotó ír-amerikai nők esetében radikális emelkedés figyelhető meg: az amerikai nyugat a 20. század első harmadában valami speciálisat kínál nekik. Azonban ennek feltárására ezek a metaadatok már nem elegendők.

A 6–9. fejezetek a *macroanalysis* másik nagy irányát, a nyers szövegen végrehajtott számítógépes vizsgálatokat mutatja be (*Style, Nationality, Theme, Influence*). Ennek ismét két fontos iránya van: az önálló jelentéssel nem bíró funkciószavak (például névmások, névelők, segédigék), illetve a saját jelentéssel bíró, úgynevezett tartalmi szavak alapján történő vizsgálatok. A kvantitatív szerzőségi vizsgálatok (stilometria) régi elképzeléseket igazolva – a nyomok bölcsészettudományi jellegének és jelentőségének ginzburgi értelmében<sup>8</sup> – a funkciószavak alapján azonosítható egyéni szerzői

<sup>7</sup> A magyar regény története kapcsán lásd például Batori Annával közös előadásunkat: BATORI Anna – LABÁDI Gergely, *Egy regényadatbázis felépítése – kérdések és lehetőségek*, Számítógép az irodalomtudományban, MTA BTK–BME, Budapest, 2015. november 24.

<sup>8</sup> Carlo GINZBURG, *Nyomok. A jel-paradigma gyökerei*, ford. FARKAS Krisztina = *Uő.*, *Nyomok, bizo-*

nyitásokat vizsgál. Jockers példáiban nem az az érdekes, hogy igazolja, szógyakorisági listák, funkciószavak (még a központosítás is ide sorolható) abszolút és relatív arányai alapján a gép helyesen csoportosítja az egy szerzőhöz tartozó műveket – az önmagában nagyon kevés, nem vagy legalábbis ritkán hoz új ismereteket, jól példázza ezt Fotis Jannidis és Gerhard Lauer korábban idézett tanulmánya –, hanem megmutatja, hogyan lehet továbblépni, és bevonni a vizsgálatokba a műfaj, a nemzetiség, a nem, a korszak problematikáját is. Az általa lefutott elemzések alapján egy szöveget leginkább a szerzője határoz meg, utána viszont meglepő módon nem a korszak, hanem a nem és a műfaj – a korszak a legutolsó. Érdekes módon még a nemzetiség kérdése is releváns lehet a funkciószavakon alapuló vizsgálatok esetében. A módszer tehát a „great unread” anyagának feldolgozását teszi lehetővé, a tartalom, a forma és a stílus összefüggéseit tárhatjuk fel vele, újra tudjuk gondolni a műfaji kategóriákat és az irodalmi korszakokról alkotott elképzeléseinket.

A tartalomelemzés számítógépes alapját jelentő modellező algoritmusok abból indulnak ki, hogy egy szöveg több lehetséges téma vegyülete, ezért a tulajdonnevek, funkciószavak kiszűrése után a lemmatizált szavakhoz próbál lehetséges témákat rendelni. Tehát meg kell adni, hány témát keressen, milyen egységekben, ezután a gép az együttes szóelőfordulásokból tematikus csoportokat hoz létre, majd az egyes dokumentumokat a tematikus csoportok alapján klaszterizálja. Ezeket a csoportokat aztán az egyes szövegekhez hozzáadott metaadatok alapján lehet elrendezni. Jockers példái a Stanford Literary Lab 3346 digitalizált 18–19. századi, angol nyelvű regényeinek tartalomelemzésén alapulnak. A kutatásokban ismét nem csak az az érdekes, hogy téma és történelmi esemény, téma és nemzetiség, téma és a szerző neve között egyértelmű összefüggések vannak, hanem hogy az egyes témák történetét, divatjait nyomon lehet követni, azaz a szoros olvasásnak elképesztően részletes makrotrendek tárhatunk fel, illetve az egyes könyvek közti távolságot is mérni tudjuk, amely nemcsak arra ad választ, hogyan függnek össze témák és stílusok az idővel, milyen a női írók beágyazottsága, hanem kiderül az is, kik a kívülállók, kik előzik meg korukat stb. A vizsgált korpuszban a hálózati elemzések alapján egyébként a *Tristram Shandy*, a *The Whirlpool* George Gissingtől és kissé váratlanul Benjamin Disraeli *Venetiája* áll a központban.

Jockers tehát meggyőzően érvel a számítógépes elemzés potenciálja mellett, végig hangsúlyozva, hogy a *macroanalysis* nem helyettesíti a szoros olvasást. Könyve végén pedig a fennálló adminisztratív korlátokra is felhívja a figyelmet. A 10. fejezetben (*Orphans*) ezért szerzői jogi kérdésekre keresi a választ, mivel ezek jelentősen akadályozzák a kutatásokat. Matthew Sag nyomán a digitalizált szövegek non-expressive használata mellett érvel, tehát olyan digitalizálásról, amely nem teszi lehetővé az eredeti reprodukcióját: például egy szógyakorisági listához fér hozzá a kutató, mint JSTOR Data for Research Service szolgáltatásában.<sup>9</sup> Hiszen hiába van például az

*nyitkok, mikrotörténelem*, szerk. K. HORVÁTH Zsolt, Kijarat, Budapest, 2010, 13–53.

<sup>9</sup> Lásd Allen Beye RIDDEL, *How to Read 22,198 Journal Articles? Studying the History of German Studies with Topic Models = Distant Readings*, 94–96. Magáról a JSTOR-szolgáltatásról: <http://about.jstor.org/service/data-for-research>

elmúlt 200 évből tízmillió oldalnyi magyar szakszöveg digitalizálva az Arcanum Digitális Tudománytárban, a jelenlegi formátum kvantitatív vizsgálatokra alkalmatlan – legalábbis amihez és ahogyan én kutatóként hozzáférek. Jockers könyve, úgy gondolom, sok továbbgondolnivalót kínál olvasójának. Akit pedig a technikai részletek is érdekelnek, bevezetésnek előveheti Jockers másik munkáját.<sup>10</sup>

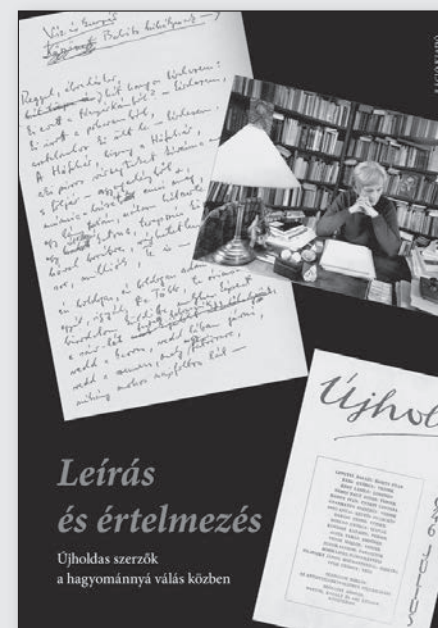
(University of Illinois Press, Chicago, 2013.)

<sup>10</sup> Matthew L. JOCKERS, *Text Analysis with R for Students of Literature*, Springer, New York, 2014.

## Leírás és értelmezés

Újholdas szerzők a hagyománnyá válás közben

2016, 492 oldal, 3900 Ft



A Nemes Nagy Ágnes halála után negyedszázaddal összeállított kötet tanulmányai tovább bővítik a Petőfi Irodalmi Múzeum Kézirattárának tulajdonában lévő, az Újhold folyóirat szerzőihez kapcsolódó szövegek vizsgálatát a filológiai, illetve az irodalomtudományi szempontok alkalmazásával. A sorozat két előző darabjában tárgyalt szerzők mellé itt felsorakozik Pilinszky János és Mándy Iván is, teljesebbé téve az újholdasok műveinek, illetve azok interpretációjának megismerését. Míg a Pilinszky testábrázolásával kapcsolatos írás a gazdag értelmezői hagyomány-

nyal rendelkező versek újraolvasására törekszik, addig a Mándyt tárgyalóban kevésbé ismert novellák felidézésére kerül sor. A költészet, ahogy az eddigiekben, e gyűjteményben is hangsúlyosabb, de Nemes Nagy Ágnes egyik fontos esszéjéről és verseinek kiadástörténetéről ugyancsak szó esik. Két tanulmány olvasható Lakatos István lírájáról: az egyik a politika és az irodalom viszonyát, Lakatos 1956-ot megelőző és a forradalom alatti tevékenységét dolgozza fel, a másik pedig megmutatja hosszabb versciklusa keletkezéstörténetének bevonásával a személyes élmény tárgyiasított irodalommal alakulását. A kötet második, nagyobb egysége szövegközléseket tartalmaz. Kuriózum Nemes Nagy Ágnes Babits Mihályról szóló, bőséges magyarázó jegyzetanyaggal ellátott színpadi játéknak közlése, amely megerősíti Babits újholdasokra gyakorolt hatását. Ezt követi Lengyel Balázs korábban csak folyóiratok oldalain megjelent, 1950 előtti kritikáinak, esszéinek sorozata. A kötetet Pilinszky János műveinek és befogadásának reprezentatív bibliográfiája zárja, amely a 2001-ben megjelent előző gyűjtést egészíti ki, azt kiterjesztve a 2015. év végéig.

A kötet megrendelhető vagy kedvezményesen megvásárolható a kiadó szerkesztőségében: Ráció Kiadó • 1072 Budapest, Akácfa utca 20.  
tel.: (1) 321-8023 • fax: (1) 321-4757 • e-mail: racio@racio.hu • www.racio.hu