

Szemes Botond

Digitális irodalomtudomány

Olvasás, számítás és ábrázolás egysége az irodalmi elemzésben

Kijárat
Budapest, 2024

Makkai T. Csilla

AZ ADATOK VONZÁSÁBAN

A tudománytörténet valójában paradigmaváltások sorozataként is felvázolható, melyek során az új felfedezések, technológiák és elméletek alakítják a gondolkodásmódot, a módszertant, egyben meg is határozzák az újabb kutatási irányokat. A bölcsészettudomány különböző területeit provokálja az elmúlt évtizedekben kiforrni látszó *digitális bölcsészet (digital humanities)* – mely területtel kapcsolatban méltán vetődik fel a kérdés, hogy képes-e beteljesíteni ígéretét? Miközben egyre több konferencia, szakmai műhely, képzés, kutatócsoport szerveződik a digitális bölcsészet jegyében, a bölcsészettudományi kutatások módszertanát, kérdéseit még nem változtatta meg gyökeresen; tehát továbbra sem dőlt el, hogy paradigmaváltó erővel bír-e hosszú távon a bölcsészettudomány területén, vagy pusztán eszközöket biztosít a gyorsabb, pontosabb, nagyobb adatmennyiségen alapuló kutatásoknak.

A diszciplína hazai intézményesülését jelzi mindenesetre a 2018-ban alapított Digitális Bölcsészet folyóirat, valamint a Helikon Irodalom és kultúratudományi Szemle a témának szentelt lapszámai (2017/2., 2020/1.). Meglátásom szerint ebbe a tendenciába illeszthető Szemes Botond átfogó, tudományos eredményeket felmutató munkája is, a *Digitális irodalomtudomány – Olvasás, számítás és ábrázolás egysége az irodalmi elemzésben* című kötet. A mű megkerülhetetlenné válik mindazon

kutatók számára, akik magyar nyelvterületen a digitális irodalomtudomány elméleti megfontolásaival, módszertani kérdéseivel és tudományos eredményeivel kívánnak foglalkozni. A kutatások részeredményei tudományos folyóiratokban független tanulmányok formájában is olvashatók, és mivel a könyv alapját a szerző doktori értekezése képezi, sokak számára ismerős lehet a kérdések tárgya és elméleti keretezése. Mindezek ellenére több okból is indokolt a Kijárat Kiadó és a szerző döntése, hogy az eredményeket könyv formában is publikálják: egyrészt hiánypótló jellege miatt különösen hasznos ez a megjelenés, másrészt pedig mindazok számára, akik figyelemmel kísérték a szerző munkásságát, végre összeáll a kép, láthatóvá válik az összefüggések hálózata.

Szemes Botond a kötetében nem taglalja túl hosszasan a tágan értett digitális bölcsészet múltját és jövőjét, érvelésének fókusza, ahogy a cím is jelzi, a *digitális irodalomtudomány*. Pontosabban az, hogy milyen ismeretekhez vezet a gépi olvasás, valamint az egyéb olvasásmódok szintézise alapján végzett értelmezői munka – a *digitális irodalomtudomány* definiálásához a szerző tehát episztemológiai dimenziót javasol. Ha már a digitális irodalomtudomány megnevezésnél tartunk, első látásra az olvasó talán kételkedik a cím ígéreteiben. A kötet viszont sikeresen teljesíti a vállalt feladatot: nemcsak a fogalom meghatározása történik meg az olvasás, számítás és ábrázolás kultúrtechnikáinak viszonylatában, hanem esettanulmányok formájában válaszokat kapunk olyan valós irodalomtörténeti kérdésekre is, mint például, hogy miként változott a magyar nyelvű regények mondathosszúsága, valamint hogy a tagmondatkapcsolatok feltárása által mit tudunk meg egyes szerzők prózanyelvéről. Míg az irodalmi szövegekkel foglalkozó, digitális bölcsészeti módszertant érvényesítő publikációk jelentős hányada megmarad az elvégzett kvantitatív vizsgálatok bemutatásánál, vagy kizárólag elméleti kérdéseket fogalmaz meg, esetleg a tudományterület egy-egy vitatott témájáról értekezik, ebben a kötetben egyszerre történik reflexió mindezekre, kiegészülve a rendelkezésre álló adatok értelmezésének mozzanatával.

A mű első felében az olvasás kultúrtechnikáinak elméleti keretezése történik: a szerző az Új Kritikához köthető *close reading* nagy karriert befutó módszerét kontextualizálja, majd kiegészíti azon új, tudományos olvasási tendenciákkal, melyek a mediális környezet változása hatására jöttek létre, olyanok, mint a *helyrehozó*, *szakadozó*, *felszíni* olvasás, amelyek közös vonása, hogy „az emberi olvasásfolyamatot digitális–online

logika szerint kívánják újragondolni" (24). A tárgy szempontjából a fejezet legfontosabb mozzanataként a távoli és gépi olvasás közötti különbségtételt jelölném meg. Szemes reflektál a Moretti-féle *distant reading* fogalomra, aki „nem kifejezetten számítógépes környezetben végzett és statisztikai számításokon alapuló szövegelemző eljárásokra alkotta meg a kifejezést" (34), hanem leginkább a nyomtatásalapú megismerés, a termelés és fogyasztás kvantitatív alapú feltárására dolgozta ki koncepcióját. A kortárs digitális bölcsészeti technikák pedig nem feltétlenül a távoli olvasás gyakorlata, mint inkább az adattudomány diszciplínája felől közelíthetők meg. Ebben a keretben a digitális szöveg maga adatként funkcionál, és a nagy adathalmazon elvégzett műveletekkel állítható elő új ismeret. Függetlenül attól, hogy az elvégzett műveletek a szövegbányászat (text mining), a természetesnyelv-feldolgozás (Natural Language Processing – NLP), a kvantitatív stilisztika vagy a stilometria módszertanát követik, közös jellemzőjük, hogy statisztikai alapon dolgoznak fel digitális szövegeket, különböző – az emberi olvasás számára elfedett – nyelvi markerek megszámlálásán keresztül. „A gépi olvasás nem állít mást, mint hogy ezek a területek, vagy legalábbis egy részük, igenis számszerűsíthetők, és az így kapott számsorok statisztikai kiértékelésével releváns irodalom- és kultúrtörténeti ismeretek hozhatók létre" (40), oly módon, hogy a gépi olvasás során nem a szavak jelentése, hanem például előfordulásuk gyakorisága bír relevanciával.

Viszont a számadat önmagában – főleg nagy adathalmaz vizsgálata esetén – nem elegendő ahhoz, hogy új következtetések szülessenek, szükséges az eredmények ábrázolása is. A szerző a diagramok, ábrák, térképek, grafikonok jelentőségét nemcsak puszta szemléltető eszközökként kontextualizálja, hanem a kultúra diszkurzív működésének kiegészítőjeként: dialektikus viszonyt tételezve diszkurzivitás és ábrázolás között. A leképezés kérdését vizsgálva Sybille Krämer vonatkozó munkáit veszi alapul, aki a médiaelmélet felől jut el a *diagrammatológia* fogalmához. Emellett Charles Sanders Peirce a logika felől közelít a kérdéshez, állítása szerint „a diagramok segítségével meghozott következtetések *minden* gondolkodási folyamat és (deduktív) következtetés modelljét (szó szerint láthatóvá tételét) jelentik" (91). Bár a 150 regényen végzett kutatások eredményeinek kommunikációját nagymértékben elősegítik a grafikonok, az ábrázolások kapcsán érdemes megemlíteni, hogy a kötet borítóján látható hálózat, amely vizuálisan erőteljes és figyelemfelkeltő, sajnos nem kap érdemi elemzést a könyvben. A háló-

zatelmélet diszciplínája talán szétfeszítette volna a dolgozat kereteit, de így mindenesetre némi félreértést kelthet az olvasóban. A megépített hálózat létrehozásáról olvashatunk a *Függelék*ben, de az összességében talán inkább esztétikai célokat szolgál, semmint hogy szerves része lenne a kötet tudományos érvrendszerének.

A *Digitális irodalomtudomány* egészét szakmai alaposság és átgondolt felépítés jellemzi. A szerző óriási ismeretanyagot mozgat (azt kutatva, hogy a diagramok hogyan és miért fontosak az ismeretek előállításában, Szemes Botond a térképészet kezdetéig kalauzol minket), és bár meggyőzően érvel amellett, hogy a téma ezen vetületeit különkülön is ismernünk kell ahhoz, hogy megértsük relevanciáját és azt, hogy milyen kiaknázatlan lehetőségek rejlenek a digitális bölcsészetben a kultúratudomány nexusában, olvasás közben mégis létrejön egyfajta várakozás, türelmetlenség arra nézvést, hogy szóba kerüljön a téma gyakorlati oldala is. Nem mintha a kötet első részében nem találkozott volna az olvasó megannyi újszerű, inspiráló gondolattal, de a feltevések „bizonyítása", a mérések prezentálása és az eredmények értelmezése csak ezután következik.

A mondathosszúság egyszerre stílusmarker, valamint egy korszak irodalmi nyelvi normájának közvetítője. A tézis, miszerint csökkenő tendencia látszik a mondathossz változásában, talán ismerősen hangzik, viszont magyar nyelvterületen ezt eleddig csak kisebb méretű korpuszon, emberi számításokkal mérték meg. Módszertanilag jó kiindulópont a tárggyal kapcsolatos, statisztikai alapokon nyugvó kvantitatív stilisztikai kutatások ismertetése és tárgyalása, viszont egyértelmű, hogy a szerző 150 regényen elvégzett mérései jelentik a témában az igazán releváns eredményeket. A kötet erőssége, hogy végrehajtja azt, amit állít az ismeretközvetítés és -előállítás diagrammatikus természetéről: a diagramok nemcsak a mérések eredményeit prezentálják, hanem új, további kutatást kívánó összefüggésekre is rámutatnak. A regények átlagos mondathosszának kapcsán tehát a csökkenő tendencián túl az 1970-es évektől meghatározóvá váló hosszúmondatos prózanyelv kiugró értékei is figyelmet érdemelnek úgy, mint az az új keletű tendencia, miszerint sokkal nagyobb szórás figyelhető meg ebben az időszakban: az extrém hosszú mondatokból építkező regények mellett (30-45 szó között), eltérő szövegekben az extrém rövid (5-6 szó) frázisok kapnak helyet (123). Nem véletlen tehát, hogy ebben a tekintetben a szerző az 1970-es, 1980-as évek regényeit elemzi kvantitatív adatokból kiindulva.

A kötet felépítését a szűkülő fókusz jellemzi. A mondathosszra vonatkozó elemzéseket a tagmondatkapcsolatok gyakoriságelemzése követi: „[…] az összetett mondatok kötőszavakkal kidolgozott, így nyelvileg hozzáférhetővé tett fajtáinak statisztikai elemzése […] árnyaltabb megközelítést tesz lehetővé" (151). Szemes ezen a ponton lép párbeszédbe az egyes regények recepciójával, és ehhez viszonyítva tárgyalja az eredményeket. Az ellentétes kapcsolatok esetében például Nádas Péter regényei kerülnek az első két helyre, melyre reagálva a szerző megállapítja, hogy ez rávilágít Nádas prózájának sajátosságára: a világ jelenségeit ellentétes viszonyokba rendezi (276). Miközben az eredmények alapján Szemes több kanonizált szerző recepcióját is bevonja az értekezésbe, sikerül elkülönítenie a tagmondatkapcsolatok vonatkozásában két, egymástól élesen elváló stílushagyományt.

Az utolsó fejezetek Mészöly Miklós és Déry Tibor prózanyelvének elemzésére vállalkoznak. Mészöly kapcsán a szerző arra keresi a választ, hogy milyen nyelvi markerek jelzik és segítik jobban a prózanyelv megértését a recepcióban szereplő „*sűrített, megfeszített, érzéki, vizuális, sugalmazó*" (193) jelzőkön túl. Szemes „az okozatiság kidolgozatlanságát" (200) bizonyítja be, továbbá azt, hogy az életművön belül különválasztható két írói technika, mely elősegíti azt. A korábbi munkák esetében „ezt a hiányt a magyarázatok hiánya tovább erősíti, vagy a magyarázó kapcsolatokat mondathatárok törik meg, addig a később született művek gyakrabban dolgoznak ki magyarázó mellérendeléseket, azok érvényességét azonban folyamatosan felülírják a rákövetkező szöveghelyek, amik szintén a tágabb összefüggések elbizonytalanodását eredményezik" (200). Ez összefüggésbe hozható az ellentétes mellérendelések használatának kiugró értékével, mely tovább erősíti azt a narratívát, mely a Nádas-prózapoétikát erőteljesen a Mészöly-prózához köthetőnek tartja.

E néhány példával próbáltam szemléltetni a kötet belső logikáját, valamint azt, hogy milyen értékes, a szöveghagyományok közötti hasonlóságokra és különbségekre rávilágító számtalan következtetést tartalmaz a munka. Ilyen értelemben a könyv mintegy iránymutatás is a jövő digitális irodalomtudományi kutatásai számára, amelyek a digitális módszertanok mélyebb megértését és szélesebb körű alkalmazását hivatottak előmozdítani a magyar tudományos diskurzusban.