

A DIGITÁLIS IRODALOMTUDOMÁNY EGYFAJTA ÉRTELMEZÉSE*

– SZEMES Botond. *Digitális irodalomtudomány: Olvasás, számítás és ábrázolás egysége az irodalmi elemzésben*. Kijárat Kiadó–I.T.E.M. Alapítvány, Budapest, 2024, 252 lap –

MARÓTHY SZILVIA

független kutató

E- mail: mthy.szilvi@gmail.com

ORCID 0000 0003 2558 9504

■ Kinek a mije a digitális vagy számítógépes bölcsészet, a digitális vagy számítógépes irodalomtudomány? Ahány írás, annyiféle megközelítés, a diszciplinárizálódás és intézményesülés, egy kutatási terület újraértelmezése vagy újrapozicionálása miatt örökké izzó téma ez a bölcsészettudományok számára. Szemes Botond is kísérletet tesz a digitális irodalomtudomány definiálására, mely könyvének mind elméleti, mind gyakorlati fejezeteiből úgy tűnik, egy szűkebb kutatási területet jelöl, ahol a nyelvstatisztikáé, stilisztikáé, nyelvtechnológiáé a főszerep. Hova lett az irodalom? – kérdezhetnénk. (Előrebocsátom: megmaradt.)

A kötet egyértelmű pozitívuma, hogy a digitális irodalomtudomány irodalomelméleti, hermeneutikai megalapozását kínálja a magyar nyelvű diskurzus számára, hiszen a kutatások e területén sokkal jellemzőbb a számítástechnológia és digitalizáció hozta új módszertanok elutasítása és ismeretének hiánya. Míg a szerző által körülbírt hivatkozott nyelvtechnológiai, stilisztikai területeken, és a szerző által alig hivatkozott irodalomtörténeti és textológiai kutatásokban tehát végbement (mondhatni többször is) ez a fordulat, mely során a kutatás módszertana, kritikai gondolkodása bővült, a számítástechnológiai ismereteket is bíró és integráló irodalomelméleti megalapozás még váratott magára. Nagy erénye ez Szemes Botond könyvének, valamint az, hogy egyenlő arányban szentel helyet az elméleti megalapozásnak és a gyakorlati alkalmazások bemutatásának. A jelen recenzióban Szemes szövegét szintetizálva előbb az elméleti, majd a gyakorlati részre térek ki.

* A publikáció az Innovációs és Technológiai Minisztérium Nemzeti Kutatási Fejlesztési és Innovációs Alapból nyújtott támogatásával, a K135631 számú, *A régi magyar költészet számítógépes metrikai és stilometriai vizsgálata* című pályázati program keretében készült.

Elméleti megalapozás és tudománytörténeti narratíva

Szemes Botond sokértelmű és titokzatos címei kiemelkednek a szaktanulmányok szürke címadási gyakorlatának óceánjából (pl. *Az ív segédegyenesei*, *Az idővonal felemelkedése*, *Kidolgozott viszonyok* - hogy csak néhányat említsek), kár, hogy ezúttal nem engedte szabadon kreativitását. Valószínűleg nem véletlenül. A könyv terjedelmes elméleti része a címében foglalt tudománytörténetet hivatott körülhatárolni irodalomelméleti nézőpontból, ennek narratíváján belül maradván történeti perspektívát nyújt, összefüggéseket világít meg, érveket vonultat fel.

Szemes definíciója szerint a digitális irodalomtudomány: „[I]rodalmi szövegek gépi olvasása – azaz a nyelvi elemek eloszlásának vizsgálata, az eredmények statisztikai kiértékelése és vizuális megjelenítése –, valamint egyéb olvasásmódok kapcsolatából kiinduló értelmezői gyakorlatok és az ennek során létrehozott ismeretek intézményesült rendszere.” (13).

Ezt a szűkebb értelmezést több helyütt alátámasztja az elméleti rész érvelésében, valamint a gyakorlati rész egészében. Előbbire példa még Moretti munkásságának értékelése (36–37), melyben sajátos értelmezést ad a *computational criticism* önmeghatározásnak. Szemes szerint „nem véletlen”, s azért így és nem digitális bölcsészettnek vagy irodalomtudománynak nevezik tevékenységüket, mert arra nem jellemzők a „statisztikai és nyelvtechnológiai eljárások [...], [melyek] a diszciplína alapjait képezik”, állítása szerint Moretti csoportja csak közeledett (!) a digitális bölcsészethez, de nem művelte azt.

Ez az értelmezés figyelmen kívül hagyja a digitális bölcsészet (korábban számítógépes bölcsészet, bölcsészinformatika és sorolhatnám) önmeghatározásának mintegy hét évtizedét, melyre egyébként Moretti is reflektál az általa választott kifejezéssel. A *computational* használata e közösségekben a számításmélettől módszertani fontosságát hangsúlyozza, a *criticism* pedig az eljárások megfontolt és reflektált alkalmazására is utal. Az, hogy a digitális bölcsészet terminusra rakódott az „igazi digitális” öntudata, merőben új keletű dolog: a kilencvenes évek végétől a kétezres évek végéig tartó bő évtizedben két nagy terminológiai vita is lezajlott a diszciplína elnevezése körül (korábban igyekeztem is összefoglalni ezeket, itt nem teszem).

A másik fő ok, amiért különbözhet a ma digitális bölcsészete, irodalomtudománya a megelőző évtizedekétől, technológiai természetű. Elképzelhetetlenül nagyobb teljesítménnyel, kapacitással bírnak a ma bárki számára hozzáférhető eszközök, még felfoghatatlanabb volumennel egyes kutatóintézetek, kutatási konzorciumok által hadra fogható (szuper)számítógépek. A tárolás sem okoz gondot, bármit digitalizálhatunk (az irodalomtudományon belül maradván ez egészen biztos). És a nyolcvanas évek óta digitalizáljuk is a bármit! Hát persze, hogy most végeznek big data elemzéseket, amikor már van nagy mennyiségű adat. És ki állította elő azt? A megannyi „analóg” bölcsész, könyvtáros és laikus mellett a megelőző időszak számítógépes bölcsészei, akik csak „közeledtek” a digitális bölcsészet eljárásaihoz.

Szemes könnyen körülhatárolhatta volna elméleti kereteit egy-két bekezdésnyi kitéréssel; tudniillik hogy mivel foglalkozik a könyve (nyelvstatisztika, stilisztika) és mivel nem (textológia, irodalomtörténet, irodalomszociológia, hálózatelemzés stb.). Ám a bevezetőt olvasva világossá válik, hogy ez némiképp szándékos a részéről: „Az alábbiakban amellet érvelek, hogy a tárgyalt technikák keveredése elsősorban a *stilisztika* területének előtérbe kerülését vonja maga után” (12, kiemelés az eredetiben). Míg a történeti áttekintésben a hazai előzményekre való rálátását bizonyítja a stilisztikai, nyelvstatisztikai kutatások terén (50–57) izgalmas példákat felvonultatva, az irodalomtörténet esetében már szinte teljesen elmarad a digitális mellől az irodalomtudomány.

A digitális irodalomtudomány körülhatárolásakor számos ponton utal rá a szerző, hogy könyve úttörő, kevés előzménye van, ez a kép azonban hamis. Néhány példát említve: a bölcsészet szerinte (illetve Sybille Krämerre hivatkozva) „mindig is távol tartotta magát az ilyen matematikai műveletektől” (40). Nos, az irodalomelmélet egyes irányzatainak képviselőire igaz lehet ez az állítás, de a bölcsészet, vagy akár az irodalomtudomány egészére biztosan nem. Számos magyar és külföldi kutatás támasztja ezt alá, csak a legismertebbeket említve: a digitális bölcsészet (bár vitatottan) atyjának számító Roberto Busa a számítógépek megjelenésének évtizedében már foglalkozik Szent Tamás szövegeinek gépi feldolgozásával, elemzésével. Horváth Iván a hetvenes években indítja útjára az első irodalomtörténeti adatbázist, mely mögött épp az a célkitűzés áll, hogy majd statisztikai elemzésekkel támogassa, egészítse ki vagy épp cáfolja a történeti és poétikai kutatásokat, a kilencvenes években pedig, ma úgy neveznénk, „born digital” tudományos szövegkiadásokat készít kutatótársaival. A *Gépeskönyv* máig az első meghatározó szintézise a számítógépes/digitális irodalomtudománynak, s mint ilyen elméleti szempontból is releváns. A gépi olvasás terminust (az irodalomtudományba bizonyosan) Labádi Gergely vezette be a magyar diskurzusba elméleti és gyakorlati szempontokat ötvöző írásaival, Szemes könyve ezt nem is említi, noha a kifejezést sűrűn használja. (További előzményeket e helyt nem sorolok, vannak azonban bőven, a szakirodalomban elérhetők.) A számítógép megjelenése előttről szintén találhatunk példákat a matematikai, kombinatorikai, statisztikai, adatbázisba rendező stb. módszertanra a bölcsészet területén (a bölcsészeti szakirodalom mellett a könyvtártörténeti is bővelkedik ilyen példákban).

A szerző több helyütt értékeli elődeinek munkásságát kezdeti próbálkozásokként a tekintetben, hogy csak kevés szövegen, vagy kézi módszerekkel végeztek kísérleteket, azonban e történeti visszatekintésekből teljesen hiányzik egy igen lényeges szempont: az egyes korok technológiai és feldolgozottságbeli ismertetése, hogy mire volt lehetősége az adott korban a kutatóknak. Az 1950–80-as években nyilván nem dolgozhatott mindenki számítógéppel, aki dolgozhatott is, csak igen kis számítási és tárhelykapacitással a maiakhoz képest. Manapság könnyű a MEK-ről automatikusan lementeni nagy HTML-állományokat és elemezni akár egy közepes teljesítményű laptopon. De hogy több tízezer vagy millió szöveg rendelkezésre álljon, ahhoz szükség volt a már lassan negyven éve tartó SGML/XML és HTML-alapú szövegfeldol-

gozásra, számos kutató, könyvtáros, laikus munkájára. Eközben ő maga (vitatható módon) eltekint a digitális filológia mint az irodalomtudomány integráns részétől, viszonyulásának ellenére pedig maga is alkalmaz kézi feldolgozási-javítási módokat, amikor nem működik megfelelően vagy nem alkalmas rá az eszköztára (pl. 207). Vajon a 2050–60-as évek Szemes Botondja hogyan ír majd erről a módszertanról?

Szemes könyve záró fejezetében így fogalmaz: „... a gyakorlati, gyakorlattani és elméleti munkák egymásra vonatkoztatására van szükség. A könyv ennek első lépéseit kívánta megtenni a két – a tudománytörténeti és gyakorlattani, valamint az esettanulmányokat bemutató – rész párbeszédbe állításán keresztül.” (231) Nem vitatva a gyakorlat és elmélet összekapcsolásának szükségességét, sem Szemes könyvének kvalitásait e téren, fontosnak tartom kiemelni, hogy korántsem az első lépéseket teszi meg a szerző. Úgy vélem, elég – és méltán kiérdemelt – dicsőség a szerzőnek, hogy a szűkebben vett kutatási területén, a digitális irodalomtudomány egyik meghatározó irányzatán belül úttörő elméleti-módszertani és gyakorlati eredményeket felmutató monográfiát hozott létre.

A digitális irodalomtudomány gyakorlata: esettanulmányok

Szemes Botond elemzéseiben alaposan és korrekt módon mutatja be a feldolgozás elméleti alapvetéseit és pontos lépéseit, több helyütt kitér a kiinduló szövegforrások és az eszközök adta korlátokra, melyek az eredményekre is hatással lehetnek. Nagy erénye az általa bemutatott kutatásoknak, hogy nem seprí szőnyeg alá ezeket a limitációkat. Példa erre az elemzendő szövegen végzett átalakításokat részletező hosszú listája (120–121), a többszöri figyelemfelhívás a 19. századi szövegek eltérő helyesírásából fakadó pontatlanságokra az automatikus morfológiai elemzés során (pl. 121), az összehasonlíthatóság érdekében végzett mintavételezési módszerek alkalmazása és ismertetése (pl. TTR-ről 139).

Számos izgalmas felismerést rejteget a kötet a modern próza változásaival és az egyes szerzők életművével kapcsolatban, ezeket azonban nem szeretném helyette elsütni, így csak néhány oldalszámot említek (126–127, 176–177, 190). Hangsúlyozom, a szerző elemzéseiben az irodalomtörténeti diskurzusokra érzékenyen, ugyanakkor attól elrugaszkodva a szövegekre koncentrálna képes elsősorban nyelvstatistikai eszközökkel figyelemre méltó eredményeket elérni, melyek hol megerősítik a korábbi kutatások állításait, hol árnyalják azokat, máskor pedig új kérdéseket fogalmaznak meg.

Mint valaha kora újkori irodalommal foglalkozó kutató, meg kell említem, hogy a középkortól a felvilágosodásig tartó korszak prózairodalmának gyorstalpaló áttekintése (128–129) megérte volna a szakmai revideálást. A szerző a modern és kortárs próza szakértője, így nem várható el tőle ezen korszakok szakirodalmának beható ismerete. A kötet egy pontján nagy ívű, kissé pozitivistá lelkesedéstől fűtött fejlődéstörténetbe ágyazza a regény, illetve a prózastílus változásait. Ez a fejezet talán megállta volna a helyét anélkül, hogy a régiség és a felvilágosodás prózairodalmáról fogalmazzon meg állításokat, összesmosva a 15–17. századok prózájának jellemzőit,

valamint az alfabetizációt a retorikaoktatással, neveléstörténeti kérdésként kezelve a retorika jelenlétének, funkciójának változásait. A szerző mentségére szóljon, hogy ezen áttekintése elsősorban Herczeg Gyula 1976-ban publikált stilisztikai tárgyú cikkén alapul.

Visszatérve a kötet erényeihez: számomra üdítő volt, hogy Szemes nyelvstatisztikai és stilometriai kutatásai irodalomtörténeti kérdésekre irányulnak, nem pedig irodalomszociológiaiakra, és pláne nem a női–férfi arányok vagy nemi stiláris jegyek kutatására, melyek önmagukban az irodalomtudományi kutatásban idegen testként vannak jelen. Dicséretre méltó az is, hogy kutatási adatainak egy részét közzéteszi a könyv mellékleteként egy GitHub repozitóriumban, így egyes elemzések eredményei, valamint a hozzá tartozó ábrák külön szemügyre vehetők, némely esetben talán más elemzések számára is felhasználhatók.

A kötet elméleti része kitér az ábrázolás irodalomelméleti-fenomenológiai meghatározására, melyre bővebben nem reflektáltam. Üdvözlendő, hogy ezen elemzési módszer számára is megágyaz a szerző jelen értelmezési keretben. Könyvének mind elméleti, mind gyakorlati részéből hiányzik azonban az információábrázolás, adatvizualizáció szakirodalma és módszertana. A szerző kutatási eszköztára tekintetében jól körülhatárolható pályán mozog: a nyelvstatisztikáét, stilisztikáét, stilometriáét használja. Ezeknek nem magától értetődő eleme az információábrázolás, mely önálló és egyre szélesedő szakággá fejlődött az információmennyiség radikális növekedésével. E helyt tehát megint csak nem róható fel a szerzőnek, hogy ezeket elméleti és gyakorlati vizsgálódásaiba nem építette be (utóbbira eklatáns példa lehet a 185–186. oldal két diagramja), pusztán az, hogy narratívájában nem ezt ígéri, nem emellett áll ki a könyv, s hamis a kép, hogy az adat ábrázolásának nem voltak előzményei (43). (A terület behatóbb megismerésére, ha csak egy könyvet ajánlhatnék, ami a nemzetközi és a hazai előzményeket és módszertant egyaránt ötvözi, akkor Bátorfy Attila szintén 2024-ben megjelent *Adatvizualizáció* című munkája lenne az.)

Izgalmas eredményeket vonultatnak fel a *Mondathosszúság* és a *Tagmondatkapcsolatok gyakoriságelemzése* című alfejezetek. Itt például hasznos lett volna, ha e két egység oszlopdiagramjainak egy része szinkronba kerül, vagy készültek volna olyan diagramok is, ahol a kétféle elemzés összevethetővé válik. Az előbbiben ugyanis öt és három évcsoport szerint láthatjuk az eredményeket, míg utóbbiban azok tizenöt évcsoportra lettek bontva (ahol öt könyv egy csoport, azaz nem évek szerint skálázott). Az egyes elemzésekben a csoportosítás megalapozott, az érveléshez illeszkedik, pusztán a két vizsgálat összevetése nem válik így lehetővé.

*

A szerző a kötet zárlatában kitér a digitalizáció előnyeire és hátrányaira, valamint néhány mondat erejéig a (nyílt) tudomány fontosságára. Utóbbit hadd idézzem: „[É]ppen a tudomány intézményei teremthetik meg azt a keretet, amelyben szabályozott módon, tudást és értéket teremtve hasznosíthatók az egyes módszerek,

hiszen ezek az intézmények nem a profitmaximalizálásban és a versenylőny megszerzésében érdekeltek, hanem a tudás ellenőrzött és nyílt terjesztésében.” (229)

Bár az elmúlt hányatott évtized ennek gyökeresen ellentmond a tudomány intézményeit illetően (átszervezések, szétszabdalások, forrásmegvonások, szabadalmi/szellemi tulajdonnal kapcsolatos kötelességek szabályozása, a kutatási intézmények direkt hasznát kereső átvilágítások, a tudományometriai és kutatásértékelési őrület, a tudomány presztízisének hanyatlása stb.), csak egyetérteni tudok abban, hogy a tudomány előrehaladását az segíti, ha annak eredményei elérhetőek a tudományos közösség és a finanszírozó szélesebb közösség (lásd még: adófizető) számára. Tiszteletre méltó hát a szerző azon törekvése, hogy kutatási adatainak egy részét közzétette hagyományos publikációi mellett.

Szemes Botond könyve azt állítja: ez „A” digitális irodalomtudomány. A fentiek tükrében inkább azt mondanám: ez *is* digitális irodalomtudomány, s mint ilyen, a kötet tudományos értéket, úttörő kutatásokat, fontos felismeréseket tartalmaz.