

Tematikus összeállítás

SZÁMÍTÓGÉPES TÁRSADALOMTUDOMÁNY

COMPUTATIONAL SOCIAL SCIENCE

VENDÉGSZERKESZTŐK: RUDAS TAMÁS, PÉLI GÁBOR

BEVEZETŐ

INTRODUCTION

Rudas Tamás¹, Péli Gábor²

¹ az MTA doktora, az MTA Társadalomtudományi Kutatóközpont főigazgatója, rudas.tamas@tk.mta.hu

² az MTA külső tagja, az MTA Társadalomtudományi Kutatóközpont főigazgató-helyettese, peli.gabor@tk.mta.hu

ÖSSZEFOGLALÁS

Bevezetőnkben elsőnek az MTA keretében a közelmúltban létrejött *Számítógépes Társadalomtudomány* témacsoport céljait ismertetjük, majd röviden áttekintjük, hogy a tematikus blokk írásai miként járulnak hozzá a célok megvalósulásához.

ABSTRACT

First, we introduce the goals of the *Computational Social Science* thematic group that have been recently established within the organizational frame of the Hungarian Academy of Sciences. Then we briefly survey how the papers of this thematic bloc contribute to the fulfillment of these goals.

Kulcsszavak: számítógépes társadalomtudomány, transzdiszciplináris kutatások, elméletépítés

Keywords: computational social science, transdisciplinary research, theory building

Az MTA Társadalomtudományi Kutatóközpontja kezdeményezésére 2017 őszen megalakult az MTA *Számítógépes Társadalomtudomány (Computational Social Science)* témacsoportja. Hivatalos bemutatkozásunkra 2017. november 14-én, a Magyar Tudomány Ünnepe keretében került sor. Tematikus összeállításunkban az akkor elhangzott előadások kibővített és szerkesztett változatai szerepelnek.

Témacsoportunk erőfeszítéseinek középpontjában a nagy teljesítményű számítástechnikai módszerek társadalomtudományi kutatásokban való felhasználásának elősegítése áll. A Magyar Tudományos Akadémia fő küldetésének megfelelően különös hangsúlyt helyezünk a feltáró jellegű alapkutatásokra. Döntő fontosságúnak tartjuk azonban, hogy ezek kiegészüljenek az alkalmazásra irányuló kutatásokkal, valamint a társadalmi és kutatói felhasználóknak szánt algoritmikus eszköztár fejlesztésével. Segítségünkre lehet az a szerencsés körülmény, hogy a magyar társadalom-, természet- és adattudományi szakemberek körében már mintegy két évtizedes hagyománya van a számítógépes támogatású társadalomkutatás egyes kérdésköreiben, leginkább a társadalmi hálózatkutatás terén való együttműködésnek. Tapasztalataink szerint ez a laza szálakból fonódó, de folytonos transzdiszciplináris kapcsolat számottevő versenyelőnyt jelent a magyar tudományosságban a nemzetközi mezőnyben. Előnyünk azonban illanékony. Hogy tartósan megőrizhessük, a kialakult tudományközi hálózat laza kapcsolódásait módszeres együttműködési rendszerré szükséges alakítani, mely egyszersmind jól azonosítható entitásként, elismert szereplőként jelenik meg a nemzetközi tudományosságban is. Ezt elősegítendő kiemelt figyelmet kívánunk fordítani a módszertani vonzatú kutatásokra, melyek átjárást biztosítanak az egyes tudományágak történetileg rögzült diszciplináris megközelítései között. Ezzel témacsoportunk a magyar kutatók egy olyan új generációjának a képzéséhez is hozzá kíván járulni, melynek tagjai képesek az egyes, ma még egymástól nagy módszertani távolságra eső szaktudományos ismeretanyagok hatékony ötvözésére. Külön öröm számunkra, hogy tematikus blokkunk sokszerzős cikkeinek hozzájárulói között nagy számban találunk tudományos pályafutásuk első harmadában-negyedében járó fiatal kutatókat. E változatos diszciplináris háttérű kutatók szakmai szocializációja szakterületeik meghatározó kutatóinak intézményi műhelyeiben folyik. Ezek az intézmények egyúttal a Számítógépes Társadalomtudomány témacsoport alapító tagjai is.

A tematikus összeállításunkat alkotó cikkek sorához Vámos Tibornak, a számítógépes társadalomtudomány hazai nesztorának írása (*Akadémiánk és az információs világ*) adja meg a hangütést. A társadalmi algoritmizáció mindenütt jelen lévő folyamata során az emberi tényező, az értelmezési kereteket kiépítő intellektus szerepe döntő. E szerep betöltése azonban egy transzdiszciplináris együttműködést lehetővé tevő közös szakmai nyelv kialakítását feltételezi. Tematikus blokkunk egy ilyen, a természet-, társadalom- és adattudományokat összekötő,

a számítógépes társadalomtudomány együttes művelését segítő közös tudásalap kialakításához kíván hozzájárulni.

A következő négy írás mindegyike valamely, a számítógépes társadalomtudomány témakörébe tartozó területtel és az ehhez kapcsolódó kutatói megközelítésmódokkal ismerteti meg. Novák Attila, Siklósi Borbála és Prószék Gábor cikke (*Segíthetnek-e a szóbeágyazási modellek a társadalomtudósoknak?*) az MTA–PPKE Magyar Nyelvtechnológiai Kutatócsoportjának szójelentés-reprezentációs kutatásait ismerteti. Neurális hálózatokon alapuló, egyben grafikus megjelenítéssel támogatott modelljeik meglepő „kreativitással” ragadják meg a vizsgált kifejezések tágabb stílári és csoportnyelvi vonatkozásait. Ács Judit, Borbély Gábor, Makrai Márton, Nemeskey Dávid, Recski Gábor és Kornai András írása (*Hibrid nyelvtechnológiák*) a hagyományos, szabályalapú nyelvészeti megközelítéseknek a nagyméretű adatállományok gépi tanulós módszerekkel történő átszűrésén alapuló módszerekkel való összekapcsolása mellett érvel, az MTA SZTAKI Nyelvtechnológiai Kutatócsoportja által követett hibrid megközelítésen alapuló kutatásokat ismerteti. Galántai Júlia, Pápay Boróka, Kubik Bálint György, Szabó Martina Katalin és Takács Károly első ránézésre meglepőnek tűnő című írása (*A pletyka a társas rend szolgálatában. Az informális kommunikáció struktúrájának mélyebb megértéséért a Computational Social Science eszközeivel*) a pletykának a kooperáció és a társadalmi normák karbantartását szolgáló szerepéről szól. A pletykakutatás a számítógépes támogatású társadalmi hálózatok kutatás egy gyorsan fejlődő irányzatának, az úgynevezett negatív kapcsolatok vizsgálatának egyik érdekes leágazása. Kertész János, Roberta Sinatra és Vedres Balázs írása (*A társadalom hálózati jelenségeinek adatvezérelt vizsgálata*) három, a CEU Hálózattudományi Központjához kapcsolódó szimulációs módszertanú vizsgálatot ismerteti. Az első a társadalmi hatások kaszkádszerű terjedését szemlélte, egyebek között a hazánkban egykor népszerű iWiW-szolgáltatás összeomlásának példáján. A másik vizsgálat a *teamek* eltérő kreativitási fokára keres magyarázatot a hálózati szerkezeti gyűrődések (structural folds) alapján, míg a harmadik a tudományos hatás életkorfüggési mintázatát tárja fel a közlemények citációs hálózatára támaszkodva.

Az ezeket követő három írás az adatinfrastruktúra alakulásának a tudomány művelésében betöltött szerepét járja körül. Lévai Péter és Telcs András hozzájárulása (*Új módszerek régi kérdések megválaszolására az Akadémiai Felhőben. Hálózatok és oksági kapcsolatok felderítése a társadalomtudományokban*) bemutatja, hogy az akadémiai felhő szolgáltatásai miként támogathatják az adatigényes társadalomtudományi kutatásokat. Noha a számítógépes felhőt az MTA Wigner Fizikai Kutatóközpontjában, illetve az MTA SZTAKI-ban hozták létre, annak szolgáltatásai valamennyi tudományterület számára nyitva állnak. A Wigner Kutatóközpontban egy újonnan létrehozott osztály kifejezetten azt hivatott elősegíteni, hogy e lehetőséggel hazánk számítástechnikai részletekben kevésbé járatos

kutatói is élni tudjanak. Kovács László, az MTA SZTAKI Elosztott Rendszerek osztályának vezetője az emberi tényező tudományterületenként változó szerepét veszi szemügyre (*Adatkezelés. A [kutatási] adatok kezelése a magyar tudományos- és memóriaintézményekben*). Az egyes diszciplínák adatkezelési szokásrendszerei egyedileg, sokszor egy jóval korábbi technikai szinthez igazodva alakultak ki. Ezeknek a meggyökeresedett rutinoknak a *state-of-the-art* adatkezelési lehetőségekkel való felváltása döntő fontosságú, bár korántsem egyszerű feladat. Simon Eszter és Váradi Tamás munkája (*Magyar nyelvtechnológiai infrastruktúra a társadalomtudományok szolgálatában*) az MTA Nyelvtudományi Intézetében kifejlesztett magyar nyelvű szövegeket feldolgozó eszközláncot – nyelvi elemző eszközök egyetlen koherens technológiai láncba rendezett együttesét – ismerteti társadalomtudományi alkalmazásokon keresztül. A különféle digitális formában előálló nyelvi korpuszok már ma is számos szociológiai, politikatudományi, lélektani kontextusú empirikus kutatás számára szolgálnak lehetséges bemenetként. Így az adatfolyamok elemzését támogató felhasználóbarát eszközlánc az adattudományokban kevésbé járatos kutatók széles köre számára teheti elérhetővé a fejlett nyelvtechnológiai eszközök kutatásaikban való alkalmazását. Végül Muraközy Balázs tanulmánya (*Gépi tanulás, predikció és okság a közgazdaságtudományban*) a metaszintet feszegeti. Hogyan változik a hagyományos oksági, elméletalapú és a gépi tanulásra építő statisztikai-predikciós fókuszú kutatások viszonya a társadalomtudományok, jelesül a közgazdaságtan területén? A szerző a két megközelítés komplementaritását és a két megközelítést hatékonyan ötvöző empirikus módszerek kikísérletezésének szükségességét emeli ki.

Ezzel visszaértünk tematikus blokkunk központi gondolatához. A számítógépes társadalomtudomány talán legnagyobb kihívása, hogy az értelmezési kapacitás lépést tud-e tartani a számítógépes módszerek *output*jainak rohamos bővülésével. Ha igen, úgy ez nagy teherbírású transzdiszciplináris értelmezési keretek kiépítését feltételezi. Olyan módszerek megalkotását, melyekkel a hagyományos oksági-analitikus kutatás „megmagyarázott szabályszerűségekké” alakítja azokat az empirikus mintázatokat, melyeket a gépi-statisztikai masinéria az adattömegben felfedez. A nagy teljesítményű számítógépes módszerek hozadéka így nem a társadalomra vonatkozó elméletalkotás kiüresítése, hanem hatékony támogatása lehet.

AKADÉMIÁNK ÉS AZ INFORMÁCIÓS VILÁG

OUR ACADEMY AND THE WORLD OF INFORMATION

Vámos Tibor

az MTA rendes tagja, professor emeritus, MTA SZTAKI
vamos@sztaki.mta.hu

ÖSSZEFOGLALÁS

Villámnézet az információs világ és akadémiánk viszonyairól, az emberi tényezőkkel kapcsolatos legfontosabb feladatokról és a robotvilágról szóló históriákról.

ABSTRACT

A short flash on the relation of our Academy and the World of Information, on the most important tasks concerning the human aspects and the world of robots hysteria.

Kulcsszavak: információs világ, emberi tényező, automatizálás, mesterséges intelligencia

Keywords: world of information, human factor, automatization, artificial intelligence

Kibernetika? Burzsoá áltudomány, ezzel kezdődött a válasz a szovjet birodalomban arra az Amerikában elindult történelmi forradalomra, amelynek jelentőségét, méreteit ott sem tudták felmérni, ahol elkezdődött, és azok sem, akik megalapozták.

A feleszmélés errefelé körülbelül egy évtized késéssel indult, a rendszert időnként rázogató reformmozgalmak során. Itthon nem sokkal 1956 első tanulságainak bátortalan és ellentmondásos emésztése idején, nem véletlenül az Akadémián és az ipari vezetés reformköreiben. Az utóbbi egy másik, többé-kevésbé jól feldolgozott történet, amely néhány nyugati számítógép importjával és az azok körül gyülekezőkkel indult, a mienk a Rákosi-börtönből szabadult Tarján Rezsővel és a csodásan forgó agyú Kalmár Lászlóval, majd egy kibernetikai kutatócsoport alakításával. Itt hamar összegyűltek a nagyra hivatott ifjú tehetségek, akik messze látóan érezték a máig kibontakozó fejlődést, így az operációkutatásban, azaz a folyamatok optimális tervezésében úttörő Prékopa András, majd Kornai János, a nyelvészetben Kiefer Ferenc, a szoftvereknél Dömölki Bálint. A későbbi évek sok

és értékes alkotócsapata kapcsolódott hozzájuk, itt töltötték tanuló éveiket. A korai szellemi pezsgést mutatja, hogy az őstörténet naiv számítógép-építési kísérlete mellett még zenei struktúraelemzés is született. Az akadémiai intézménylétesülés a SZTAKI (Számítástechnikai és Automatizálási Kutatóintézet) és a KFKI (Központi Fizikai Kutatóintézet) kereteiben teljesedett ki.

Meg kell jegyeznünk, hogy ezeknek az új diszciplínáknak a helyei Akadémiánkon még mindig nem teljesen elfogadottak, ezért is fontos a most induló interdiszciplináris kezdeményezés, amelyben a történet feldolgozott anyagainak gyűjtése és újraértékelése fontos feladat a lemaradásaink és további lehetőségeink vizsgálatában.

Szerintem ezekben a ma és főleg a közel-holnap legnagyobb kérdése az emberi szerep. Ma ez két póluson jelentkezik, a szakemberhiányban és a ma már igényesebb feladatok elvégzésére alkalmatlan *underclass* milliós tömegében. A holnap, azaz a közvetlenül következő néhány évtized kegyetlenebb és kegyesebb lehet. Az automatizálás feltartóztathatatlan technológiai és minden tevékenységet érintő folyamata kiszorítja a társadalmak többségét mai munkakörülményeiből, és közben változó feladatokat teremt egy egyfelől szórakoztató és szórakozó világnak, másfelől a mély tartalmú humán szolgáltatásoknak. Mindez már ma meg kell hogy mozgassa a természettudományok és a társadalomtudományok újonnan alakuló egészét.

Ahogy a politikában, úgy ebben is felmerülnek figyelemelterelő nézetek. Sokszor igen tekintélyes, alkotó tudósok is beleesnek a tömegcsábító érdekesség és híresség csapdájába. A rém az embereket először csak helyettesítő, majd felülmúló, pusztító robot és annak mesterséges intelligenciája. A mítoszok és egyéb fantazmagóriák eljövendő világát nehéz logikai módszereinkkel kizárni. Részben elegendő, ha ezeket a bizonyíthatatlanság agnosztikus válaszával határoljuk el valós teendőinktől. Van sok ellenérvünk is, hiszen az ember és a gép két nagyon különböző fejlődés eredménye, az egyik az evolúció sok millió éves, izgalmasan összetett, jóval, rosszal kevert és azokat őrző szülőtte, a másik egy korának válaszoló célszerkezet. Az agy kutatás majdnem naponta tár fel újabb és újabb olyan bonyolultságokat, amelyek a belátható időben elérhető gépi reprezentációk számosságát messze meghaladják. Ezekben a számszerűségeken, a bonyolultságok becsléseiben is elengedhetetlen és szerencsére megállíthatatlan a kutatói kíváncsiság és eszközfejlődés.

A mai, útválasztásokat tükröző mítoszoknak is ősieik a gyökerei. Az egyik a Gólem, a félelmetes (a horrorfilmekben), a másik az Afrodité-kultusz körül alakult ideálkép, Pygmalion, Páfosz és Galatea története, a művészi alkotás csodájáé, a szépségé és a szerelemé, meg a máig élő kultúrvárosé. A mítosz Gólemét a bölcsesség egy mozdulattal megsemmisítette, a maiakat nehezebb lesz.

Mindennek feltétele a közös nyelv. Ma még egy-egy diszciplínán belül is nyitott probléma, akadály a együttműködéseknek, a részeredmények kölcsönös felhasználásából építkező tudásnak. Találkozunk megértéssel!

SEGÍTHETNEK-E A SZÓBEÁGYAZÁSI MODELLEK A TÁRSADALOMTUDÓSOKNAK?

CAN WORD EMBEDDING MODELS HELP SOCIAL SCIENTISTS?

Novák Attila¹, Siklósi Borbála², Prószéky Gábor³

¹PhD, Pázmány Péter Katolikus Egyetem Információs Technológiai és Bionikai Kar,
MTA–PPKE Magyar Nyelvtudományi Kutatócsoport, novak.attila@itk.ppke.hu

²PhD, Pázmány Péter Katolikus Egyetem Információs Technológiai és Bionikai Kar,
MTA–PPKE Magyar Nyelvtudományi Kutatócsoport, siklosi.borbala@itk.ppke.hu

³az MTA doktora, Pázmány Péter Katolikus Egyetem Információs Technológiai és Bionikai Kar, MTA Nyelvtudományi Intézet
proszeky.gabor@itk.ppke.hu

ÖSSZEFOGLALÁS

A nyelvtechnológiában az utóbbi néhány évben előtérbe kerültek az olyan disztribúcióalapú szójelentés-reprezentációs modellek, amelyek a szavak jelentésének a szűken vett grammatikai és szemantikai dimenzióin túl a tágabb stiláris, illetve szociolektális (csoportnyelvi) dimenzióit is meglepő pontossággal megragadják. Ezért ezek a mesterséges neurális hálózatokon alapuló szóbeágyazási modellek nemcsak a nyelvtechnológusoknak, sőt nem is csak a nyelvészeknek érdekesek, hanem mindazon tudományágak képviselőinek gazdag tudásforrást jelenthetnek, akik számára a szövegek alapvető nyersanyagként szolgálnak.

A társadalomtudósok számára különösen érdekesek lehetnek azok a szövegek, amiket a különféle közösségi oldalak felhasználói vagy akár az online sajtóhírekhez fűzött hozzászólások szerzői generálnak. Rengeteg ilyen szöveg áll rendelkezésre ma már elektronikus formában, és ez lehetővé teszi, hogy jó minőségű modelleket hozzunk létre a korábban említett technológia felhasználásával, és azokat különböző dimenziók mentén kereshetővé tegyük. Ízelítőként bemutatunk néhány példát a modell által megfogható jelenségek köréből.

ABSTRACT

Distributional models of word meaning have recently become ubiquitous in language technology. These models represent in remarkable detail the meaning of words encompassing not only the narrow grammatical and semantic but also the wider stylistic and sociolectal dimensions. Thus these word embedding models created using artificial neural networks are not only interesting for NLP researchers or linguists, but they can be rich sources of knowledge also for social scientists, for whom texts serve as essential research material.

Texts generated by users of social media sites and comments on articles published on-line at news portals may be of special interest for social scientists. A great amount of such text is available in a digital form, and this makes it possible for us to create high-quality models using the technology mentioned above, and to make them searchable along various dimensions. As a showcase, the paper presents some examples of the phenomena tackled by the model.

Kulcsszavak: szóbeágyazási modellek, neurális hálózatok, disztribúciós szemantika, csoportnyelv, regiszter

Keywords: word embedding models, neural networks, distributional semantics, sociolects, register

DISZTRIBÚCIÓS MODELLEK

A strukturalista nyelvészek az 1930-as években azt az álláspontot fogalmazták meg, hogy a nyelvi tudás elsődleges forrása a szavak és morfémaák disztribúciója. Ennek bizonyítására azonban csak napjaink új tudományos eredményeinek felhasználásával adódott lehetőség. Napjaink digitális társadalma nagyon nagy mennyiségben állít elő újabb és újabb szöveges tartalmakat, melyekben a nyelv alakulása, illetve a nyelvhasználat különböző rétegei jól tetten érhetők.

A disztribúciós szemantika a strukturalisták által korábban megfogalmazott elvet olyan formában fogalmazza újra, hogy a szavak jelentése szorosan összefügg azzal, hogy milyen kontextusban használjuk őket (Firth, 1957). Az egészen a közelmúltig egyeduralgó hagyományos számítógépes disztribúciós szemantikai modellek létrehozásakor az egyes szavakhoz tartozó reprezentáció ténylegesen az adott szó előre meghatározott méretű környezetében előforduló szavak egy nagy korpuszból számított előfordulási statisztikáit tartalmazta. Ezek a modellek – annak ellenére, hogy bizonyos eredményeket értek el – nem igazán váltották be a hozzájuk fűzött reményeket. Ezzel a fajta reprezentációval az az egyik fő probléma, hogy a legtöbb szó környezetében a legtöbb másik szó soha nem fordul elő, ezért az együttes előfordulásokat ábrázoló mátrix „ritka”, ugyanis legtöbb pozíciójában 0 áll.

Az áttörést napjaink nyelvtchnológiai kutatásainak egyik kurrens módszere hozta, amely a szövegek alapján mesterséges neurális hálózatok alkalmazásával folytonos vektortérbeli tömör reprezentációkat, ún. *szóbeágyazásokat* (word embedding) hoz létre. Az alap gondolatot Yoshua Bengio és munkatársai vetették fel a 2000-es évek elején (Bengio et al., 2003), de a hatékony gépi háttér igazán csak a 2010-es években tette lehetővé az igazán nagy méretű modellek betanítását. A tanítás során az egyes szavak fix méretű környezetében szereplő többi szót vesszük figyelembe, az ezekből álló vektor azonban egy neurális hálózat bemenete. A környezetben álló szavak összességét reprezentáló vektorokat használja a hálózat arra, hogy megjósolja az adott környezetben legvalószínűbb célszót. Szemben a hagyományos számolásalapú módszer milliányi dimenziós ritka mátrixaival, az így létrehozott pár száz dimenziós vektorok mindegyik pozíciójában egy -1 és $+1$ közötti, szinte minden esetben 0-tól különböző szám szerepel. Az egyes dimenzióknak nincs saját jelentésük, hanem a hálózatot alkotó mesterséges

idegsejtek közötti kapcsolatok erősségét reprezentálják. A tanítás során a rendszer összehasonlítja a hálózat által a környezet alapján jósolt szót az ott ténylegesen szereplővel, és a hiba visszaterjesztésével, illetve ennek megfelelően a környezetet reprezentáló vektorok frissítésével jön létre a tanítás végén a célszót helyesen megjósoló súlyvektor, ami a neurális hálózat megfelelő rétegéből közvetlenül kinyerhető. Mivel a hasonló szavak hasonló környezetben fordulnak elő, ezért a szöveggörnyezetre optimalizált vektorok a hasonló jelentésű szavak esetén hasonlóak lesznek.

Ebben a rendszerben a lexikai elemeket egy valós vektortér egyes pontjai reprezentálják, melyek konzisztensen helyezkednek el az adott térben, azaz az egymáshoz szemantikailag és/vagy morfológiailag hasonló szavak egymáshoz közel, a jelentésben eltérő elemek egymástól távol esnek. Mindemellett vektoralgebrai műveletek is alkalmazhatók ebben a térben, tehát két elem szemantikai hasonlósága a két vektor távolságaként meghatározható, illetve a lexikai elemek pozícióját reprezentáló vektorok összege jó közelítéssel azok jelentésének összegét határozza meg (Mikolov et al., 2013a, 2013b). A módszer hátránya csupán az, hogy önmagában nem képes a poliszémia, illetve homonímia kezelésére, tehát egy többjelentésű lexikai elemhez is csupán egyetlen jelentésvektort rendel, azonban a szakirodalomban erre a problémára is találunk sikerrel alkalmazott módszereket (Banea et al., 2014; Iacobacci et al., 2015; Trask et al., 2015).

A szóbeágyazási modellek hatékonyan ragadják meg a szövegekben megjelenő szemantikai információkat, sőt jelentős mennyiségű világismereti tudást is (Mikolov et al., 2013a). Ezek a beágyazási modellek magyar nyelvre is jó eredményre működnek kellő méretű és elemzett tanítóanyag alkalmazása esetén (Siklósi–Novák, 2016; Siklósi, 2018).

A KORPUSZ ELŐKÉSZÍTÉSE ÉS A MODELLEK LÉTREHOZÁSA

Egy nagyméretű, több mint egymilliárd szavas, a webről gyűjtött korpuszból hoztunk létre szóbeágyazási modelleket. A korpuszt automatikusan egyértelműsített morfológiai elemzéssel láttuk el. A modell építésekor nem a ragozott szavakat, hanem a szótöveket tartottuk meg, melyek után külön elemként szerepeltek a morfológiai elemző által generált címkék. Mivel ezek a címkék az aktuális szó környezetében maradtak, az általuk hordozott szintaktikai információ továbbra is szerepet kapott az egyes szavakat reprezentáló vektorok létrehozásában. Azonban mivel a modellt csak szótöveket tartalmaz, így robusztusabb modell jön létre, mint ha közvetlenül a szövegben szereplő felszíni szóalakokból építenénk a modellt, mert egy-egy szó reprezentációjának kiszámításához annak minden ragozott alakja hozzájárul. Ez a ritka szavak esetében jelentősen javítja a modell minőségét.

NYELVI RÉTEGZŐDÉS

A modellből lekérdezhető a benne szereplő szavakhoz legközelebb elhelyezkedő további szavak listája, az adott szótól való távolság szerint rendezve. Ezt a műveletet a már megjelenített elemek egy részhalmazán folytatva feltérképezhető az adott régió szókincse. A nyelvi rétegek és a rétegnyelvek példátlan gazdagságban és árnyaltságban jelennek meg a lexikai térben, kezdve az *online* játékok rajongóinak zsargonjától a *fanfiction* irodalmat felvonultató fórumok látogatóinak speciális szóhasználatán keresztül a szemészeti szaknyelv rétegein át egészen a vasúti irányítórendszerek szakterminológiájáig. Az így létrejött szólistán automatikus klaszterezési eljárást alkalmazva további tematikus osztályozást végezhetünk, illetve kiszűrhetjük az oda nem illő elemeket. Az 1. táblázatban a fenti kategóriákból választott *kempel*, *ficc*, *macula* és *balíz* szavakhoz kérdeztük le a modellből a hozzájuk legközelebb eső első néhány szót. A terjedelmi korlátok miatt itt csak a listák elejét van lehetőségünk bemutatni, azonban általánosan elmondható, hogy az ilyen listáknak akár még a többszázadik elemei is releváns kifejezéseket tartalmaznak, amelyek természetesen adott esetben már lazább kapcsolatban állnak az eredeti szóval. A vektortérben olyan típusú kategóriák is elkülönülnek, amilyen típusú megkülönböztetés semmilyen létező szótárban nem szerepel, és sokszor megfelelő elnevezést sem könnyű találni az adott kategória számára. A 2. ábrán látható például, hogy világosan elkülönülnek a férfi keresztneveken belül az „avított dzsentrinevek”, a hagyományos keresztnevek, az átlagos gimnáziumi osztálynévsorban fellelhető trendi fiúnevek és a roma kiskorúak divatos angolszász–újlatin keresztnevei.

1. táblázat. A különböző rétegnyelvekből való *kempel*, *ficc*, *macula*, *balíz* szavak és a hozzájuk legközelebb eső néhány szó a vektortérben

kempel	ficc	macula	balíz
wowozik	fic	sárgafolt	balízcsoport
farmol	fici	degeneratio	vezérlőjel
fearless	fanfic	atrophia	főjelző
healel	törid	glaukóma	transzponder
VF-ezik	ficu	látóidegfő	vágányút
hackel	drarry	szürkehályog	vezérlőegység
maxol	fanfiction	makula	EVC
castol	sztory	ideghártya	jelsorozat
turret	snarry	látóhártya	menetengedély
leöl	SSHG	zöldhályog	kijelzés
sentry	oneshot	centralis	DMI
questel	feji	látóideg	vezérlőközpont
betámad	függővég	glaucoma	riasztóközpont
lewarezol	manga	naevus	komparátor
limpel	dorama	erythema	nyugtázás



2. ábra. Néhány férfinév elrendeződése a vektortérben



3. ábra. A többértelmű *reggeli* szó és környezete a vektortérben



4. ábra. A többértelmű *vár* szó környezete a vektortérben

2. táblázat. Néhány kultúraspecifikus szó képéhez legközelebb eső szavak az angol szóbeágyazási modellben

busó	pörc	cigó
reveler	bacon	thug
reveller	dough	strikebreaker
parade	sauce	racist
re-enactor	sliced	troublemaker
clown	gravy	Palestinians
townspeople	soup	rioter
carnival	curd	hoodlum
festival-goer	steak	Tutsis
townsfolk	stew	Jew
villager	pastry	Arab
onlooker	tortilla	bigot
festivity	lard	whites
mummer	butter	fascist
maypole	flatbread	drunk
procession	mayonnaise	bookie

A webről gyűjtött korpusz gazdagon tartalmaz olyan a „nép” által írt szövegeket, amelyek a különböző webes fórumokon és a cikkekhez írt hozzászólásokban jelennek meg. Ezekben a szövegekben – és következésképpen a szemantikai vektortérben – a szókincs olyan rétegei jelennek meg (vagy egy épp ebből a rétegből vett kifejezéssel: *figyelnek be*), amelyek nyomtatott szótárakban nem szerepelnek. A modell ezeknek a szavaknak az adott szociolektális közegben szokásos jelentését is megragadja, így alkalmas lehet az adott rétegnyelv vizsgálatára, az abban való elmélyülésre (lásd az *1. ábrát*).

DOMÉNADAPTÁCIÓ ÉS -SZELEKCIÓ

Ahhoz, hogy jó minőségű modellek jöjjenek létre, a rendszernek nagy mennyiségű tanítóanyagra van szüksége. Az általunk vizsgált korpusz több milliárd szóból áll. Ha egy adott réteg- vagy szaknyelv szókincsét szeretnénk vizsgálni, akkor nem feltétlenül elegendő a modell betanításához csak az adott nyelvi réteget reprezentáló korpusz, hanem a nagyobb általános korpuszon kapott modellből kiindulva a rendszert az adott szakkorpuszon tovább tanítva létrehozható egy olyan lexikális reprezentáció, amelyben a köznyelvben dominánsan az adott rétegnyelvtől eltérő jelentésben használt szavak reprezentációja a rétegnyelvben domináns jelentéshez közelít. A rendszer tehát arra is használható, hogy egy nagyobb vegyes korpuszból egy adott rétegnyelvet reprezentáló részkorpuszt válasszunk ki annak a rétegnyelvre jellemző lexikai elemei alapján. Ehhez kiindulásként elegendő a jellemző

terminológiának csak néhány elemét megadni, majd az adott vektortérrégió közeli elemeiből automatikusan egy bővebb szakterminológiai szókinccset összeállítva és ezt lekérdezve az egész korpuszból kiválaszthatjuk a releváns részkorpuszt.

MATEMATIKAI TRANSZFORMÁCIÓK A VEKTORTÉREN

A disztribúciós modellbeli távolságmérték önmagában általában nem választja el egymástól a hasonló jelentésű, de különböző polaritású elemeket, mint például *jó-rossz*, *szép-csúnya*, illetve ezek hasonló jelentésű társait, azonban az ellentétpárokra adott példák alapján általában definiálható egy olyan transzformáció a téren, amely olyan forgatást végez, amelyet alkalmazva a vektortér valamelyik dimenziója mentén az ellentétes polaritású elemek szétválnak. Tehát bár az eredeti vektortérmodellben az egyes dimenziókhoz általában nem rendelhető semmiféle jelentés, megfelelő transzformáció után a transzformált vektortérben egy adott dimenzió specifikus jelentést nyerhet.

Egy másik probléma a homonim alakok kezelése. Bár a vektortérmodell a többjelentésű elemekhez egyetlen reprezentáló vektort rendel, ez nem feltétlenül jelenti azt, hogy ne lenne kinyerhető a modellből az egyetlen vektorban reprezentált jelentéshalmaz megfelelő gépi tanulási algoritmusok alkalmazásával. Problémát csak azok az esetek jelentenek, amikor túl sok különböző jelentése van egy szónak, illetve amikor valamelyik jelentés nagyságrendekkel gyakoribb, mint a többi. A 3. ábra azt szemlélteti, hogy a modell a *reggeli* szónak mind az 'étkezés', mind a 'napszaki' jelentését megragadja, ugyanakkor a *vár* igei használata annyival gyakoribb, mint a főnévi, hogy a főnévi jelentés alig jelenik meg a modellben (4. ábra). Az utóbbi problémára ugyanakkor megoldást jelent, ha morfológiailag annotált korpuszból építjük a modellt: ekkor két különálló vektor reprezentálja a szó igei, illetve főnévi használatát.

TÖBBNYELVŰSÉG

További érdekes lehetőségek nyílnak annak a ténynek a kiaknázásával, hogy a különböző nyelveken készített szóbeágyazási modellek topológiája általában hasonló, ezért akár néhány ezer fordítási szópár megadásával viszonylag pontos leképezés definiálható két különböző nyelvhez készült modell között. Ez lehetővé teszi egyrészt a két nyelv „rokon” lexikális mezői közötti leképezést és az egyik oldalról kiindulva a másik oldal felfedezését, másrészt a kultúraspecifikus szavaknak (például: *busó*, *pörc*, *cigó* stb.) a másik nyelven megfelelő terület megvizsgálását. Emellett a leképezés azt is lehetővé teszi, hogy az egyik nyelven hozzáférhető (akár kézzel, sok munkával létrehozott) lexikai erőforrás a másik nyelven is használha-

tóvá váljon. A 2. táblázatban a *busó*, *pörc*, *cigó* szavaknak megfelelő vektorok által meghatározott pontokhoz legközelebb eső angol szavak láthatóak az angol *Wikipédiából* létrehozott szóbeágyazási modellben. Látható, hogy a modell megragadja és leképezi a busójárás fesztiváli hangulatát, vidéki látványosság jellegét, a *pörc* szóról pedig megtudhatjuk, hogy denotátuma étel, míg a *cigó* szó leképezésével kapott listában megjelennek mind a bűnözéssel kapcsolatos, illetve az etnikai intoleranciára utaló szavak, mind a kurrens etnikai ellentétekkel kapcsolatban gyakran felmerülő nemzetiségnevek. Ugyanakkor az angol *Wikipédiából* készült korpusz nemigen tartalmaz olyan jellegű csoportnyelvi elemeket, amilyenek közé a magyar *cigó* szó tartozik, ezért az adott esetben a magyar szó és az angol modellbeli képe között nincs pontos regiszterbeli megfelelés. Egy általános angol nyelvű webkorpuszból készült modell esetén azonban nem állna fenn ez a probléma.

ÖSSZEFOGLALÁS

Írásunkban bemutattunk néhány olyan lehetőséget, amelyet a nagyméretű korpuszokból neurális hálózatok segítségével épített szóbeágyazási modellek a szövegekre alapozott kutatásokat végző társadalomtudósok számára megnyitnak. Megpróbáltuk néhány példával illusztrálni, hogy ezek a modellek igen árnyalt módon képesek megragadni a szavak és a hozzájuk kapcsolódó fogalmak tágabb értelemben vett jelentésével kapcsolatos nyelvi szinten tetten érhető tudást, beleértve a stílár, rétegnyelvi, szakterületi jellemzőket. Az ígéretes lehetőségeknek egy része még csak most körvonalazódik, hiszen a bemutatott megoldások csak néhány éve jelentek meg. Egészen pontosan: a matematikai módszerek nagy része korábban is megvolt, csak a hatékony működtetésükhöz szükséges számítástechnikai háttér nem volt meg. A cikkünkben vázolt modellek alapfogalmai, tehát a vektoros reprezentáció, a neurális hálók vagy a mélytanulás napjainkban a legtöbb területen, így a társadalomtudományi kutatások területén is új lehetőségeket nyitnak. Ezek kiaknázásához időszerű a tanuláselmélet, a nyelvtechnológia és azon társadalomtudományi területek kutatóinak összefogása, ahol a szövegekben megbújó tudás efféle feldolgozása egyre újabb és egyre hasznosabb tudományos megoldások kialakítását teszi lehetővé.

KÖSZÖNETNYILVÁNÍTÁS

A cikkünkben bemutatott eredmények részben az FK 125217 és a PD 125216 számú projekt keretében a Nemzeti Kutatási Fejlesztési és Innovációs Alapból biztosított támogatással az FK 17 és a PD 17 pályázati program finanszírozásában megvalósuló kutatások keretében születtek meg.

IRODALOM

- Banea, C. – Chen, D. – Mihalcea, R. – Cardie, C. – Wiebe, J. (2014): Simcompass: Using Deep Learning Word Embeddings to Assess Cross-level Similarity. In: *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. Dublin: ACL, 560–565. <https://pdfs.semanticscholar.org/4b7b/10ffe383addfc134fb5b10000d085ffd9709.pdf>
- Bengio, Y. – Ducharme, R. – Vincent, P. – Jauvin, C. (2003): A Neural Probabilistic Language Model. *Journal of Machine Learning Research*, 3, 1137–1155. <http://www.jmlr.org/papers/volume3/bengio03a/bengio03a.pdf>
- Firth, J. R. (1957): A Synopsis of Linguistic Theory, 1930–1955. *Studies in Linguistic Analysis*, 1–32. <http://annabellelukin.edublogs.org/files/2013/08/Firth-JR-1962-A-Synopsis-of-Linguistic-Theory-wfih5.pdf>
- Iacobacci, I. – Pilehvar, M. T. – Navigli, R. (2015): Senseembed: Learning Sense Embeddings for Word and Relational Similarity. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Beijing: ACL, 95–105. <http://www.aclweb.org/anthology/P15-1010>
- Mikolov, T. – Chen, K. – Corrado, G. – Dean, J. (2013a): Efficient Estimation of Word Representations in Vector Space. *CoRR*, abs/1301.3781, <https://arxiv.org/pdf/1301.3781.pdf>
- Mikolov, T. – Yih, W. – Zweig, G. (2013b): Linguistic Regularities in Continuous Space Word Representations. In: *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Atlanta: ACL, 746–751. <https://www.aclweb.org/anthology/N13-1090>
- Siklósi B. (2018): Using Embedding Models for Lexical Categorization in Morphologically Rich Languages. In: Gelbukh, A. (ed.): *Computational Linguistics and Intelligent Text Processing: 17th International Conference CICLing 2016*, Springer, Cham, 115–126. https://link.springer.com/chapter/10.1007/978-3-319-75477-2_7
- Siklósi B. – Novák A. (2016): Beágyazási modellek alkalmazása lexikai kategorizációs feladatokra. In: *A XII. Magyar Számítógépes Nyelvészeti Konferencia*, Szeged: SZTE, 3–14.
- Trask, A. – Michalak, P. – Liu, J. (2015): sense2vec - A Fast and Accurate Method for Word Sense Disambiguation in Neural Word Embeddings. *CoRR* abs/1511.06388, https://www.researchgate.net/publication/284476537_sense2vec_-_A_Fast_and_Accurate_Method_for_Word_Sense_Disambiguation_In_Neural_Word_Embeddings

HIBRID NYELVTECHNOLÓGIÁK

HYBRID HUMAN LANGUAGE TECHNOLOGIES

Ács Judit¹, Borbély Gábor², Makrai Márton³, Nemeskey Dávid⁴, Recski Gábor⁵, Kornai András⁶

¹tanársegéd, Budapesti Műszaki és Gazdaságtudományi Egyetem Automatizálási és Alkalmazott Informatikai Tanszék

²tanársegéd, Budapesti Műszaki és Gazdaságtudományi Egyetem Algebra Tanszék

³tudományos segédmunkatárs, MTA Nyelvtudományi Intézet

⁴informatikus mérnök, MTA SZTAKI Nyelvtechnológiai Kutatócsoport

⁵PhD, tanársegéd, Budapesti Műszaki és Gazdaságtudományi Egyetem Automatizálási és Alkalmazott Informatikai Tanszék

⁶az MTA doktora, az MTA SZTAKI Nyelvtechnológiai Kutatócsoport vezetője, tudományos tanácsadó, a Budapesti Műszaki és Gazdaságtudományi Egyetem Algebra Tanszék professzora, kornai@sztaki.hu

ÖSSZEFOGLALÁS

Az elmúlt harminc év nyelvészetét a „racionalista” (szabályalapú, szimbólumkezelő) és az „empirista” (statisztikai alapú, gépi tanulás) nyelvészeti modellek harca jellemezte. Míg a nyolcvanas években még egyértelműen a racionalista paradigma volt az uralkodó, mára ez, különösen az utolsó néhány év mélytanulós forradalmának köszönhetően megfordult, és egyértelműen az empirista paradigma lett a domináns. Az MTA SZTAKI nyelvtechnológiai csoportja elsősorban a hibridizáció kérdéseivel foglalkozik, azzal, hogy miképp találhatjuk meg a diszkrét, szimbolikus struktúrát a folytonos, zajos adatokban, illetve hogyan tudjuk a struktúráról való ismereteinket hatékonyabb algoritmusok építésében kamatoztatni.

ABSTRACT

In the last thirty years linguistics was characterized by a debate between the “rationalist” (rule-based, symbol-manipulating) and the “empiricist” (statistics-based, machine learning) camps. Back in the 1980s clearly the rationalist paradigm had the upper hand, but by now the situation is reversed, and thanks to the deep learning revolution of the past few years, today the empiricist paradigm dominates. The human language technology group at MTA SZTAKI focuses on issues of hybridization, in particular on finding the discrete symbolic structure in the continuous (and noisy) data, and on leveraging our knowledge of structure in building more efficient algorithms.

Kulcsszavak: nyelvtechnológia, gépi tanulás, mélytanulás, hibrid rendszerek

Keywords: human language technology, machine learning, deep learning, hybrid systems

BEVEZETÉS

Az MTA SZTAKI Nyelvtechnológiai (Human Language Technology, HLT) Kutatócsoportjának előzményei az origo.hu (Origo) és a Northern Light Technologies (NLT) közti együttműködés időszakára nyúlnak vissza. Ma az Origo csupán egy a számtalan webes portál közül, de 2002-ben, amikor az együttműködés az addig használt AltaVista (AV) keresőtechnológia tarthatatlansága miatt szükségessé vált, az Origo még úgy uralkodott a magyar weben, mint a 19. században Britannia a habok felett: látogatottsága nagyobb volt, mint az őt követő két legnagyobb portálé együttvéve. Az NLT, melynek akkoriban Kornai András volt a tudományos vezetője, 1999-ben nőtt nagyobbra, mint az AltaVista (Yahoo), és kettejük versenyében (melyet végül a nevető harmadik, a Google nyert meg) már tetten érhető volt az a szemléletbeli különbség a racionalista és az empirista megközelítések közt, amelyet pár évvel korábban már igen markánsan jelzett Judith L. Klavans és Philip Resnik (1996).

Míg az AV (web yahoo-*knak* nevezett) szerkesztők százait foglalkoztatta, akik szabályalapon kézzel sorolták be a weblapokat eleinte néhány tucat, később több ezer, hierarchikusan elrendezett tartalmi kategóriába, addig az NLT statisztikai módszerekkel alakította ki az egyes kategóriák modelljeit, és mivel a besorolás teljesen automatikus volt, nem volt szükség a szerkesztői gárdának a web robbanásszerű növekedését követő bővítésére (mely végső soron a Yahoo/AV vesztét is okozta). A magyar tematikus hierarchia úttörője Ungváry Rudolf (Országos Széchenyi Könyvtár) volt, az Origóban használt rendszert az ő munkáját továbbfejlesztve dolgozta ki Kárpáti András és Halácsy Péter (ma a Pécsi Tudományegyetem, illetve a Prezi, akkoriban az Axelero, a mai T-Online munkatársai). Az NLT az általuk készített katalógus mint tanulóadat alapján építette fel a saját modelljeit gépi tanulással [Kornai, 2003 EACL]¹. Mint ismeretes, a gépi tanulást (machine learning) máig a címkézett adatokon alapuló ún. felügyelt tanulás (supervised learning) dominálja. A nyers adatokon alapuló felügyeletlen (unsupervised) tanulás nagy erővel kutatott terület, ahol komoly eredményekről csak az utóbbi tíz évben beszélhetünk (Erhan et al., 2010), és az igazi áttörés, a felügyeletlen struktúratanulás, még várat magára.

Ebben a cikkben a kutatásoknak a Műegyetemen otthont adó Média Oktató és Kutató Központtal (MOKK) nem tudunk annak jelentőségéhez mérten foglalkozni, bár kétségkívül ez volt a számítógépes társadalomtudomány első multidiszciplináris műhelye hazánkban, ahol a számítógépes nyelvészet csupán egy volt a digitális szerzői joggal, kulturális termeléssel, a digitális térrel és annak

¹ A munkacsoport azon cikkei, melyek hivatkozásai [...] közt szerepelnek, elérhetőek a HLT honlapján: <https://hlt.bme.hu/en/publications>, a (...) közti hivatkozásokat lásd a cikk végén lévő irodalomjegyzékben.

szociológiájával, a formális cselekvéssel, az új médiával, *peer to peer* hálózatokkal stb. foglalkozó kutatások közül. Reméljük, hogy a nemrég a Preziben Babarczy Eszter, Bodó Balázs, Csígy Péter, György Péter, Halácsy Péter, Kacsuk Zoltán, Szakadát István, Varga Dániel és Vályi Gábor részvételével megrendezett MOKKtóber találkozó anyagai megeremtik az alapot e műhely történetének és máig érezhető hatásának alaposabb feltárásához.

Az akkori nyelvtechnológiai munkák közül megemlíjtjük az első magyar szabadon letölthető korpuszt (WebKorpusz), az első párhuzamos magyar–angol korpuszt (hunglish.hu), és a Hun* eszközláncot, melyek az első nyílt forráskódú (open source) magyar nyelvi szoftverek közt voltak. Ebbe az eszközláncba épült be az eredetileg Németh László által külön fejlesztett HunSpell helyesírás-ellenőrző is, mely azóta is a szabad világ vezető helyesírás-ellenőrzője (ugyanaz a szoftver-keret több mint száz nyelvhez lett adattal feltöltve, és ma megtalálható a Thunderbird, FireFox, LibreOffice sok millió példányában); a Simon Eszter által épített HunNER névelem-felismerő [Simon, 2013; Nemeskey–Simon, 2012], és még sok más eszköz, melyekről az alábbiakban részletesen lesz szó. A mokk.bme.hu és a nyelvtechnológiai vonalon ezt továbbvivő hlt.bme.hu máig a nyílt forráskódú nyelvtechnológia egyik vezető képviselője, azzal a fontos különbséggel, hogy az elmúlt másfél-két évtizedben megfordult a széljárás, és az egykor ignorált, majd kinevetett, majd ellenségnek tekintett nyílt forráskódú megközelítés mára uralkodóvá vált.

KÉTFÉLE SZEMLÉLET

Tudományszociológiai szemszögből a racionalista és az empirista kutatási modellek közti különbség lényege a felülről vezérelt (top down) és az alulról kiinduló (bottom up) keresési stratégia. Előbbi klasszikus példája a Manhattan Project, amely a fizikusok elismert vezetőjének, Albert Einsteinnek az elnökhöz intézett levele alapján indult be: legfelül pár tucat elméleti fizikus, alattuk több száz mérnök és kísérleti fizikus, akik alatt munkások ezrei dolgoztak. A nyelvészetnek is megvolt a maga elismert vezetője, Noam Chomsky, aki nagyon is határozott irányú kutatásokat kezdeményezett. Annak az egyszerű, de előtte kevésbé hangsúlyozott ténynek az alapján, hogy a kisgyermekek viszonylag gyorsan, néhány év alatt lényegében tökéletesen megtanulják anyanyelvüket (és bármely nyelvi környezetbe helyezük a csecsemőt, az ottani nyelvet képes ilyen szinten megtanulni), arra a következtetésre jutott, hogy ennek a tanulási képességnek kizárólag az lehet a magyarázata, hogy a gyermek fejében a tudásanyag egy nagy része, az univerzális grammatika, már örökletesen ott van.

Bár kezdettől voltak ennek az elméletnek komoly ellenzői, például Jean Piaget (Chomskyval való vitájának hiteles összefoglalóját adja Piattelli-Palmarini

et al., 1980), nyugodtan elmondhatjuk, hogy a fentebb idézett nagy hatású publikációktól kezdve a modern nyelvészeti kutatások fővonalát a 20. században Chomsky jelölte ki [Kornai, 2010 HRP], és nem kevesek számára máig az ő felfogása szolgál irányítúként. De a *Zeitgeist* megváltozott, a bölcs vezetők kora lejárt, és ami a legfontosabb: a predikciók nehezen megfoghatóan bizonyultak, specifikus nyelvtani struktúrákat/géneket nem sikerült azonosítani a szótan és mondattan területén. A kudarc annál is fájóbb volt, mert a hangtanban frappáns csecsemőkísérletek sora (összefoglalásukat lásd Werker–Tees, 1984) nyilvánvalóvá tette, hogy Chomskynak igaza van: az egyes nyelvek hangtanának kisgyermekkorú elsajátítása nem magyarázható univerzális fonetika tételével.

Ez a megváltozott *Zeitgeist* tette lehetővé, hogy a természetlennek bizonyult elméleti megfontolásokból nagyrészt kiábrándult nyelvészek egyre komolyabban vegyék a lentről, a kutatás lövészárkaiból érkező empirikus anyagot. Egyre nagyobb és nagyobb egy- és többnyelvű korpuszt lehetett számítógépes elemzés alá vetni. A bevezetőben már érintettük azokat a korpuszfejlesztési munkálatokat, melyeket a HLT-csoport végzett. Ezek jelentősége nem pusztán abban áll, hogy az addigi nagyon komoly és szakmailag jól megalapozott korpuszokat, mint például a Magyar Nemzeti Szövegtár akkori változata (Váradi, 2002) vagy az elemzett (és ezért természetesen jóval kisebb) Szeged Korpusz (Vincze et al., 2014) nyíltabbá, jobban elérhetővé tette (ezt inkább a megváltozott *Zeitgeist*nek, mint a Webkorpusz és a Hunglish megjelenésének tudjuk be), hanem abban, hogy elődeiknél lényegesen nagyobbak voltak.

A modern számítógépes elemzés legfontosabb alapanyagát a milliárdszavas (gigaword) korpuszok adják. Azok az elemzési technikák, melyek ma a kutatást uralják, kisebb anyagokon egyszerűen nem működnek jól. A legfontosabb elméleti újítás, mely az utóbbi öt-tíz évben áttörést hozott számos olyan területen, mint a képek és nyelvi leírásuk (caption) közti szemantikai kapcsolat gépi tanulása (Karpathy et al., 2014), a szóvektorok (embedding) bevezetése volt. Minden szóhoz egy véges (általában pár száz) dimenziós vektort rendelünk úgy, hogy a hasonló kontextusokban szereplő szavak vektorai egymáshoz hasonlóak (euklideszi térben közeli) legyenek. Az első áttörést Ronan Collobert és szerzőtársai (2011) hozták el, akik egyszerre, ugyanazon vektorok felhasználásával, tudtak javítani több olyan klasszikus feladat addigi legjobb eredményén, mint a szófaj szerinti címkézés (part of speech tagging), a névelem-felismerés (named entity recognition), a sekély mondattani elemzés (tehát a mondatok pszichológiailag releváns darabokra, például főnévi csoportokra bontása [chunking]) és a szemantikai szerep felismerése (semantic role labeling). A kulcsmomentum itt az, hogy Colloberték nem egy új feladatot oldottak meg az új reprezentációval, hanem már régről ismert, nehéz, kutatók százai által vizsgált feladatokra (melyek többségével csoportunk is foglalkozott, például a HunTag szekvenciális címkéző [Halácsy

et al., 2006 LREC] vagy a sekély mondattani elemzés, mely máig aktív témánk [Recski, 2014]) értek el az eddigieknél jobb eredményeket.

A szemantika területén, ahol régen, évtizedekig előre hatóan a vezető kutatók, Chomsky és Richard Montague jelölték ki a kutatás fő irányát, ma a kutatók többsége egy olyan jelenséggel foglalkozik, amelyet egy brünni műegyetemista, Tomas Mikolov fedezett fel: a szövektorok lineáris struktúráját mutatnak, például $v(\textit{king}) - v(\textit{man}) + v(\textit{woman}) \approx v(\textit{queen})$ (Mikolov et al., 2013b). Csoportunk a vektoros szófordítás (lineáris fordítás, lásd Mikolov et al., 2013a) módszerét alkalmazta közép-európai nyelvekre [Makrai et al., 2013], olyan ritkábban vizsgált lexikai relációk felé általánosítottuk az analógia vektoralgebrai megfogalmazását, mint a jó-rossz (peace-war, pleasure-pain) vagy a fönt-lent (tall-short, rise-fall), Makrai Márton [2014 MSZNY] pedig oksági párok (például sérül-fáj) geometriáját elemezte. Új módszereket vezettünk be többjelentésű beágyazások (multi-sense embeddings) szemantikai felbontóképességének mérésére [Borbély, 2016 RepEval]. Ezekben a reprezentációkban egy-egy szóalakhoz több vektor is tartozhat, melyek elvileg a szó különböző jelentéseinek felelnek meg. A gyakorlatban azonban a jelentésvektorok között nem mindig figyelhető meg fogalmi különbség, egy-egy általánosabb vektor több jelentést is lefed, és fölösleges vektorok is lehetnek, melyek a modellnek egy alkalmazásban való hasznosságát ronthatják.

Utólag természetesen megtalálható a szövektorok használatának elméleti megalapozása: a kontextus nyilvánvalóan fontos, és a gondolat, hogy egy szó jelentését a használati kontextuson keresztül érdemes megragadni, kétségkívül jelen van már a nagy brit strukturalista, John Rupert Firth munkáiban is, aki azt írta, „a word is characterized by the company it keeps” (a szavakat a társaságuk jellemzi). Ugyanakkor világosan kell látni, hogy Firth (akinek a prozódiaira vonatkozó felfogása is újra életre kelt a modern fonológiában, lásd Goldsmith, 1990) éppen ahhoz az iskolához tartozik, amely ellen Chomsky egész életében harcolt. A nagy tömegű adat viszont minden területen a strukturalistákat, nem pedig az elsősorban szellemes anekdotikus példákra és nyelvi intuícióra alapozott chomskyánus megközelítést látszik igazolni.

HIBRID MODELLEK

A fentiek után talán meglepően hangzik, de korunkban az egész nyelvészet Chomsky programját követi két alapvető tekintetben is. Az egyik a már Chomsky (1965) által középpontba állított magyarázó adekvátság (explanatory adequacy) elve, mely szerint a nyelvelmélet nem állhat meg a tények leírásánál, hanem arra is magyarázatot kell adnia, hogy a kisgyermek hogyan sajátítja el a nyelvet, a másik az univerzálék (minden nyelvre egyaránt igaz állítások) keresése, melynek

Joseph Greenberg (1963) után szintén Chomsky fentebb vázolt programja adott új lendületet.

A legfontosabb különbség nem a generatív felfogásban, hanem az univerzális metaelméletet konkrétan realizáló nyelvtanok technikai apparátusában van. A szintaxis területén ez azt jelenti, hogy a környezetfüggetlen mélyszerkezeten és az ezt mozgató faátalakításokon alapuló transzformációs grammatika helyét átvette egy másik, szintén a strukturalista korszakból átvett formalizmus, a függőségi grammatika (Tesnière, 1959). Ebben az elméleti keretben ma már ötven nyelvhez találunk komoly, elemzett fabank (treebank) korpuszokat, jelenleg hetvenet, de számuk egyre nő (URL2). Ezek egységesített (univerzális) szófaj- és függőségtipológián alapulnak, és ezzel nagyban elősegítik a minden emberi nyelvre kiterjedő univerzálékutatást. Az empirikus alapok kiterjesztésére mindig is megvolt a szándék: már Greenberg is harminc nyelvvel dolgozott, de nyersanyagául nyelvtani leírások, nem pedig a direkt empirikus adatok szolgáltak. Tekintve, hogy mintegy hat-hétezer emberi nyelvről tudunk (bár ezekből gigaword korpuszra és fabankra a digitális nyelvhalál miatt legfeljebb háromszáznál számíthatunk [lásd Kornai, 2013 PLoS]), az univerzális grammatika kutatása még sok évtizedre fog programot adni a nyelvészetnek.

Az új technikai apparátusra való áttérés egyébként a fonológiában is végbe ment, ahol a környezetfüggő, szekvenciális szabályrendszereket egy véges automatákkal megfogalmazható elmélet, az optimalitás elmélete váltotta fel (Prince–Smolensky, 1993; Karttunen, 1998). A technika megváltozása jelentős átalakulást hozott a szemantikában is, ahol a logikai formán (első- vagy magasabb rendű predikátumkalkuluson) alapuló reprezentációkat egy egyszerűbb, a függőségi fákkal egyenértékű függvényargumentum-szerkezet váltotta fel. Ezt tekinthetjük az ún. generatív szemantikához (Huck–Goldsmith, 1995) való visszatérésnek, de valójában sokkal régebbre, egészen az első formalizált nyelvtanig, Pāṇini *Aṣṭādhyāyī*-jéig (i. e. 450 körül) megy vissza.

Ebben a szellemben dolgoztuk ki 2009 és 2012 között a 4lang formalizmust [Kornai, 2010 MoL, 2012 LNCS; Kornai–Kracht, 2015; Kornai megj. alatt], mely a természetes nyelvi jelentést fogalmak irányított gráfjaként reprezentálja. Megalkottuk a text_to 4lang szoftvert [Recski, 2016 LREC], mely nyers angol és magyar szövegekhez automatikusan rendel ilyen reprezentációkat; ezeket sikerrel alkalmaztuk lexikális ontológiák építéséhez [Recski, 2016 LREC], és a fentebb Mikolov kapcsán már említett analógiás feladatok megoldásában [Recski et al., 2016 RepLearn]. Megemlítjük néhány a 4lang jelentés-reprezentációs rendszerhez [Kornai et al., 4th JCLCS] kapcsolódó kutatásunkat: az igei szerepek vizsgálata [Makrai, 2014 MSZNY], a definiáló szókincs [Kornai et al., 2015 MOL] és az aktivációterjedés [Nemeskey et al., 2013] kapcsán.

A magyarázó adekvátság tekintetében is ugyanez a folyamat játszódott le: az eszme győzedelmeskedett, de a technikai apparátus gyökeresen szembe megy a

Chomsky és Lasnik (1993) által javasolttal. Kicsi, néhány tucat diszkrét (bináris) paraméter beállításán alapuló döntési fák helyett nagy, sok százezer (gyakran sok millió) folytonos paraméter gradiens-módszerrel való tanulása vált uralkodóvá. Az ilyen sokparaméteres rendszerek tanulása a beszéd- és írásfelismerés terén indult be az ún. Rejtett Markov Modellek (Hidden Markov Model, HMM) felhasználásával: itt kapott először fontos szerepet a valószínűségi nyelvmodellezés (language modeling). Csoportunk mind a hagyományos (szó-n-eseken alapuló, n-gram), mind a mélytanulásban elterjedt rekurrens neurális háló alapú modelleket kutatja. Foglalkozunk a terület mind általános, mind a magyar nyelvre specifikus problémáival is [Nemeskey, 2017 MSZNY]. A természetes nyelvi mondatok hosszára valószínűségi, generatív modellt alkottunk, ami magyarázni tudja a mondatok empirikusan mérhető hosszeloszlását.

A magyar nyelv agglutinatív voltából fakadóan a szavak sok felszíni formában lehetnek jelen, ami az angol nyelvben jól működő szóalapú módszereknek komoly kihívást jelent. Vizsgálataink egyik fókuszja annak megállapítása, hogy morfológiai eszközök mennyiben tudják ezt a problémát enyhíteni. OTKA-pályázat keretében vizsgáljuk a szavak és morféimák (legkisebb önálló jelentéssel rendelkező nyelvi egységek, például tárgyrag) neurális hálózatokkal történő azonosítását. A morfológiai elemzés számos nyelvtechnológiai feladat elengedhetetlen része, amit hagyományosan nyelvészek hosszas munkájával összeállított szabályok segítségével végeznek, azonban ezek a szabályok csak a világ nyelveinek töredékéhez állnak rendelkezésre. Kutatásunk célja olyan módszerek kidolgozása, amelyek pusztán nyers szövegből képesek ezeket a szabályokat felismerni. Bár ez a rendszer még nincs kész, előmunkálatai közül említést érdemelnek az automatikus szótárépítéssel [Ács et al., 2013, 2014] és ékezet-visszaállítással [Ács–Halmi, 2016] foglalkozó rendszereink.

Foglalkozunk a szövektorok általánosításaival mátrix- és projektívter-modellekre. A szokásos szó-vektoralapú beágyazások szisztematikus hibája (Pennington et al., 2014), hogy antonima-párok hasonló vektorokkal reprezentálódnak, például *good* \approx *bad*. Ennek egy megoldását kínálja a projektív tér, ahol egy gömbön az antipodális pontok azonosítva vannak. Egy erre épülő célfüggvénnyel sikerült javítanunk a vektorbeágyazások által elért eredményt a Simlex999-adaton (Hill et al., 2014). A mátrixbeágyazások esetében egy szóhoz nem egy vektort, hanem egy mátrixot rendelünk. Ezzel egy nem-kommutatív általánosítását adjuk a szó-vektoroknak, melyek alkalmasak nyelvmodellezésre és speciális véges automaták tanítására is. A hibrid modellek diszkrét komponensei a mátrixmodellek, illetve az ezekkel szoros formai kapcsolatban álló véges automaták, melyek tanítása súlyozott nyelveken [Kornai et al., 2013] a szimbolikus és a probablisztikus modellezésnek az eddigieknél mélyebb hibridizációját készíti elő.

ÖSSZEFOGLALÁS

A racionalista és az empirista megközelítések nem kizárják, hanem támogatják egymást. A modern gépi tanulás alapvető sikerkritériumai messze túlmennek a leíró adekvátságon (descriptive adequacy). A terület egyik legsikeresebb kutatócsoportja, Yoshua Bengio, Aaron Courville és Pascal Vincent (2013) külön kiemeli, hogy „In good high-level representations, the factors are related to each other through simple, typically linear dependencies” (a jól működő magas szintű reprezentációkban a tényezők egyszerű, tipikusan lineáris kapcsolatban állnak). Ez alól, úgy tűnik, a nyelvtan sem kivétel: a sikeres modellek mögött egyszerű lineáris nyelvtanokat (véges automatákat, véges transzducereket), illetve ezek olyan egyszerű általánosításait találjuk, mint a rejtett Markov-modellek vagy az Eilenberg-gépek (Eilenberg, 1974). A jövő útja, úgy véljük, az ilyenek automatikus tanulása, és ehhez, úgy tűnik, nincs semmilyen speciális, az ember általános kognitív képességein túlmutató eszközre szükség.

Azt gondoljuk, hogy a nyelvészeti vizsgálatok a számítógépes társadalomtudományok más területei számára is szolgálhatnak ilyen általános tanulságokkal, hiszen ezekben is az egyik fő cél a mögöttes struktúra feltárása, és ezekben is egyre inkább elérhetővé válik az a hatalmas tömegű adat, amelynek alapján e struktúra algoritmikus módszerekkel megragadható.

IRODALOM

- Bengio, Y. – Courville, A. – Vincent, P. (2013): Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35, 8, 1798–1828. <https://arxiv.org/pdf/1206.5538.pdf>
- Chomsky, N. (1965): *Aspects of the Theory of Syntax*. MIT Press, <https://faculty.georgetown.edu/irvinem/theory/Chomsky-Aspects-excerpt.pdf>
- Chomsky, N. – Lasnik, H. (1993): Principles and Parameters Theory. *Syntax: An International Handbook of Contemporary Research*. (ed. Jacobs, J.) 1. Berlin: de Gruyter, 505–569.
- Collobert, R. et al. (2011): Natural Language Processing (Almost) from Scratch. *Journal of Machine Learning Research (JMLR)*, 12, 2493–2537. <http://www.jmlr.org/papers/volume12/collobert11a/collobert11a.pdf>
- Eilenberg, S. (1974): *Automata, Languages, and Machines*. Orlando, FL: Academic Press
- Erhan, D. et al. (2010): Why Does Unsupervised Pre-training Help Deep Learning? *Journal of Machine Learning Research*, 11, 625–660. <http://www.jmlr.org/papers/volume11/erhan10a/erhan10a.pdf>
- Goldsmith, J. A. (1990): *Autosegmental and Metrical Phonology*. Cambridge, MA: Blackwell
- Greenberg, Joseph H. (1963): Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. *Universals of Human Language*. (ed. Greenberg, J. H.) MIT Press, 73–113. <http://pkdas.in/JNU/typo/lu.pdf>
- Hill, F. – Reichart, R. – Korhonen, A. (2014): Simlex-999: Evaluating Semantic Models with (Genuine) Similarity Estimation. *Computational Linguistics*, 41, 4, 665–695. <https://arxiv.org/pdf/1408.3456.pdf>

- Huck, G. J. – Goldsmith, J. A. (1995): *Ideology and Linguistics Theory: Noam Chomsky and the Deep Structure Debates*. London: Routledge
- Karpathy, A. – Armand, J. – Fei Fei, L. (2014): Deep Fragment Embeddings for Bidirectional Image Sentence Mapping. *Advances in Neural Information Processing Systems*, 27. (ed. Ghahramani, Z. et al.) Curran Associates, Inc., 1889–1897. <https://cs.stanford.edu/people/karpathy/nips2014.pdf>
- Karttunen, L. (1998): The Proper Treatment of Optimality in Computational Phonology: Plenary Talk. *Proceedings of the International Workshop on Finite State Methods in Natural Language Processing*. Association for Computational Linguistics, 1–12. <https://arxiv.org/pdf/cmp-1g/9804002.pdf>
- Klavans, J. L. – Resnik, P. (eds.) (1996): *The Balancing Act – Combining Symbolic and Statistical Approaches to Language*. MIT Press
- Mikolov, T. – Le, Q. V. – Sutskever, I. (2013a): Exploiting Similarities among Languages for Machine Translation. arXiv:1309.4168. <https://arxiv.org/pdf/1309.4168.pdf>
- Mikolov, T. – Yih, W. – Zweig, G. (2013b): Linguistic Regularities in Continuous Space Word Representations. *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2013)*. Atlanta, Georgia: Association for Computational Linguistics, 746–751. <https://www.aclweb.org/anthology/N13-1090>
- Pennington, J. – Socher, R. – Manning, C. (2014): GloVe: Global Vectors for Word Representation. *Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*. 1532–1543. <https://www.aclweb.org/anthology/D14-1162>
- Piattelli-Palmarini, M. – Piaget, J. – Chomsky, N. (1980): *Language and Learning: The Debate between Jean Piaget and Noam Chomsky*. Routledge
- Prince, A. S. – Smolensky, P. (1993): *Optimality Theory: Constraint Interaction in Generative Grammar*. (Rutgers University Center for Cognitive Science Technical Report 2.) Piscataway, NJ: Rutgers University DOI:10.1002/9780470759400
- Tesnière, L. (1959): *Éléments de syntaxe structurale*. Paris: Klincksieck, <https://archive.org/details/LucienTesniereElementsDeSyntaxeStructurale>
- Váradi T. (2002): The Hungarian National Corpus. *Proceedings of the Third International Conference on Language Resources and Evaluation*, 385–389. https://www.researchgate.net/publication/228608174_The_Hungarian_National_Corpus
- Vincze V. et al. (2014): Szeged Corpus 2.5: Morphological Modifications in a Manually POS-tagged Hungarian Corpus. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. (eds). Nicoletta Calzolari (Conference Chair) et al. Reykjavik, Iceland: European Language Resources Association (ELRA), 1074–1078. <http://publicatio.bibl.u-szeged.hu/4736/1/szk.pdf>
- Werker, J. F. – Tees, R. C. (1984): Cross-language Speech Perception: Evidence for Perceptual Reorganization during the First Year of Life. *Infant Behavior and Development*, 7, 49–63. DOI: 10.1016/S0163-6383(84)80022-3, <https://bit.ly/2IJwyt8>

URL1: hlt.bme.hu

URL2: universaldependencies.org

A PLETYKA A TÁRSAS REND SZOLGÁLATÁBAN – AZ INFORMÁLIS KOMMUNIKÁCIÓ STRUKTÚRÁJÁNAK MÉLYEBB MEGÉR- TÉSÉÉRT A *COMPUTATIONAL SOCIAL SCIENCE* ESZKÖZEIVEL¹

GOSSIP IN SERVICE FOR SOCIAL ORDER – USING THE TOOLS OF COMPUTATIONAL SOCIAL SCIENCE FOR A DEEPER UNDERSTANDING OF THE STRUCTURE OF INFORMAL COMMUNICATION

Galántai Júlia¹, Pápay Boróka², Kubik Bálint György³, Szabó Martina Katalin⁴, Takács Károly⁵

¹tudományos segédmunkatárs, MTA Társadalomtudományi Kutatóközpont „Lendület” RECENS Kutatócsoport

²tudományos segédmunkatárs, MTA Társadalomtudományi Kutatóközpont „Lendület” RECENS Kutatócsoport

³kutató, MTA Társadalomtudományi Kutatóközpont „Lendület” RECENS Kutatócsoport

⁴tudományos segédmunkatárs, MTA Társadalomtudományi Kutatóközpont „Lendület” RECENS Kutatócsoport,
egyetemi tanársegéd, Szegedi Tudományegyetem Bölcsészettudományi Kar Orosz Filológiai Tanszék

⁵PhD, az MTA Társadalomtudományi Kutatóközpont „Lendület” RECENS Kutatócsoport vezetője

takacs.karoly@tk.mta.hu

ÖSSZEFOGLALÓ

A pletykálás, amelynek során másokról értékelő módon a hátuk mögött beszélünk, az emberi kommunikáció jelentős hányadát teszi ki. Miközben a mindennapi életben pejoratív értelemben beszélünk róla, a társadalomtudományi kutatások arra világítanak rá, hogy a pletykának jelentős szerepe van a társas rend biztosításában, a normák fenntartásában és a kooperáció biztosításában a különböző társas kontextusokban. Miközben ezeket a társadalomtudományi alapkérdéseket eddig főleg kvalitatív, önkitöltős kérdőíves vagy absztrakt kísérleti körülmények között vizsgálták, cikkünkben azt kívánjuk bemutatni, hogy miként gazdagíthatóak ezek a hagyományos elemzési módszerek a *computational social science* eszköztárának felhasználásával. Leírjuk azt a kvantitatív stratégiát, amellyel elemezzük a pletykát és alaptulajdonságait egy élő, spontán beszédet tartalmazó korpuszban. Bemutatjuk azokat a morfológiai, téma szerinti, szóhasználat szerinti, kontextuális és nem szövegalapú (például időzítés, beszélőváltás, nevetés és más hangok) elemzési lehetőségeket, amelyek a pletyka természetének és funkcióinak mélyebb társadalomtudományi megértését segítik.

¹ A kutatást az Európai Kutatási Tanács (European Research Council), az Európai Unió Horizont 2020 kutatási és innovációs programjának keretében (ERC_CoG_2014_648693 sz. szerződésben) támogatja, a kutatás vezetője Takács Károly.

ABSTRACT

Gossip, as an evaluative talk about others who are not present, constitutes a major portion of human communication. Meanwhile we talk about gossip in a pejorative way in everyday language, research in social sciences highlights that gossip might play an important role in the establishment of social order, and in the maintenance of social norms and cooperation in various social contexts. These fundamental questions have been analyzed in the social sciences almost exclusively by qualitative methods, self-reports, and in abstract laboratory experiments. In this paper, we would like to illustrate, how these traditional research methods could be supplemented with the application of the tools of computational social science. We describe the quantitative strategy for the analysis of gossip and its characteristics in a large corpus of spontaneous conversations. We portray the morphological, topic-based, vocabulary-based, contextual, and non-textual (e.g., timing, turn-taking, laughing, and other sounds) opportunities of analysis that could improve the scientific understanding about the nature and functions of gossip.

Kulcsszavak: pletyka, társas normák, spontán beszéd, szövegtörzset, törzsetépítés, kvantitatív szövegelemzés

Keywords: gossip, social norms, spontaneous speech, text corpus, corpus building, quantitative text analysis

BEVEZETÉS

Az emberi informális kommunikáció jelentős részét, egyesek szerint kétharmadát, jelen nem lévő, más személyekről folytatott értékelő tartalmú beszélgetés teszi ki (Dunbar, 1996, 2004; Foster, 2004). Az ilyen beszélgetéseket, amelyekben legalább egy értékelő és egy hallgató vesz részt, tekintjük pletykának (Kurland–Pelled, 2000; Ellwardt, 2011). Jó pletykát hallani és pletykálni mindenki szeret, mégis magához a pletykához a köznapi értelemben pejoratív konnotációkat fűzünk, elítéljük azt. Miért létezik akkor, és miért olyan elterjedt a pletyka? Miért használunk ki szinte minden alkalmat mások hátuk mögött történő kibeszélésére?

Ezeknek a kérdéseknek a megválaszolásához elengedhetetlen elsőként annak igazolása, hogy tényleg ilyen léptéket ölt-e az emberi társas kommunikációban a pletyka. Az erre vonatkozó ismeretanyag viszonylag szerény, és kvalitatív tapasztalatokra hagyatkozik. Nagyon ritka a spontán informális beszélgetéseket tartalmazó törzset, amely ennek becslésére objektív lehetőséget adna. A magyar nyelvre vonatkozóan ilyen törzset korábban nem létesült.

A pletyka gyakoriságának leírásán túlmutatóan szükség van arra, hogy megértsük, miért pletykálunk másokról, és miért van szükség ilyen mértékben mások értékelésére. Kutatási projektünk kiinduló hipotézise szerint azért, mert a pletykának pozitív közösségi funkciója van. A pletyka egy olcsó eszköz, amely

biztosítja a közösség szereplőinek a reputációs kontrollját, értesít azok esetleges normaszegéseiről, és így hozzájárul a társas normák fenntartásához, a közösségi rendhez, és elősegíti az együttműködést.

Mindezeket a tartalmakat a beszélt nyelvben a pletyka beazonosítása után lehet vizsgálni. Változtatja-e a hallgató a pletyka tárgyról alkotott véleményét? Például: tesz-e olyan előjelű értékeléseket az adott személyről, amelyek egybevágnak a pletyka előjelével? Tartalmaz-e a pletyka utalást konkrét normaszegésre, kapcsolja-e a hallgató a pletykát az uralkodó közösségi normákhoz? Óvatosabban viszonyul-e a későbbiekben a hallgató a negatív pletyka tárgyát jelentő személyhez? Kerüli-e esetleg az együttműködést vele?

Ezen kérdések megválaszolását azonban számos lépés előzi meg, tele elméleti és gyakorlati buktatókkal. Már a pletyka azonosítása sem egyszerű feladat, a legtöbb felvetett kérdés vizsgálatához pedig az adott közösség társas viszonyainak hosszabb távú és emellett mély megfigyelésére van szükség. A jelen tanulmányban bemutatjuk, hogy miként lehet egy mélyebb megfigyelésből származó, nagy mennyiségű, élő spontán beszédet tartalmazó szövegtörzset a computational social science eszköztárával felhasználni ezeknek a kérdéseknek a vizsgálatához. Elsőként érvelünk amellett, hogy a mindennapi nyelv elemzésének és a korpuszok használatának óriási jelentősége lehet a társadalomtudományban alapvető kérdések vizsgálatakor. Ezt követően bemutatjuk a HuTongue korpusz létrehozásának legfontosabb lépéseit. Felvázoljuk a készülő adatbázis tervezett szerkezetét, a főbb változókat, majd az elemzési irányokat. Végül a továbblépési lehetőségek között tárgyaljuk azokat a további társadalomtudományi kérdéseket, amelyek vizsgálata megvalósulhat a HuTongue korpusz létrejöttével.

A MINDENNAPI NYELV ÉS A KORPUSZOK JELENTŐSÉGE A TÁRSADALOMTUDOMÁNYBAN

A különböző, így többek között a társadalomtudományi és az alkalmazott nyelvészeti (például pragmatikai) tárgyú kutatások egyik legfontosabb vizsgálati eszközét a számítástechnikai eszközökkel elemezhető formátumú szövegtörzsek jelentik. A korpuszok három legfontosabb tulajdonságát a következőkben határozhatnánk meg: mindenekelőtt a korpusz ténylegesen előforduló írott vagy leiratozott beszélt nyelvi adatok gyűjteménye (Oravecz et al., 2014). Mindenképpen olyan adatokból áll tehát, amelyek a nyelvhasználat folyamán keletkeztek, nem pedig a kutató maga állította elő azokat a vizsgálat céljából, ún. introspektív adatként (vö. McEnery, 2005). A korpusz e sajátsága kétségkívül új távlatokat nyit a tudományos vizsgálatokban, hiszen annak tartalma – szemben például a nyelvészeti kutatásban egyébként gyakran alkalmazott int-

rospektív adatokkal – a nyelvi valóságot, a tényleges emberi nyelvhasználatot tükrözi, annak kvalitatív és kvantitatív sajátásaival együtt. Ha a problémát a társadalomtudományi kutatás oldaláról nézzük, még nyilvánvalóbbá válik a korpuszadat jelentősége: szociológiai vizsgálatokat ugyanis kizárólag valós, a természetes kommunikáció folyamatában létrejött nyelvi produktumokon lehet végezni. A szövegtörzsek másik fontos előnye, hogy nagy mennyiségű szöveget foglalnak magukban, ezáltal általános érvényű állítások empirikus igazolását teszik lehetővé. Végül, az utóbbi sajátással összefüggésben, a szövegtörzsek elektronikus formájúak, amelynek köszönhetően a bennük lévő nagy mennyiségű adat elemzése nem csupán manuális, de automatizált módszerekkel is lehetséges, így egy-egy sajátosság feldolgozása gépi megoldással gyorsan és költséghatékonyan elvégezhető.

A Magyar Nemzeti Szövegtár (MNSZ2, URL1; Oravecz et al., 2014) honlapja alapján a korpuszban „a szövegeket valamilyen szempont szerint válogatják és rendezik. Nem feltétlenül egész szövegeket tartalmaz, és nemcsak tárháza a szövegeknek, hanem tartalmazza azok bibliográfiai adatait, bejelöli a szerkezeti egységeket (bekezdés, mondat).” Bár a definíció nem mutat rá, a szövegtörzsek a legtöbb esetben valamilyen manuális vagy automatikus feldolgozási folyamaton (másképpen: annotáción) esnek át. Ráadásul ennek a feldolgozási folyamatnak a sajátosságait a korpusz jövőbeli felhasználási céljai határozzák meg. A korpuszok annotációja egy rendkívül értékes sajátosság, ugyanis ez teszi lehetővé a korpusz sok szempontú, géppel támogatott lekérdezését. Az annotáció olyan annotálási jelek (számítógépes nyelvészeti terminussal: tagek) alkalmazását jelenti, amelyeket a korpuszban levő szövegekre visznek fel a korpusz építői (másképpen annotátorai). Ezek a jelek hivatottak explicitté tenni a nyelvi adatokban már meglévő, azonban addig implicit formájú információt (vö. McEnery, 2005). Amennyiben a korpusz egyáltalán vagy az aktuális tudományos kutatásnak megfelelő annotációt nem tartalmaz, úgy lekérdezése, azaz a benne szereplő adatok automatikus kinyerése rendkívüli mértékben korlátozott, jobbára csupán kézi megoldással lehetséges.

A korpuszok annotálása automatikus, félautomatikus, valamint manuális munkával valósítható meg (vö. McEnery, 2005). Számos feladat (például a szótövezés vagy a szófaji egyértelműsítés) ma már olyan hatékonyággal végezhető el automatikus módszerrel, hogy ezekben a feladatokban nem szükséges humán annotátorokat alkalmazni. Amennyiben azonban az annotálás – annak jellege miatt – automatikus módszerrel egyáltalán nem végezhető el, és nem támogatható, úgy a teljes munkát humán annotátoroknak kell elvégezniük. Ezt a megoldást alkalmazzák a legtöbbször olyan esetekben, amikor a korpuszt valamely szemantikai vagy pragmatikai jelenség vizsgálatára kívánják felhasználni a jövőben. Ilyenkor ugyanis olyan nyelvi sajátosságokat kell annotálni a korpusz szövegeiben, amelyek az automatikus feldolgozására jelenleg nincsenek ren-

delkezésre álló eszközeink. Tekintettel arra, hogy ezeknek a korpuszoknak az elkészítése jelentős költséget igényel, a nagyméretű, kézzel annotált adatbázisok száma kimondottan csekély.

A fentebb elmondottakkal összefüggésben a manuális módszerrel feldolgozott és sokrétű annotációval ellátott korpuszok két alapvető okból is a legértékesebb kutatási és fejlesztési eszközök között tartandók számon. Egyrészt tudományos felhasználásuk új, eddig ismeretlen összefüggések feltárását teszi lehetővé. Másrészt a felhasználásukkal lehetségessé válhat újabb típusú annotációra is képes automatikus nyelvfeldolgozó eszközök fejlesztése és tesztelése.

A korpuszok között írott és beszélt nyelvi szövegekörpuszokat is találunk, azonban a legtöbb létező korpusz az írott nyelvet reprezentálja (McEnery, 2005). Ennek talán a legfontosabb oka az, hogy a beszélt nyelvi anyag feldolgozására jelenleg sokkal kevesebb eszköz áll a rendelkezésünkre, mint az írott nyelvi adatok kezelésére. Különösen csekély a magyar beszélt nyelvet reprezentáló korpuszok száma, és ezek is többségükben olvasott szövegeket tartalmaznak (vö. Gósy et al., 2012). Spontán, beszélt magyar nyelvet tartalmaz az adatközlői történetmeséléseket tartalmazó Kivi korpusz (Kugler, 2015), az ún. Magyar spontán beszéd adatbázis (BEA) (Gósy et al., 2012), és egy különösen lévő spontán beszédet tartalmazó korpusz, a Budapesti Egyetemi Kollégiumi Korpusz (BEKK, URL2). A BEA korpusz létrehozóinak fő célja az volt, hogy fonetikai, és nem szemantikai vagy pragmatikai vizsgálatokat tegyen lehetővé, ennek megfelelően alakították ki a korpuszban alkalmazott annotációt. Az elmondottak okán a BEA-korpusz társadalomtudományi tárgyú kutatásokra csupán korlátozottan alkalmazható. A Bodó Csanád kutatócsoportja által épített BEKK Budapesten élő fiatalok egymás közötti nyelvi interakcióit tartalmazza, és eleve társadalomtudományi céllal készül. A társalgáselemzés keretében lehetővé teszi majd a társas identitások létrejöttének és változásának vizsgálatát, különösen a társadalmi nem és a szexualitás konstrukciói mentén.

Míg a BEKK esetében az interakciókat a résztvevők saját telefonjaikon rögzítették, ezért szelektív társalgásokat tartalmaz, és nem reprezentálja tökéletesen a teljes élőbeszédet, az általunk létrehozandó HuTongue korpusz hiánytalanul tartalmaz majd egy hosszabb időszakból magyar nyelvű spontán beszélgetéseket. A nagyméretű és megfelelően annotált korpusz lehetővé teszi a pletyka eddig ismeretlen természetének a feltárását, mélyebb megismerését. Emellett az adatbázis, reményeink szerint, különböző célú további szociológiai és nyelvészeti tárgyú kutatáshoz és fejlesztéshez nyújt majd táptalajt a jövőben.

A HUTONGUE KORPUSZ LÉTREHOZÁSA

A hanganyag előfeldolgozása

A HuTongue korpusz alapja egy hozzávetőlegesen ezerórás hanganyag, amelyet a nap 24 órájában rögzítettek nyolc napon keresztül, zárt környezetben.² A nyolc résztvevő mindegyike saját mikrofonnal rendelkezett, ezzel segítve elő a hangfájlok jó minőségű rögzítését. A hanganyagot kiváló minőségben tartalmazó tömörítetlen audiofájlok összmérete mind a hanganyag exportálását, tárolását, mind a feldolgozását megnehezítette, így eleve kisebb méretre hozott .wav kiterjesztésű audiofájlokkal dolgoztunk. Az audiofájlok tárolására és feldolgozására az MTA Felhőben létrehozott virtuális gépen kaptunk lehetőséget.

A hanganyag a résztvevők ébren töltött idejének és a hosszú csendek, nem beszédhangok (mosogató, háttérzene stb.) szűrése után összesen közel ötszáz órányi leiratozható hangfelvételt tartalmaz. Az előfeldolgozás során először eltávolítottuk azokat a részeket (néhány esetben egész órákat), melyek közel teljes csendet tartalmaztak, egy hangerősségbeli küszöbérték felhasználásával. A szűrés következő lépcsője során egy gépi tanulást alkalmazó módszerrel azonosítottuk azokat az egységeket, amelyek nem minősülnek csendnek, avagy kellően hangosak voltak, és koherens szegmenseket képeztek. Mivel az azonosítás sok esetben kis terjedelmű (néhány tizedmásodperces) szegmenseket eredményezett, ezeket a bennük hallható emberi hang (a Voice Activity Recognition technika alkalmazásával) és időbeli eloszlásuk alapján (egymáshoz képest 15 másodperces távolságban) nagyobb méretű szegmensekké egyesítettük. A nagyobb egységeket aztán darabonként körülbelül egyórás audiofájlokban fűztük össze, melyekben a szegmenshatárokat jól elkülöníthetően egy dallammal jeleztük. A szűrés egyes lépéseit manuálisan ellenőriztük, hogy az előfeldolgozás eredménye koherens egységeket alkosson, és hogy kivehető emberi beszédet a szűrőmechanizmus semmilyen esetben se távolítson el. Az ellenőrzés a módszer magas pontosságára engedett következtetni.

A gépi leiratozás próbái sajnos nem hoztak megbízható eredményeket, így a HuTongue magyar nyelvű, spontán nyelvi korpusz kézi leiratozással és annotálással történő építése mellett döntöttünk.

² A spontán nyelvi hanganyagot, amelyet egy szórakoztatóipari cég rögzített, kizárólag tudományos célokra adták át, és használjuk fel, teljes titoktartási kötelezettségvállalás mellett. A hanganyag a résztvevők nyolc napra vonatkozó összes beszélgetését tartalmazza. A felvételen részt vevő önkéntesek teljes körű tájékoztatásban részesültek a hangfelvételek elkészüléséről.

Annotáció

A hanganyagot a különböző időszegmensek visszakereshetővé tétele és összefűzése érdekében meghatározott tagolás után az annotálók időbélyeggel látták el. Az így kapott nyersanyag azonkívül, hogy megőrzi a hanganyaggal való kapcsolatát, további időintervallum alapú mérési eszközök kidolgozására ad lehetőséget. A kézi annotálás lehetőséget biztosít ahhoz, hogy kiszűrjük a szövegből azokat az értékelő párbeszédeteket, amelyek során a diskurzus tárgya nincs jelen, tehát az adott beszélgetést pletykaként értékelhetjük. A pletyka megjelenésekor azonosíthatóvá válik a küldő és fogadó fél, hiszen az annotátorok kódokkal jelzik a résztvevők nevét és a pletyka tárgyát is, amennyiben a résztvevők közül kerül valaki említésre. A diskurzus közben beazonosíthatunk olyan további személyeket is, akik a beszélgetés során csendben maradtak, de jelenlétük az annotáló által érzékelhető.

A korpusz kialakításakor arra törekedtünk, hogy mélyebb betekintést nyerjünk a pletyka spontán beszédhelyzetekben való megnyilvánulásairól, ezért annotációs jeleket használtunk a beszélők érzelmi megnyilvánulásának rögzítéséhez. A kézi annotálás nem szövegszerű kódok rögzítésére is lehetőséget ad. Az így használt annotációs kódok egyik csoportja az emóciós hanghatások lejegyzéséhez fűződik, hiszen arra is kíváncsiak vagyunk, hogy melyek azok az érzelmek, amelyek a mindennapi, informális kommunikáció során megjelenhetnek. Az ilyen annotációs jelek például a nevetés, sírás, torokköszörülés, sóhajtás, gúnyos nevetés stb. azonosítására alkalmasak (Szabó–Galántai, 2017). Így vizsgálhatóvá válhat, hogy milyen típusú érzelmek jelennek meg olyan mindennapi, társas szituációkban, amelyekben pletyka hangzik el.

További annotációs jelek beazonosíthatóvá teszik a beszélőváltást, a beszéd közbeni szünetek hosszát és gyakoriságát, az egy időben történő beszédet és az egymás szavába vágás előfordulásának jellemzőit is, amelyek a beszélők közötti erőviszonyok feltárását is lehetővé teszik.

A korpusz minőségének és értelmezhetőségének biztosítása érdekében értelmező annotációs kódokat is alkalmazunk, melyek az érthetetlen, az azonosíthatatlan vagy nem a diskurzusban részt vevő személyektől érkező beszédet jelölik (részletesebben: Szabó–Galántai, 2017).

Minőségbiztosítás

Az így kapott korpusz megbízhatóságát és összeegyeztethetőségét többféle módszerrel és több dimenzióban mérjük. Mindez a kvalitatív, szűrőpróbaszerű ellenőrzés mellett gépi eszközökkel történik. A leiratozást és annotálást végzők munkáját több dimenzióra bontva hasonlítjuk össze, egyrészt egymáshoz képest, másrészt egy referenciagépelőhöz viszonyítva. Súlyos minőségi kifogások ese-

tén az adott szöveganyagot újra leiratoztatjuk és annotáltatjuk. A leiratozóknak és annotálóknak az egyes minőségbiztosítási dimenziókban egyéni visszajelzést adunk, és munkájuk minőségjavulását ellenőrizzük. Az összehasonlítás a szövegegyezés, az annotálás, a szereplők azonosítása és jelölése, és az időbélyeg használatának fő dimenzióiban történik, amelyeket a pontosabb visszajelzés érdekében részdimenziókra bontunk. Ezeket a korpuszépítés alatt folyamatosan ellenőrizzük, és mérőszámokkal dokumentáljuk. A szövegegyezés főbb mérőszámaiként a tisztított szegmenseken számolt Levenshtein-távolságot (Levenshtein, 1965) és a koszinusz hasonlóságot (cosine similarity) használjuk.

Az adatbázis szerkezete

A nagy szövegtörzs eltárolása egy olyan adatbázisban történik, amely hatékonyan képes nagy mennyiségű szöveget rögzíteni és kereshetővé tenni (Elasticsearch). A szöveges keresőmotor képes a teljes korpuszunk megbízható tárolására és gyorskereső, összegző és akár elemzési műveletek végrehajtására is.

Az általunk gyűjtött információmennyiség több különböző (az Elasticsearch terminológiáját használva) index és típus alatt tárolódik az adott adattípus jellemzőitől függően. Két fő adatsort különböztetünk meg. Az első adatbázist az egyedi (egy szereplőtől egy időegységben származó) megszólalások időbélyegekként (a megszólalás pontos ideje az adatfelvétel időintervallumán belül) ellátott korpusza adja. Az egyes megszólalásokhoz az időbélyegeken túl számos egyéb attribútumot társítunk, melyek a beszélő azonosítóját, az egyes annotációs kódok jelenlétét, a szöveget rögzítő gépelő kilétét, és még sok más információt tartalmaznak. A teljes szövegtörzs adatbázisát kiegészítendő létrehozunk egy olyan adatsort is, amely a különálló szövegrészek minden egyes szavához olyan attribútumokat társít, mint a szófaja, a mondatrész vagy a szó lemmatizált formája. Ezen adatbázis minden eleme egyértelműen megfeleltethető a korpusz minden szavának. Ez a struktúra a kutatók számára jelentősen megkönnyíti a kereshetőséget és az elemzést, különösen az NLP (Natural Language Processing) problémákra vonatkozóan.

Az adatbázis továbbá lehetővé teszi, hogy a beszélők hangfájlaiból leiratozott beszédek párbeszéddé illesszük össze. Ezzel a párbeszéd is elemzési egységgé válhat a későbbiekben. Mivel minden beszélőnek saját hangfelvétel készüléke volt, ezért sok esetben ezek leiratai ugyanazon párbeszéd különböző részeit tartalmazzák. Mivel minden leiratozott sor saját időbélyeggel rendelkezik, az azonos időben történő beszédek egymás mellé tesszük, és az illesztést horgonyszavak segítségével pontosítjuk. Ezt segíti, hogy minden leirat tartalmaz több beszélőt, és tartalmazza, hogy kik vettek részt az adott diskurzusban. A pletyka azonosításának szempontjából zárt közösségről lévén szó, ugyancsak fontos, hogy ennek következtében az is behatárolható, hogy kik *nem* voltak jelen az adott beszélgetés során.

A PLETYKA ELEMZÉSÉNEK A COMPUTATIONAL SOCIAL SCIENCE ÁLTAL KÍNÁLT IRÁNYAI

A korpusz tervezett automatikus feldolgozási lépései

A manuálisan gépelt és annotált korpuszt különböző automatikus megoldásokkal is fel kívánjuk dolgozni. Ahogyan azt a korpusz feldolgozása kapcsán ismertettük (lásd fentebb), a létrejövő spontán nyelvi adatbázis tartalmaz szövegszerű és nem szövegszerű változókat is. Az automatikus feldolgozási megoldásokkal közülük a szövegszerű adatokra szeretnénk további információkat felvinni az elkészült adatbázisban. E munka célja, hogy a korpusz minél részletesebb annotációval rendelkezzen a szövegek grammatikai, szemantikai, valamint pragmatikai sajátosságairól (Szabó–Galántai, 2017).

A szövegszavak tokenizálásához, morfológiai, valamint szófaji elemzéséhez a magyarlanc elemző eszközt kívánjuk használni (Zsibrita et al., 2013). Ez a feldolgozási lépés két okból is fontos a számunkra. Egyrészt úgy véljük, hogy a különböző grammatikai sajátságok kvantitatív adatai a pletykaszövegek azonosításában is hasznosíthatóak lesznek (lásd lentebb). Másrészt, a magyarlancsal kapott kimenet megfelelő bemenetként szolgálhat a további automatikus szemantikai feldolgozási lépésekhez.

A szöveg szemantikai tartalmát illetően elsősorban névelem-felismerést, szentiment- és emócióelemzést, valamint topikmodellezést tervezünk végrehajtani a kész korpuszon. A szentiment- és emócióelemzéshez a természetesnyelv-feldolgozás (NLP) eszköztárából a szótárillesztés módszert alkalmazzuk, amely például a gépi tanulás, a szintaktikai elemzésen alapuló mintaillesztés mellett egyszerűbb és költséghatékonyabb információkinyerési módszer (Drávucz–Szabó, 2017). A szentimentelemzéshez olyan lexikonra van szükségünk, amely szótári formába rendezi a lexikai szinten pozitív vagy negatív értékelő tartalommal rendelkező nyelvi elemeket. A szövegszintű emóciók felcímkezéséhez pedig olyan szótárat kell alkalmaznunk, amely a különböző érzelmek nyelvi realizációit tartalmazza, illetve minden olyan elemet, amely valamely emóció meglétére utalhat (Szabó–Morvay, 2015). E két elemzés segítségével fel tudjuk tárni a szövegekben megbúvó negatív és pozitív értékítéleteket és a különböző emóciókat.³

A korpuszban megjelenő témákat két megoldással kívánjuk feltárni. Annak érdekében, hogy a diskurzusok témáit leíró módon meghatározhassuk, és ezt követően tovább elemezhessük, az elkészült korpusz segítségével saját nyelvi szótárakat is készítünk. A szótárak a tervezetteknek megfelelően egy- és többszavas kifejezéseket egyaránt tartalmaznak majd. A szótárakon kívül topikmodellek segítségével tervezzük a nagyméretű szöveghalmazban a különböző témaegységek elkülönítését (például: pletyka, politika, sport, időjárás stb.).

³ Erre a célra például a PrecognoX Informatikai Kft. által fejlesztett szótárak használhatóak.

Ahhoz, hogy a korpusz szövegeinek pragmatikai sajátosságait még mélyrehatóbban feltárhassuk, a korpuszt a nyelvi bizonytalanság különböző típusú jelölőinek a szótárával, valamint a diskurzusjelölők szótárával is fel kívánjuk dolgozni. A bizonytalanságot jelölő kifejezések automatikus azonosítása napjaink nyelvtechnológiai kutatásainak egyik fontos problémaköre (Vincze, 2014). A bizonytalanságot jelölő elemek ugyanakkor a beszélői szubjektivitás fontos indikátorai lehetnek, ezért a nyelvi bizonytalanság detektálásával lehetőségünk nyílt a vizsgált szemantikai tartalmak szubjektivitási értékeinek a megismerésére is (Drávucz–Szabó, 2017). A feldolgozás során egy bizonytalanságjelölő elemeket tartalmazó szótárt (Vincze, 2014) illesztünk a kész korpuszra. Mindez a pletyka értékítéletet tartalmazó tulajdonságát hivatott azonosítani.

Végezetül, a diskurzusjelölők annotálása lehetőséget teremt arra, hogy a diskurzusok e fontos szövegszintű kapcsolóelemeit is lekérdezzük, és azok kvantitatív és kvalitatív sajátosságait a spontán nyelvhasználatban feltárjuk. A diskurzusjelölőknek nincs lexikális értelemben vett jelentésük, ehelyett procedurális (műveleti) jelentéssel bírnak. Színesítik, árnyalják a mondandót, illetve emocionális tartalmakról és beszédtervezési folyamatokról tanúskodnak.

A pletyka azonosítási lehetőségei

A fentebbieknek megfelelően, a manuálisan és gépi megoldásokkal egyaránt feldolgozott adatbázis számos elemzési módot kínál a kutatási kérdéseink megválaszolására. A pletyka általunk vázolt és az irodalomban általánosan elfogadott definíciója alapján az azonosításhoz több feltétel szükséges. Egyrészt, szükséges annak megállapítása, hogy a beszélő egy harmadik személyről beszél. Másrészt, szükséges annak megállapítása, hogy a harmadik személy nincs jelen a beszélgetésnél. Harmadrészt, szükséges annak meghatározása, hogy az említés értékelő tartalommal jár együtt. Az azonosítás lépései többféleképp történhetnek, ezért pragmatikusan megkülönböztethetünk pletykát szűkebb és tágabb értelemben is.

A legegyszerűbb általunk használt módszer az annotálás során használt jelekre és a névemlítés személyes és utaló névmásokkal kiegészített összekapcsolására épül. Az azonosítást követően a gazdagon annotált korpusz lehetővé teszi számunkra, hogy azonosítsuk a pletyka megjelenésében potenciálisan szerepet játszó tényezőket, és mérjük ezek fontosságát. Az elemzés nagyrészt a kvantitatív szövegelemzés eszközeivel történik.

A pletyka és nem pletyka jellegű szövegeket arányaiban és karakterisztikáiban is el tudjuk majd különíteni egymástól. A részletesen bemutatott manuális és automatikus feldolgozási megoldások alapján, prediktív modellek segítségével képesek leszünk megvizsgálni, hogy milyen tényezők valószínűsítik a pletykát egy adott szövegben. Ilyen a különböző nyelvi kifejezések jelenléte (például: a

szeniment- és emóciókifejezések, a nyelvi bizonytalanság jelölői vagy a diskurzusjelölők), a lexémák morfológiai és szófaji tulajdonságai, valamint a különböző, a szövegekben megjelenő nem szövegszerű hanghatások. Mindemellett a topikmodellezéssel feltárt témák sajátosságai rámutathatnak, hogy a pletykához milyen témák kapcsolódnak, és melyek azok a témák, amelyek tipikusan nem jelennek meg a pletykaként azonosított szövegek közvetlen környezetében.

A topikmodellezés célja a beszélgetések tartalmában a közösségi normákra, együttműködésre utaló tartalmak azonosítása is. Ez hozzájárulhat annak a kiinduló hipotézisnek a teszteléséhez, hogy a pletyka pozitív szerepet játszik a közösségi normák és a kooperáció fenntartásában.

A vizsgált jelenség sajátosságai okán elsősorban szövegcsoportosításra alkalmazott klasszifikációs modellekkel fogunk dolgozni. A szövegkategorizáció a nyelvi elemek előre meghatározott kategóriákhoz való hozzárendelése a tartalmuk figyelembevételével. A pletyka szempontjából lényeges tartalmak azonosítása egy csoportosítási problémának tekinthető, amely megvalósítható neurális hálózatok modelljeivel, *support vector machines* (SVM) vagy naiv bayesi klasszifikációs modellek segítségével.

Az elkészült adatbázisban lehetséges lesz a beszélők informális kommunikációs hálózatának kvantitatív elemzése. A kapcsolatok alakulását és a hálózat változását időben is vizsgálni tudjuk kapcsolati háló elemzési módszerekkel, valamint azonosítható és vizsgálható lesz konkrét tartalmú értékítéletek terjedése az adott kisközösségben.

A pletyka mint diskurzus

Az adatbázis lehetővé teszi, hogy az egyes diskurzusok között időbeni összekapcsolásokat végezzünk. Az összekapcsolás a gyakorlatban azonban egyáltalán nem triviális. Megnehezíti egyrészt az egy időben, de egymástól távol zajló beszélgetések azonos időbélyege és ugyanazon beszélgetés két oldalról történt egymástól eltérő leírása. Az összekapcsoláshoz az időbélyegek használata mellett ezért a beszélők és a beszélőpartnerek annotált jelzéseit és a szövegben előforduló horgonyszavakat és horgonyszólancokat is használjuk.

Az összeillesztett párbeszédben lehetőség nyílik diskurzuselemzésre, ahol megvizsgálhatjuk, hogy milyen stratégiákat alkalmaz a beszélő, és hogyan vonja be a fogadó felet a pletykálás folyamatába. Azonosíthatókká válnak a fogadó fél tipikus reakciói és saját értékítéletének a változása.

Mivel a korpusz mellé a hangfájlok is rendelkezésre állnak, a későbbiekben a megfelelő elemzési módszerek fejlődésével és az annotált leirattal való összekapcsolással a hangok karakterisztikáit is megvizsgálhatjuk, mint például, hogy milyen sűrűn, milyen hangosan vagy milyen intenzitással beszélnek az azonosított pletykahelyzetekben a beszélők.

ÖSSZEZÉS

Általánosabban arra szeretnénk választ kapni, hogy a mindennapi élet során hogyan jelenik meg a pletyka? Kézzelfogható eszközeink vannak arra, hogy olyan alapállapításokat is felülvizsgáljunk, amelyeket eddig a szakirodalomban készpénznek vettek, de megnyugtató módon empirikusan nem támasztottak alá, mint hogy az emberek a beszélgetéseik kétharmadát pletykálással töltik, vagy hogy a pletyka alapvető funkciója a normaszegők kibeszélése.

A pletyka operacionalizálásához kapcsolódó kérdésünk, hogy miként azonosítható be kvantitatív szövegelemzési eszközökkel az élő, spontán szövegben a pletyka, hogyan lehet azt elválasztani a diskurzus más elemeitől és témáitól, és milyen morfológiai jellemzőkkel írható le.

IRODALOM

- Drávucz F. – Szabó M. K. (2017): A beszélői szubjektivitás vizsgálata szentiment- és emóciókorpuszokon. In: Ludányi Zs. (szerk.): *Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből 2017. XI. Alkalmazott Nyelvészeti Doktoranduszkonferencia*. Budapest, 39–49. http://www.nytud.hu/alknyelvdok17/proceedings/Dravucz_Szabo.pdf
- Dunbar, R. I. (1996): *Grooming, Gossip and the Evolution of Language*. Cambridge, MA: Harvard University Press
- Dunbar, R. I. (2004): Gossip in Evolutionary Perspective. *Review of General Psychology*, 8, 2, 100–110. DOI:10.1037/1089-2680.8.2.100, <http://allegatific.unipv.it/ziorufus/Dunbar%20gossip.pdf>
- Ellwardt, L. (2011): *Gossip in Organizations. A Social Network Study. (ICS Dissertation Series)* Groningen. https://www.researchgate.net/publication/254821799_Gossip_in_organizations_a_social_network_study
- Foster, E. K. (2004): Research on Gossip: Taxonomy, Methods, and Future Directions. *Review of General Psychology*, 8, 2, 78. DOI:10.1037/1089-2680.8.2.78, <https://pdfs.semanticscholar.org/8b2f/3c70bd2346b2218b743a765766e5a80a1718.pdf>
- Gósy M. – Grácsi T. E. – Gyarmathy D. et al. (2012): *Magyar spontán beszéd adatbázis = Hungarian Spontaneous Speech Corpus*. OTKA kutatási beszámoló, <http://real.mtak.hu/12552>
- Kugler N. (2015): *Megfigyelés és következtetés a nyelvi tevékenységben*. Budapest: Tinta Kiadó
- Kurland, N. B. – Pelled, L. H. (2000): Passing the Word: Toward a Model of Gossip and Power in the Workplace. *Academy of Management Review*, 25, 2, 428–438. DOI: 10.5465/AMR.2000.3312928, <http://www.csun.edu/~nkurland/PDFs/AMR%20Gossip%202000.pdf>
- Levenshtein, V. I. (1965): Двоичные коды с исправлением выпадений, вставок и замещений символов. *Доклады Академии Наук СССР*, 163, 4, 845–848. <http://www.mathnet.ru/links/50defca5677a80b2d88d3dc027ac4173/dan31411.pdf>, Angolul: Levenshtein, V. I. (1966): Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. *Soviet Physics Doklady*, 10, 8, 707–710. <https://nymity.ch/sybilhunting/pdf/Levenshtein1966a.pdf>
- McEnery, T. (2005): Corpus Linguistics. In: Mitkov, R. (ed.): *The Oxford Handbook of Computational Linguistics*. Oxford University Press, Oxford, 448–463.
- Oravecz Cs. – Váradi T. – Sass B. (2014): The Hungarian Gigaword Corpus. *Proceedings of LREC, Reykjavik, European Language Resources Association (ELRA)*, 1719–1723. http://real.mtak.hu/20143/1/681_Paper.pdf

- Szabó M. K. – Galántai J. (2017): Egy magyar nyelvű spontán beszélt nyelvi korpusz (HuTongue) létrehozásának tapasztalatai. In: *XXVI. MANYE Kongresszus konferenciakötete*. Pécs
- Szabó M. K. – Morvay G. (2015): Emócióelemzés magyar nyelvű szövegeken. In: Gecső T. – Sárdi Cs. (szerk.): *Nyelv, kultúra, társadalom*. Budapest: Tinta Kiadó, 286–292.
- Vincze V. (2014): Uncertainty Detection in Hungarian Texts. In: *Proceedings of COLING 2014*. Dublin. 1844–1853. <http://www.aclweb.org/anthology/C/C14/C14-1174.pdf>
- Zsibrita J. – Vincze V. – Farkas R. (2013): magyarlanc: A Toolkit for Morphological and Dependency Parsing of Hungarian. *Proceedings of Recent Advances in Natural Language Processing*, Hissar. 763–771. <https://www.aclweb.org/anthology/R/R13/R13-1099.pdf>

URL1: <http://clara.nytud.hu/mnsz2-dev/>

URL2: bekk.elte.hu

A TÁRSADALOM HÁLÓZATI JELENSÉGEINEK ADATVEZÉRELT VIZSGÁLATA

DATA DRIVEN INVESTIGATION OF NETWORK PHENOMENA OF SOCIETY

Kertész János¹, Roberta Sinatra², Vedres Balázs³

¹ az MTA rendes tagja, Kerteszl@ceu.edu

² PhD, SinatraR@ceu.edu,

³ PhD, VedresB@ceu.edu

Közép-európai Egyetem (CEU) Hálózattudományi Központ

ÖSSZEFOGLALÁS

A CEU Hálózattudományi Központja társadalmilag releváns kérdések adatokon alapuló, hálózati szempontú vizsgálatával foglalkozik. Néhány példa: Milyen az emberi kapcsolatok szerkezete társadalmi méretekben a kommunikációs adatok tükrében, és hogyan változik mindez az új médiumok hatására? Hogyan terjednek a járványok, az információ vagy egy innováció a társadalomban? Hogyan lehet optimalizálni egy team összetételét a kreativitás szempontjából? Mi a nemek szerepe a munkacsoportok hatékonyságában? Hogyan lehet mérni és előre jelezni a sikert? Ebben a cikkben röviden áttekintünk néhány eredményt, amelyeket a társadalmi terjedési jelenségek, a teamek kreativitásának hálózati vonatkozásai és a siker társadalmi elmélete területeit kutatva értünk el.

ABSTRACT

The Center for Network Science of the Central European University (CEU) deals with data-based, network related investigation of socially relevant problems. Some examples: What is the structure of human interactions at the societal scale in the light of communication data and how does it change due to the new media? How do epidemics, information or an innovation spread in society? How can the composition of a team be optimized from creativity's point of view? What is the role of genders in the efficiency of working teams? How can success be quantified and predicted? In this article we briefly survey some results we have obtained by researching social contagion phenomena, the network aspects of team creativity and the social theory of success.

Kulcsszavak: hálózattudomány, adatvezérelt kutatás, társadalmi terjedési jelenségek, munkacsoportok kreativitása, a siker mérése és modellezése

Keywords: network science, data driven research, social contagion, creativity of working teams, measuring and modelling success

ADATÖZÖN, KOMPLEXITÁS ÉS HÁLÓZATI SZEMPONT A TÁRSADALOM VIZSGÁLATÁBAN

A társadalom a komplex rendszerek paradigmatis példája: nagyszámú kölcsönható egyedből épül fel, visszacsatolások, nem-linearitások jellemzik, és úgynevezett emergens jelenségek lépnek fel benne. Az ilyen rendszerekre érvényes Philip W. Anderson Nobel-díjas fizikus híres mondása: „A több másmilyen” („More is different”), amivel a természettudományokat évszázadokon keresztül meghatározó redukcionizmus korlátaira mutatott rá. Nem elég alaposan megérteni az alkotóelemek – a társadalom esetében: az emberek – közötti kölcsönhatásokat (és persze az alapos megértéstől távol vagyunk), az alkotóelemek nagy száma új minőséget hozhat létre. Ilyenek lehetnek a nagy társadalmi változások, az új struktúrák, intézmények megjelenése, de akár egy innováció létrejötte és elterjedése is. Stephen Hawking szerint a 21. század a tudományban a komplexitás évszázada lesz. Ennek a jóslatnak a beteljesüléséhez új módszerekre, új megközelítésekre és diszciplínákon átívelő együttműködésre van szükség.

A társadalom az egyik legkomplexebb rendszer, mégis, az új tudományos erőfeszítések jelentős része rá irányul. Ennek oka részben abban rejlik, hogy a globális világunkat érő kihívásoknak úgy tudunk megfelelni, ha jobban megértjük a társadalom szerkezetét és dinamikáját. Mindehhez az ezredfordulóra értek meg azok a technikai feltételek, amelyek a tudomány gyors fejlődését lehetővé tették.

Az elmúlt két évtized drámai változást hozott a társadalmi jelenségek kvantitatív vizsgálatában. Az infokommunikációs technológia fejlődése, a digitális megoldások általánossá válása a legkülönbözőbb területeken hihetetlen mennyiségű adatot eredményez. Ma már az emberi aktivitás önkéntelenül is folyton adatokat hoz létre az élet szinte minden területén, legyen szó kommunikációról, egészségügyi vizsgálatokról, kereskedelmi tevékenységről vagy közlekedésről. Ezek a „digitális lábnyomok” alkalmasak az egyéni viselkedés részletekbe menő, „mikroszkopikus” tanulmányozására, ráadásul óriási, társadalmi méretű („makroszkopikus”) mintákon. Olyan ez, mintha egy gáz viselkedését nem a hőmérséklete, sűrűsége és nyomása segítségével, hanem az egyes molekulákat nyomon követve próbálnánk leírni. A gázok esetében tudjuk, hogyan jönnek létre a makroszkopikus állapotjelzők a mikroszkopikus kölcsönhatásokból és a mozgástörvényekből, így a molekuláris szintű leírás gyakran fölösleges. A társadalom esetében a helyzet bonyolultabb: a mikroszkopikus mozgástörvényeket és a kölcsönhatásokat nem vagy alig ismerjük, így felbecsülhetetlen értékű információkhoz juthatunk az adatözön által biztosított, új lehetőségek révén.

Az ezredforduló körüli forradalmi változásokat a társadalom, és általában a komplex rendszerek vizsgálatában nemcsak az adatözön (Big Data), hanem egy attól nem független szemléleti megújulás is okozta. Ennek elindítója két elméleti dolgozat volt: Duncan J. Watts és Steven H. Strogatz (1998) kisvilág-hálózatokról, valamint Barabási Albert-László és Albert Réka (1999) skálamentes hálózati

modellről szóló cikke. A figyelem a komplex rendszerek váza, a mögöttük rejlő hálózatok felé fordult, és kiderült, hogy tanulmányozásuk elengedhetetlen a komplex rendszerek megértéséhez. Az egyik legmeglepőbb felismerés, hogy a komplex rendszerek sokfélesége ellenére hálózataik számos univerzális sajátosságot mutatnak. Ilyen jellemzők a hálózatok méretéhez képest kis átlagos távolság (lépésszám), a csomópontok erősen heterogén kapcsolódásai (fokszám) és a moduláris szerkezet. Mára a hálózattudomány önálló diszciplínává vált, és kiváló tankönyvek tárgyalják a komplex rendszerek hálózatainak tulajdonságait, a rajtuk zajló folyamatokat és modellezésüket.

A Közép-európai Egyetemen (CEU) 2008-ban alakult Hálózattudományi Központ célul tűzte ki a társadalmilag releváns kérdések adatokon alapuló, hálózati szempontú vizsgálatát. A fentiekből is érzékelhető, hogy a kutatómunka első számú feltétele a megfelelő adatokhoz való hozzáférés, vagy megfordítva: azok a kutatások válnak lehetségessé, amelyekhez sikerül adatokat szerezni. Ilyen, rendelkezésünkre álló adatok például a mobiltelefon-forgalomra vonatkozó számlázási adatok, a Skype internetes rendszer hálózati felépülési adatai, az iWiW internetes szociális hálózat adatai, a *Web of Science* adatai, az amerikai dzsesszenészek együtteseire és lemezfelvételeire, valamint a számítógépes játékok fejlesztő *team*-jeire és a termékek sikerességére vonatkozó adatok. Az ilyen mélységű adattömeg aranybánya a tudomány számára, mindazonáltal használata különleges körülményt igényel. A projekteket ezért a kutatóhelyek etikai bizottságai felügyelik, és biztosítják, hogy a személyiségi jogok ne kerüljenek konfliktusba a kutatással, és így van ez a CEU esetében is.

A továbbiakban néhány példát mutatunk be a CEU adatvezérelt kutatásaiból. Először a társadalmi terjedési jelenségekkel foglalkozunk innovációk terjedésének példáján. Utána néhány eredményt mutatunk be azzal kapcsolatban, hogy a csapatok összetétele hogyan befolyásolja a kreativitásukat. Végül a társadalmi sikeresség új elméletét tárgyaljuk röviden a tudományos eredményesség ismert mérőszámaiból kiindulva. A cikket egy rövid kitekintéssel zárjuk.

TÁRSADALMI TERJEDÉSI JELENSÉGEK

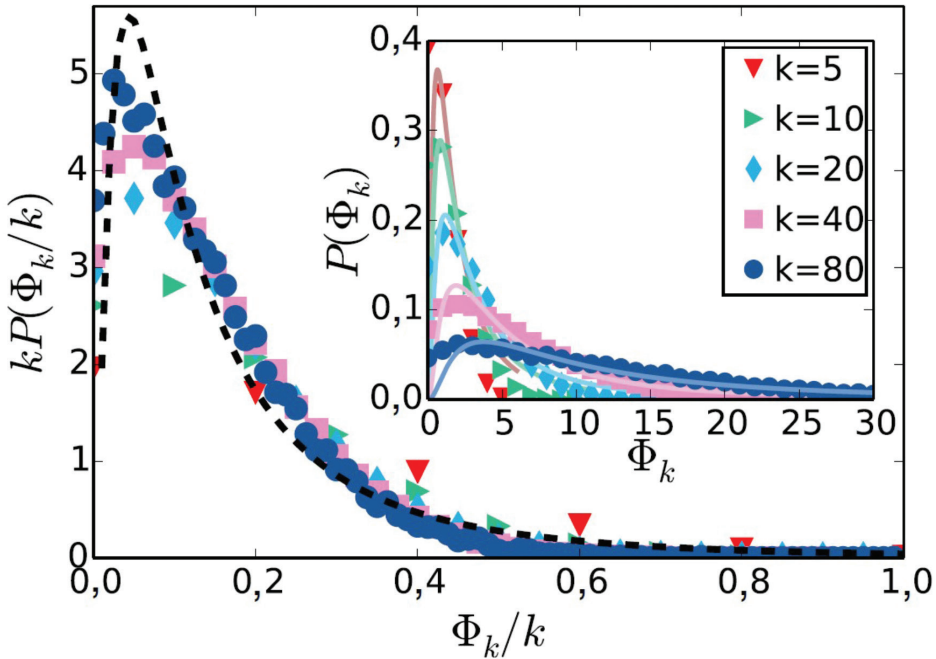
Az emberi kapcsolatok hálózatán számos terjedési jelenséget figyelhetünk meg. A hírek, pletykák, viselkedési formák vagy innovációk jelentős mértékben ilyen módon érik el a társadalom egyre nagyobb hányadát, de a járványok terjedése is ide sorolható. Az információ és a járványok terjedése közötti analógiára már a hatvanas években felhívták a figyelmet. Az adattudomány és a hálózatok kutatás együttes erőfeszítéseinek köszönhetően a járványterjedés leírásában, sőt az előrejelzésében is jelentős haladást sikerült elérni az elmúlt években (Chovell et al., 2017). Az említett társadalmi jelenségek terjedésének leírása azonban bizonyos

értelemben még nehezebb. Itt ugyanis nem valamilyen anyagi hordozók (mikrobák) párkölcsönhatásokban megnyilvánuló átadásáról van szó, hanem döntésekről és azokat megelőző hatásokról. Elhiszem-e és továbbadom-e a pletykát? Átveszem-e az adott viselkedési formát? Megveszem-e az új technikai berendezést? Ezek a döntések általában a környezet, az ismerősök állapotától, hozzáállásától függenek. Az innováció példájánál maradva: ha ismerőseim jelentős része egy új típusú telefonkészüléket használ, és jó véleménnyel van róla, én is hajlok rá, hogy beszerezzek egyet.

Az ilyen komplex terjedési folyamatok (complex contagion processes) leírására vezette be Mark S. Granovetter a küszöbmodellt, amit Duncan J. Watts (2002) öntött matematikailag jól kezelhető formába. A modell lényege a következő: a személyek egy hálózat csúcspontjai, akik társadalmi kapcsolataikkal vannak összekötve. Minden i egyed rendelkezik egy ϕ_i küszöbértékkel, ami az innovációt adaptáló szomszédok kritikus hányadát jelzi. Vagyis, ha $k_i^{(a)}/k_i \geq \phi_i$, akkor az i -edik egyed is adaptálni fogja az innovációt, ahol $k_i^{(a)}$ az adaptáló, k_i pedig az összes szomszédok száma, a fokszám. Az innovációt egy véletlenül kiválasztott csúcsponttól indítjuk. (Itt az innováció helyettesíthető más, társadalmilag terjedő jelenséggel.) Ha egy csúcspont körül az adaptáló szomszédok aránya túllépi a küszöbértéket, akkor a csúcspont maga is adaptálónak válik, ezáltal egy addig nem adaptáló szomszédjánál következhet be a küszöb túllépése, és így tovább. Ilyen módon adaptációs kaszkádok, vagyis egymást előidéző követők füzerei jöhetnek létre.

A modellben a hálózat jellege, az átlagos küszöbérték és a fokszám függvényében két fázist lehet megkülönböztetni, egy globális és egy lokális terjedési fázist. A paraméterek egy tartományában globális kaszkádok alakulhatnak ki – ez felel meg a gyors terjedésnek. A másikban a kaszkádok elhalnak, globális terjedés nem lehetséges. Ez az egyszerű modell számot ad arról az általános megfigyelésről, hogy bizonyos innovációk futótűzként elterjednek a társadalomban, míg mások kudarcot vallanak.

A Watts-féle küszöbmodell azonban nem alkalmas a megfigyelési adatok közvetlen értelmezésére. Igaz, a társadalmi terjedési jelenségekről csak a legutóbbi időben sikerült részletes adatokhoz hozzáférni. Kutatócsoportunk a Skype „Voice over Internet” rendszer adatait vizsgálta (Karsai et al., 2016). Rendelkezésünkre álltak az anonimizált felhasználók regisztrációs adatai, a visszajelzett kapcsolatok kiépítésének és az utolsó belépésnek az időpontja – az ingyenes és a „fizetős” szolgáltatási platformokon egyaránt. Ez utóbbi jelentősége, hogy a fizetős szolgáltatások mindig az ingyenes hálózaton terjednek, így a háttérhálózat ebben az esetben ismert – ellentétben azzal a helyzettel, amikor az innováció az egyébként ismeretlen társadalmi kapcsolatok hálózatán terjed. Ilyen fizetős szolgáltatások tanulmányozása révén lehetett a küszöbérték-eloszlásról információt szerezni (1. ábra).



1. ábra. Betét: A szolgáltatást már használó szomszédok Φ_k számának eloszlása a szolgáltatás adaptációjakor különböző fokszámok esetében, amelyek eltérő szimbólumokkal vannak jelölve.

Nagy ábra: Ugyanaz, de a szolgáltatást már használó szomszédok arányának (Φ_k/k) függvényében. A folytonos, illetve szaggatott vonalak illetett lognormál eloszlásokat mutatnak (Karsai et al., 2016 nyomán)

Azzal, hogy a nagy ábrán a görbék jó közelítéssel egymásra esnek, sikerült bebizonyítani az elmélet korábbi vitatott feltételezését, hogy a küszöbérték megfelelő változója nem az adaptáló szomszédok száma, hanem aránya (Karsai et al., 2016).

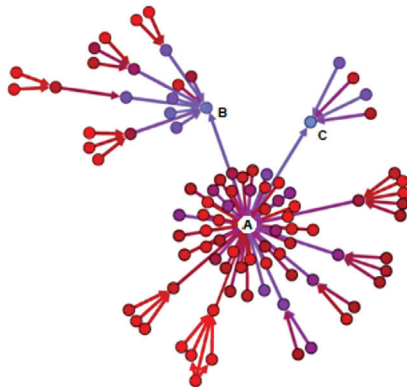
Az adatok arra is alkalmasak, hogy a terjedés mechanizmusába mélyebben betekintsünk. Megállapítható volt, hogy – ellentétben a Watts-modell által sugalmazott képpel – úgy is létrejöhet globális terjedés, hogy nem alakul ki globális kaszkád. Nyilvánvaló, hogy a Watts-modellből hiányoznak bizonyos fontos komponensek. Az egyik kézenfekvő: a terjedést nemcsak a szomszédok viselkedése, illetve a tőlük érkező információ befolyásolja, hanem külső hatások is, például hírek a médiában vagy reklámkampányok. Ilyen hatások révén „spontán” adaptálók léphetnek fel, akik aztán új kaszkádok kiindulópontjai lehetnek. Ez a mechanizmus globális terjedést eredményez kis kaszkádok sorozatán keresztül, ahogyan az az adatokon is megfigyelhető.

Egy másik, az eredeti küszöbmodellekből hiányzó vonás az innovációtól teljesen elzárkózók figyelembevétele. Vannak, akik sem az ismerőseik, vagyis há-

lózati szomszédjaik példáját követve, sem pedig külső hatásra nem hajlandók az innovációt adaptálni. Ennek számos oka lehet, kezdve azon, hogy elégedettek egy korábbi technikai megoldással, addig hogy ideológiai (például környezetvédelmi) szempontok miatt nem hajlandók kipróbálni az új terméket. Az ilyenek beleszámítanak az adaptáló szomszédok arányába, és lassítják a terjedési folyamatot.

A fenti két jelenséget sikerült egy általánosított küszöbmodellbe beépíteni, és a modellt bizonyos típusú hálózatokon jó közelítő módszerekkel megoldani (Ruan et al., 2015). A modell paramétereit a megfigyelésekből lehetett származtatni, kivéve az elzárkózók arányát, amire vonatkozó adatunk természetesen nem áll rendelkezésre, úgyhogy ezt a paramétert a modell kalibrálása során illesztettük. A modell igazolását a bemenő adatoktól nem közvetlenül függő mennyiségeknek a megfigyeltekkel való összehasonlítása jelenti, így például a kaszkádok eloszlása (Karsai et al., 2016).

A fent bemutatott mechanizmus nemcsak innovációk terjedését írja le, hanem alkalmazható például az *online* szociális hálózatok összeomlásának modellezésére is (Török–Kertész, 2017). Ismeretes, hogy a korábban Magyarországon rendkívül népszerű iWiW-szolgáltatás nagyon rövid idő alatt lényegében összeomlott. A Facebooknál korábban, 2002-ben alapított iWiW-et végül 2014 júniusában végképp leállították. Egy online szociális hálózat elveszti vonzerejét, ha az ismerősök jelentős hányada már máshol aktív – ez az a kollektív hatás, amelyet a küszöbmodellek vesznek figyelembe. Ebben a rendszerben az erős vetélytársról, a Facebookról érkező hírek jelentik a külső hatást. Annak ellenére, hogy itt a „makacs ragaszkodók” aránya jóval kisebb volt az előző példában bemutatott r értéknél, itt is véges kaszkádok alakultak ki, amelyeket a spontán távozók váltottak ki. A 2. ábra egy ilyen véges kaszkádot mutat be.



2. ábra. Az iWiW online szociális hálózatról távozók egy véges kaszkádjá. A nyilak a befolyást jelzik, vagyis a korábban távozóról a később távozóra mutatnak. Először A egy kaszkád eredményeképpen távozik, hozzájárulva B és C későbbi távozásához.

(Török–Kertész, 2017 nyomán)

MUNKACSOPORTOK KREATIVITÁSÁNAK HÁLÓZATI VONATKOZÁSAI

Az innováció terjedését meg kell hogy előzze annak létrehozása, ami egy kreatív folyamat. Ez legtöbbször munkacsoportokban, teamekben zajlik, így kézenfekvő a kérdés: hogyan befolyásolja a teamek kreativitását az összetételük és a tagjaik közötti kapcsolatrendszer? Az innováció terjedésének irodalmához képest az innováció létrehozásának hálózati feltételeiről meglepően kevés munka szól. A kérdés hálózati irodalmában az uralkodó paradigma szerint a hálózatok szerepe kettős: az újításokhoz egyrészt szükség van gyenge, de távoli hálózati régiókba mutató kötésekre, és erős, csoporton belüli kötésekre. E paradigma szerint a gyenge és diverz kötések jelentősége abban áll, hogy ötleteket csatornázzanak be kívülről egy munkacsoportba, míg az erős és homogén kötések az ötletek megvalósításában játszanak fontos szerepet (Aral–Alstyn, 2011; Obstfeld, 2005).

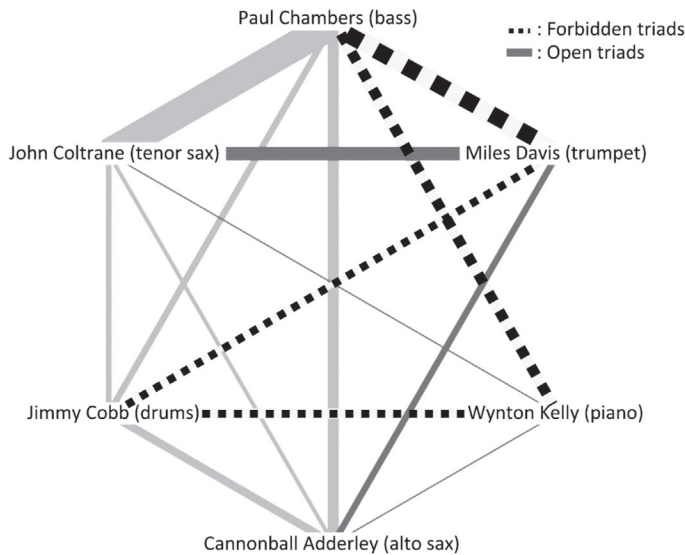
A kreativitás gyenge és erős kötésekre építő irodalma azonban éppen a kreativitás alapkérdését hagyja megválaszolatlanul: hogyan járulnak hozzá kapcsolataink az új ötletek keletkezéséhez? A gyenge kötések, amelyek sokféle társadalmi miliőhöz kapcsolhatnak, valóban alkalmasak arra, hogy eladdig ismeretlen ötletekhez adjanak hozzáférést, de mi magyarázza ezeknek az ötleteknek a keletkezését?

Korábbi munkáinkban erre a kérdésre válaszul alkottuk meg a strukturális gyűrődések fogalmát (De Vaan et al., 2015; Vedres–Stark, 2010). A strukturális gyűrődések egy kontextuson (például munkacsoporton) belüli átfedések kohézív csoportok között. Ezek az átfedések hozzájárulnak ahhoz, hogy a munkacsoport tagjai egyrészt felfedezzenek lehetséges új kombinációkat, másrészt (a már meglévő kohézióknak köszönhetően) meg is tudják valósítani az új kombinációkat mint új terméket, kutatási tervet, műalkotást.

A strukturális gyűrődések hipotézise átfedő csoportok esetében egyedibb termékeket feltételez, illetve sikeresebb termékeket (amennyiben a mező kulcsértékei közé tartozik a kreativitás). Átfedő csoportokat azonban nem könnyű azonosítani: ehhez egyrészt szükség van egy megfelelően nagy csoportra, ahol „van hely” két alcsoport számára. (Ha az elemi kohézív csoport három főből áll, akkor legalább ötfős csoportokra van szükség, hogy két háromfős alcsoportot azonosítsunk, amelyek egy csomóponton átfednek.) Másrészt arra van szükség, hogy a munkacsoport belső hálózata ne legyen túl sűrű, mert ebben az esetben két alcsoport nem lesz azonosítható.

Ezen kihívásokat orvosolandó, javasoltuk a tiltott triádok sűrűségét mint a strukturális gyűrődések elemi mutatóját. A tiltott triádok olyan nyitott triádok, ahol a két meglévő él súlya nagy (Granovetter, 1973). Két átfedő kohézív csoport átfedése körül számos tiltott triád található: az élék súlya nagy (azaz a kötések erősek), mivel minden él kohézív kontextusban szerepel, ugyanakkor számos hiányzó él is található (azon csomópontok között, amelyek különböző kohézív csoportba tartoznak).

A dzsesszvilág egy olyan mező, ahol a kreativitás kifejezetten központi érték. Ezt a világot vizsgáltuk meg a *Tom Lord Discography* adatai alapján (Lord, 2010), amely a dzsessz legteljesebb diszkografikus adatbázisának mondható. Ezen adatok alapján a dzsessz teljes történetét elemeztük, 1896 és 2010 között, 175 064 lemezfelvételt, 42 929 zenekartól, összesen 187 784 zenész részvételével. Minden lemezfelvétel esetében rekonstruáltuk az abban szereplő zenészek korábbi együttműködési hálózatát. A 3. ábra Miles Davis és zenésztársainak súlyozott együttműködési hálózatát mutatja a *Kind of Blue* első *session*-je idején.



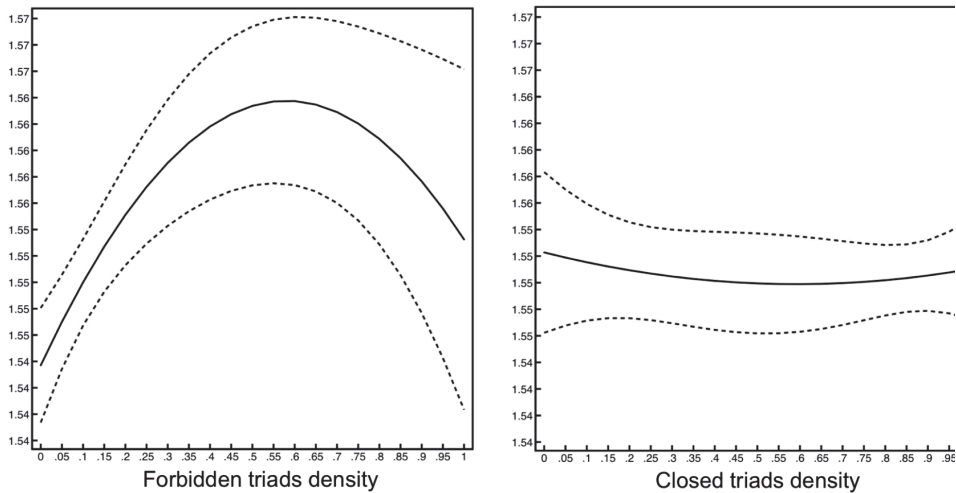
3. ábra. Miles Davis és zenésztársainak együttműködési hálózata a *Kind of Blue* felvételekor (1959. március 2.) (Vedres, 2017 nyomán)

A 3. ábrán látható hálózat meglehetősen sűrű, mégis szerepel benne egy hiányzó él: Miles Davis és Wynton Kelly korábban soha nem játszottak együtt stúdióban, holott mindkettőjüknek vannak közös és erős kötésekkkel kapcsolatos zenésztársaik. Az ábrán szaggatott vonallal jelölt triádok Davis és Kelly körül tiltott triádok.

A sikeresség és a tiltott triádok sűrűsége között pozitív kapcsolatot feltételezhetünk a dzsessz esetében, bár a túlzott sűrűség már feltétlenül jelent sikert. Túlzottan sok tiltott triád esetén már a csoport koordinációja válik nehézkessé.

A várt összefüggést statisztikai modellekkel vizsgáltuk. A sikerességet a lemezfelvétel anyagából megjelenő kiadások számával mértük. Miles Davis *Kind of Blue* felvételéből például 186 kiadás született a későbbiek során (és ezzel ez a legsikeresebb dzsessz lemezfelvétel). Mivel a kiadások száma diszkrét eseményadatnak tekinthető, és a megoszlása szignifikánsan ferdebb, mint a Pois-

son-megoszlás, negatív binomiális regressziós modellt használtunk a becsléshez. A 4. ábra mutatja az eredményeket. A részletes modellek Vedres Balázs cikkében (2017) található. Ezen az ábrán a tiltott triádok és a zárt triádok marginális hatásait tüntettük fel, mialatt az összes kontrollváltozó az átlagánál volt rögzítve.



4. ábra. A tiltott és zárt triádok sűrűsége és a siker közötti összefüggés negatív binomiális regressziós modell alapján (Vedres, 2017 nyomán)

Az eredmények arra mutatnak, hogy a tiltott triádok hozzájárulnak a sikerességhez a dzsesszben, míg a zárt triádok nem. Azok a felvételek, ahol egy zenész több összeszokott zenésztársát is szerepelteti, akik még nem játszottak együtt, sikerebbek lehetnek, mint azok a zenészek, akik ugyanabban a felállásban többször is játszottak már együtt.

A SIKER MÉRÉSE ÉS MODELLEZÉSE

Mitől függ, hogy egy tudományos dolgozat, egy művészeti alkotás vagy egy innováció sikeres lesz-e? A *teljesítmény* az egy területen kifejtett aktivitás teljes, objektív eredményeinek összessége, ami például egy tudós publikációiban vagy egy sportoló, illetve egy csapat elért csúcsaiban, helyezéseiben jut kifejezésre, és egyéni tevékenységeken alapul. Ezzel szemben a hírnév, a népszerűség, a befolyás és a láthatóság által jellemzett *siker* kollektív jelenség, ami a közösségnek az egyén vagy egy csoport teljesítményére adott reakcióját, illetve annak elfogadását testesíti meg (Uzzi, 2008; Yucesoy–Barabási, 2016; Ke et al., 2015). Ezen két jelenség között a kapcsolatot gyakran nyilvánvalónak tekintik, pedig azt távolról

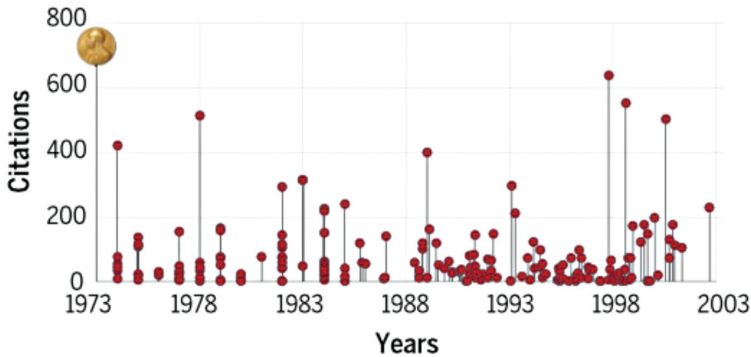
sem értjük, és gyakran ellentmondásokba ütközünk. Valóban, még a legalapvetőbb tudományos felfedezés is észrevétlen marad, ha a tudományos közösség nem ismeri el a fontosságát diszkussziók és hivatkozások formájában. A siker szociológiájának kiinduló feltételezése a teljesítmény és a siker közötti különbség felismerésén alapul, és azon, hogy kivételes hatás előrejelzése nem szükségképpen feltételezi rendkívüli képességek észlelését. Abból indulunk ki, hogy megérthetjük, kvantitatív módon leírhatjuk, és akár előre is jelezhetjük a sikert csupán a sikert vezérlő kollektív társadalmi jelenségek feltárása alapján.

Annak ellenére, hogy a probléma régóta az érdeklődés középpontjában áll, még szakemberek is rendszeresen tévednek a siker, a népszerűség és a hatás hosszú távú előrejelzésében. A siker azonban egyszer csak előre jelezhetővé válik, ha nem egyéni, hanem kollektív tevékenységnek tekintjük: ahhoz, hogy valami sikeres legyen, nem elég újnak és vonzóknak lennie, hanem széles körű egyetértésre van szükség abban, hogy dicséretre érdemes. Ha elfogadjuk a siker kollektív természetét, a hálózat- és adattudomány segítségével fel lehet tárni kvantitatív jellemzőit a számos rendelkezésre álló adatból (Clauzet et al., 2017).

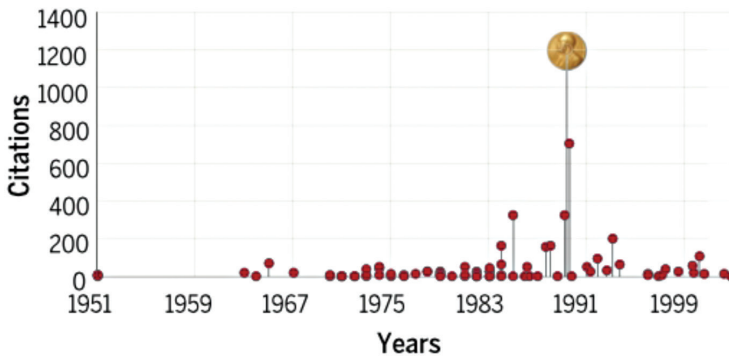
A teljesítmény és a siker dichotómiájának eredete az, hogy az emberi tevékenységek széles körében nehéz megkülönböztetni a teljesítményt és a sikert (Murray, 2003). Valóban, idézetséget, díjakat és más mértékeket használunk egy tudományos felfedezés értékének mérésére, ugyanakkor hiányoznak a teljesítmény objektív mérőszámai, amelyek egy tudományos dolgozat vagy tudós innovációs mértékét vagy tehetségét jellemeznék. Kutatásainkban elsősorban azon mechanizmusok kvantitatív leírására összpontosítottunk, amelyek a tudományos munka sikerét hajtják, mivel ezen a területen kitűnő példákat találunk azon kérdésekre, amelyeket a siker szociológiája meg tud válaszolni.

Egyik legutóbbi munkánkban az egyéni tudományos hatás fejlődését vizsgáltuk (Sinatra et al., 2016). Az emberi teljesítmények legtöbb területén, a sporttól a mérnökségig, a nagyobb eredményhez vezető út meredek tanulási szakaszból és hosszú gyakorlásból áll. A tudomány sem igazán különbözik: a kiemelkedő felfedezéseket gyakran kevésbé emlékezetes hatású dolgozatok előzik meg. Azonban annak ellenére, hogy egyre növekszik az igény az ígéretes fiatal tudósok korai azonosítására, a tudományos kiválóság kiemelkedésére jellemző pályafutás időbeli mintázatát nem sikerült felfedni. Hogyan változik a hatás és a produktivitás egy tudományos pályafutás során? Követ a hatás valamilyen előre jelezhető mintázatot? Előre tudjuk jelezni egy tudós kiemelkedő eredményének idejét? Képesek vagyunk tudományos pályafutásokat modellezni mérhető és előre jelezhető mennyiségek segítségével? Ezek a kérdések motiváltak, amikor elhatároztuk, hogy tudományos karrierük ezreinél számszerűsítjük a hatás és produktivitás fejlődését. Negyvenmillió tudományos közlemény és hivatkozásaik felhasználása tette lehetővé, hogy hét diszciplína tudósainak publikációs jegyzékét rekonstruáljuk, és azonosítsuk a dolgozatoknak a tudományos közösségre gyakorolt hosszú

távú hatását az idézettségi mutatókkal. Azt találtuk, hogy a legnagyobb hatású dolgozat egy tudós pályájának véletlenszerű időpontjában jelentkezik. Más szóval, a legnagyobb hatású munka egyforma valószínűséggel található a tudós dolgozatainak sorozatában – lehet az az első publikáció, felbukkanhat a pályafutás közepén, vagy lehet a tudós utolsó dolgozata (5. ábra).



Frank A. Wilczek
Physics Nobel,
2004



John B. Fenn
Chemistry Nobel,
2002

5. ábra. A véletlen hatás szabály. Két Nobel-díjas tudós, Frank A. Wilczek (fizikai Nobel-díj, 2004) és John B. Fenn (kémiai Nobel-díj, 2002) publikációinak története illusztrálja, hogy a legnagyobb hatású dolgozat azonos valószínűséggel lehet akárhol a tudós publikációinak sorozatában. A függőleges vonalak egy-egy tudományos cikknek felelnek meg. A vonalak hossza a cikk hatását mutatja, vagyis azt, hogy hány hivatkozást kapott a cikk tíz év után. Wilczek a Nobel-díjat a legelső publikált dolgozatáért kapta, míg Fenn a Nobel-díjjal jutalmazott cikkét karrierje kései szakaszában írta, azután, hogy nyugdíjba küldték a Yale-ről.

(Az ábra átvétel a Sinatra et al., 2016 cikkből)

Ez a véletlen hatás szabály érvényes különböző diszciplínák tudósaira, akik különböző hosszúságú életpályákat futottak be, különböző évtizedekben dolgoztak, és egyedül vagy teamben publikáltak. Ez a szabály lehetővé tette, hogy kifejlesszünk egy kvantitatív modellt, amely szisztematikusan szétválasztja a produkti-

vitás és a szerencse szerepét a tudományos karrierben. A modell minden tudóshoz egyértelműen hozzárendel egy paramétert, ami leírja a tudós képességét arra, hogy nagy hatású dolgot hozzon létre, és független a tudós életpályájának szakaszaitól. A modell analitikus kifejezéseket szolgáltat a hatás szokásos mértékeire, így a *h*-indexre, vagy a hivatkozások teljes számára, és lehetővé teszi számunkra, hogy előre jelezzük az egyes tudósok időbeli fejlődését, beleértve a független elismeréseket, mint a Nobel-díj előrejelzését is (Sinatra et al., 2016). Mindez együtt arra vezet, hogy nem tudjuk megjósolni, hogy egy karrieren belül *mikor* születik meg a legnagyobb hatású cikk, de előre tudjuk jelezni, hogy *kinek* lesz nagy hatása, és mi lesz az egyes pályafutások hatásának a fejlődése. A „szerencse” és az egyéni képességek szétválasztása, ami az egyes pályafutásokon mérhető, lehetővé teszi, hogy minden tudósra számszerűen meg tudjuk becsülni a legnagyobb hatású munkájának az *impaktját* és az *impakt* indikátorok *dinamikáját* akár évtizedekre előre.

Néhány további példa a CEU-s csoport további eredményeiből a siker szociológiájának területén: dolgozatok sikerességének hosszú távú, akár harminc évre vonatkozó előrejelzése (Wang et al., 2013), az elismerés allokálása egyes tudóshoz többszerzős cikkeknel (Shen–Barabási, 2014) az interdiszciplinaritás számszerűsítése a hatás alapján és kapcsolata tekintélyes díjakkal, mint a Nobel-díjjal (Szell–Sinatra, 2017), valamint a teljesítmény és a népszerűség kapcsolatának megértése egyes esetekben, amikor objektív mérték áll rendelkezésre mindkettőre, mint a sportban (Yucesoy–Barabási, 2016).

Mindezen megközelítések szigorú, kvantitatív keretet adnak a siker számszerűsítéséhez és előrejelzéséhez. Az így elért megértés segíthet a teljesítmény jobb megítélésében, és hozzájárulhat, hogy felismerjük a korlátait annak, ha az emberi teljesítményt a siker alapján akarjuk számszerűen megítélni. Hozzájárulhat ahhoz is, hogy korai életszakaszukban azonosítani lehessen személyeket, akik nagy teljesítmények elérésére képesek, és tanácsokkal szolgáljon a döntéshozóknak a támogatásukhoz.

ÖSSZEFOGLALÁS

A fenti példákkal betekintést kívántunk nyújtani azon kutatások egy részébe, amelyek a számítógépes társadalomtudomány keretében a CEU-n folynak. Az ötletek sikerre vitelének különböző szempontjait próbáltuk megvilágítani, a kreatív csapatok kialakítási feltételeitől az innováció terjedési mechanizmusának leírásán keresztül a siker elméletének a tudományos eredményekre történő alkalmazásáig. Nem állítjuk, hogy máris sikerült egységes képbe foglalni ezt a szerteágazó és fontos témakört, de az adatok feldolgozása egészen újszerű megközelítéseket tett lehetővé, illetve közvetlenül tesztelhetővé tette az elméleti modellezést. Ezeket az előnyöket kihasználva, a jövőben további adatokat bevonva tervezzük folytatni a

kutatásokat. A cél, hogy kvantitatív elméleti leírást tudjunk adni a folyamatokra, így lehetővé téve az előrejelzést, aminek óriási gazdasági-társadalmi haszna lehet.

KÖSZÖNETNYILVÁNÍTÁS

A kutatást az OTKA (K112713), továbbá a H2020, az AFOSR és az NSF támogatta.

IRODALOM

- Aral, S. – Van Alstyne, M. (2011): The Diversity-Bandwidth Trade-Off. *American Journal of Sociology*, 117, 90–171. DOI: 10.1086/661238, http://www.jstor.org/stable/10.1086/661238?seq=1#page_scan_tab_contents
- Barabási A-L. – Albert R. (1999): Emergence of Scaling in Random Networks. *Science*, 286, 509–512. DOI: 10.1126/science.286.5439.509, <http://barabasi.com/f/67.pdf>
- Chovell, G. et al. (2017): Perspectives on Model Forecasts of the 2014–2015 Ebola Epidemic in West Africa: Lessons and the Way Forward. *BMC Medicine*, 15, 42, DOI: 10.1186/s12916-017-0811-y, <https://bmcmmedicine.biomedcentral.com/track/pdf/10.1186/s12916-017-0811-y>
- Clauset, A. – Larremore, D. B. – Sinatra, R. (2017): Data-driven Predictions in the Science of Science. *Science*, 355, 477–480. DOI: 10.1126/science.aal4217, <http://science.sciencemag.org/content/355/6324/477.full>
- De Vaan, M. – Stark, D. – Vedres B. (2015): Game Changer: The Topology of Creativity. *American Journal of Sociology*, 120, 4, 1144–1194. http://www.personal.ceu.hu/staff/Balazs_Vedres/papers/devaan.vedres.stark.games.pdf
- Granovetter, M. (1973): The Strength of Weak Ties. *American Journal of Sociology*, 78, 6, 1360–1380. DOI: 10.1086/225469, https://www.jstor.org/stable/2776392?seq=1#page_scan_tab_contents
- Karsai M. – Iníguez, G. – Kikas, R. et al. (2016): Local Cascades Induced Global Contagion: How Heterogeneous Thresholds, Exogenous Effects and Unconcerned Behavior Govern Online Adoption Spreading. *Science Reports*, 6, 27178. DOI: 10.1038/srep27178, <https://www.nature.com/articles/srep27178>
- Ke, Q. – Ferrara, E. – Radicchi, F. – Flammini, A. (2015): Defining and Identifying Sleeping Beauties in Science. *Proceedings of the National Academy of Sciences of the USA*, 112, 7426–7431. DOI: 10.1073/pnas.1424329112, <http://www.pnas.org/content/112/24/7426>
- Lord, T. (2010): *The Jazz Discography*. <https://www.lordisco.com/>
- Murray, C. (2003): *Human Accomplishment: The Pursuit of Excellence in the Arts and Sciences, 800 BC to 1950*. New York: Harper Collins
- Obstfeld, D. (2005): Social Networks, the Tertius Iungens Orientation, and Involvement in Innovation. *Administrative Science Quarterly*, 50, 100–130. DOI: 10.2189/asqu.2005.50.1.100, https://www.jstor.org/stable/30037177?seq=1#page_scan_tab_contents
- Ruan, Z. – Iníguez, G. – Karsai M. – Kertész J. (2015): Kinetics of Social Contagion. *Physical Review Letters*, 115, 218702 DOI: 10.1103/PhysRevLett.115.218702, <https://arxiv.org/pdf/1506.00251.pdf>
- Sinatra, R. – Wang, D. – Deville, P. et al. (2016): Quantifying the Evolution of Individual Scientific Impact. *Science*, 354, aaf5239, DOI: 10.1126/science.aaf5239, https://www.researchgate.net/publication/309692079_Quantifying_the_evolution_of_individual_scientific_impact

- Shen, H. W. – Barabási A-L. (2014): Collective Credit Allocation in Science. *Proceedings of the National Academy of Sciences of the USA*, 111, 34, 12325–12330. DOI: 10.1073/pnas.1401992111, <http://www.pnas.org/content/111/34/12325>
- Szell, M. – Ma, Y. – Sinatra, R. (2018): Interdisciplinarity: A Nobel Opportunity. (Preprint)
- Török J. – Kertész J. (2017): Cascading Collapse of Online Social Networks. *Science Reports*. 7, 16743. DOI:10.1038/s41598-017-17135-1, <https://www.nature.com/articles/s41598-017-17135-1.pdf>
- Uzzi, B. (2008): A Social Network's Changing Statistical Properties and the Quality of Human Innovation. *Journal of Physics A: Mathematical and Theoretical*, 41, 224023. DOI: 10.1088/1751-8113/41/22/224023, <https://bit.ly/2KzwPj9>
- Vedres B. – Stark D. (2010): Structural Folds: Generative Disruption in Overlapping Groups. *American Journal of Sociology*, 115, 4, DOI: 10.1086/649497, <http://www.jstor.org/stable/pdf/10.1086/649497.pdf?refreqid=excelsior%3A99619439d35b35221326a91c4a44f95f>
- Vedres B. (2017): Forbidden Triads and Creative Success in Jazz: The Miles Davis Factor. *Applied Network Science*, 2, 31, <https://arxiv.org/ftp/arxiv/papers/1709/1709.03787.pdf>
- Wang, D. – Song, C. – Barabási A-L. (2013): Quantifying Long-term Scientific Impact. *Science*, 342, 127–132. DOI: 10.1126/science.1237825. <http://science.sciencemag.org/content/342/6154/127.full>
- Watts, D. (2002): A Simple Model of Global Cascades on Random Networks. *Proceedings of the National Academy of Sciences of the USA*, 99, 5766–5771. DOI: 10.1073/pnas.082090499, <http://www.pnas.org/content/99/9/5766>
- Watts, D. – Strogatz, S. H. (1998): Collective Dynamics of 'Small-world' Networks. *Nature*, 393, 440–442. DOI:10.1038/30918, https://www.researchgate.net/publication/13660517_Collective_Dynamics_of_Small_World_Networks
- Yucesoy, B. – Barabási A-L. (2016): Untangling Performance from Success. *EPJ Data Science*, 5, 17, <https://link.springer.com/article/10.1140/epjds/s13688-016-0079-z>

ÚJ MÓDSZEREK RÉGI KÉRDÉSEK MEGVÁLASZOLÁSÁRA AZ AKADÉMIAI FELHŐBEN – HÁLÓZATOK ÉS OKSÁGI KAPCSOLATOK FELDERÍTÉSE A TÁRSADALOMTUDOMÁNYOKBAN

NEW METHODS IN THE ACADEMIC CLOUD TO ANSWER OLD QUESTIONS – EXPLORING NETWORKS AND CAUSAL RELATIONS IN SOCIAL SCIENCES

Lévai Péter¹, Telcs András²

¹ az MTA rendes tagja, főigazgató, MTA Wigner Fizikai Kutatóközpont
levai.peter@wigner.mta.hu

² az MTA doktora, megbízott osztályvezető, MTA Wigner Fizikai Kutatóközpont, MTA–Pannon Egyetem Budapest
Rangsor Kutatócsoport, Budapesti Műszaki és Gazdaságtudományi Egyetem Számítástudományi és Információelméleti Tanszék
telcs.andras@wigner.mta.hu

ÖSSZEFOGLALÁS

Cikkünkben röviden bemutatjuk, milyen drámai változáson megy keresztül az informatika fejlődésének hatására a kutatók napi munkája. Részletesen kitérünk annak ismertetésére, milyen, eddig soha nem volt lehetőségek nyílnak meg az asztali és a számítógépes felhőben rendelkezésre álló kapacitások révén a társadalomtudományok (és más területek) kutatói előtt. Példaként röviden ismertetjük, hogyan járulhat hozzá az akadémiai felhő, a nagy kapacitású számítógépek alkalmazása a nemzetközi akadémiai tér, illetve rejtett oksági kapcsolatok feltárásához.

ABSTRACT

In this paper we briefly review the dramatic changes in the researchers' daily practice owing to the development of information technology. We look at it in details how desktop and cloud computational capacities open opportunities never existed before for researchers of social sciences (and practitioners of other fields as well). We present two examples how cloud computing, high performance computing can be applied in the analysis of the global academic space and how hidden causal factors can be revealed.

Kulcsszavak: Moore-törvény, akadémiai felhő, számítógéppel segített kutatás, társadalmi hálók, oksági kapcsolatok

Keywords: Moore law, academic cloud, computer-aided science, social networks, causal relations

BEVEZETÉS

Korunk társadalomtudósai életében egyre gyakoribb, hogy olyan feladatokkal találkoznak, amelyek megoldásához egyrészt a komplex rendszerek és hálózatok területén elért legfrissebb eredmények felhasználása, az ott alkalmazott új algoritmusok alkalmazása szükséges, másrészt jelentős számítási igénnyel lépnek fel, hatalmas számítógépes memóriára és hosszú számolási időre van szükség. Ilyen paradigmaváltáson korábban már átment a fizika, kémia, csillagászat, meteorológia, és ez éri el napjainkban a közgazdaság-tudományt, szociológiát és más társadalom- és humán tudományokat. Ez az új kihívás megkövetelte, hogy újragondoljuk a tudományos szakemberek számítógépes háttérének, igényeinek biztosítását. A Magyar Tudományos Akadémia ezt felismerve teremtette meg a Közép-Európában jelenleg egyedülálló akadémiai felhő szolgáltatást, amely a kutatók számára lehetővé teszi, hogy versenyben maradhassanak a legjobb informatikai háttérrel élvező társaikkal.

Cikkünkben bemutatjuk az Akadémiai Felhő nyújtotta lehetőségeket és néhány példát arra, hogyan is lép magasabb szintre egy társadalomtudományi kérdés vizsgálata és megértése az akadémiai felhő segítségével. Konkrét példaként bemutatjuk, hogyan alkalmazhatjuk ezt az újféle számítási diszciplínát az európai egyetemek rangsorolására, a rangsor finomhangolására, idősorok kauzális kapcsolatainak feltárására.

SZÁMÍTÁSI KAPACITÁSOK EGYKORON ÉS NAPJAINKBAN

Nehéz dolgunk van, ha érzékeltetni szeretnénk napjaink számítógépeinek teljesítményét. Az Olvasónak a Moore-törvény juthat először eszébe, miszerint a processzorok kapacitása másfél-kétévente megduplázódik. Hogy ez mit is jelent, azt legegyszerűbb egy személyes példán érzékeltetni. Az 1982-ben megjelent, és egy nyugat-európai magánút során, 1985-ben, leárazás után beszerzett japán CASIO FX-700P programozható számológép már pont elegendő kapacitást nyújtott egykoron, hogy egy fizikus diplomamunka alapszámolásait el lehessen rajta otthon végezni, és ezzel időt nyerjen a tulajdonosa. A CASIO-ban rendelkezésre állt egy 455 kHz-en ketyegő processzor, 2 kB memória és 12 kB tárhely. Egy évtizeddel később, 1992-ben, már egy ötvenszer gyorsabb processzort (25 MHz, azaz 25 000 kHz) és kétezerszer nagyobb memóriát és tárolóhelyet (2 MB memória és 25 MB tároló) tartalmazó személyi számítógép feküdt az asztalon, hogy egy kandidátusi dolgozat kerülhessen ki a billentyűzet alól. Ez a cikk pedig már egy olyan hordozható laptopon íródott, amelyben 2,5 GHz-es sebességgel 6 db processzormag zakatol, 8 gigabyte (GB) memória és 1 terabyte (TB) tárolóegység mellett. A harmincöt év alatt 33 ezerszeresére nőtt számítási kapacitás 2,3 évet ad

ki a Moore-törvényre, de hozzátehetjük, hogy ma már 128 vagy 256 processzormagot tartalmazó chipkártyák is vannak, úgyhogy érvényes a kétéves duplázódás törvénye. Ha pedig a memóriát és tárolókapacitás növekedését nézzük, akkor inkább másfél évenkénti duplázódás tanúi lehetünk. Ne felejtjük el, hogy mindhárom példában új áron 1500 dolláros gépekkel számoltunk.

Ha ilyen gyorsan fejlődött az informatika, akkor felmerül a kérdés, hogy szükséges-e egyáltalán követnünk ezt a teljesítménynövekedést, nem lenne-e elegendő mindaz, amit ma magunkkal hordozunk, vagy akár csak az asztalunkon tartunk? Sőt, talán célszerűbb lenne csak egy kisebb kapacitást magunknál tartani, ami éppen elég a feladataink végrehajtására. Ez ma már meg is valósult, gondoljunk csak az „okos” mobiltelefonjainkra vagy az iPad-ekre, könnyű, kisképernyős laptopjainkra, amelyeket elegendő kétnaponta feltölteni. Ma már egyes okostelefonokban négy processzor ketyeg, s ha akarjuk, akkor a kutatóberendezéseinket, mérőműszereinket is tudjuk távolról irányítani, leolvasni, újraindítani. Ma már nincs technikai akadálya annak, hogy egy-egy új ötletnek szinte azonnal utánajárjunk, egy ismert adatsort új szempontok szerint újraelemezhessünk, akár a buszon ülve is. Egy nehézség azonban jelentkezik: manapság már mindenkinek lehet okostelefonja, iPad-je, sőt laptopja is. Vagyis azon ötletek területén, amelyeket a magunkkal/magunkon hordható infokommunikációs eszközeinkkel megoldhatunk, nagyon nagy a túlekedés: eredményre éhes diákok és kutatók ezrei indulnak ugyanabból a pozícióból. Tulajdonképpen az nyer, akinek hamarabb jut eszébe egy nagyszerű ötlet, és hamarabb tudja ezt hordozható eszközein megoldani, feldolgozni. Úgy is fogalmazhatunk, hogy segédeszközeink, mankóink támogatása mellett megint a szürkeállomány válik fontossá, az ötlet, a koncentrációképesség, a fókuszáltság.

Ugyanakkor létezik egy másik irány is: ha a Moore-törvény továbbra is érvényben marad (és úgy tűnik, hogy igen), akkor a tovább növekedő számítógépes kapacitás lehetővé teszi olyan új kérdések megválaszolását, amelyekre eddig nem is mertünk gondolni, fel sem merült bennünk, mert azt éreztük, hogy ennek még nem jött el az ideje. Azokról az esetekről van szó, amikor az adatállományunk már nem terabyte nagyságú, hanem inkább ezernyi terabyte, s így inkább már petabyte-ban (PB) kell gondolkoznunk. A Big Data Science az az új terület, ahol óriási adatmennyiségeket tudunk átvizsgálni, elemezni, hogy érdekes és új eredményekre juthassunk. Jó példát mutatnak a CERN¹ részecskefizikusai, akik 5000 trillió proton-proton ütközésben begyűjtött 15 000 terabyte-nyi, azaz 15 petabyte adatot rostáltak át ahhoz, hogy háromszáz olyan eseményt találjanak, amely tisztán mutatja a Higgs-bozon megjelenését (és így kísérletileg megalapozhatták a Nobel-díj jogos odaítélését Peter Higgsnek és François Englertnek, akik ötvenöt éve vártak erre a pillanatra). A CERN-ben jelenleg kb. 200 ezer processzormag

¹CERN: European Organization for Nuclear Research, Európai Nukleáris Kutatási Szervezet.

dolgozik a részecskefizikusoknak, hogy a Higgs-bozonnál is nagyszerűbb, új felfedezések születhessenek. Ehhez még hozzájárul ugyanennyi processzormag a társintézményeknél, ahol szintén adatanalízist végeznek. Ne felejtjük el azonban, hogy ezek a processzormagok továbbra is 2,5–3,5 GHz órajel mellett működnek! Az előrelépés ott történt, hogy a gyártók egyre több magot integrálnak egyetlen chipre (már a 8, 16 vagy 24 magú CPU²-k sem ritkák), s ezekre a CPU-kra vonatkozik a Moore-törvény.

Amint láttuk, valójában nem az egyedi processzorsebesség növekszik, hanem a bonyolultsági szint emelkedik. Ami jelentősen megnöveli a megkövetelt szakértelem színvonalát is, hogy ezek az új (és meglehetősen drága) egységek optimálisan kerüljenek kihasználásra. Korábban a kutatók (például a fizikusok) végezték a személyi számítógépeik üzemeltetését, majd rábízták azt a doktoranduszaikra, érdeklődő diákjaikra. De ma már ez nem elegendő. Ma már ismét magasan iskolázott, kiválóan képzett, széles körű tapasztalatokkal rendelkező információtechnológiai szakemberek (IT-mérnökök és IT-technikusok) felügyelik a legkiválóbb rendszereket, biztosítják a felhasználók számára a megígért jelentős kapacitást – és egyúttal versenyelőnyt is biztosítanak.

Korábban az órajelek és adattovábbítási sebességek növelése területén elől álló gépek képviselték a „High Performance Capacity” (HPC) gépek igencsak szűk csoportját. Sokáig ezek a gépek uralták a TOP-500 rangsor elejét, közepét és végét. Az utóbbi években azonban kiderült, hogy a csúcskategóriás, méregdrága HPC-gépek mellett a nagyon alacsony ár/érték arányt képviselő elemekből is hatalmas, optimálisan kihasználható, úgynevezett „High Throughput Capacity” (HTC) egységeket lehet létrehozni. A HTC-egységek megjelenése lehetővé tette, hogy egy adott kérdés megválaszolásánál ne lineárisan, hanem párhuzamosan gondolkodjunk: mit nyerhetünk azzal, ha ezer, vagy akár 20–40 ezer párhuzamosan indítható lépésre bontjuk szét a megoldandó feladatot? Ha időt tudunk nyerni, akkor máris előnyösebb helyzetből indulunk!

De hogyan tegyünk szert több tízezer processzormagra? Hiszen ha gyorsan elvégezzük a feladatunkat, akkor utána nekünk már nem kell az egység. S akkor kapcsoljuk le a gépünket, bocsássuk el a szakembereinket? Mi lesz, ha néhány hónap múlva megint lesz egy szuper ötletünk, és megint szükségünk lesz, mondjuk tízezer processzormagra? Vagy csináljuk azt, hogy üresen üzemeltetjük a rendszert addig, amíg újra szükségünk lesz rá? Hogy tudjuk ezt megfizetni?

A megoldást a felhő (Cloud Computing) jelenti. A felhő maga egy hatalmas számítógépes potenciál több ezer processzormaggal, nem okvetlenül egyetlen helyszínen. A tárterület petabyte nagyságrendű, legtöbbször mágneses szalagra író-olvasó egységgel kiegészítve, hogy az időlegesen nem használt állományok ne foglalják a drága diszkterületet. Meg kell oldani, hogy a felhőt használó *userek*

² CPU: central processing unit, központi feldolgozó egység, a köznyelvben processzor.

úgy érezzék magukat, mintha a „saját” gépükbe léptek volna be, hogy a felhőt bármikor és bárhol elérhessék, és ne vesszen el semmi. Ha kevesebb kapacitás is elég, akkor csökkenteni tudják a számukra elkülönített erőforrás nagyságát. Ha pedig néhány napra nagyobb kapacításra van szükségük, akkor megkaphassák azt – előbb vagy utóbb. Napjainkban így születnek újjá a régi feladványok a felhőkben, így kereshetünk új feladványokra még újabb megoldásokat.

A WIGNER-FELHŐ ÉS AZ AKADÉMIAI FELHŐ

Az MTA Wigner Fizikai Kutatóközpont (FK) néhány éve már szembesült azzal a problémával, hogy a megoldandó feladatok bonyolultsága egyre nőtt, egyre nagyobb számítógépes kapacításra volt szükség, a kutató kollégáknak egyre több ideje, energiája ment el a technikai háttér biztosítására. Mire sikerült pénzügyi forrást szerezni, majd megvásárolni és beállítani az új számítógépeket, addigra az eredeti feladat sok esetben már el is vesztette fontosságát – jó esetben találtak a kutatók egy új feladatot, amelynél a kutatócsoport által kiépített rendszer felhasználható volt.

Amikor az intézet megépítette a Wigner Adatközpontot, és 2013 januárjában elindult a CERN Kutatóközponttal való együttműködés, lehetőség nyílt arra, hogy modern megoldás szülessen az intézményi szintű problémákra. Először is a Wigner munkatársai a CERN-es kollégákkal együttműködve nagy gyakorlatot szereztek a GRID-hálózatban összekötött számítógépek üzemeltetésében. A CERN számítógépeinek közel fele, mintegy 80 ezer mag és 80 PB tárolókapacitás segíti Budapestről a CERN TIER-0 központi rendszerének működését, szorosan összekapcsolva a genfi székhelyen található 120 ezer maggal.

Először létrejött a különálló Wigner-felhő, amely kb. 1000 processzormaggal és 0,5 PB tárolókapacítással állt a Wigner FK munkatársainak rendelkezésére. A 2015-ös próbaév után, 2016 elejétől a Wigner-felhő már üzemszerűen működött.

Az így szerzett tapasztalatok és a CERN *know-how* alapján a Wigner vezetősége az MTA SZTAKI-val együttműködve meggyőzte az MTA bizottságait és vezetőit, hogy az elszigetelt fejlesztések helyett MTA-szinten egy akadémiai felhő kialakítása lenne a megoldás, hogy a kutatóhálózatban tevékenykedő tudósok megelőzhessék a gazdagabb infrastruktúrával rendelkező országok kutatóit, vagy legalábbis felzárkózhassanak hozzájuk. A SZTAKI-ban és a Wigner Adatközpontban kiépített kapacitás 2016-ban sikeresen teljesítette a próbaidejét, és 2016. októbertől az MTA közösségének rendelkezésére áll a szintén 1000 processzormagból és közel 1 PB tárterületből álló kapacitás. Az akadémiai felhő jelenleg egyedülálló szolgáltatás Közép- és Kelet-Európa kutatási környezetében.

Az akadémiai felhőről részletesebb információk találhatóak az URL1 internetes oldalon. Az oldalon közel negyven projekt összefoglalója is megtalálható, ezzel segítve az érdeklődők kezdeti tájékozódását (URL2).

Az alábbiakban az egyetemi rangsorok elemzése és a kauzális kapcsolatok feltárása szolgál arra példaként, hogy a nagyméretű számítási kapacitás miként teszi lehetővé egy régi probléma újszerű vizsgálatát, új megoldások felfedezését. Ezzel szeretnénk érzékeltetni, hogy a nagy számítógépes kapacitás felhasználása ma már nemcsak a fizikusok, agykutatók és meteorológusok igénye, fontos eszköze, hanem igen hasznos lehet számos különböző társadalomtudományi területen is.

EGYETEMI RANGSOROK

Az egyetemi rangsorok elég hosszú ideje már a közbeszéd, a szakmai vizsgálatok és tudományos kutatás tárgyai. A rangsorok kialakítására irányuló törekvések mögött többek között az az igény húzódik, hogy a társadalom visszajelzést kapjon arról, hogy a felsőoktatási intézmények (továbbiakban: egyetemek) mennyire töltik be a velük szemben támasztott társadalmi elvárásokat, milyen szerepet játszanak a társadalom fejlődésében. Ennek az elvárásnak a tudományos igényű kielégítésére írt ki tematikus kutatócsoportos pályázatot a Magyar Tudományos Akadémia 2017-ben, melynek elnyerése után a Pannon Egyetemen alakult meg a csoport (Budapest Ranking Research Group, BRRG) jelen cikk második szerzője, Telcs András vezetésével.

A csoport munkatársai a hallgatók jelentkezési adatai alapján már korábban is vizsgálták a magyar egyetemek vonzerejét (Telcs et al., 2013). Évente nagyjából százezer diák körülbelül négyszázezer jelentkezési lapot nyújt be, megadva milyen sorrendben preferálja az egyetemeket (karokat, szakokat), amelyekre jelentkezik. A kutatás kiindulásául szolgáló adatbázis tizenöt év rekordjait öleli fel. Ez az adatmennyiség már olyan nagy, hogy egy erősebb személyi számítógép (PC) segítségével is legfeljebb csak egyszerű statisztikákat készíthetünk belőle. A csoport azonban mélyebb elemzéseket is el kívánt végezni, amelyek PC-n megvalósítva soknapos futtatást jelentettek volna. A memória- és teljesítménykorlátokat figyelembe véve, az eredeti szándékhoz képest erősen redukált feladatokat tűztek ki, de még ezek végrehajtása is egy-két napot vett igénybe a rendelkezésre álló számítógépeken.

Ilyen feladat volt a jelentkezések alapján a weblapok rangsorolásánál is alkalmazott PageRank módszer, genetikus algoritmus vagy kimerítő keresés megvalósítása. Ezek, az egyetemek közötti páros összehasonlítást tartalmazó, hálózaton elvégzendő számítások igen erőforrás-, illetve időigényesek. Hasonlóan gépet próbáló feladat a hálózat aggregálása és klaszterezése, a kistérségi preferenciák időbeli és térbeli alakulásának bemutatása (Kosztyán et al., 2015). A vizsgálat közben sok ábra készült, amelyek számítógépes megjelenítése sokat segített a jelenségek megértésében, ugyanakkor további memóriát és kapacitást igényelt hardver oldalról.

Minden eddiginél jobb megoldást kínál a csoport számára az akadémiai felhő és a Wigner-intézet partnerénél, a CERN-ben Jean-Marie Le Goff csoportja által kifejlesztett Collaborativ Spotting (CollSpot, URL3) rendszer. Ezen keretek között tervezzük kialakítani az Academic Space Explorer (ASE) hálózatmegjelenítő és -elemző eszköz modelljét. A BRRG a CollSpotot a rangsorkutatás során három módon is hasznosítani tudja.

Az első, nagyon fontos előnye a nagy mennyiségű adat hálózatos tárolása. A CollSpot a szokásos „Excel-táblázatos”, illetve relációs adatbázis helyett az adatokat eleve csúcsok és közöttük futó élek formájában, azaz gráfszemléletben tárolja. A csúcsok és élek attribútumokkal láthatók el. Az így kapott gráf csúcsai objektumosztályokat alkothatnak, amelyek igény szerint klaszterezhetőek, aggregálhatóak, különböző hálózatelemzési eljárásokat lehet elvégezni rajtuk. Az egyetemkutatáshoz ez a gráfszemléletű adattárolás igen szerencsés. Természetes módon adódik, hogy az egyetemet mint csúcsokat ábrázoljuk, és a szokásos indikátorok lehetnek ezek attribútumai. Ilyen indikátor például a hallgatói létszám vagy az egyetem, illetve a kar kutatói által írt publikációk, a maguk szakterületi besorolásával és a megjelenés adataival (a folyóirat és annak impakt faktora, a megjelenés éve stb.). Már ez a példa is jól mutatja a tárolandó adatmennyiség nagyságát és az adatok között meglévő és megjelenítendő hatalmas és bonyolult kapcsolatrendszert. Mint említettük, a CollSpot lehetővé teszi nemcsak a csúcsok, de a kapcsolatokat reprezentáló élek definiálását is, szintén a maguk attribútumaival. Ez a funkció kiválóan alkalmazható lesz az egyetemi publikációk idézettségi kapcsolatainak a tárolására azok jellemzőivel együtt. Ehhez hasonlóan az ERASMUS-program keretében létrejött diák-, illetve oktatói és dolgozói utazások is jól reprezentálhatóak a hálózat szemléletű adatbázisban.

Mekkora adatmennyiségről is van szó? Ha csak az európai egyetemekre szorítunk, az is közel kétezer intézményt jelent. Ezek a vizsgálatba vont tizenöt év alatt évi sokmillió publikációt hoznak létre, azokra tízmilliónyi hivatkozás keletkezik. Hasonló módon évente sok ezer utazás jön létre az ERASMUS keretében. A publikációk, a hivatkozások és az utazások is számos jellemzőjükkel kell hogy tárolásra kerüljenek, hogy csak egyet említsünk: a szakterülettel. Ahhoz, hogy a kutatás során a csoport tagjai az aktuális feladatnak megfelelően gyorsan, egymás zavarása nélkül hozzáférjenek mindezen adatokhoz, megfelelő tároló szerverre van szükség. Ugyanez igaz a mintarendszer működtetésére is. Amennyiben ez az akadémiai felhőben valósul meg, a tárolás és használat költsége töredéke egy saját szerver beszerzési és főleg üzemeltetési költségének.

A hálózatos adattárolás mellett a kutatók egy, az elemzést támogató eszköz is kapnak a CollSpot révén. A rendszer a tárolt hálózatot képes megjeleníteni, azon kívánság szerint műveleteket végrehajtani. A CollSpot számos megjelenítést készen ajánl fel, és továbbiakat is be tud fogadni. A hálózatmegjelenítés egyes opciói menüből érhetők el, másokat magán a megjelenített hálózaton navigálva

lehet kezdeményezni (szép demonstráció nézhető meg a Collaboration Spotting honlapján, URL4).

A hálózatot transzformálhatjuk, aggregálhatjuk, klaszterezhetjük, egyes részeit kiemelhetjük, és áttekinthetjük időbeli fejlődését. Természetesen a rendszer a kiválasztott objektumokat a kért attribútum szerint sorba is rendezi és megjeleníti, ami már az egyetemi rangsorok létrehozásához visz közelebb bennünket.

Végezetül, a kidolgozott új elemzési, megjelenítési és rangsorolási eredmények a rendszerbe beépíthetők, és remekül mutatathatók be. Hogy csak egy példát említsünk az utóbbira: az egyetemek közötti idézettségi kapcsolatokból kiindulva általános vagy akár szakterületspecifikus, a tudományos teljesítmény szerinti rangsor készíthető a PageRank algoritmus, illetve annak megfelelő általánosítása segítségével.

A CallSpot tehát a nagy mennyiségű adat hálózatos formában való tárolása mellett a hálózat elemzését és bemutatását is elősegíti, mindehhez azonban, tekintettel az adatmennyiségre és az adatok közötti kapcsolatok komplexitására, erős számítógépes kapacitásra van szükség. A CallSpot kényelmes és kutatóbarát szolgáltatásai olyan, erősen számításgényes feladatokat eredményeznek, amelyek az akadémiai felhőben lényegesen hatékonyabban végezhetőek el, mint saját eszközön.

KAUZÁLIS KAPCSOLATOK FELTÁRÁSA

Az oksági kapcsolat a tudomány szent grálja. Világunk megértése oksági kapcsolatok feltárásán keresztül történik, és történelmünk kezdetéig nyúlik vissza. Valóban a papok imádsága hozza el a Nílus áradását biztosító életadó esőt, a nap és a csillagok állása határozza meg, mekkora lesz az áradás? Egy jelenség vizsgálata során alapvető a kérdés: mi okozza? Hasonlóan gyakori kérdés, hogy a vizsgált jelenség egyik, illetve másik komponense az ok, illetve az okozó-e. Arisztotelész négy oksági kapcsolatot határozott meg. Francis Bacon kettéválasztja a gondolkodást fizikára és metafizikára, lehetővé téve a természettudományos oksági kapcsolat definiálását és vizsgálatát. A modern természet- és társadalomtudomány az ilyen jellegű kérdéseket már igyekszik minél racionálisabban, azaz adatokra alapozva megválaszolni. Norbert Wiener majd Clive Granger (vö. URL5) vezették be a statisztikus kauzalitás fogalmát, amely két alapelven nyugszik:

1. az ok megelőzi az okozatot, és
2. az ok figyelembevételével az okozat jobban jósolható, mint nélküle.

Granger maga az idősorok ARIMA-modelljére dolgozta ki a nevét viselő kauzalitás vizsgálatát. Ezt számos általánosítás, illetve alternatív módszer kidolgozása követte, amelyeket a nemlineáris, entrópiaalapú elemzések széles családjába lehet besorolni.

A kauzalitásról beszélve mindenképpen kiemelendőnek tartjuk a Floris Takens időeltolásos beágyazási elméletén (Takens, 1981) alapuló, George Sugihara által kidolgozott „converging cross mapping” módszert (Sugihara et al., 2012). Sajnos, abban az esetben, amikor a Granger- vagy Sugihara-módszer által kimutatott ok-sági kapcsolat áll fenn, e módszerek segítségével nem lehet kizárni azt, hogy az A és B jelenségeknek létezik egy rejtett közös oka is.

Somogyvári Zoltán és Telcs András csoportjukkal olyan új módszert dolgoztak ki, amely képes a rejtett ok meglétét kimutatni annak közvetlen megfigyelése, megfigyelhetősége nélkül. A módszer kiindulópontja Takens elmélete. Ennek lényege, hogy a megfigyelt idősorokból külön-külön és együttes megfigyelésükből, időeltolásos beágyazással egy-egy geometriai alakzat, sokaság készíthető. Ennek dimenziója jellemzi az idősor információtartalmát. Ha A, B a két megfigyelt idősor, jelölje dimenziójukat $d(A)$, $d(B)$ és az együttesüket $d(A,B)$. Az alábbi táblázat dimenziók közötti kapcsolatok és a kauzális összefüggések ekvivalenciáit tartalmazza.

dimenziók kapcsolata	kauzális kapcsolat
$d(A) < d(B) = d(A,B)$	A okozza B-t
$d(B) < d(A) = d(A,B)$	B okozza A-t
$d(A) = d(B) = d(A,B)$	A és B kölcsönösen okozzák egymást
$d(A), d(B) < d(A,B) = d(A)+d(B)$	A és B független
$d(A), d(B) < d(A,B) < d(A)+d(B)$	A-nak és B-nek közös oka van

A módszer mintavételi adatokra épül, ezért inherensen statisztikus természetű, a belőle levonható következtetés is az. Ennek megfelelően a módszer a fenti öt esethez bayesi gondolatmenettel valószínűségeket rendel.

Eltételezve az adatok előfeldolgozásától a módszer alkalmazásához egy legalább ötdimenziós paraméterterben kell a jó paramétereket meghatározni. Ez csak a módszer egyenkénti újrafuttatásával lehetséges, ami nyilvánvalóan sokszoros párhuzamos futtatással valósítható meg ideálisan. A módszer alkalmazására álljon itt egy egyszerű példa.

Tekintsük a New York-i tőzsdén forgalmazott részvények kellően hosszú idősorait. Döntsük el melyek a „fontos” papírok, amelyek mozgása „okozza” a többi árváltozását. A feladat igen hasonló a sokváltozós regresszióelemzésből ismert modellválasztási feladathoz. A feladat megoldása során egyes változókra vagy változócsoportokra újra és újra le kell futtatni az elemzést. A feladattól függően n idősor esetén ez a szám n -től akár 2^n -ig változhat. Nyilvánvaló, hogy az utóbbi értékhez közeli esetekben a számítás párhuzamosítás nélkül elképzelhetetlen.

Párhuzamosítással, az akadémiai felhőben a feladat tíz-tizenöt értékpapírra még észszerű idő alatt elvégezhető.

Példáink rávilágítanak arra, hogy új tudományos eredmények eléréséhez nemcsak a természet- és élettudományok területén, hanem a társadalomtudományok területén is szükséges a nagy számítógépes kapacitás és tárolás. Az akadémiai felhő nyújtotta lehetőségek kiaknázása lényegesen javíthatja a nemzetközileg is versenyképes eredmények létrehozását. Az akadémiai felhő, valamint a Wigner Kutatóközpont és annak Komputációs Tudományok Osztálya nyitott a társadalomtudományok művelői felé. Célunk az, hogy igényeikhez, kutatási problémáikhoz igazodó szakmai és erőforrás-támogatást adjunk, ezzel is hozzájárulva kutatásuk sikeréhez.

IRODALOM

- Kosztzán Zs. T. – Telcs A. – Török Á. (2015): Felsőoktatásba jelentkezők preferenciáinak térbeli és időbeli szerkezete, teljesítményfüggése. *Statisztikai Szemle*, 93, 10. 917–942. http://www.ksh.hu/statszemle_archive/2015/2015_10/2015_10_917.pdf
- Telcs A. – Kosztzán Zs. – Török Á. (2013): Hallgatói preferencia-sorrendek készítése az egyetemi jelentkezések alapján. *Közgazdasági szemle*, LX, március, 290–317. <http://www.kszemle.hu/tartalom/letoltes.php?id=1371>
- Sugihara, G. – May, R. – Ye, H. et al. (2012): Detecting Causality in Complex Ecosystems. *Science*, 338, 6106, 496–500. DOI: 10.1126/science.1227079, https://www.researchgate.net/publication/230895543_Detecting_Causality_in_Complex_Ecosystems
- Takens, F. (1981): Detecting Strange Attractors in Turbulence. *Lecture Notes in Mathematics*, 898, 1, 366–381. <http://www.crcv.ucf.edu/gauss/info/Takens.pdf>

URL1: <https://cloud.mta.hu/>

URL2: <https://cloud.mta.hu/projektek>

URL3: <http://collspotting.web.cern.ch/>

URL4: <http://collspotting.web.cern.ch/sites/collspotting.web.cern.ch/files/HTML/maps/technogram.html#technogram.json>

URL5: http://www.scholarpedia.org/article/Granger_causality

ADATKEZELÉS – A (KUTATÁSI) ADATOK KEZELÉSE A MAGYAR TUDOMÁNYOS ÉS MEMÓRIAINTÉZMÉNYEKBEN

DATA HANDLING – RESEARCH DATA MANAGEMENT IN HUNGARIAN SCIENTIFIC AND MEMORY INSTITUTIONS

Kovács László

osztályvezető, MTA SZTAKI Elosztott Rendszerek Osztály
laszlo.kovacs@sztaki.mta.hu

ÖSSZEFOGLALÁS

Az MTA Könyvtár és Információs Központ, az MTA Társadalomtudományi Kutatóközpont Kutatási Dokumentációs Központja, valamint az MTA Számítástechnikai és Automatizálási Kutatóintézet *Kutatási adatok kezelése* címmel műhelykonferenciát rendezett 2015. május 14-én. A műhelykonferencia főbb, csak kéziratban rendelkezésre álló megállapításai mentén mutatjuk be a kutatási adatok kezelésének néhány kérdését, a kutatási szféra elmúlt húsz éve adatkezelési gyakorlatának jellemző vonásait. A tudomány és a kulturális örökség szféráiban lévő (memória) intézmények hasonló adatkezelési problémái és gyakorlata miatt megállapításaink többsége a memóriaintézményekre is érvényes lehet.

ABSTRACT

Some institutions of MTA – the Hungarian Academy of Sciences – as the MTA KIK, the Library and Information Centre, the MTA TK, the Centre for Social Sciences and the MTA SZTAKI, the Institute for Computer Science and Control organized a workshop on 14 May 2015 under the title of *Handling Research Data*. Actual issues of research data handling and the major characteristics of data management practice of the last 20 years of the research sphere are presented here via the yet unpublished notes of the workshop. Similarity between problems and practices of data handling within the research and the cultural heritage spheres leads us to conclusions that may be valid for the memory institutions as well.

Kulcsszavak: kutatási adatok digitális kezelése, DMP, adatkezelési terv, repozitórium, adatsiló, digitalizálás, FAIR-követelmények, adatinfrastruktúra, tartalominfrastruktúra, kapcsolt adatok, hosszú távú digitális megőrzés

Keywords: digital research data handling and management, DMP, Data Management Plan, repository, data silo, digitalisation, FAIR requirements, data infrastructure, content infrastructure, linked-data, long term digital preservation

BEVEZETÉS

Az adatvezérelt tudományos kutatási tevékenységek széles körű elterjedése, a tudományok adatigényének drasztikus növekedése oda vezetett, hogy az adatkezelés és feldolgozás a különféle diszciplínákban szignifikánssá, több diszciplína esetében pedig a kutatási tevékenység alapvető meghatározójává vált.

Az MTA Könyvtár és Információs Központ, az MTA Társadalomtudományi Kutatóközpont Kutatási Dokumentációs Központja, valamint az MTA Számítástechnikai és Automatizálási Kutatóintézet *Kutatási adatok kezelése* című műhelykonferenciájának (Kovács et al., 2015) részt vevői közös problémákat és tendenciákat véltek felfedezni a kutatási adatok magyar kezelésével kapcsolatosan.

TUDOMÁNYOS ADAT KELETKEZÉSE

A kutatási és memóriaintézményekben jelentős mennyiségű (tudományos) adat keletkezik. A kezelendő adatok között a nyers adatoktól kezdve a különféle feldolgozottságú adatokon keresztül egészen a publikálásra kerülő adatállományokig mindenfajta és -féle adat megtalálható. Az adatállományok mérete, milyensége, formátuma, többek között, függ az egyes tudományágakban járatos adatkezelési szokásrendszerektől, a konkrét kutatási tevékenységektől és/vagy az alkalmazott mérő és regisztráló rendszerektől.

Az elmúlt hús évben jelentős elmozdulás volt tapasztalható az analóg adatoktól a már eleve digitálisan keletkező (born digital) adatok felé, a digitális adatkezelési igény manapság már szinte egyeduralmukodóvá vált világszerte.

DIGITALIZÁLÁS

Az igény a korábban analóg formában rögzített adatok digitalizálására ugyanakkor nem csökkent, a digitalizálás folyamata eddig, bár különféle intenzitással, de folyamatos volt, és e folyamat mind a mai napig nem fejeződött be hazánkban. Ennek elsődleges oka a digitalizálás mint tevékenység támogatásának hazai forráshiánya, de mint később látni fogjuk, a digitalizáláshoz szorosan kapcsolódó egyéb tevékenységek, a nem elégséges munkaráfordítás és a szakmai hozzáértés hiányosságai is visszatartják a digitalizálás folyamatát.

A digitalizálás motivációja tekintetében a szakmai közvélekedés három szempontot szokott említeni: az értékmegőrzést (A), a digitális adatkezelés egyszerűségét és olcsóbb voltát (B), valamint a digitális adatterjesztés univerzális lehetőségét (C).

A) Az analóg formátumú adatok, pontosabban az analóg adathordozók (például filmszalagok) kezelését az idő előrehaladtával egyre nehezebb fenntartani, a fizikai romlás látható jeleinek prognosztizálható hatására az adatkezelők előszeretettel választják a digitalizálást mint adatmentési, adatmegőrzési mechanizmust. Teszik mindezt abban a (hamis) tudatban, hogy a digitális térben az adatok megőrzése egyszerűbb, olcsóbb, hatékonyabb.

Mint később látni fogjuk, ez távolról sincs így, e (hamis) tudatot elsősorban a mindennapi életünket lehetővé tévő, mindenütt jelen lévő digitális számítástechnikai és kommunikációs eszközök milliárdjainak megléte és mindennapos használata hozza létre, ugyanakkor megalapozott informatikai/gazdasági megfontolások kevésbé játszanak itt szerepet.

B) A digitális adatkezelés finanszírozhatóságát lényegesen befolyásolja az informatikai piaci verseny árcsökkenő hatása, ugyanakkor a közbeszerzések pontosan az ellentétes mozgásokat indukálják. A szférában kezelendő adatmennyiség exponenciális növekedése (eltekintve a *Big Data*-tárolás és -feldolgozás speciális problémakörétől), az adathordozók, tárolók áralakulása révén, időben meglepően állandó finanszírozási igényt mutat.

C) A digitális adatok és információk átvitele, szállítása nagyságrendekkel olcsóbbá vált, köszönhetően az új kommunikációs lehetőségeknek és kiépült kapacitásoknak. A web általánossá válása a világméretű hálózaton keresztüli adatterjesztést forradalmasította. A digitalizálás és a digitális entitások interneten történő közzététele valóban jelentős motivációs erő, mely előmozdítja a digitalizálás folyamatát.

ADATKÖZLÉS WEBEN KERESZTÜL

Az állagmegőrzés, illetve az univerzális digitális terjesztés motiválta magyar digitalizálási gyakorlatok során azonban legtöbbször nem vették figyelembe azt, hogy mekkora a digitális információk felhasználásának tervezett kiterjedése, a potenciális felhasználói bázis prognosztizálható mérete, és az adat- és információfelhasználásnak milyen tovaryűrűző hatása keletkezik a tudományban magában vagy a gazdaság, az oktatás stb. szféráiban.

Hatáselemzések hiányában számos kielégítetlen adatigény mellett jelentős számosságú adatközlő mű kapott finanszírozást (web-honlap, -portál formájában) a 2000-es években. E webszolgáltatok létrehozása egyrésztől valóban megteremtette az univerzális és globális adathozzáférést hálózaton keresztül, ugyanakkor katasztrofális következményekkel járt, ha ezen webszolgáltatok fenntarthatóságát tekintjük.

A digitális információk webes (például honlapon keresztüli) közlése az informatikailag sokszor alulképzett döntéshozók számára a digitális adattárolás és

adatközlés együttes, egy csapásra történő megoldásának potenciális lehetőségét jelentette. A modern honlap-/portáltechnológiák (a korai honlapoktól eltérően) ténylegesen tartalmaznak adattárolásra szolgáló adatbázis-kezelőt, hozzáférést adatbázis-kezelőkhöz és ezzel párhuzamosan, a webes kiszolgálásra szolgáló adatközlő modulokat is. A beépített adatbázis-kezelők azonban egyértelműen a webszolgáltatás sajátos, informatikamotiválta igényeit szolgálták/-ják ki, és csak igen kevéssé a tárolt adatok kezelésének, tisztításának, megőrzésének stb. céljait. Egyszóval az adatközlést, nem pedig az adatkezelést támogatják.

A portáltechnológiák alkalmazásának illetően félreértése nemzetgazdasági szinten vezetett veszteségekhez, ugyanis a fent nevezett honlapalapú projektek elkészülte után pár évvel azzal a problémával szembesültek a fenntartók, hogy a webvilág technológiai változási sebessége az egyik leggyorsabb a világon.

A portálok esetében három-öt évenként biztosan cserélni-módosítani-javítani szükséges a meghajtó portálmotorokat, különben a webszolgálat használhatósága szignifikánsan romolhat, vagy szélsőséges esetben az adatok hálózati elérhetősége akár meg is szűnhet. A webböngészők folyamatos változása, a felhasználói felületek újabb kiszolgáló technológiái és az új interakciós lehetőségek fejlődése (és néha persze a technológiai divat) állandó fejlesztési készenlétet és/vagy beavatkozást igényel. Mindez folyamatos és szignifikáns finanszírozási feladatot jelent.

Összefoglalóan az történt, hogy a digitális adatkezelés/tárolás/közvetítés webes kontextusba helyezésével egyidejűleg egy folyamatos technológiakövetési feladatot is magukra vettek a döntéshozók anélkül, hogy ennek műszaki feltételrendszerét, szervezeti és munkaigényét, finanszírozhatóságát előre látták és biztosíthatták volna. A webszolgálatok fenntarthatatlansága pedig nem csupán az adatelérhetőséget, de magát az adatok létének fennmaradását is veszélyeztette, vagy ténylegesen meg is szüntette, jelentős nemzetgazdasági károkat okozva ezzel.

A tudományos adatok és információk (mint értelmezett adatok) webes közlése tehát nemhogy megoldotta volna az analóg adattárolás során felmerülő problémákat, de egy újabb problémával tetézte azokat. Ebből a nézőpontból az elkésett magyar fejlődésnek, a digitalizálásban észlelhető, a fejlett országokhoz képesti lemaradásunknak (mely kb. tíz-tizenöt évre becsülhető) kicsiny pozitív hozadéka lehet (ha az analóg adathordozók fizikai állapota, illetve tervezett élettartama megengedi) e csapdahelyzet (részleges) kikerülése és a 2010-es évek elején-közepén kezdődő, szakmailag igenelhető adatkezelési gyakorlat művelése.

REPOZITÓRIUMOK MAGYARORSZÁGON

A digitális repozitóriumok magyarországi elterjedésének kezdete a 2010-es évek elejére datálható, ami szintén legalább egy évtizednyi lemaradást jelent a fejlettebb országok gyakorlatához képest. A repozitóriumok felállítására vonatkozó

döntések a közvetlen felhasználási igényeken túlmenően többek között azon alapulnak, hogy az adatkezelés, az adatkuráció műszakilag, technológiailag és tevékenység szempontjából így célszerűen elválasztható az adatközlési feladatoktól. Adattárolók, repozitóriumok, digitális könyvtárak és archívumok létesítésével egyidejűleg az adatkezelési feladatok (tárolás, tisztítás, válogatás, hosszú távú megőrzés stb.) funkcionális szétválasztása és az eltérő technológia támogatása így megvalósulhat. A repozitóriumokban a tárolt adatokhoz társított metaadatok sémája egyértelműsíthető, jól meghatározható, ami által a későbbi visszakereshetőség és a (szemantikusan) helyes felhasználhatóság alapja teremődik meg.

Magyarországon a digitális repozitórium fogalmán egy, elsősorban a digitális dokumentumok (például publikációk) avagy multimédia entitások (hang- és videóanyag stb.) tárolására szolgáló rendszert értenek, a repozitórium sokkal kevésbé jelent (ma még) adattároló (data repository) rendszert. Míg a magyar tudományosságban ténylegesen kiépült a publikációk nyílt hozzáférését előíró MTA-rendelet indukálta országos repozitóriumhálózat (az MTA REAL és a kutatóintézetekhez, egyetemekhez kötött intézményi repozitóriumok hálózati rendszere) (URL1), addig szándékosan és célzottan a csupán tudományos adattárolás célú intézményi repozitóriumok létrehozására még csak az első lépések történtek meg.

Létrejtek az első, célzottan kutatási adatokat tartalmazó hazai adatrepozitóriumok: MTA KRTK Adatbank (URL2), MTA TK KDK (Micsik–Gárdos, 2014), (URL3). A kutatók és adatállományok azonosítására alkalmas azonosítók elérhetők Magyarországról is: ORCID (URL4), DOI (URL5). A REAL alkalmas DOI-val ellátható adatok elhelyezésére is, bár erre még kevés a konkrét példa. Az MTMT (URL6) adathivatkozást is képes kezelni. Létrejött a hazai repozitóriumokban tárolt entitások közös országos keresője (URL7).

Az adatpublikálás problémáiról és az azzal párhuzamos Open Data mozgalomról itt és most nem ejtünk több szót, megtették azt már mások korábban (Micsik–Gárdos, 2014; Holl, 2015, 2016), csak a repozitórium rendszerek létrehozását és fenntarthatóságát elemezzük.

A repozitóriumokat – Magyarországon is – sokkal inkább intézmények, mintsem kutató közösségek hozzák létre és tartják fenn; elsősorban a saját munkatár-saik kiszolgálását célozva. A tárolók létrehozásához szükséges források megszerzése, a tárolók fenntartása intézményi keretek között tervezhető, és a hosszú távú fennmaradás esélye is nagyobb.

A magyar intézményi repozitóriumok és azok országos hálózatának stabilitását segíti elő a repozitóriumok minősítési rendszerének bevezetése (URL8), mely (nagyon helyesen) leginkább a repozitóriumok hosszabb távú fenntartásának feltételrendszerét kéri számon.

Határozott előrelépés lehetne az országban egy, a REAL-hoz hasonló, de célzottan adattárolásra alkalmas, országosan megszervezett adatrepozitórium- és/

vagy adatsilóhálózat felállítása, valamint a projektek befejeztével a kutatási adatállományok megkövetelt elhelyezése ezekben a silókban, legalább a közpénzen finanszírozott kutatások esetében. E koncepció azonban pontosításra szorul a legújabb technológiai fejlemények hatására.

ADATSILÓK LÉTREHOZÁSA

Az individuális, egy adott intézményhez kötött, egy-egy adatállományt biztosító adattárolók, repozitóriumok mellett az adatsilók megjelenése okozott nagyobb változást. Az adatsiló definíció szerint leginkább az intézmények közötti, ha tetszik intézményfüggetlen adattárolás céljait szolgáló rendszer, amelynek fenntartása akkor is biztosított, ha az intézmények megszűnnek, átalakulnak, avagy belső infrastrukturális változtatás miatt a korábban üzemeltetett intézményi repozitórium működése esetleg veszélybe kerül. Ezzel ellentétben az adatsilók fenntartása legtöbbször nagyvállalatok, egyetemek, kutatóintézetek vagy azok kisebb részlegei hatáskörébe tartoznak, és ezért nem szükségszerűen biztosítják az intézményfüggetlenségi elvárást.

MULTIDISZCIPLINÁRIS ADATSILÓK

Az adatsilók tartalmuk szerint lehetnek egy tudományágot támogatók vagy multidiszciplinárisak. A tudományban az adatsilók létrehozásának gyakorlata eltér az egyes diszciplínák között. Diszciplínához kötött, tematikusan homogén adatsilók mellett a multidiszciplináris adatsilók megjelenése új szintre emeli az adatfelhasználhatóságot, ugyanis az az inhomogén, eltérő sémákkal rendelkező, tematikusan, diszciplinárisan eltérő adatok adattársításának lehetőségét segíti elő.

A multidiszciplináris adatsiló biztosítani tudja a mostanában kiemelkedő adat-tudomány (data science) analitikai eszközeinek hatékony felhasználhatóságát, közvetlen és egyszerű hozzáférést engedve e társítható, esetleg eltérő sémákkal rendelkező adatállományokhoz. Mint később látni fogjuk, az adattársítás direkt és közvetlen módszerei ma még beláthatatlan kihatással kecsegtetnek, és a tudomány globális fejlődésének újabb forradalmát eredményezhetik.

ADATKEZELÉSI GYAKORLAT

Világszerte jelentős eltérés tapasztalható az egyes kutatási intézmények, sőt azon belül az egyes kutatási projektek adatkezelésével, annak céljaival, minőségével, szervezettségével kapcsolatosan. A különbségek leginkább a kis és közepes adat-

állományok kezelése tekintetében jelentősek, míg a nagytömegű adatkezelés (Big Data), a jól kialakult adatkezelési szokások és az adatkezelés, -felhasználás és adatanalízis nagy, infrastrukturális, időben stabil komplex rendszereinek nyomatéka miatt (például elemi részek fizikája, CERN adatkezelése) jobban szervezett. A különbségek másik, el nem hanyagolható része az eltérő diszciplínákban járatos eltérő adatkezelési eljárásokból és technológiákból származik.

Az adatkezelés funkcionális tartalma a tudományban legtöbbször a következőket jelenti: az adatok forrásainak és az adatelérés módozatainak meghatározása, interfészek, kommunikációs formák és kommunikációs szolgálatok, technológiák és rendszerek, hozzáférési kommunikációs protokollok (például OAI-PMH (URL15) adataratási logikák) meghatározása. Az adatokon végzett transzformációs műveletek összessége, így például az adatok begyűjtése, felvétele, rögzítése, rendszerezése, szűrése, válogatása, tárolása, megváltoztatása, formátum transzformációi, az adatok felhasználása, importálása, exportálása, szállítása, továbbítása, nyilvánosságra hozatala, publikálása. Ide tartozik még az adatok összekapcsolása, zárolása, törlése és megsemmisítése, az adatok további felhasználásának biztonságos megakadályozása. Az adatvesztés kiküszöbölése és az adatokhoz való hozzáférés szabályozása, az adatvédelem ugyancsak integráns része e funkcionális fogalmi kiterjedésnek.

Mint látjuk a tudományos adatok kezelése rendkívül összetett és szerteágazó feladatot jelent, és mint ilyen, digitális adatkezelési szakértelmet követel meg. E szakértelem elméleti alapjait az informatika- és a könyvtártudomány, gyakorlati megvalósulását pedig az informatika gyakorlata teremti meg. E szakértelem tehát vagy magukban a kutatókban, vagy a kutatók és informatikusok/digitális könyvtárosok kooperációjában testesülhet meg.

A műhelykonferencia résztvevőinek szinte egyöntetű véleménye alapján e téren jelenleg nagyfokú szakértelemhiány mutatkozik a magyar tudományosság szinte minden szférájában és diszciplínájában. A kutatók informatikusokkal való együttműködési igényének kielégületlenségét jelen sorok szerzője saját, az MTA kutatóintézeti hálózatából származó közvetlen tapasztalataival tudja megerősíteni.

A modern digitális adatkezelési informatikai technológiák, módszerek, gyakorlatok ismeretének hiánya megdöbbentően elmaradott adatkezelési, adattárolási rendszerek meglétét és aktuális használatát jelenti Magyarországon. Tapasztalatunk alapján a magyar bölcsészet- és társadalomtudományok művelői jelentős hátrányban vannak ilyen tekintetben a természet- és műszaki tudományok művelőihez képest. A magyar memóriaintézmények jelenlegi helyzete pedig tragikus, ezen intézményekben mindenhol (minőségi) informatikushiány mutatkozik. Mindez az alkalmazott adatkezelési rendszerek minőségében, technológiai fejlettségében csapódik le, pontosabban a nem alkalmazott fejlett technológiák és a *state-of-the-art* ismeretének hiányában.

ADATKEZELÉS KÖLTSÉGE

A magyar kutatási projektek előkészítése során rendszerint hiányzik a projektek adatkezelésének megtervezése, így általában nem tervezik ennek költségeit sem. Az adatkezelési feladatokat legtöbbször a kutatók végzik, az (informatikai) eszközök beszerzésének terve mellett az explicit adatkezelési költség- és tevékenységbecslés ritka.

Jelentős probléma, hogy hiányzik a rendszerek és bennük az adatszolgáltatások hosszú távú, a projekt befejezése utáni fenntartási költségeinek tervezése is. A magyar finanszírozó szervezetek legtöbbször nem gondolkoznak a projekt futamidején túlmenően, alig követelik meg a kutatási projektek utáni követési, fenntartási, hasznosítási feladatok keretében az adatállományok túlélésének biztosítását és persze e feladatok reális finanszírozásának megteremtését és/vagy támogatását. Ennek súlyos következménye a projektekben keletkező vagy az ott kezelt (és ezáltal jelentős mennyiségű élőmunkát, értéket hordozó) adatállományok, adatszolgáltatások továbbélésének, hasznosulásának, egyáltalán fennmaradásának veszélyeztetése.

Hiányzik egy, legalább ágazati szintű elvárásrendszer, adatkezelési stratégia, szabályzat, útmutatás arról, hogy hogyan óvjuk meg ezeket az adatállományokat a projektek befejezte után. Nemzetközi pályázatoknál (például Horizon 2020) ugyanakkor elterjedt az adatkezelési terv (DMP – Data Management Plan) megkövetelése. Ilyen projektek esetében az adatkezelési terv a projektek során begyűjtött, feldolgozott és/vagy létrehozott kutatási adatok kezelésének teljes életciklusára ki kell hogy terjedjen, információkat biztosítva a FAIR-követelmények tervezett megvalósulásáról.

FAIR-KÖVETELMÉNYEK

A FAIR-követelmények (Findable, Accessible, Interoperable and Re-usable) a kutatási adatok projektek futása közbeni és utáni megtalálhatóságát, a széles körű hozzáférés biztosítását, az adatok csereszabotosságát, illetve az újrahasonosításhoz szükséges feltételek és metaadatok meglétét követeli meg (Wilkinson–Dumontier, 2016). Mindez azt jelenti, hogy az adatkezelési tervben egyértelműen definiálni kell azt, hogy a projekt során milyen adatokat gyűjtenek be, azokat hogyan, milyen módszerrel dolgozzák fel, valamint milyen új vagy származtatott adatok keletkeznek a projekt során. Az adatformátum szabványoknak való megfelelés, a használt szabványok egyértelmű meghatározásán túlmenően az adatkezelési terv tartalmazza az adatok közzétételének, megosztásának tervezett módozatait, a hozzáférés biztosításának módszereit, jogi, műszaki, szervezeti feltételeit. Információ szükséges arról is, hogy a kutatási

adatok feldolgozási folyamata során milyen adatkurátori (válogatás, szűrés, aggregáció stb.) munkát terveznek végezni, és a projekt befejezte után hogyan fogják az adatokat megőrizni és/vagy újrahasznosítani, különféle időintervallumokat feltételezve, és azt, hogy az újrahasznosítást milyen származási/nyomkövetés jellegű metaadatok (provenance metadata) támogatják. Az adatkezelési terv, elvárás szerint, foglalkozik az adatkezelés tervezett költségeivel, a projektek lezárulása után felmerülő, hosszú távon jelentkező fenntartási költségekkel is.

Bár az adatkezelési terv megkövetelése magyar viszonyok között nem jellemző, ugyanakkor diszciplínaspecifikus adatkezelési mintatervek kidolgozása és elterjesztése jelentősen segíthetné a jelenlegi magyar adatkezelési gyakorlat javítását.

A magyar tudomány jelenleg ugyancsak kevésbé alapozhat a tudományos adat- és tartalomkezelés olyan nagy léptékű modellkísérleteire, amelyek mint „best practice”, megfontolandó, esetleg átvehető mintaként szolgálhatnak az egyes kutatóhelyeken.

Hiányoznak vagy hiányosak a kutatási adatok és digitális tartalmak kezelését lehetővé tévő funkcionális, architekturális, technológiai, működési (és egyes esetekben üzleti) modellek és rendszerek, szabványok, szabályzatok, jogok, adat- és információszolgáltatások, regiszterek és repozitóriumok (adatbázisok, adattárak, adattárházak, digitális gyűjtemények), valamint ezek interoperábilis rendszere, tehát mindaz, amely a tudományos adatkezelés nemzeti és intézményi szintű rendszerkontextusát, tágabb értelmű infrastruktúráját adná.

ADAT- ÉS TARTALOMINFRASTRUKTÚRA

Az adatkezelés infrastrukturális megközelítése, mint új fejlemény, azon a felismerésen alapul, hogy az adatszolgáltatások egyedileg, önmagukban nem életképesek, hanem más adatszolgáltatások egymáshoz harmonikusan, informatikailag és szemantikusan is illeszkedő rendszerében, egy tervezett hálózatban tudnak csak rendesen létezni, működni. Egy könnyen átlátható példa a magyarországi névterek problematikája.

A memóriaintézmények regisztereiben az intézményközi névtérkezelés azonban már legalább húsz éve megoldatlan az országban, annak ellenére, hogy ez alatt szinte folyamatos (volt) a nevekkal, névterekkel való foglalkozás, névtérállomány-építés és -kezelés, egyéni és intézményi szinteken is.

A névtérkezeléshez szükséges, országos szinten jelentkező globális szervező, megvalósító, finanszírozó tevékenységeket eddig sem az állam, sem pedig valamilyen intézményi önszerveződés nem tudta fenntartható módon megvalósítani. Ennek következménye a memóriaintézményekben jelentkező, feleslegesen párhu-

zamos névkezelési munkák miatti erőforrás-pazarlás, egyben a névkezelés-minőség optimumának elérhetetlensége, mely gátolja az oktatási, kutatási, kulturális, sőt még a kormányzati szféra különféle tevékenységeit is.

Az Európai Unióban országokon keresztülnyúló hálózatok, hálózati infrastruktúrák támogatják a digitális tudományos adatkezelés legújabb, adatfelhő alapú megközelítéseit (például nemzetközi DARIAH-infrastruktúra a digitális bölcsészettudományok területén [URL9]), sőt azon túlmenően a tudomány elektronikus művelése, az eScience-funkciók teljes vertikumát. Magyarországon ugyanebben az időben a legfelső szinten kérdőjelezzik meg a digitális bölcsészet diszciplináris létét, az azt támogató felhőalapú digitális infrastruktúrák létesítésének szükségességét, az ilyen célú projektek támogathatóságát, lásd például (MTA BTK–MTA SZTAKI–DE–ME, 2016) GINOP-pályázat, a nemzetközi főáramtól való leszakadást indukálva e tudományok területén.

A magyar tudományban jellemző, hogy széles körben hiányzik az adatkezelés infrastrukturális megközelítése, az, hogy a digitális kutatási adatok létrehozását, feldolgozását, megtalálását, tárolását, szállítását, felhasználását, megőrzését stb. – egyszóval a digitális adat- és tartalomkezelést lehetővé tévő rendszereket – egységes digitális adattartalom-infrastruktúráknak tekintsük, és mint ilyeneket (országos vagy ágazati szinten) központilag tervezzük, létrehozunk és fenntartsuk. A nemzetközi, hasonló célú kezdeményezésekbe, infrastruktúrákba való belépésünk, csatlakozásunk, az adatkapcsolati szintű kapcsolatunk (adat import/export) csak akkor lehetséges, ha rendelkezünk ilyen célú, stabilan működő hazai adatinfrastruktúrákkal. Míg a kommunikációs és internethálózatokat a közfelfogás infrastruktúrájának tekinti, az ezekre szervesen épülő adat- és tartalominfrastruktúrák fogalma ma még nem elterjedt a közbeszédben.

Ezekben az adatinfrastruktúrákban, az adatszolgáltatások között különleges szerepet kapnak az interoperabilitást elősegítő szolgáltatások (protokollregiszterek, egymással interoperábilis névtér- és ontológiakezelők, a metaadatsémák, felhasználási profilok (application profile) regiszterei, a szótárak, szöszedetek és egyéb nyelvi szolgáltatások stb.), egyszóval azok a központi adatszolgáltatások, melyek egységes rendszerre fűzik fel az egyedi, kutatás célú adatszolgáltatásokat. Az adatszolgáltatások és azok megbízható interoperabilitásának létrehozása szabványos informatikai megoldásokat, egységes mérnöki tervezést, megvalósítást és persze szabályozást igényel.

Az ilyen nagy nemzeti infrastruktúrák ugyan létre tudnak jönni projektfinanszírozási logikával, de hosszú távú (akár több évtizedre szóló) fenntartásuk csak az erre a célra létrehozott speciális intézmények keretében, jól kidolgozott üzleti-finanszírozási és/vagy működési modellek alapján lehet reális. Ezeket az intézményeket meg kell alapítani, létre kell hozni, működtetni és persze finanszírozni kell.

ADATKEZELÉS MINT KUTATÓI ERŐFESZÍTÉS

Az adatkezelést lehetővé tévő hardver- és szoftvereszközök használata esetén felmerül az a kérdés, hogy milyen szakértelemre van szükség az adatfeldolgozási folyamatban, annak teljes életciklusában, a különféle munkafázisokban? Megvan-e a szükséges szakértelem az adatkezelést legtöbbször *de facto* végző kutatókban, illetve a kutató-informatikus, a kutató-könyvtáros együttműködési viszonylatokban? Az eltérő diszciplínák eltérő adatkezelési szokásrendszerei hogyan viszonyulnak a state-of-the-art adatkezelési lehetőségekhez? E felmerülő kérdésekre más vizsgálatok adhatnak pontos válaszokat, itt és most csak egyetlen kapcsolódó aspektusra kívánjuk felhívni a figyelmet, nevezetesen a kutató szerepére az adatfeldolgozási folyamatban.

A hivatkozott *workshop* résztvevői annak a gyakorlati tapasztalatuknak adtak hangot, hogy a képviselt diszciplínák (leginkább társadalom- és bölcsészettudományok) legtöbb kutatása esetében a kutató élő, közvetlen és napi kapcsolata a tudományos adatfeldolgozás különféle fázisaihoz elengedhetetlen. A feldolgozási folyamatban az adatkurátori munkákat ugyanis csak diszciplináris tudással bíró kutató tudja szakszerűen, a diszciplína általános és az adott kutatási projekt specifikus elvárt követelményei és céljai ismeretében elvégezni. Minőségi tudományos adatok előállítás és kezelése tehát a kutató közvetlen hatása és munkája nélkül elképzelhetetlen.

Amennyiben a kutató napi tevékenységének szignifikáns, netalántán túlnyomó részét az adatkurátori és persze az ehhez társuló klasszikus adatfeldolgozási (gyakorlati) munka alkotja (ez ma már nem csupán a memóriaintézményekben, de a kutatóhelyeken is megjelenő jelenség), akkor a kutatói lét megalapozását és általános teljesítménymérését jelenleg kvázi egyetlen paraméterben mérő publikációs tevékenység csorbát szenvedhet. Leginkább azért, mert a (magyar) tudományosságban nem alakult ki a tudományos adatközlés és adatpublikálás, tudományon belüli, jutalmazási és elismerési rendszere. A minőségi tudományos adatok publikálása mint olyan nem érvényesíthető tudományos teljesítményként, sem egyénileg, sem pedig intézményi szinten. Az adatvezérelt tudományok jelenlegi felívelő korszakában ez nyilvánvalóan felülvizsgálendő kérdés, melyet a végletekig feszít majd az újabb keletű publikációs formák (nanopublikáció [URL10], kapcsolt adatok [URL12] publikálása) elterjedése.

A korrekt adatpublikálás az adat-újrafelhasználás előfeltétele. Az adatpublikálás helyes végrehajtásához a publikálásra kerülő adatállományokhoz részletes eredetinformációk társítása (provenance metaadatok) szükséges. E metaadatok előállítása *de facto* multi- és interdiszciplináris szaktudást követel meg. Az adateredet és -feldolgozás történetiségének metaadatulása ugyanis, a forrástól a végfelhasználásig nem csupán a mérő és regisztráló eszközök működésének és beállított paramétereinek ismeretét (műszaki ismeretek) és pontos regisztrálását,

az adatfeldolgozási lépések és (informatikai) módszerek feltárását és rögzítését, de a korrekt (könyvtárosi, archiválási) metaadatolást, osztályozást stb. is igényli, vagyis egy interdiszciplináris szaktudást, melyben a diszciplináris tudás és a fenti szakterületeken történő jártasság együttes jelenléte elengedhetetlen.

Az adatfeldolgozási folyamat legtöbbször mint önálló rész nem választható el mechanikusan a kutatási *workflow*-tól. Vannak olyan diszciplínák (például digitális bölcsészet) ahol az adatfeldolgozási workflow informatikai és szaktudományi ismeretrendszere annyira összeolvad, hogy szétválasztásuk nem is lehetséges, vagyis a kutatónak egy személyben kell ismernie az informatikai és a szak- (példánk esetében bölcsészet) tudományi ismereteket lényegében teljes mélységben. A kutatási és az adatfeldolgozási workflow egyetlen integráns egységet képez.

Mindkét esetben (adatpublikációt előállító kutató és az integrált kutatási-adatfeldolgozás workflow alapján dolgozó kutató esetében) a publikálás jelenlegi nehézségei elrettenthetik a kutatót e területektől, mely visszavetheti a tudomány fejlődését ezen a ma még marginálisnak látszó, de a fősodor felé navigáló területen.

KAPCSOLT ADATOK

A géppel értelmezhető és automatikusan feldolgozható adatok világméretű hálózatának, a szemantikus web gondolatának mint vízióknak korai felvetése (Berners-Lee, 1998), a kialakított szemantikus web-architektúra és technológiai szabványrendszer rendkívüli komplexitása nem tette vonzóvá az idea gyakorlati megvalósulását, és mindeddig nem vált a fősodor részévé. Az utóbbi évek leglátványosabb fejlődése azonban a kapcsolt adatok (linked data) kvázi mint 'a szemantikus web' megjelenése és rapid elterjedése (Bizer et al., 2009).

Kapcsolt adatok esetén explicit ábrázolásra kerül (URL11) két adatentitás közötti reláció, mely valamilyen tudásközösségben közmegegyezéssel létrehozott, ontológiákkal pontosan meghatározott viszony. A tudásábrázolás ilyen atomi szintű megközelítése teszi lehetővé azt, hogy a pusztán adatok kezelése és/vagy publikálása az adott tudásközösségen, diszciplínán túlmenő adatfelhasználás esetén is szemantikusan helyesen történhessen meg, egyben a technológia egyszerűsége a gyors és széles körű elterjedés alapfeltétele. Ennek a folyamatnak vagyunk manapság tanúi.

Míg korábban egy adatállomány publikálásánál az adatállományhoz társított származási (provenance) adatok hordozták leginkább az adatok értelmezési keretét és kontextusát, mely legtöbbször csak az adott diszciplína, tudásközösség kutatói/tagjai számára volt pontosan értelmezhető, addig a kapcsolt adatok publikálása során minden egyes adatrészecske szemantikája pontosan megadható nyilvánosan elérhetővé tett ontológiák segítségével. Tehát egy alacsonyabb szintű granuláció és explicit tudásábrázolás váltja fel a korábbi nagy léptékű tudás-

granulációt és implicit, hallgatólagos tudáskövetelményt. A közösségi, közmeg-
egyezett tudás (ontológia) explicit ábrázolása és hálózati publikálása, valamint a
közösség kutatási adatainak ugyanilyen módú, nyílt (például Linked Open Data)
nyilvánosságra hozatala, valamint e kettő szerves kapcsolatának hosszú távú,
mechanikus fenntartása biztosíték arra, hogy a közösség tudása egyrészt be-
épülhet a tudomány egészébe, másrészt úgy épülhet be (szemantikusan interp-
retálva), ahogy azt a közösségi tudáslétrehozás/felhalmozás során a hozzáértők
feltárták.

A kutatási adatok interoperábilis, multi- és interdiszciplináris felhasználásá-
nak az alapfeltétele, a helyes értelmezés lehetőségének univerzalitása valósul meg
ezáltal, mely új utakat nyit meg az adatok idegen diszciplínákban történő helyes
felhasználására, egyben egy tágabb kontextusban a korábban inkább elveszni lát-
szó univerzális globális tudásközösség újrafelépíthetőségét alapozhatja meg.

Manapság a tudástárolás/tudásmegosztás felhőalapú technológiai terjednek.
A korábbi adatrepozitórium, adatsiló megközelítés intézményi szinten ugyan
megmarad, de egy olyan informatikai adatmegosztó rendszer-réteg mögé kerül
elrejtésre, mely a silókban tárolt adatállományokból közvetlenül konvertál kap-
csolt adatokat, és teszi azt elérhetővé az interneten például SPARQL (URL13)
nyelvű keresőfelület segítségével. A SPARQL-kereső működésének folyamatos
fenntartása révén a külső adatfelhasználók számára mindez úgy jelentkezik,
hogy a siló adatállománya állandóan rendelkezésre áll egy alacsony granulációs
szinten, mintegy virtuális adatfelhőt létrehozva az interneten. Az adatfelhő ada-
tain, akár következtető (szoftver) gépek segítségével, bonyolult (logikai) adat-
feldolgozások hajthatók végre, új felismeréseket, új adatfelhasználási eseteket
hozva létre.

Az adatfelhő az elektronikus tudományművelés egyik fontos infrastrukturális
alapeleme, melyhez társítva az adatok feldolgozását, valamint a kutatók minden-
napos tevékenységét, kommunikációját, kollaborációját stb. támogató szoftver-
eszközöket és hálózati szolgáltatókat, feltehetően a tudomány elektronikus műve-
lésének 21. századi új dinamikáját hozza el.

HOSSZÚ TÁVÚ DIGITÁLIS MEGŐRZÉS

A digitális adatok hosszú távú megőrzése korunk egyik égető kérdése. A hosz-
szú távú megőrzés egyrészt az adatállományok (és persze adathordozók) fizikai
megőrzését, másrészt a digitális objektumok eredeti szemantikájának, az ada-
tok korrekt értelmezhetőségének a hosszú távú (100+ évre vonatkozó) megőrzését
jelenti. A probléma forrása itt is az informatika gyors fejlődése, ahogy azt már
korábban láttuk, ez a web világában is együtt járt a technológiai fejlődés követé-
sének problémájával.

Ez esetben a digitális állományok formátumának gyors avulásával kell megküzdeni. A korábbi szoftververziók által létrehozott fájlformátumok nem használhatóak hosszú távon. Egy idő után a korábbi fájlformátumok fenntartása nem válik lehetségessé vagy kívánatossá. A (szoftver-) rendszerek fejlődése ugyanis magával hozza azt is, hogy a régi adatformátumok már nem adnak elegendő lehetőséget az új, összetettebb értelmezési keretek ábrázolására, ezért új, gazdagabb adatformátumokat definiálnak, és kezdenek el használni szélesebb körben. A mérő- és regisztráló eszközök technológiai fejlődése, szofisztikáltságának fokozódása is ez irányba mutat. A régi fájlformátumok avulása az adatállományok elvesztésének rémével fenyeget. Ha az adatállomány tulajdonosai és/vagy felhasználói nem lépnek időben, akkor egy idő után, még ha az adatállomány fizikailag rendelkezésre áll is, megfelelő szoftver hiányában az nem vagy csak korlátozottan lesz értelmezhető/felhasználható.

A hosszú távú digitális megőrzés egy, a gyakorlatban használható megoldását, a megőrzés rendszer- és tevékenységmodelljének kidolgozását és szabványosítását tűzte ki célul az OAIS, később ISO szabvány (URL14). A megőrzési folyamat lényeges részei a következők. Figyelni szükséges (obszervatórium-modell segítségével) az adott közösségben alkalmazott fájlformátumok aktuális használatát, és amint egy-egy adott fájlformátum használata kezd leáldozni a közösségben (vagy akár globálisan) akkor az archívumokban, az abban a formátumban tárolt fájlokat vagy transzformálni (migrálni) kell az újabb keletű fájlformátumokba, vagy pedig a formátumokat helyesen értelmező szoftverrendszerek túlélését kell valamilyen, például emulációs technikával biztosítani. Bármelyik módozatot is választjuk, az aktív, legtöbbször élőmunkával társuló beavatkozást igényel, és mint ilyen, jelentősen erőforrás-igényes. A fájlformátum-transzformációk elvégzése vagy az emulátorok programozása egy-egy nagyobb adatrepozitórium, adatsiló esetében ráadásul jelentős időt is vehet igénybe, és mint ilyen, előzetesen tervezni és finanszírozni szükséges.

A hosszú távú digitális megőrzés, beleértve a fizikai megőrzés folyamatát is, erőforrás-igényessége miatt, szokás szerint, a felhalmozott digitális állományok szűrésével, válogatásával, selejtezésével csökkenti a megőrzésre kerülő állományok számát, méretét. Az adatsелеjtezés felelősségteljes tevékenysége ugyancsak nem lehetséges (inter)diszciplináris tudás nélkül.

A digitálisan keletkezett tudományos adatok hosszú távú megőrzésének problémafelvetése hazánkban még csak most kezdődött el, csak néhány korai kezdeményezésről, projektről van tudásunk (lásd az Országos Levéltár, az MTA SZTAKI működő, hosszú távú tárolói, az Országos Széchényi Könyvtár rekonstrukciós projektjének céljai). Így a jelen hazai helyzet a meglévő digitális kutatási adatok jelentős mennyisége elvesztésének rémével fenyeget, akár már középtávon (tízéves távlatban) is.

ZÁRSZÓ

Európai kontextusban ezenfelül olyan kérdésekkel kellene foglalkoznunk, mint az Open Science és/vagy a Science 2.0 adatkezelési trendjei, az RDA (Research Data Alliance) és tevékenysége egy globális adatinfrastruktúra felé, a nyílt adatinfrastruktúrák létrehozásának, az adatmegosztáson túlmenően a kutatási workflow megosztásának kérdései, a kutatási adatok metaadatolásának részletei, az adathivatkozások módozatai és szabványos megoldási javaslatok, a kutatási adatok becsomagolásának módszertana, a kutatási objektumok létrehozásának, kutatási kontextus felismerési/tárolási képességének, a kutatási objektumok felhasználásának módozatai, az adatkarbantartás automatikus lehetőségei és újabb technológiái, vagy akár a kutatási adatok nyílt hozzáféréseinek hatása a tudomány művelésének egészére.

E rövid cikk azonban csak e korábbi, hiánypótló hazai műhelykonferencia hiányzó beszámolójának egyfajta utólagos pótlására vállalkozhatott.

IRODALOM

- Berners-Lee, T. (1998): *Semantic Web Road Map*. September, <https://www.w3.org/DesignIssues/Semantic.html>
- Bizer, Ch. – Heath, T. – Berners-Lee, T. (2009): Linked Data – The Story So Far. *International Journal on Semantic Web and Information Systems*, 5, 3, 1–22. DOI – 10.4018/jswis.2009081901, <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>
- Holl A. (2015): Kutatási adatok kezelésének nemzetközi trendjei. *Tudományos és Műszaki Tájékoztatás*, 62, 5, 177–180. http://real.mtak.hu/24531/1/201505Holl_cikk_TMT.pdf
- Holl A. (2016): Tudományos kommunikáció a XXI. században – Open Science. *Magyar Tudomány*, 177, 3, 307–316. <http://www.matud.iif.hu/2016/03/08.htm>
- Kovács L. – Gárdos J. – Holl A. (2015): *Kutatási adatok kezelése az MTA intézményeiben*. Memorandum. Verzió: 0.76, 2015. június 9. Kézirat
- Micsik A. – Gárdos J. (2014): Tudományos repozitóriumok az MTA-ban: a KDK és a SZTAKI tanulságai. In: *Informatika a felsőoktatásban 2014*. Debreceni Egyetem Informatikai Kar, <http://real.mtak.hu/25200/1/if2014micsikgardosdkd.pdf>
- MTA BTK – MTA SZTAKI – DE – ME (2016): *Nemzeti digitális bölcsészeti kiválósági központ*, GINOP 2.3.3-15-2 pályázat. Kézirat
- Wilkinson, M. D. – Dumontier, M. et al. (2016): The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data* 3, Article number: 160018, DOI:10.1038/sdata.2016.18, <https://www.nature.com/articles/sdata201618>

URL1: Repository of the Academy's Library <http://real.mtak.hu>

URL2: MTA KRTK Adatbank <http://adatbank.krtk.mta.hu/nyito>

URL3: MTA TK KDK Repository <http://openarchive.tk.mta.hu>

URL4: ORCID <https://orcid.org>

URL5: DOI <https://www.doi.org>

URL6: Magyar Tudományos Művek Tára <https://www.mtmt.hu>

- URL7: Repozitóriumi Közös Kereső <http://oakereso.sztaki.hu/kereso/index.php?type=0>
- URL8: Repozitóriumminősítő Szakbizottság <https://www.mtmt.hu/repozitoriumminosito-szakbizottsag>
- URL9: DARIAH – Digital Research Infrastructure for the Arts and Humanities <http://www.dariah.eu>
- URL10: Nanopub.org <http://nanopub.org/wordpress/>
- URL11: RDF Resource Description Framework <https://www.w3.org/RDF/>
- URL12: Linked Data <https://www.w3.org/standards/semanticweb/data>
- URL13: SPARQL Query Language for RDF <https://www.w3.org/TR/rdf-sparql-query/>
- URL14: CCSDS – OAIS model (2012) <https://public.ccsds.org/pubs/650x0m2.pdf>
- URL15: OAI-PMH – Open Archives Initiative Protocol for Metadata Harvesting <https://www.openarchives.org/pmh/>

MAGYAR NYELVTECHNOLÓGIAI INFRASTRUKTÚRA A TÁRSADALOMTUDOMÁNYOK SZOLGÁLATÁBAN

HUNGARIAN LANGUAGE TECHNOLOGY INFRASTRUCTURE FOR THE SOCIAL SCIENCES

Simon Eszter¹, Váradi Tamás²

¹tudományos főmunkatárs, MTA Nyelvtudományi Intézet
simon.eszter@nytud.mta.hu

²tudományos főmunkatárs, MTA Nyelvtudományi Intézet
varadi.tamas@nytud.mta.hu

ÖSSZEFOGLALÁS

Tanulmányunkban bemutatjuk, hogy mivel foglalkozik a nyelvtechnológia, és hogy melyek azok a területek, ahol – jellemzően nagyobb rendszerekbe beépítve – nyelvtechnológiai eszközöket használunk. Ismertetjük az *e-magyar* szövegfeldolgozó eszközláncot, amelyet hatékonyan tudnak használni a társadalomtudományok területén tevékenykedő kutatók is, illetve bárki, aki magyar nyelvű szöveget szeretne feldolgozni további kutatási céljainak megfelelően. A nyelvtechnológiai fejlesztések használhatóságát három alkalmazás bemutatásával igyekszünk megvilágítani. Hangsúlyozzuk, hogy az MTA Nyelvtudományi Intézet koordinálásával, a HunCLARIN égisze alatt a magyar nyelvtechnológia kész és képes hatékonyan segíteni a társadalomtudományi kutatásokat.

ABSTRACT

The present paper aims to introduce the field of human language technology and the areas where language technology can be usefully applied, typically as embedded in a complex system. We present the *e-magyar* language processing pipeline, which can be used with ease by researchers in social sciences and indeed by anyone who wants to process large amounts of Hungarian text material. As examples of how to use language technology tools, we describe three applications. As a take-away message, we emphasize that mobilizing the resources and services of the HunCLARIN research infrastructure the Hungarian language technology community with the coordination of the Research Institute for Linguistics of the Hungarian Academy of Sciences is ready and able to efficiently support research in the social sciences.

Kulcsszavak: nyelvtechnológia, nyelvi erőforrások, társadalomtudományok, szövegfeldolgozás, kutatási infrastruktúra, HunCLARIN

Keywords: language technology, language resources, social sciences, text processing, research infrastructure, HunCLARIN

BEVEZETÉS: MI A NYELVTECHNOLÓGIA?

A nyelvtechnológia célja, hogy a gépi rendszereket felruházza azzal a páratlan képességgel, nyelvi intelligenciával, amellyel mi emberek rendelkezünk, és amit anyanyelvünk esetében olyan rendkívüli könnyedséggel használunk. Ezt a mentális kapacitást természetesen a gépekről nem feltételezhetjük, de távlati célként erre vagy legalábbis az emberi nyelvtudás olyan szintű gépi modellálására lenne szükség, ahol a gépi rendszerek (itt nemcsak a számítógépekre, hanem egyre inkább az okostelefonokra gondolunk) az emberhez hasonló gyorsasággal, könnyedséggel és intelligenciával kezelik a nyelvet. Azaz megértik a beszédet és az írott szöveget, és azt a tárgyi tudást, amellyel rendelkeznek, természetes emberi nyelven tudják kommunikálni.

Ebben az értelemben a nyelvtechnológia a mesterséges intelligencia azon területe, ami a gépi rendszerek nyelvi készségét, intelligenciáját hivatott kiépíteni. E távoli jövőbe nyúló kutatásfejlesztési program nem jelenti azt, hogy a nyelvtechnológia a sci-fi világába tartozó terület lenne. Ellenkezőleg, bár nincs nagyon a köztudatban, a nyelvtechnológia eredményeit szinte naponta használjuk az alkalmazások egész sorában. Elég csak a helyesírás-ellenőrzést, a karakterfelismerést, a gépi szövegfelolvasást, a diktálást vagy a gépi fordítást említeni.

Bár a nyelvtechnológia nagy fejlődésen ment keresztül az utóbbi években, az emberi nyelv komplexitásának köszönhetően távolról sem tekinthető megoldottnak a nyelvfeldolgozás minden lépése. A feladat egyik fő nehézségét az adja, hogy az ember az értelmezés során számos nehezen formalizálható információt is figyelembe vesz, amelyeket egy gép számára csak korlátozott módon lehet elérhetővé tenni. Ilyenek többek között a megnyilatkozás körülményei (hol, mikor, kikkel), valamint azok többletjelentése (például: ígéret, fenyegetés), amely szintén hatással van arra, hogy hogyan értelmezünk egy üzenetet. A nyelvtechnológia feladata azonban egyelőre nem az ilyen jellegű többletinformáció figyelembevételére, hanem csakis a szövegfolyamban detektálható releváns információ adott célnak megfelelő feldolgozása. Mint minden, a nyelvben tárolt információ – bizonyos fokú – megértését magában foglaló, tehát szemantikai célkitűzéssel bíró feladat, ez is rengeteg, önmagában is kihívást jelentő részfeladatot tartalmaz.

A nyelvtechnológiai fejlesztések tipikusan nagyobb alkalmazásokba beépítve jelennek meg, segítve bennünket, amikor például információt keresünk internetes keresővel, helyesírást és nyelvtant ellenőrzünk szövegszerkesztőben, termékajánlásokat olvasunk *online* vásárláskor, egy navigációs szoftver szóbeli utasításait hallgatjuk, vagy *online* szolgáltatással fordítunk weboldalakat. A nyelvtechnológia segítségével elérhetővé válik az automatikus fordítás és tartalom-előállítás, az információfeldolgozás és a tudásmenedzsment – akár több nyelven is. Emellett elősegíti az intuitív, természetes nyelv alapú interfészek fejlesztését a háztartási

elektronika, a gépészet, a járműgyártás, a számítástechnika és a robotika területén is. Nagy lehetőségek rejlenek a nyelvtechnológiának az oktatásba és a szórakoztatóiparba, például játékokba, oktatóprogramokba, szimulációs környezetekbe való integrálásában is. A számítógéppel támogatott nyelvtanulás, az *e-learning*, az önellenőrző eszközök és a plágiumszűrő szoftverek csak kiragadott példák arra, hogy hány helyen játszik fontos szerepet a nyelvtechnológia. A közösségi alkalmazások terjedésével felmerül az igény a kifinomultabb nyelvtechnológiai alkalmazásokra is, amelyek figyelemmel követik a bejegyzéseket, összegzik a vitákat, ajánlásokat tesznek, kiszűrik az érzelmi tartalmú válaszokat, szerzői jogi szabálytalanságokat vagy visszaéléseket.

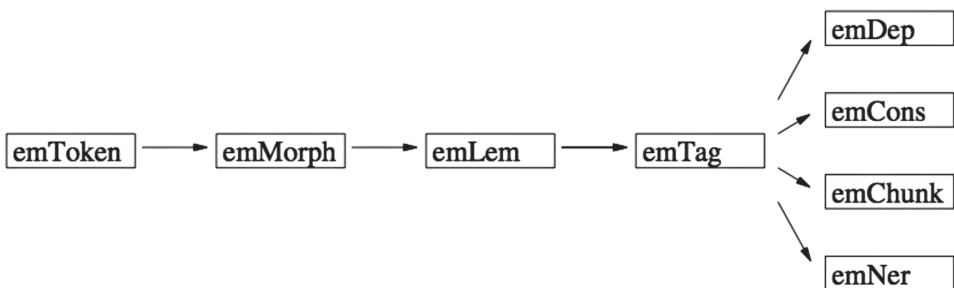
A nyelvtechnológia egyik legfontosabb módszere, hogy a nyelvi információt hordozó digitális adatfolyamokat (nyomtatott oldalak beszkenelt képét, írott szöveget, beszédet tartalmazó hangfelvételt, videót) automatikus módszerekkel feldolgozva olyan – az eredeti anyagban explicit formában nem szereplő – további információval lássa el, amely lehetővé teszi a szövegben kódolt információ, illetve tudás minél többféle szempont szerinti megtalálását (az intelligens keresést), strukturált adatbázisokba szervezését és a felhasználó számára optimális prezentációját. Az optimális prezentálás magában foglalja többek között a legrelevánsabb információ kiemelését és annak a felhasználónak leginkább megfelelő modalitásban (írás, beszéd, esetleg animált szájmozgással kísérve, jelelés stb.) és nyelven (például automatikus fordítással) való visszaadását. A nyelvi tartalmak hatékony szűrése, a lényeges információ megtalálása és kiemelése az adattengerből, a nehezen érthető információnak a felhasználó számára jobban értelmezhető formában való visszaadása alapvető fontosságú feladatok, melyek megoldásával a nyelvtechnológia nélkülözhetetlen háttér-infrastruktúrát ad a többi tudományterületnek. Az intézmények, a vállalkozások számára a (szöveges-, hangzó- és videóanyagokból automatikusan kinyert) tudásbázisok komoly versenyelőnyt jelentenek információs társadalmunkban, és az állampolgárok számára is az élet megjobbításának alapvető eszközeivé válhatnak. Kiemelkedő szerepet játszhatnak különböző hátrányos helyzetű csoportok (siketek, gyengénlátók, baleset következtében beszédképességüket elvesztők, idegen nyelveket nem tudók) életminőségének javításában.

Jelenleg a sokrétű előfeldolgozási feladatoknak a megoldása áll a magyar nyelvtechnológia homlokterében: a szöveg alkotóelemeinek azonosítása (mondatokra és szavakra bontás), különböző szó- és mondatszintű nyelvtani információk hozzárendelése (morfológiai elemzés és egyértelműsítés, szintaktikai elemzés) már megoldottnak tekinthető, de folyamatban van a mondatok közötti összefüggések felismerése, a világról szóló tudásunkat reprezentáló ontológiák építése, valamint az érzelmek detekciója is.

AZ E-MAGYAR NYELVFELDOLGOZÓ ESZKÖZLÁNC

Tanulmányunkban ismertetjük az *e-magyar* rendszert (Váradi et al., 2017), amely a fent említett alapvető szövegfeldolgozási lépéseket valósítja meg a munkában részt vevő műhelyekben eddig előállított különböző eszközök továbbfejlesztésével, egységesítésével és egyetlen koherens technológiai láncba való szervezésével. Az *e-magyar* rendszer a Magyar Tudományos Akadémia támogatásával készült a 2015-ben kiírt infrastruktúra-fejlesztési pályázat keretében. A munkálatok a pályázat kedvezményezettje, a Nyelvtudományi Intézet koordinálásával széles körű együttműködés keretében folytak, melyben részt vett a hazai nyelvtechnológia számos vezető kutatófejlesztő műhelye.

A központi gondolat az *e-magyar* kialakításában az integráció volt. A magyar nyelvtechnológiai közösség külön-külön, de az utóbbi évtizedben egyre inkább együttműködve számos erőforrást és eszközt hozott létre. Ezek a Nyelv- és Beszédtechnológiai Platform, illetve a META-NET (URL1) hálózaton keresztül is publikálásra kerültek. Az eszközök egy része nyílt forráskódú (ilyen például a *hun** eszközcsalád [URL2]), mások csak bináris formában érhetők el kutatófejlesztési célokra (ilyen például a Humor morfológiai elemző [Novák, 2003]). A mostani infrastruktúra interoperábilissá tette ezeket az eszközöket abban az értelemben, hogy az infrastruktúra egyes eszközei modulárisan egymásra épülnek, vagyis önállóan is működnek, de olyan elemzési láncba is szervezhetők, amelyben zökkenőmentesen halad az adat a különböző eszközökön át. Ez azt jelenti, hogy a nyers szövegből kiindulva az *e-magyar* szövegfeldolgozó eszközlánca elvégzi a szöveg elemeinek a szegmentálását (*emToken*), megállapítja az egyes szavak tövét és teljes morfológiai elemzését (*emMorph*, *emLem* és *emTag*), majd ezek után megadja a mondatok összetevőit (*emCons*), valamint függőségi elemzését (*emDep*); de ha csak egy sekély elemzésre van szükségünk, felismeri a mondatban szereplő frázisokat (*emChunk*), továbbá a szövegben előforduló tulajdonneveket (*emNer*). Az eszközök egymásra épülése az 1. ábrán látható.



1. ábra. Az *e-magyar* szövegfeldolgozó lánc elemeinek egymásra épülése

Fontos megemlíteni, hogy magyarra már létezik egy szövegfeldolgozó eszköz-lánc, a *magyarlanc* (Zsibrita et al., 2013), amely szintén megvalósítja ezt a moduláris architektúrát, de egy zárt rendszeren belül. Az *e-magyar* fejlesztésénél fontos szempont volt, hogy nyílt rendszer legyen, vagyis, hogy az infrastruktúra egésze és annak minden eszköze külön-külön is elérhető, letölthető, világos licenccel publikált és kutatásfejlesztési célra, de adott esetben üzleti felhasználásra is ingyenesen használható legyen.

A láncban részt vevő szoftverek licence GNU GPLv3 vagy GNU LGPLv3, a nem-szoftver elemekre vonatkozó licenc pedig Creative Commons Attribution-ShareAlike 4.0 (CC BY-SA), kivéve az *emMorph* morfológiai elemző alatt működő adatbázist, amelyre az üzleti felhasználást kizáró Creative Commons Attribution-NonCommercial-ShareAlike 4.0 (CC BY-NC-SA) licenc vonatkozik.

Az *e-magyar* nem csak a nyelvtechnológiai szakma vagy a nyelvtechnológiát használó ipari fejlesztők igényeit kívánja szolgálni. A szakmabeli felhasználók mellett támogatjuk a számítógépes eljárások iránt fogékony, de a nyelvtechnológiában nem jártas diákok és kutatók körét is: a bölcsészet- és társadalomtudományok művelőit, valamint akár az érdeklődő nagyközönséget is. Nekik két szinten kínálunk támogatást. Egyrészt egyedi igények alapján különböző elemzőláncokat állíthatnak össze az *e-magyar* eszközeinek a felhasználásával a GATE szövegelemző rendszer (Cunningham et al., 2013) keretein belül. Másrészt igénybe vehetik az *e-magyar.hu* oldal webszolgáltatását, amely rövidebb szövegek azonnali elemzését végzi.

Az eszközlánc négyféleképpen használható. A honlapon keresztül egy rövidebb szöveget egyszerűen bemásolva kipróbálhatjuk az eszközláncot. Szövegelemzéshez, társadalomtudományi vagy digitális bölcsészeti kutatáshoz a GATE-rendszer GATE Developer nevű grafikus felhasználói felületét ajánljuk, amelybe az *e-magyar* lánc egyszerűen telepíthető. A telepítés leírása a teljes rendszerrel együtt elérhető az URL3 oldalon. Itt lehetőség van a rendszer továbbfejlesztésére, vagyis az elemzőláncokhoz saját készítésű modulok is hozzáadhatók. Nagyobb korpuszok feldolgozásához a GATE parancssori hozzáférést ajánljuk, ennek használata szintén az említett honlapon található, szükséges hozzá a *github* repozitórium használata. Negyedik módszerként használatba vehető az ún. *gate-server* is, ez szintén parancssori technológia, és ez az egyébként, amely a honlap mögött is üzemel. Az integrációról és a rendszer használatáról részletesebben lásd még Sass Bálint és szerzőtársai ismertetését (2017).

A projekt célkitűzései között szerepelt, hogy az elemzőlánc hozzáférhető és érdemben használható legyen olyan felhasználók körében is, akik nem feltétlenül járatosak az informatika területén. Ennek az igénynek igyekszik megfelelni az *e-magyar.hu* webes szövegelemző szolgáltatása (URL4), amely lehetővé teszi, hogy egy webes interfészen keresztül bárki egyszerűen kipróbálhassa az egyes

elemző modulokat vagy akár a teljes elemzőláncot anélkül, hogy ehhez a böngészőn kívül bármilyen egyéb szoftvert használnia kellene.

A szövegelemző egy olyan webszolgáltatásra épül, amely a GATE-es könyvtárakat használja, bemenetként az elemzett szöveget és a futtatni kívánt elemző modulok listáját várja, kimenetként pedig a GATE által generált, az annotációkat tartalmazó XML-t adja vissza. A weboldal a visszakapott XML-t feldolgozza, és a kinyert adatokat megjeleníti egy könnyen értelmezhető, vizualizált formában. Az elemzés eredményét a felhasználó le is töltheti magának további felhasználásra.

TÖVÁBBI ERŐFORRÁSOK ÉS FELHASZNÁLÁSI LEHETŐSÉGEK

A nyelv- és beszédtechnológia területén sikerrel alkalmazható módszerek és eljárások jellegéből következik, hogy magas színvonalú kutatási eredmények és alkalmazások nem jöhetnek létre a megfelelő erőforrások, írott és beszélt nyelvi adatbázisok, alapvető sztenderdizált feldolgozó eszközök nélkül; ezek a nyelvtechnológia elengedhetetlen szükségletei a fejlesztésben és az elért eredmények kiértékelésében is. A magyarországi nyelv- és beszédtechnológia fennállása óta a különböző műhelyekben szép számú adatbázist és szövegfeldolgozó eszközt fejlesztettek ki.

Ezek a műhelyek, melyek sokáig elszigetelten működtek, az utóbbi években felismerték az együttműködés fontosságát, és annak szükségét, hogy egységes, mindenki számára elérhető és könnyen kiterjeszthető kutatási infrastruktúrákat hozzanak létre. Ennek a törekvésnek az eredménye a magyarországi nyelv- és beszédtechnológiai műhelyeket tömörítő Nyelv- és Beszédtechnológiai Platform, amely az akadémiai/egyetemi kutatásfejlesztés vezető műhelyei és ipari partnerek stratégiai szövetségéből jött létre. A platform kidolgozott egy középtávú stratégiai kutatási tervet (URL5), valamint egy megvalósítási tervet (URL6). A platform kutatásfejlesztő központjai részt vettek a META-NET kiválósági hálózatban, ahol a CESAR- (URL7) projekt keretében értékes nyelvi erőforrásokat és eszközöket tettek egységes alakban elérhetővé a META-SHARE- (URL8) hálózaton keresztül. A Nyelvtudományi Intézet egyik alapítója volt a CLARIN (URL9) európai kutatási infrastruktúra-hálózatnak. A HunCLARIN, a CLARIN magyar hálózata, összhangban a jelenleg húsz országot összefogó európai szervezet céljaival, küldetésének tekinti a bölcsészeti- és társadalomtudományi kutatások támogatását a nyelvtechnológia, a nyelvi erőforrások könnyen elérhetővé tételével.

Az MTA Nyelvtudományi Intézete számos olyan nyelvi erőforrást és nyelvfeldolgozó eszközt fejlesztett az elmúlt években, amelyek hatékonyan tudják támogatni a társadalom- és bölcsészettudományban dolgozó kutatók munkáját. Az alábbiakban bemutatunk néhány példaalkalmazási területet, társítva hozzájuk azokat az erőforrásokat, amelyek jól használhatók ezeken a területeken.

Információkinyerés és trendelemzés a munkaerőpiacon

Az elektronikus formában rendelkezésre álló feldolgozatlan állományok sok hasznos információt rejtnek, amelyekből szövegbányászati technikák segítségével statisztikák gyűjthetők, vagy következtetések vonhatók le. Emellett éppoly fontos az ontológiák, tudástárak építése is, amelyekkel a világismereti és nyelvi tudásunkat szimulálhatjuk. Ezekre támaszkodva megvalósíthatóvá válnak a jelenleginél jobb eredményt nyújtó trendelemző szoftverek, amelyek segítségével, különböző aggregált statisztikákra támaszkodva az érdeklődők reális képet kaphatnak olyan speciális területekről, mint például a munkaerőpiacon.

Az álláshirdetések és önéletrajzok számos értékes információt rejtnek, de ezek nem mindig érhetőek el strukturált formában, így gépi feldolgozásukhoz jelentős támogatást nyújthat bizonyos nyelvtechnológiai eredmények felhasználása. Szakontológiák és szinonimaszótárak beépítésével, valamint információkinyerő eszközök alkalmazásával könnyebben elérhetőek lehetnek olyan, egyes iparágakra, szakterületekre jellemző adatok, amelyek nagyban megkönnyítik az álláskereső dolgát. Többek között választ kaphatnak az elvárható fizetések nagyságáról, vagy arról, hogy melyik régióban keresett egy adott szakterület. Lehetőség nyílik a különböző végzettségek összehasonlítására, így nem utolsósorban az egymással versengő felsőoktatási intézmények rangsorolására is kiváló eszköz lehet. Másrészt a humán erőforrás-szakemberek munkáját jelentősen segítheti egy, az önéletrajzokból automatikusan felépített adatbázis, amely igen nagy mértékben megkönnyíti a megfelelő jelöltek kiválasztását.

Ehhez a feladathoz szükség van szakontológiákra és szinonimaszótárakra, melyek közül a legismertebb a WordNet (Miller, 1995). A Magyar WordNet (Miháلتz et al., 2008) egy szemantikailag strukturált, általános célú fogalomtár a magyar nyelvre, amely bizonyos szavak közötti lexikai relációkat ír le, így reprezentálva a világ bizonyos elemei között fennálló kapcsolatokat. A WordNethez hasonló szemantikai fogalomháló kinyerésére az elmúlt években elsősorban neurális hálókra épülő, ún. mély tanulást megvalósító technikákat használnak.

A magyar nyelvváltozatok feltérképezése és adatbázisba szervezése

A magyar nyelv különböző, még élő változatainak digitális rögzítése fontos feladat, hiszen a ritkább nyelvváltozatok beszélői kiöregednek, és a vidéki lakosság városokba áramlásából fakadóan egyre inkább megfigyelhető a dialektusok eltűnése, illetve a határon túli nyelvváltozatok esetében a nyelvvesztés folyamata. A magyar nyelvváltozatok kutatása túlnyomórészt nyelvtechnológiai támogatás nélkül zajlott, az elmúlt években azonban a szociolingvisztikai és szociológiai kutatások számos területén kezdődött meg az együttműködés különböző kutatócsoportok között. A nyelvtechnológiai eszközökkel segített értékőrző tevékenység nagy része

az írott szövegek digitalizálására terjed ki, de a beszélt nyelv hangzó formában való rögzítése és kereshetővé tétele is nagyon fontos. A magyar nyelv különböző nyelvváltozatainak rögzítésével digitális lenyomatot készíthetünk a már eltűnőben lévő dialektusokról, megnyitva az utat ezzel a szociolingvisztikai vizsgálatok széles spektruma előtt. A felvett anyagok hangzó és szövegesen lejegyzett változata egyaránt fontos, hiszen más és más vizsgálatok alapjául szolgálhatnak.

A Nyelvtudományi Intézet számos olyan kutatásban vett részt, amelynek céljai egybeesnek a fent megfogalmazottakkal. Az egyik ezek közül a Budapesti Szociolingvisztikai Interjú (BUSZI) (Várad, 2003), amely egy nagyszabású felmérés a magyar nyelv Budapesten beszélt változatairól. Az 1987 óta több fordulóban felvett adatokat tartalmazó adatbázis segítségével megbízható adatok és elemzések kaphatók az élő nyelvhasználat és a szociológia számos fontos kérdésére. A hanganyagok digitalizálása után került sor az anonimizálásra, a hangzó anyag szöveges lejegyzésére, a morfológiai elemzésre és egyértelműsítésre, a szóalak fonetikai reprezentációjának hozzáadására, mindennek strukturált adatbázisba szervezésére és a lekérdező rendszer kifejlesztésére, mely utóbbin keresztül az adatbázis hozzáférhetővé vált az érdeklődő kutatók számára. Ehhez és minden megelőző lépéshez szükség volt nyelvtechnológiai támogatásra, amely nélkül az információ nem lenne explicit, egyértelmű, számítógéppel egyszerűen és hatékonyan kiolvasható és feldolgozható formában tárolva, vagyis nem lenne alkalmas további kutatásokra.

Egy másik projekt, amely a magyarországi és a határon túli nyelvváltozatok feltérképezését segíti, a Magyar Nemzeti Szövegtár (MNSz) (Oravecz et al., 2014) építése volt. Az MNSz – szándékai szerint – reprezentatívan tartalmazza a mai magyar nyelv jellegzetes megnyilvánulásait. A korpusz nagyobb része magyarországi forrásokból származik, de jelentős mennyiségű szlovákiai, kárpátaljai, erdélyi és vajdasági szöveget is tartalmaz. Az MNSz lényegi tulajdonsága, hogy minden szó mellett feltünteti a szótövet, a szófajt és a szó morfológiai elemzését is. A szótó, szófaj és elemzés megállapítása és az elemzések egyértelműsítése automatikus gépi eszközökkel történt. A rendszer megbízhatósága kb. 97,5%-os, így az összes szóalak kb. 2,5%-a van hibásan elemezve. Ennél jobb eredményt csak a kézi elemzés biztosíthatna, ami ekkora méretű anyag esetén megvalósíthatatlan. Az MNSz aktuális verziója 1,04 milliárd szövegszót tartalmaz.

Pszichodiagnosztikai vizsgálatok nyelvtechnológiai támogatása

A pszichodiagnosztika és a nyelvtechnológia integráns összefüggésének logikai alapja az, hogy az egyének és a csoportok pszichológiai állapotai és folyamatai (érzelmek, gondolkodásmód, szándékok stb.) nem csupán a fizikai, hanem a verbális viselkedésben is kódolódnak. E kódok nyelvi markerek formáját öltik. Ekképp az elektronikusan rögzített kommunikáció tartalomelemzése révén az

egyének és a csoportok pszichológiai folyamatai diagnosztizálhatók, ezek időbeli változásai statisztikailag mérhetők és feltérképezhetők, továbbá az egyes egyének és csoportok egymással kvantitatíven összehasonlíthatók. A nyelvi markerek és a pszichológiai állapotok és folyamatok összefüggéseit mintegy hatvan éve kutatják. Klasszikus példa, hogy az egyes szám első személyű igék, névmások és a tagadószavak együttes túlzott használata a depresszió nyelvi tünete lehet, illetve a veszélyes küldetést teljesítő legénységek csoporton belüli konfliktusa, illetve a távoli irányító személyzettel való szembefordulása a kommunikációból előrejelezhető.

Az MTA Nyelvtudományi Intézete többéves együttműködés keretén belül segíti az MTA Pszichológiai Intézet munkatársainak kutatásait, melyeknek célja automatikus tartalomelemző módszerek fejlesztése, melyek részben kiválthatják a személyiség- és klinikai pszichológiában jelenleg használatos tesztekét, továbbá lehetővé teszik célvezérelt kiscsoportok távoli, automatikus pszichodinamikai monitorozását.

ÖSSZEFOGLALÁS

Tanulmányunkban bemutattuk, hogy mivel foglalkozik a nyelvtechnológia, és hogy melyek azok a területek, ahol – jellemzően nagyobb rendszerekbe beépítve – nyelvtechnológiai eszközöket használunk. Prezentáltuk az *e-magyar* szövegfeldolgozó eszközláncot, amelyet hatékonyan tudnak használni a társadalomtudományok területén tevékenykedő kutatók is, illetve bárki, aki magyar nyelvű szöveget szeretne feldolgozni további kutatási céljainak megfelelően. A nyelvtechnológiai fejlesztések használhatóságát három példaalkalmazással igyekeztünk megvilágítani. Hangsúlyozzuk, hogy az MTA Nyelvtudományi Intézet koordinálásával, a HunCLARIN égisze alatt a magyar nyelvtechnológia kész és képes hatékonyan segíteni a társadalomtudományi kutatásokat.

IRODALOM

- Cunningham, H. – Tablan, V. – Roberts, A. – Bontcheva, K. (2013): Getting More Out of Biomedical Documents with GATE's Full Lifecycle Open Source Text Analytics. *PLoS Computational Biology*, 9, 2: e1002854. DOI:10.1371/journal.pcbi.1002854, <http://tinyurl.com/gate-life-sci/>
- Miháltz M. – Hatvani Cs. – Kuti J. et al. (2008): Methods and Results of the Hungarian WordNet Project. In: Tanács A. – Csendes D. – Vincze V. et al. (eds.): *Proceedings of the Fourth Global WordNet Conference GWC 2008*. Szeged: University of Szeged, 310–320. http://www.inf.u-szeged.hu/projectdirs/gwc2008/GWC2008_Proceedings_Final.pdf
- Miller, G. A. (1995): *WordNet: A Lexical Database for English*. *Communications of the ACM*, 38, 11, 39–41. DOI: 10.1145/219717.219748, <http://nlp.cs.swarthmore.edu/~richardw/papers/miller1995-wordnet.pdf>

- Novák Attila (2003): Milyen a jó Humor? In: *I. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: SZTE, 138–144. http://www.morphologic.hu/downloads/publications/na/2003_mszny_Humor_na.pdf
- Oravecz Cs. – Váradí T. – Sass B. (2014): The Hungarian Gigaword Corpus. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*. Reykjavik: European Languages Resources Association. http://www.lrec-conf.org/proceedings/lrec2014/pdf/681_Paper.pdf
- Sass B. – Miháltz M. – Kundráth P. (2017): Az e-magyar rendszer GATE környezetbe integrált magyar szövegfeldolgozó eszközlánc. In: Vincze V. (szerk.): *XIII. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi Tudományegyetem. 79–90. <http://real.mtak.hu/72474/1/kotet.pdf>
- Váradí T. (2003): A Budapesti Szociolingvisztikai Interjú. In: Kiefer Ferenc (szerk.): *A magyar nyelv kézikönyve*. Budapest: Akadémiai Kiadó, 339–360. <http://www.nytud.hu/oszt/elonyelv/adat/buszi.pdf>
- Váradí T. – Simon E. – Sass B. et al. (2017): Az e-magyar digitális nyelvfeldolgozó rendszer. In: Vincze V. (szerk.): *XIII. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi Tudományegyetem, 49–60. <http://real.mtak.hu/72361/>
- Zsibrita J. – Vincze V. – Farkas R. (2013): magyarlanc: A Toolkit for Morphological and Dependency Parsing of Hungarian. In: *Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013*. Shoumen, Bulgaria: INCOMA Ltd., 763–771. <http://www.aclweb.org/anthology/R13-1099>

URL1: META-NET <http://www.meta-net.eu>

URL2: hun* eszközcslád <http://hlt.bme.hu/en/resources/hun-toolchain>

URL3: <https://github.com/dlt-rilmta/hunlp-GATE>

URL4: az *e-magyar.hu* webes szövegelemző szolgáltatása <http://e-magyar.hu/parser>

URL5: a Nyelv- és Beszédtechnológiai Platform Stratégiai Kutatási Terve nih.gov.hu/download.php?docID=19926

URL6: a Nyelv- és Beszédtechnológiai Platform Megvalósítási Terve <http://nkfih.gov.hu/download.php?docID=23049>

URL7: Central and South-East European Language Resources <http://cesar.nytud.hu>

URL8: META-SHARE <http://www.meta-share.org>

URL9: CLARIN <https://www.clarin.eu>

GÉPI TANULÁS, PREDIKCIÓ ÉS OKSÁG A KÖZGAZDASÁG-TUDOMÁNYBAN

MACHINE LEARNING, PREDICTION AND CAUSALITY IN ECONOMICS

Muraközy Balázs

PhD, tudományos főmunkatárs, MTA Közgazdaság- és Regionális Tudományi Kutatóközpont Közgazdaságtudományi Intézet
murakozy.balazs@krtk.mta.hu

ÖSSZEFOGLALÁS

Az empirikus közgazdaság-tudomány módszertanának fókuszában az oksági hatások becslése áll, miközben a gépi tanulás fő erőssége az előrejelzés vagy predikció. Az eltérő célok miatt a két módszertan nem helyettesíti automatikusan egymást. A közelmúlt kutatásai rámutatnak, hogy legalább három területen inkább kiegészítő a két módszer: i) a gépi tanulás adatokat generálhat az oksági elemzés számára; ii) számos közgazdasági kérdés valójában prediktív és nem oksági; iii) az oksági elemzés bizonyos lépései prediktívek.

ABSTRACT

Empirical economics research focuses on estimating causal effects, while the main strength of machine learning is prediction. Because of the different aims the two approaches are unlikely to substitute each other automatically. Recent research has uncovered three promising areas when these two methods may strongly complement each other: i) machine learning can generate useful information for causal analysis; ii) many questions in economics are predictive in nature; and iii) some steps of causal analysis are, in fact, predictive.

Kulcsszavak: gépi tanulás, predikció, oksági elemzés, közgazdaságtan, ökonometria

Keywords: machine learning, prediction, causal analysis, economics, econometrics

ÖKONOMETRIA ÉS GÉPI TANULÁS

Mint ez a tematikus szám rámutat, a társadalomtudományokat alapvetően változtatja meg a rendelkezésre álló adatok és számítási kapacitások minőségi növekedése. Csábító azt gondolni, hogy a Big Data és a gépi tanulás automatikusan megoldja ezeknek a tudományoknak számos ügyét: segítenek előre jelezni a közvetkező nagy válságot, megmutatják, hogy milyen szakpolitikai intézkedéseket kell bevezetni, és, végső soron, a gépek „kitanulják”, mi egy-egy iparág vagy

ország gazdasági működésének modellje. A közgazdászok munka nélkül maradnak. Ez persze nem pont így van – ebben az esszében a céloom az, hogy árnyaljam a képet.

Mivel is foglalkoznak a közgazdász kutatók? A fő cél a gazdaság működésének megértése, aminek legfőbb eszköze – gyakran formalizált – logikai modellek építése és tesztelése. Az előrejelzés vagy a gazdaságpolitikai tanácsadás ezeken a modelleken alapul. A teszteléshez kulcsfontosságú a változók közötti oksági kapcsolatok felderítése, amelyek általában egyben a modell paramétereit is. Például egy modellparaméter azt mutatja, hogy mennyivel többen tudnak elhelyezkedni egy újfajta foglalkoztatáspolitikai eszköz bevezetésének hatására, és nem azt, hogy mennyivel többen helyezkedtek el abban az évben, amikor bevezették az eszközt (és számos más dolog is történt). Az előbbi hasznos a modellteszteléshez és a szakpolitika támogatásához, az utóbbi viszont önmagában kevésbé érdekes.

Az oksági hatások kiszámításának legfőbb nehézsége a társadalomtudományban az, hogy a jelenségek ritkán vizsgálhatók kísérletekkel; így a becsléseknek megfigyelési adatokra kell támaszkodniuk. Az empirikus társadalomkutatás – és annak közgazdasági ága, az ökonometria – alapproblémáját legtöbbször éppen ez jelenti. Az orvosi kísérletekkel ellentétben nem véletlenszerű, hogy kik részesülnek egy-egy „kezelésben” (például járnak jobb iskolába vagy nyernek támogatást), hanem nagyobb valószínűséggel veszik igénybe azok, akik több hasznot remélnek belőle – más szóval önszelekcióra kerül sor. (Az itt leírt szemlélet az elmúlt egy-két évtizedben jellemzi markánsan az empirikus közgazdaságtan főáramát: korábban sokkal kisebb hangsúly helyeződött arra, hogy a becslött kapcsolatok oksági jellegűek-e.)

Ezért nem áll meg az a feltevés, hogy a kezelték (például a támogatásban részesültek) és a kontrollcsoport csak a kezelés tényében különbözne egymástól, hanem általában más jellemzőikben is eltérnek. Például könnyen lehet, hogy a támogatást kapók jobban informáltak, és támogatás nélkül is jobban boldogulnának. E miatt a két csoport egyszerű összehasonlítása egyszerre tartalmazza a kezelés hatását és az önszelekciót. Az empirikus közgazdaságtan és az ökonometria kutatói az elmúlt években számos olyan módszert dolgoztak ki, amelyek a megfigyelési adatokban olyan mintázatokat keresnek, amelyek a kísérletekhez hasonlónak, természetes vagy kvázi-kísérletként értelmezhetőek.

Ezek a megközelítések a kezelésben részesült csoport mellé valamilyen ötlettel olyan kontrollcsoportot keresnek, amelyik minden fontos (akár megfigyelhető, akár nem megfigyelhető) jellemzőjében a lehető legjobban hasonlít a kezeltre. Ez a szemlélet olyan súlyt kapott az elmúlt évtizedek empirikus közgazdasági kutatásában, hogy egyenesen egy „hihetőségi forradalomról” beszélhetünk (Angrist–Pischke, 2010).

Egy példa ezekre a kvázi-kísérleti módszerekre a szakadós regresszió (regression discontinuity), amelyet egy példán keresztül a legkönnyebb megérteni.

Egy alkalmazás például azt vizsgálta, hogyan hat az emberek jövőbeli keresetére az, ha egy adott nagy presztízsű iskolába járhatnak. Természetesen, ha egyszerűen összehasonlítjuk az iskola diákjainak a bérét azokkal, akik jelentkeztek az iskolába, de nem vették fel őket, akkor túlbecsüljük az oksági hatást, mert akiket felvettek, várhatóan eleve jobb képességűek. A szakadásos regresszió abból becsli meg a hatást, hogy összehasonlítja azokat a diákokat, akiket éppen hogy felvettek (épp a ponthatár fölött voltak) azokkal, akiket pont nem vettek fel. Ezek a diákok feltehetőleg nagyon hasonló képességekkel rendelkeznek, de csak egy részük részesült kezelésben (Abdulkadiroğlu et al., 2014). Ugyanez a kérdés megbecsülhető akkor is, ha túljelentkezés esetén sorsolással választják ki a hallgatókat, vagyis tényleg egy természetes kísérletre kerül sor (például Angrist et al., 2012).

A modern empirikus közgazdaságtan fő célja tehát az oksági hatások becslése. A gépi tanulás célja ezzel szemben a predikció, vagyis hogy a rendelkezésre álló magyarázó (vagy prediktor) változók segítségével minél pontosabban előre tudja jelezni egy eredményváltozó alakulását. A siker kritériuma az, hogy az előrejelzésre használt mintán kívül is minél pontosabban tudjon előre jelezni a modell. Az optimális modell megtalálja az egyensúlyt az alulillesztés (túl kevés változó bevonása a modellbe) és a túlillesztés (a túl sok változó nemcsak a mintát, hanem a zajt is leképezi) között. A cél a gyakorlatban általában az, hogy az eljárás a sok lehetőség közül kiválassza azokat a változókat (és a megfelelő függvényformákat), amelyek javítják az előrejelző erőt; ezt az eljárást nevezzük regularizációnak (például Hastie et al., 2009).

A gépi tanulás különösen hatékony a Big Data környezetben, vagyis olyankor, amikor nagyon sok megfigyelés és változó szerepel az adatbázisban. Egy másik lényeges jellemzője az, hogy maguk a modellek nagyon eltérő szerkezetűek, amelyek a döntési fáktól a regressziókon át a neurális hálókig terjednek. A különböző módszerek eltérő körülmények között hatékonyak, viszont ha ezekben a nagyon eltérő struktúrákban hasonló eredmény születik, az igencsak meggyőző.

Az alapvető célok és megközelítések eltérése több lényeges különbséghez vezet az ökonometria és a gépi tanulás között. Miközben az ökonometriai modellek alapvetően olyan egyenleteket becsülnek, amelyek együtthatói közvetlen kapcsolatban vannak az elméleti modellekkel, ez a gépi tanulásnál nem így van. Először is, sok gépi tanulási modell nem is néhány egyenletről áll, hanem fákból vagy egymásra épülő egyenletek neurális hálójából. Közel sem világos, hogy ezek hogyan feleltethetők meg a közgazdasági szereplők viselkedését leíró egyenletek paramétereinek. Másodsor, ha ki is lehet számolni ilyen paramétereket, akkor sincs garancia arra, hogy ezek oksági kapcsolatokat írnának le. Lehet, hogy éppen egy olyan változó rendelkezik a legnagyobb előrejelző erővel, amely valójában egy harmadik változó hatását mutatja. Harmadszor, a gépi tanulási modellekből kinyert együtthatók, még ha léteznek is, gyakran instabilak. A nagyon hasonló előrejelző tulajdonságú (és mély struktúrájú) modellek is gyakran eltérő változó-

kat tartalmaznak, így nagyon eltérő közgazdasági modellek lennének felírhatóak belőlük (Mullainathan–Spiess, 2017).

Ezek a különbségek vezetnek ahhoz, hogy a Big Data és a gépi tanulás nem veszi majd át automatikusan az ökonometria szerepét. Bizonyos alkalmazásokban a két módszer közötti helyettesítés erősebb, másokban viszont ezek a módszerek erősen kiegészítik egymást. Erre következik néhány példa.

ADATOK AZ INTERNETRŐL ÉS MŰHOLDAKRÓL

A Big Data és a gépi tanulás eszközeivel számos esetben valódi társadalomtudományi kísérlet végezhető el, ami miatt kevésbé kell tartani az önszelekció torzító hatásaitól – vagyis olyan empirikus módszertan alkalmazható az adatok elemzésekor, mint akár a kísérleti természettudományokban. Például Ali Ahmed és Mats Hammarstedt (2008) a diszkriminációt vizsgálta a svéd bérlakások piacán, olyan módon, hogy három eltérő fiktív személy nevében kerestek véletlenszerűen albérletet. Az arab nevű férfi lényegesen kevesebb visszahívást kapott, mint a svéd nevű albérletkeresők, míg a svéd nevű nő több visszahívást kapott, mint a svéd nevű férfi. Az internetes platformon pontosan, alacsony költséggel és teljesen kísérleti módszertannal vizsgálható a diszkrimináció szerepe. Azonban, mint látható, ez az adat nem magától jött létre, hanem a kutatók – egy nagyon alkalmas platformon – maguk hozták létre az információkat. Nagyon hasonló kísérletet korábban is el lehetett végezni úgy, hogy a kutatók levélben válaszoltak állásajánlatokra, amelyben ugyanazt a választ írták különböző bőrszínű emberek nevében (Lavergne–Mullainathan, 2004).

Miközben az előző példában is lényegesen egyszerűbb az interneten kísérletezni, az interneten sok esetben nagyon egyszerűen lehet nagy mennyiségű kísérleti adatot generálni különösen az e-kereskedelemmel kapcsolatos kérdések esetében. Könnyen megbecsülhető például az, hogy mennyivel vesznek többet az emberek egy-egy termékből, ha 10 százalékkal csökken az ára, ha több millió véletlenszerűen kiválasztott potenciális vásárlónak véletlenszerűen 10 százalékkal alacsonyabb árat ajánl fel egy e-kereskedő. Ezeknek a kísérleteknek az egyszerűsége és alacsony költsége ahhoz vezetett, hogy a legnagyobb internetes cégek folyamatosan ehhez hasonló kísérleteket végeznek termékválasztékuk, áraik vagy oldaluk kinézetének optimalizálása érdekében. Az így létrejött adatok sok közgazdasági paraméter becsülését könnyítik meg, és számos üzleti alkalmazásuk is van.

Hozzá kell azonban tenni, hogy az interneten létrejött Big Data nagy része nem tudatos kísérletezés eredménye, hanem típusát tekintve ugyanúgy megfigyelési adat, mint amelyeket korábban az ökonometria eszközeivel elemeztek a közgazdászok. Vagyis elemzésükhöz szükség van ökonometriai módszerekre, de egyben tömöríteni is kell az információt elemzés előtt. Például az ökon-

metria alapvetően arra a feltevésre épít, hogy eleve adott valamennyi – nem túl sok – magyarázó változó, és azokból épít a kutató valamilyen elméleti megfontolásokból modellt. Amennyiben a potenciális magyarázó változók száma több ezerre nő, akkor ez az elképzelés nem áll meg, és szükség van arra, hogy a változók kiválasztása vagy tömörítése, valamilyen automatikus eljárással – vagyis gépi tanulással – történjen.

A gépi tanulás segíthet olyan változókat létrehozni, amelyeket magyarázó vagy függő változóként lehet felhasználni az ökonometriai elemzésben olyan kérdések megválaszolására, amelyeket korábban nem vagy csak nagyon korlátozottan lehetett vizsgálni. Például a múholdas információkból rendkívüli földrajzi részletességgel mérhető, hogy mennyi fényt bocsát ki éjszaka az adott térség. Ez az információ – gépi tanulásra támaszkodva – elég pontosan közelítheti a gazdasági növekedést. Különösen fontosak az ilyen mérőszámok a fejlődő országok esetében, ahol a gazdasági aktivitás nagy része rejtve maradhat a hivatalos statisztikák előtt. Ezeknek az újfajta változóknak a segítségével sokkal pontosabban vizsgálható, hogy milyen tényezők hatnak egy-egy térség gazdasági növekedésére (Henderson et al., 2012; Donaldson–Storeygard, 2016).

Talán még sokrétűbb közgazdasági és társadalomtudományi lehetőségeket rejt magában a Google Trends-adatok felhasználása. Ezek azt mutatják meg, hogy mennyien kerestek rá bizonyos kifejezésekre, továbbá gazdag területi és időbeli bontásban érhetők el (Varian, 2014; Stephen-Davidowitz, 2017).

SOK KÉRDÉS VALÓJÁBAN PREDIKTÍV

A bevezetőben kiemeltem, hogy a legtöbb közgazdaság-tudományi empirikus elemzés célja az, hogy oksági hatásokat, valamiféle empirikus paramétereket becsüljön meg. Vannak azonban olyan mély közgazdasági kérdések, amelyek lényegüket tekintve prediktívek: például olyan elméleteket tesztelnek, amelyek azt állítják, hogy egy bizonyos változó *nem* rendelkezhet prediktív erővel. Ezeket korábban ökonometriai megközelítésben vizsgálták, de igazából a gépi tanulás sokkal jobban illik az ilyen hipotézisek teszteléséhez. Ha a gépi tanulási modellek legnagyobb erőfeszítésük dacára sem tudnak érdemben előre jelezni, a változónak gyaníthatóan tényleg nincs prediktív ereje.

Például a pénzügyi közgazdaságtan egyik alapmodellje a hatékony piacok elmélete, amelynek különböző formái azt állítják, hogy múltbeli információk alapján nem lehet előre jelezni a jövőbeli részvényhozamokat. Benjamin Moritz és Tom Zimmermann (2016) amellett érvel, hogy – mivel rendkívül sok változó jöhet szóba egy ilyen vizsgálatnál – érdemes gépi tanulást használni ennek az elméletnek a teszteléséhez. Megmutatják, hogy az amerikai tőzsdeindexet előre lehet jelezni múltbeli adatokból, ami ellentmond a hatékony piacok elméletének.

Miközben a gazdaságpolitika közgazdasági támogatásában legtöbbször fontos az oksági becslés (Hogyan hat a diákok tudására, ha még egy tanárt felvesznek?), vannak olyan területek, ahol a szakpolitikai kérdés prediktív választ igényel (Melyik tanárt érdemes felvenni a rendelkezésre álló információk alapján?) (Chalfin et al., 2016). Ezek az elemzések erősen kiegészítik az oksági vizsgálatokat: a tanárok kiválasztásának módszere befolyásolja a hatást, és a becsült hatás nagysága mutatja meg, hogy mennyi tanárt érdemes felvenni.

AZ OKSÁGI ELEMZÉSEKBEN IS VANNAK PREDIKTÍV LÉPÉSEK

Az utóbbi évek egy fontos gondolata az, hogy az ökonometriai elemzés során számos esetben szerepelnek prediktív lépések. A becslés hatásosabbá tehető akkor, ha felismerjük ezeket a lépéseket, és prediktív módszereket használunk becslésükre. Az, hogy melyek ezek a lépések, és pontosan milyen módszert kell használni becslésükhöz, egy nagyon aktív kutatási terület.

Például a kontrollcsoport létrehozásának egyik gyakran használt módszere a párosítás (*matching*, például Imbens–Rubin, 2015, Part III). Ennek alapötlete az, hogy minden kezelt egyénnek keresünk egy (vagy több) olyan nem kezelt párt, aki a legjobban hasonlít rá megfigyelhető, kezelés előtti jellemzői (például kora, neme, betegségének típusa, lakóhelye) szempontjából. Amennyiben sok megfigyelt jellemző van, akkor a legtöbb kezelt egyénnek nehéz volna olyan párt találni, amely minden egyes változójában eléggé hasonlít rá – túl sok dimenziós a probléma. Ilyenkor információtömörítésre van szükség: a sok magyarázó változóból egyet (vagy keveset) készítünk, és ez alapján végezzük a párosítást. A legtöbbször használt módszer az, hogy a megfigyelt változók alapján megbecsüljük, hogy melyik egyént milyen valószínűséggel kezelik (ezt a valószínűséget nevezzük *propensity score*-nak), és olyan párt keresünk, amelyik a legjobban hasonlít a kezelés valószínűségében. A kezelés hatását úgy kaphatjuk meg, ha összehasonlítjuk minden kezelt egyén kezelés utáni egészségi állapotát a párjával – vagyis egy olyan betegével, aki a kezelés előtt a legjobban hasonlított hozzá.

Ez a módszer tehát a következő lépésekből áll – egy orvosi kezelés példáján. Első lépésben megbecsüljük a megfigyelhető jellemzők (kor, nem, korábbi betegségek) alapján annak valószínűségét, hogy az adott ember részesül-e a kezelésben. A második lépésben minden kezelt egyénnek keresünk egy olyan nem kezelt párt, aki a legjobban hasonlít hozzá – vagyis esetében ugyanakkora a kezelés becsült valószínűsége. A kontrollcsoport ezekből a párokból áll. A harmadik lépésben pedig (általában regressziós elemzéssel) összehasonlítjuk a kezelt csoport kimenetét (mennyi idő alatt gyógyult meg) a kontrollcsoportéval. Ez adja az oksági becslést.

Nagyon lényeges meglátás az, hogy az első lépés valójában tisztán prediktív: potenciálisan nagyon sok változó alapján szeretnénk előre jelezni a kezelés valószínűségét. Nincs szó ebben a lépésben oksági hatásokról, és nem fontos az sem, hogy melyik változónak mi az együttthatója. A gépi tanulás tökéletesen alkalmas ennek a lépésnek a végrehajtására. A legtöbb közgazdász azonban – már csak megszokásból is – a „hagyományos” módszereket használja, vagyis végiggondolja, hogy mely változók befolyásolják a kezelés valószínűségét, és azokkal futtat egy regressziót.

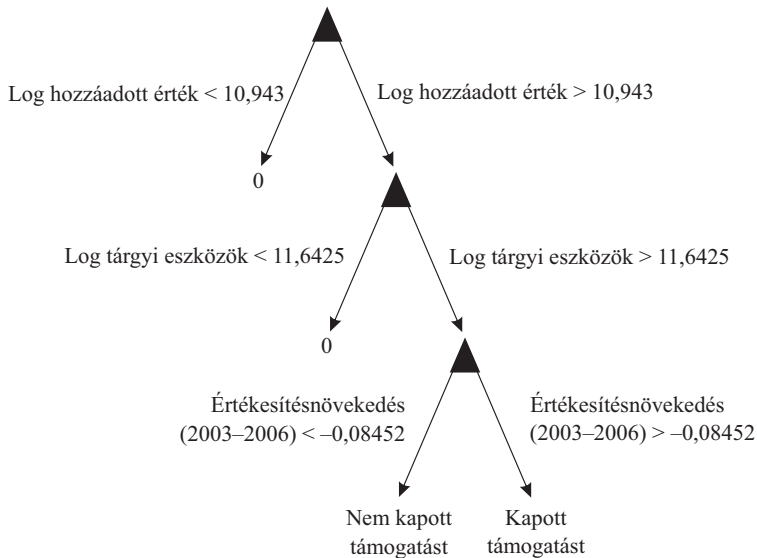
ALKALMAZÁS: A VÁLLALATI EU-TÁMOGATÁSOK HATÁSA

Ebben a fejezetben az előzőekben leírtakra mutatok egy leegyszerűsített példát. A kutatási kérdés az, hogy milyen módon befolyásolta a vállalatok teljesítményét (a dolgozók számát és termelékenységét) az, hogy a vállalat részesült-e valamiféle EU-forrásból származó vállalati támogatásban. Az egyszerűség kedvéért itt csak egy év (2007) támogatásait vizsgáltam.

Az adatok jellege miatt itt párosításos módszert érdemes alkalmazni. Ennek az első lépése, amely alapvetően prediktív, abból áll, hogy minden vállalatra különböző gépi tanulós módszerekkel megbecsüljük a támogatás megnyerésének valószínűségét. Az előrejelzéshez csak a 2007 előtti információkat kell felhasználni, hogy a támogatás hatása ne jelenjen meg a magyarázó változók között. Ezt követően minden támogatott vállalatnak párokat kell választani (a legközelebbi szomszéd módszerrel), majd pedig kiszámolni, hogy mennyiben viselkedett eltérően a kezelt csoport a párokból álló kontrollcsoporttól a támogatást követő időszakban (2007 és 2010 között).

Az első módszer egy döntési fa volt, eredményét az *1. ábra* mutatja. E szerint azok a vállalatok kapnak támogatást, amelyeknek nagy a hozzáadott értékük és a tőkeállományuk, valamint a múltban is növekedtek. Ez megegyezik a közgazdasági intuícióval is: a nagyobb vállalatok inkább hajlandóak kifizetni a pályázás költségeit, és a múltban is növekvő vállalatoknak inkább vannak olyan növekedési terveik, amelyekhez fel tudják használni az ingyenes forrásokat. Ez a modell 71 százalékban jelzi helyesen előre, hogy melyik vállalat pályázik.

A *boosting* eljárás (lásd például Hastie et al., 2009) több tucat fát épít fel, amelyek „szavazással” döntenek el, hogy mi legyen az előrejelzés: minden megfigyelés esetében megvizsgálják, hogy melyik fa milyen előrejelzést ad, és a *boosting* modell előrejelzése az, amit a fák többsége jelez előre. Ezek a módszerek gyakran hatékonyabban jeleznek előre, mint egy-egy fa. Ez esetünkben is így van: ez a modell 72,1 százalékban jelez előre helyesen. Összehasonlításképp ugyanezekkel a változókkal egy logit modellt is lefuttattam, amelyben a változókat a gépi tanulás algoritmus választotta ki. A logit modell egy olyan nemlineáris regresszió,



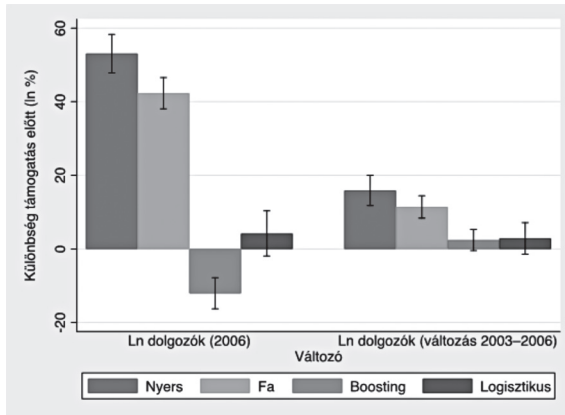
1. ábra. Döntési fa

amely egy fektetett S alakú görbét illeszt a pontokra, így különösen alkalmas az olyan elemzésekre, amikor a kimenet kétértékű (akkor vesz fel nulla értéket, ha a vállalat nem kapott támogatást, és akkor 1-et, ha kapott). Ennek előrejelző ereje 70,8%.

A párosításos eljárásban a kontrollcsoport minőségének egyik fő tesztje a kiegyensúlyozottsági (balancing) teszt, amely egyszerűen megvizsgálja, hogy mennyiben tért el a múltban, kulcsjellemzőit tekintve, a kezelt csoportja a kontrollcsoporttól. Ha már a múltban is eltérően alakult a teljesítményük, akkor feltehetőleg támogatás nélkül is eltérő mértékben növekedtek volna. A kiegyensúlyozottsági teszt nagyon jó példa a gépi tanulás és az ökonometria közötti kiegészítésre: ez nem a gépi tanulás prediktív erejét méri, hanem azt, hogy közgazdasági szempontból hiteles kontrollcsoport jött-e létre.

Ezt a tesztet szemlélteti a 2. ábra, amelyben a számok azt mutatják, hogy mennyiben tért el a támogatást megelőzően a kezelt és a kontrollcsoport átlaga.

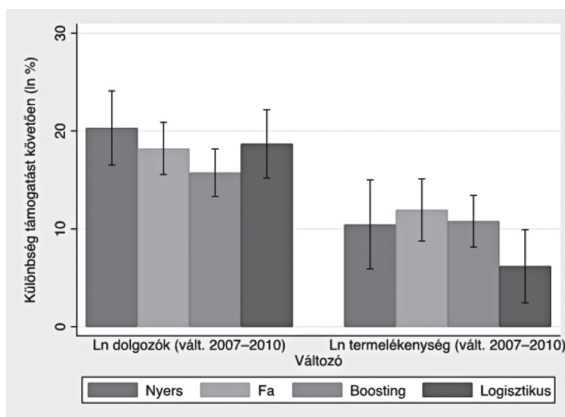
Viszonyítási pontként az első oszlop azt mutatja, hogy mennyiben tért el az eredeti (párosítatlan) mintán a kezelt és a nem kezelt csoport. A kezelt vállalatok lényegesen nagyobbak (több mint 50 százalékkal több dolgozóval rendelkeznek), és gyorsabban nőttek (15 százalékponttal gyorsabban nőtt a dolgozóik száma 2003–2006 között). Ezek a különbségek mindegyik párosítási eljárás után csökkennek. Azonban az eltérő módszerek különböző kontrollcsoportokat generálnak. A döntési fa által generált kontrollcsoport lényegesen nagyobb volt, és gyorsabban is nőtt, mint a kezelt csoport, míg a boosting és a logisztikus regresszió



2. ábra. Kiegyensúlyozottsági teszt: a kezelt és kontrollcsoport közötti különbség a kezelés előtt

kontrollcsoportja nagyon hasonlít a kezelt csoporthoz. Fontos következtetés, hogy nem feltétlenül a legnagyobb prediktív erejű eljárással készült kontrollcsoport hasonlít legjobban a kezeltre.

Végül, a 3. ábra tartalmazza a hatás becslését, vagyis azt, hogy átlagosan mennyivel növekedtek gyorsabban a támogatott vállalatok 2007–2010 között, mint a kontrollcsoport cégei. Az látható, hogy a különböző kontrollcsoportok alapján kapott eredmények nem térnek el egymástól 5 százalékos szignifikanciaszinten. A kép nagy vonalakban azonban az, hogy a támogatás segítette a vállalatok növekedését, és bizonyos mértékben a termelékenységük fellendítését is. Ez az eredmény mindegyik kontrollcsoport mellett megmarad, nem függ attól, hogyan jelezzük előre a kezelés valószínűségét.



3. ábra. Becsült hatás: a kezelt és kontrollcsoport közötti különbség a kezelést követően

A példa erősen leegyszerűsített. Valójában figyelembe kellene venni azt is, hogy melyik vállalat kapott máskor is támogatást, illetve jobban kellene modellezni a vállalati heterogenitást is. Kiterjedt ökonometriai irodalom – elsősorban a panelökonometria – foglalkozik ezeknek a problémáknak a kezelésével. Ezekből általában azt az eredményt kapjuk, hogy a támogatás hatására a vállalatoknak legfeljebb a méretük nő, de a termelékenységük nem.

KÖVETKEZTETÉSEK

Ennek a tanulmánynak a fő üzenete az, hogy a gépi tanulás és a Big Data nem teszi szükségtelenné az ökonometriát, hanem a két módszer erősen kiegészíti egymást. A megfigyelési adatok elemzésekor szükség van az ökonometria alapszemléletére, de a nagy adatbázisok hatékony kezeléséhez alapvető fontosságú a gépi tanulás eszközeinek alkalmazása.

A rendelkezésre álló új adatok és a gépi tanulás prediktív szemlélete befolyásolhatja a közgazdászok kérdésválasztásait is. Nagyobb szerepet kaphatnak a prediktív elméletek és hipotézisek, illetve az olyan vizsgálatok, amelyek az új adatok nélkül nem is lettek volna megvalósíthatóak. Aktív kutatások folynak a két módszertant egyszerre alkalmazó empirikus eszközök továbbfejlesztésére.

IRODALOM

- Abdulkadiroğlu, A. – Angrist, J. – Pathak, P. (2014): The Elite Illusion: Achievement Effects at Boston and New York Exam Schools. *Econometrica*, 82, 1, 137–196. DOI: 10.3386/w17264, <http://www.nber.org/papers/w17264>
- Ahmed, A. M. – Hammarstedt, M. (2008): Discrimination in the Rental Housing Market: A Field Experiment on the Internet. *Journal of Urban Economics*, 64, 2, 362–372. DOI: 10.1016/j.jue.2008.02.004, https://www.researchgate.net/publication/222011648_Discrimination_in_the_Rental_Housing_Market_A_Field_Experiment_on_the_Internet
- Angrist, J. D. – Dynarski, S. M. – Kane, T. J. et al. (2012): Who Benefits from KIPP? *Journal of Policy Analysis and Management*, 31, 4, 837–860. DOI: 10.3386/w15740, <http://www.nber.org/papers/w15740>
- Angrist, J. D. – Pischke, J. S. (2010): The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics. *The Journal of Economic Perspectives*, 24, 2, 3–30. DOI: 10.3386/w15794, <http://www.nber.org/papers/w15794>
- Chalfin, A. – Danieli, O. – Hillis, A. et al. (2016): Productivity and Selection of Human Capital with Machine Learning. *The American Economic Review*, 106, 5, 124–127. http://www.hbs.edu/faculty/Publication%20Files/PredictiveHiring_4a3d9f9f-62e9-4dad-ac58-e3c3571d3995.pdf
- Donaldson, D. – Storeygard, A. (2016): The View from Above: Applications of Satellite Data in Economics. *The Journal of Economic Perspectives*, 30, 4, 171–198. DOI: 0.1257/jep.30.4.171, http://dave-donaldson.com/wp-content/uploads/2016/10/Donaldson_Storeygard_JEP.pdf

- Hastie, T. – Tibshirani, R. – Friedman, J. (2009): Overview of Supervised Learning. In: *The Elements of Statistical Learning*. New York: Springer, 9–41. DOI: 10.1007/b94608_2, https://www.springer.com/cda/content/document/cda_downloadaddocument/9780387848570-cl.pdf?S-GWID=0-0-45-733855-p173883504
- Henderson, J. V. – Storeygard, A. – Weil, D. N. (2012): Measuring Economic Growth from Outer Space. *The American Economic Review*, 102, 2, 994–1028. DOI: 10.3386/w15199, <http://www.nber.org/papers/w15199>
- Imbens, G. W. – Rubin, D. B. (2015): *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press
- Lavergne, M. – Mullainathan, S. (2004): Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *The American Economic Review*, 94, 4, 991–1013. DOI: 10.3386/w9873, <http://www.nber.org/papers/w9873>
- Moritz, B. – Zimmermann, T. (2016): *Tree-based Conditional Portfolio Sorts: The Relation between Past and Future Stock Returns*. SSRN. DOI: 10.2139/ssrn.2740751, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2740751
- Mullainathan, S. – Spiess, J. (2017): Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, 31, 2, 87–106. DOI: .1257/jep.31.2.87, <https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.31.2.87>
- Stephens-Davidowitz, S. (2017): *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are*. New York: HarperCollins
- Varian, H. R. (2014): Big Data: New Tricks for Econometrics. *The Journal of Economic Perspectives*, 28, 2, 3–27. DOI: 10.1257/jep.28.2.3, <https://pubs.aeaweb.org/doi/pdf/10.1257/jep.28.2.3>

Tanulmányok

A DIKTATÚRA LÉTREHOZÁSA ÉS A HUMÁN TUDOMÁNYOS ELIT (1946–1948)

ESTABLISHMENT OF DICTATORSHIP AND THE HUMAN SCIENCES' ELITE 1946–1948

N. Szabó József

csc, dr. habil, professor emeritus, Nyíregyházi Egyetem
enikoszakos@gmail.com

ÖSSZEFOGLALÁS

1947 tavaszától a demokráciával szakító Magyar Kommunista Párt (MKP) elhatározta, hogy az egyetemekről és a tudományos intézményekből eltávolítja az ellenforradalmi reakció képviselőit. Az 1947-es augusztusi választások után a monopolisztikus hatalomgyakorlást megvalósítani szándékozó kommunisták számára a bölcsészkarok (BTK) konzervatív professzorai is elfogadhatatlanná váltak. 1947-ben több támadás történt a humántudományi elit ellen, de a legnagyobb kihívás az MKP 1948. június 8-án elkészített, de nyilvánosságot nem kapott felsőoktatási reformtervében éri a bölcsészeket. A tervzetben a párt az egyetemekről több tudományosan magas színvonalon álló bölcsészprofesszort el kívánt távolítani, és helyettük kommunista oktatókat szándékozott kinevezni. A néhány szakmailag kimagasló kutató mellett az 1948-ban kinevezett egyetemi tanárok között csak egyetlen professzor volt, akit nem a párt delegált. A kis számban habilitált bölcsészek többsége ugyanakkor magas tudományos színvonalat képviselt. Hasonlóan magas színvonalon álltak az 1947–1948-as tanév egyetemeinek bölcsész rektorai és a dékánok. A szakmailag kiváló tudósok politikai „kötődése” igen eltérő volt, illetve egyetlen politikai irányzathoz vagy párthoz sem tartoztak. Már a kommunisták hegemonisztikus hatalomgyakorlása idején, 1947 júniusában megtartott tagválasztáson az MTA egy levelező és hét rendes tagot választott a humán tudományok képviselői közül. Mindannyian magas tudományos színvonalat képviseltek. Mindösszesen egy akadémikusnak volt pártkötődése. A formálisan is kiépült totalitárius diktatúra alatt 1948 júliusában ülésező MTA tizenöt bölcsész akadémikust választott. Voltak közöttük olyanok, akik a kommunista tudománypolitika delegáltjai közé tartoztak, egy akadémikus viszont világszínvonalon állt. Tudományos teljesítmény nélkül azonban senki sem került az akadémia tagjai közé.

ABSTRACT

Breaking with democracy, the Hungarian Communist Party decided to remove all the representatives of the “counter-revolutionary movement” from universities and scientific institutions. Following the general elections of 1947, the more conservative professors of each Faculty of Arts

were considered to be intolerable in the eyes of the communists, who sought to monopolize state power. In 1947 there were several attacks against the elite of human sciences, but the most provoking of them came the following year, with the planned higher education reform package, completed on 8th June 1948. The core of the plan was to remove some highly professional and esteemed professors from the Faculty of Arts and replace them with newly appointed professors loyal to the communist regime. For the time being the situation of the Hungarian Academy of Sciences was different. The new regime had already been institutionalized, when – in July 1948 – the Academy, in session, elected fifteen scholars into membership. Some of them were representatives of the scientific communism, but one of them was an internationally recognized scholar. At this point in time, however, no one could get an Academy membership without real scientific achievements.

Kulcsszavak: totalitárius diktatúra, akadémiai és egyetemi „reform”, akadémiai tagválasztások, professzori kinevezések, habilitációk

Keywords: totalitarianism, academic and higher education reform, election of the members of the academy, appointed professors, habilitation

A DEMOKRÁCIÁVAL VALÓ SZAKÍTÁS ÉS A HUMÁN TUDOMÁNYOK

1946 őszén a magyar politikatörténetben „fordulópontot” jelentő III. kongresszuson a kommunisták kinyilvánították: „ki a nép ellenségeivel a koalícióból” (Rákosi–Szabó, 1967, 272.). Ezzel a deklarációval a párt gyakorlatilag szakított a demokrácia eddig általa is többé-kevésbé betartott játékszabályaival. A kommunista politika radikális megváltozása ugyanakkor nem tükröződött a kongresszusnak a tudománnyal kapcsolatos felfogásában, olyannyira nem, hogy a pártértekezlet tudománypolitikai kérdésekkel gyakorlatilag nem is foglalkozott. Kivételt jelentett a Moszkvából 1945 augusztusában hazatért Lukács György felszólalása, aki heves kirohanást intézett az egyetemek és a Magyar Tudományos Akadémia ellen, amelyeket a reakció fellegráinak tartott (Gergely–Izsák, 2000, 295.).

A neves kommunista filozófus véleménye elsősorban a bölcsészekkel szemben fogalmazódott meg. Lukács támadása az egyetemek nyitottságát illetően erősen túlzó volt, mert Hajnal István bölcsészkarri dékán a Pázmány Péter Tudományegyetem rektorához és rektori tanácsához 1945. augusztus 20-án írt jelentésében már javasolta Lukács György kinevezését az esztétikai tanszék élére. Sokat árul el az 1945-ös bölcsészkarri viszonyokról, hogy Thienemann Tass Tivadar egyetemi tanár a kari tanács 1945. szeptember 1-jei ülésén szintén indítványozta a filozófus meghívását. A kari tanács 18 igenlő, 5 nemleges és 2 érvénytelen szavazattal elfogadta a kiváló filozófus meghívását az esztétikai és kultúrfilozófiai tanszékre (Urbán, 1985, 96.).

A tudománypolitikában „visszafogott” III. kongresszus ugyanakkor döntött a hároméves tervjavaslat kidolgozásáról. A tervjavaslatot a Politikai Bizottság

1946. december 19-én tárgyalta, a Központi Vezetőség 1947. január 11-i ülésén pedig elfogadta. A terv legfontosabb tudománypolitikai célkitűzése az volt, hogy a tudományt az ország termelőerői szolgálatába állítsa, és szoros kapcsolatba hozza a gazdasági tervezéssel (Rákosi–Szabó, 1967, 385.). A tervezetben elsősorban a műszaki és agrártudományok terén fogalmazódtak meg tudománypolitikai preferenciák, de kiemelt célként jelent meg a társadalomtudományok támogatása is. A humán tudományok fontosságának hangsúlyozására utal, hogy az elképzelések között szerepel az egyetemektől független társadalomtudományi kutatóhálózat kiépítésének gondolata. Az MKP a felállítandó Országos Társadalomtudományi Intézettől, az Általános Neveléstudományi Intézettől, valamint a Világpolitikai és a Duna-táji Intézettől a társadalomtudományi kutatások fellendülését remélte (Rákosi–Szabó, 1967, 386.).

1947 első felében a magyar tudománypolitika központi kérdése a kutatások finanszírozása, a kutatási célkitűzések iránya, valamint a tudományos életben a II. világháború után keletkezett nehézségek felszámolása volt. A megoldandó témákat az 1946/1947-es költségvetés vitájában tárgyalták. A parlamenti vita elsősorban a természettudományokról, illetve azok támogatásáról szólt (Nemzetgyűlés Naplója, 1952, 146–150., 608., 613., 637., 648.).

A tudománypolitikai polémiákban a humán elit is érintett lett, mert bölcsészlelens vélemények hangzottak el a kommunisták részéről. A tudományos élet fejlesztéséről szóló javaslatában az MKP szónoka, Rudas László a budapesti egyetemen a természettudományi kar felállítását azért szorgalmazta, mert megítélése szerint a Vallás- és Közoktatásügyi Minisztérium (VKM) a humán tanszékeket és tudományokat fejleszti a természettudományokkal szemben (Nemzetgyűlés Naplója, 1952, 663.).

A kommunista vádat a VKM miniszter elutasította. Keresztury Dezső egyetértett azzal, hogy bizonyos területeken az oktatási rend megreformálásra szorul, de cáfolta, hogy a minisztérium elhanyagolja a természettudományokat. A fejlesztés elmaradása szerinte nem szándék, hanem kényszerűség (Nemzetgyűlés Naplója, 1952, 743.).

A Szociáldemokrata Párt (SZDP) tudománypolitikája ekkor elsősorban az egyetemek szemléletét kívánta átalakítani. Szalai Sándor a hallgatói szemléletalakítás érdekében kezdeményezte a budapesti szociológiai tanszéknek a legrövidebb időn belül való kibővítését és a vidéki egyetemeken, Szegeden és Debrecenben legalább egy szociológiai előadó állás szervezését (Népszava, 1947. január 12.).

1947 tavaszától a Magyar Kommunista Párt kiemelt feladatának tartotta a tudományos kutatás tervszerű vezetését és szervezését, valamint a hároméves terv célkitűzéseivel való összhangba hozását. A bölcsészek eddig azonban az MKP elképzeléseit illetően nem mutattak affinitást. Ennek megváltoztatására ezért a kommunisták bizonyos területeken megengedhetőnek tartották diktatórikus eszközök használatát is. A siker érdekében elhatározták, hogy az egyetemekről és a

tudományos intézményekből eltávolítják az „ellenforradalmi” reakció képviselőit. Tudománypolitikai célként ugyanakkor „még csak” azt fogalmazták meg, hogy a marxizmus a tudományban és a felsőoktatásban foglalja el azt a helyet, amely tudományos és társadalmi jelentőségének megfelel (PTI. Arch. 274-24/2). Az MKP-hez hasonlóan az SZDP is fontosnak tartotta a marxizmus tudományos életben való térhódításának elősegítését (Népszava, 1947. június 6.).

A Magyar Kommunista Pártnak a hároméves terv kidolgozására tett különböző erőfeszítéseit a Tudományos Akadémia tagjainak többsége továbbra sem akceptálta. Az Akadémia 1947. májusi gyűlésén viszont egyes bölcészek részéről már kezdeményezés történt a kommunista elképzelések támogatására. Beke Ödön baloldali politikai múlttal rendelkező nyelvész például indítványt tett arra, hogy az MTA dolgozzon ki egy hároméves tervet. Beke akadémikus kezdeményezésének azonban még nem sok támogatója akadt a testület tagjai között, mert Németh Gyula turkológuson kívül senki más nem állt az indítvány mellé (Szabad Nép, 1947. június 10.). Az MKP vezetése a tudósok közötti rossz pozíciója megváltoztatása érdekében ezért találkozót kezdeményezett a tudományos elittel. Az összejövetelre június 11-én sor is került (PTI. Arch. 274-24/14.). A pártvezetést Rákosi Mátyás, Rajk László, Molnár Erik, Révai József, Horváth Márton és Orbán László képviselte. A kommunisták által kezdeményezett találkozón elsősorban a humán tudományok egyes képviselői vettek részt. Ott volt Németh Gyula, a Pázmány Péter Tudományegyetem 1947/1948-as tanévre megválasztott rektora és Hajnal István bölcészeti dékán, továbbá Zsirai Miklós, Pais Dezső és Moravcsik Gyula nyelvész akadémikusok. A baloldali, illetve kommunista bölcészeket Mérei Ferenc, Fogarasi Béla, Rudas László, Andics Erzsébet, Szimonidesz Lajos, Waldapfel József és Imre, Komjath Aladár, Vígh Károly, Tolnai Gábor és Mátrai László képviselte. A humán elitnek a politikához nem kötődő tagjai között találjuk Gyóni Mátyást, Lakó Györgyöt, illetve Makkai Lászlót (Szabad Nép, 1947. június 12.).

Az egyetemi vezetők részéről az első két felszólaló Németh Gyula és Hajnal István volt, akik a tudósok legfőbb gondjának a túrhetetlen anyagi helyzetet és a mostoha kutatási körülményeket tartották. A pártvezetés részéről a problémákra reagáló Rákosi Mátyás főtitkár akceptálta a tudósok felvetését, és megígérte, hogy javítanak anyagi helyzetükön, és vállalják a tudomány támogatását. Révai József kultúrpolitikus pedig arról beszélt, hogy a hároméves tervben nagy szerepük van a természettudományoknak, de az emberek nevelése legalább olyan fontos, mint a gazdasági színvonal emelése. A párt fő ideológusa bölcészeket foglalkoztató tudománypolitikai kérdéseket is érintett hozzászólásában, amikor kijelentette, hogy a párt nem törekszik arra, hogy a marxisták kizárólagosságát biztosítsa az egyetemeken, ugyanakkor megjegyezte, hogy ott a helye a marxizmusnak is (Szabad Nép, 1947. június 15.).

A párt legfelső vezetőinek a tudósokkal való találkozója után a kommunista irányítású Gazdasági Főtanácsnál elérte, hogy augusztus 1-től 50%-os tudomá-

nyos pótlékot kapjanak az egyetemi tanárok és a tudományos intézeti vezetők, továbbá az egyetemi és főiskolai tudományos személyzetnél is biztosítsák a fizetés 40%-át (Szabad Nép, 1947. július 15.). A pótlék bevezetésétől az MKP a tudósok megnyerését és pozíciói megerősödését remélte, ami több esetben eredményesnek is mondható. A tudósoknak a párttal szembeni eddigi tartózkodásában, illetve elutasító magatartásában a fizetésemelés bizonyos csoportoknál változást hozott. Jól érzékelteti ezt két vélemény: Zsirai Miklós budapesti bölcsészkaros dékán szerint a karon hároméves terv készült, amelyet hatvanhét intézet és hetven tudós állított össze (Szabad Nép, 1947. augusztus 31.). A pozitív hozzáállást mutatja Németh Gyula rektor kinyilatkoztatása is, mely szerint a kommunisták segítségével a tudomány örökké hálás lesz (Szabad Nép, 1947. augusztus 31.). Mindkét tudományosan kiváló nyelvészről tudni kell, hogy a háború után rövidesen „felismerték” az MKP politikai-tudománypolitikai szerepét. Természetes, hogy a rektori és dékáni hivatalokból fakadó „kötelességük” miatt nem térhettek ki bizonyos állásfoglalások elől, de esetenként az alkalmazkodásuk a szükségést meghaladta.

Annak ellenére, hogy a humán elitek kevésbé kapcsolódnak a politikához, mint a jogászok vagy a közgazdászok, a baloldal, de különösen a kommunisták számára 1945–1946-ban fontos volt, hogy a népi demokratikus átalakulás követelményeinek megfelelően alakuljon a bölcsészkarok professzori garnitúrája. Mivel a magyar egyetemeken a professzori körben néhány szélsőjobboldali tanártól eltekintve a konzervatív gondolkodású oktatók domináltak, ezért az igazolások és bélisták idején a bölcsészek között viszonylag kevés esetben mondták ki a leg súlyosabb ítéleteket. A polgári demokrácia talaján álló politikai erők a szélsőjobb eltávolításával egyetértettek, de további tisztogatásokkal nem azonosultak. A fordulat meghirdetésétől, de különösen az 1947-es választások után a monopolisztikus hatalomgyakorlást megvalósítani szándékozó kommunisták számára a konzervatív egyetemi tanárok is elfogadhatatlanokká váltak. Az MKP úgy látta, hogy a népi demokrácia megvalósítását veszélyeztető professzorokat el kell távolítani az egyetemekről.

1947-ben több támadás érte a humántudományi elitet, de a legnagyobb kihívás az MKP 1948. június 18-án elkészített, de nyilvánosságot nem kapott felsőoktatási reformtervében érte a bölcsészeket. A probléma fontosságát jelzi, hogy a párt a pedagógusképzésre vonatkozóan „csak” a bölcsészkarok átalakítására tett javaslatot, a természettudományi oktatással nem is foglalkozott. Ezt a leszűkítést a párt azzal indokolta, hogy a bölcsészkarokon olyan tanszabadság érvényesül, amely a szakképzést teljesen dezorganizálja. A bölcsészképzésre fókuszált reformot az MKP azzal is magyarázta, hogy a tanárképzés a tudományos kutatás viszonyának a tisztázatlansága, kaotikus állapotokat teremtett, és ez mindkét feladat megoldását akadályozza. A dolgozat a bölcsészélet kérdéseivel foglalkozik, ezért nem tekinti feladatának a pedagógusképzés elemzését. Egy rövidesen megjelenő tanulmányomban a feltárom ezt a problémát is. A kommunista reform szervezeti

javaslata az volt, hogy az egységes Bölcsészettudományi Kar helyett létesüljön Bölcsészettudományi és Természettudományi Kar. A kötött tanmenetek bevezetése a kommunisták szerint a bölcsészkaron a Tanárképző Intézetet feleslegessé teszi. Az MKP-tervezet az „ad personam” létesített „felesleges” tanszékeket meg kívánta szüntetni. Példaként említette a III. sz. Olasz Tanszéket és a II. sz. Archeológiai Tanszéket, valamint azt, hogy a külön Összehasonlító Irodalomtudományi Tanszék mellett Összehasonlító Irodalomtörténeti Tanszék is van. A tanárképzés új, szocialista szükségletének megfelelően ugyanakkor közgazdasági és társadalomtudományi tanszékek felállítását is tervezte. A hallgatók alapos és rendszeres képzése érdekében a kommunista tervezet egy relatív tanszabadság biztosítása mellett kötött tanmenetet kívánt bevezetni. A színvonalas képzés garantálásával a tervezet gondoskodni kívánt arról, hogy a hallgatók két idegen nyelvet is tanuljanak. Az MKP felfogása szerint a reform csak személyi változásokkal valósítható meg. A párt a következő bölcsészprofesszorokat kívánta eltávolítani a budapesti bölcsészkarról: Schwartz Elemért, Gerevich Tibort, Huszti Józsefet, Alszeghy Zsoltot, Lukinich Imrét és Lassovszky Károlyt. Az MKP ugyanakkor be kívánta vinni az egyetemre Molnár Eriket és Andics Erzsébetet.

A szegedi egyetemről Mester Jánost, a debreceniről pedig Mészáros Edét kívánták elmozdítani. A szegedi egyetemre a párt két kommunista oktatót javasolt professzornak: Waldapfel Imrét és Halász Elődöt. Az eltávolítandó professzorok helyére az elképzelés kivétel nélkül magasabb színvonalú tanárokat akart kinevezni (PTI. Arch. 274-21/74.).

Az elvi célkitűzéssel kapcsolatban a problémák azáltal jelentkeztek, hogy a lecserélendő professzorok valamennyien magas tudományos színvonalat képviseltek. Politikailag esetleg egy személlyel kapcsolatban „lehetett” volna kifogása a pártnak. Gerevich Tibor kiváló művészettörténész a Gömbös-kormány idején művészetpolitikusként valóban szerepet játszott az olasz *novecentismo* törekvések magyarországi megvalósításában. A többiek viszont politikailag nem „kompromittálódtak”. Magas szakmai szinten állt Huszti József klasszika-filológus és Lukinich Imre történész, valamint Lassovszky Károly csillagász, továbbá Mester János és Mészáros Ede, akiknek szakmai-tudományos munkájához szintén nem fért kétség.

A bölcsészkarokra a párt által bevinni szándékozott professzorok közös jellemzője az volt, hogy valamennyien MKP-tagok voltak. Egyesek közülük bevállalták a párt célkitűzéseinek minden további nélküli megvalósítását. Volt olyan is, aki nem mondható ugyan pártkatonának, de a párt tudományos elvárásaihoz egyértelműen igazodott. Olyan is akadt közöttük, aki illegális múltja ellenére szakmaiságot is megtestesített. Egy olyan támogatóval is találkozunk, aki inkább volt professzor, mint kommunista párttag.

A június 8-án elkészült egyetemi reformelképzelések nem kerültek sem az MKP IV. kongresszusa elé, de nem tárgyalta a javaslatokat az SZDP XXXVII.

kongresszusa sem. Az MDP I. kongresszusa sem foglalkozott a reformmal kapcsolatos elképzelésekkel (Rákosi–Szabó, 1967, 581–582., 585., 600–601.). A kongresszusok után több pártfórumon és bizottsági ülésen is értékelték és elfogadták a tervezetet (Ladányi, 1989, 34.), nyilvánosságot a kormányrendelet a *Köznevelés* 1948. október 15-i számában kapott (1948, 20.). Ha összehasonlítjuk a nyilvánosságot nem kapott júniusi tervet az október 15-én megjelent kormányrendelettel, azt látjuk, hogy az egyes szakok, valamint a képzésre szánt idő részletes ismertetése mellett az anyag koncepcionálisan nem tűnik kidolgozottnak. Kutatási témánk szempontjából fontos, a bölcsészéletet érintő professzori kinevezésekkel a kormányrendelet csak az általánosság szintjén foglalkozik. A publikált anyag egy része a bölcsészkarok „színvonaltalanságát”, illetve ideológiai zűrzavarát tárja fel, hangsúlyozva, hogy a professzorok a népi demokráciával ellenséges tanokat hirdetnek, a marxizmus–leninizmus képviselői alig találhatók meg az egyetemeken.

PROFESSZORI KINEVEZÉSEK ÉS HABILITÁCIÓK A FORDULAT UTÁN

1946 őszen néhány orvos- és jogi kari professzori kinevezés mellett csak két bölcsész egyetemi tanári felterjesztés történt. Novemberben a Pázmány Péter Tudományegyetem régészeti tanszékére nevezték ki Nagy Lajost, a művészettörténetre pedig Genthon Istvánt (*Magyar Közlöny*, 1946, 242–243.). Megállapíthatjuk, hogy mindkét professzor kiváló tudós és oktató volt. Nagy Lajos Budapest környéki római császárkori és őskeresztény emlékek régészeti feltárása és dokumentálása, művészettörténeti elemzése terén alkotott kiemelkedőt. Genthon István művészettörténész kivételes műveltségű tudós pedig a magyar művészet nemzetközileg is elismert kutatója volt. A Bölcsészettudományi Kar új professzorai megérdemelten lettek az egyetem tanárai. A kinevezéseknek nem volt sem politikai, sem sajtóvisszhangjuk.

1946 decemberében azonban már a baloldali pártok és a kommunista irányítású Magyar Pedagógusok Szakszervezetének érdeklődésébe kerültek a professzori kinevezések. A szakszervezet Balla Antalnak a Közgazdasági Kar Gazdaságtörténeti Tanszékére való egyetemi tanári kinevezése ellen tiltakozott. Ugyanakkor viszont támogatta Berlász Jenő kinevezését (*Szabad Nép*, 1946. december 5.). Egyetérthetünk a szakszervezetnek azon véleményével, hogy Berlász tudományosan felkészült kutató volt, de Balla Antal akadémikus tudományos kvalitásaihoz sem férhetett volna kétség.

1947 első felében mindösszesen egy bölcsészprofesszori kinevezésre került sor. Áprilisban a Debreceni Tudományegyetem Angol Tanszékének vezetésére kapott megbízást a szakmailag kiváló Ország László (*Köznevelés*, 1947, 7.). Augusztusban viszont rendkívüli egyetemi tanár lett Rubinyi Mó-

zes (Köznevelés, 1947, 16.). Ugyancsak augusztusban lett rendkívüli egyetemi tanár Barnóthy Jenő (Köznevelés, 1947, 24.).

1947 őszéig sem a politikai pártok, sem a szakszervezetek részéről nem nyilvánítottak véleményt a professzori kinevezésekkel kapcsolatban. Kivételt jelentett a Nemzeti Parasztpárt lapja, a *Szabad Szó*, amely szeptember 7-i számában a szegedi egyetem bölcsész tanszékeinek munkájával foglalkozott. A parasztpárti újság nem volt megelégedve az egyetem pedagógiai és történelem tanszékeinek működésével, de szegényesnek tartotta az irodalom tanszéken folyó munkát is. Nem értett egyet azzal sem, hogy Sík Sándor távozása óta a magyar irodalom tanszék nincs betöltve, és hogy mindösszesen két irodalmi kollégiumot hirdettek meg: Rubinyi Mózesét és Baróti Dezsőét (Szabad Szó, 1947. szeptember 7.). Újabb kinevezések csak ősszel, a választások után voltak. Szeptemberben kapott egyetemi tanári kinevezést a Pázmány Péter Tudományegyetem Ókortörténeti Tanszékére a kiváló klasszika-filológus Marót Károly akadémikus (Köznevelés, 1947, 18.). Novemberben pedig Kondor Imre lett a Debreceni Egyetem filozófia professzora. Ugyancsak novemberben nevezték ki a szegedi egyetemre a kiváló néprajztudóst, Bálint Sándort (Köznevelés, 1947, 22.).

A professzori kinevezések terén, az egyetemek átalakításával összefüggésben a Magyar Kommunista Párt szónoka, a volt illegális kommunista orvos Weil Emil, az Orvos-Egészségügyi Dolgozók főtítkára 1947. december 10. parlamenti interpellációjában éles támadást intézett a Pázmány Péter Tudományegyetem két egyetemi tanára, Thienemann Tass Tivadar akadémikus németprofesszor és Déry József, a középkor történet egyetemi tanárai ellen (Országgyűlés Naplója, 1948, 174.). Ekkor azonban már mindketten külföldön tartózkodtak, ezért gyakorlati következményei a kritikának nem lehetett. 1948 januárjában mindösszesen egy professzori kinevezés történt a bölcsészek között. A debreceni egyetem olasz tanszékére ekkor került egyetemi tanárnak Gaetano Trombatore (Köznevelés, 1948, 13.).

A bölcsészprofesszorokkal, az új kinevezésekkel kapcsolatos újabb politikai aktivitás 1948 elején kezdett kibontakozni. Rudas László az MKP egyik legharcosabb ideológus-filozófus képviselője 1948. február 24-i parlamenti felszólalásában kifejtette, hogy a magyar egyetemek sem színvonalban, sem demokratikus tekintetben nem ütik meg egy demokratikus ország követelményeit. Negatív véleménye elsősorban a szellemi tudományokra vonatkozott. Ennek megszüntetése érdekében azt követelte, hogy teremtsenek összhangot a demokrácia szükséglete és az egyetem elzárkózottsága között (Országgyűlés Naplója, 1948, 642.).

Hiteltelen volt az igen szerény tudományos kvalitásokkal rendelkező Rudas számonkérése az egyetemi színvonallal kapcsolatban, ugyanígy hiteltelen volt a demokrácia szükségletének hangsúlyozása annál a képviselőnél,

aki a parlamentben mindenkor diktatórikusan „lépett fel”, és demokrácia-felfogása igencsak bonyolult volt. A kommunista professzorok előterésére utal, hogy áprilisban Fogarasi Béla filozófust, tudománypolitikust kinevezték a Pázmány Péter Tudományegyetemre egyetemi tanárnak. Áprilisban lett egyetemi tanár, ugyancsak a budapesti egyetem pedagógiai tanszékén Várkonyi Hildebrand Dezső (Köznevelés, 1948, 4.). A proletárdiktatúra formálisan is teljes kiépülése után a Magyar Dolgozók Pártja (MDP) létrejöttével „megteremtődtek” a feltételek a demokrácia és a „színvonal” szinkronba hozására azáltal, hogy több, az MKP-ben, illetve az illegális kommunista mozgalomban szerepet játszott pártmunkásból egyetemi tanár lett. Augusztusban kapott professzori kinevezést a szegedi tudományegyetem ókortudományi tanszékére a Magyar Kommunista Párt által támogatott Waldapfel Imre (Köznevelés, 1948, 15.). 1948 októberében lett professzor Andics Erzsébet, a budapesti bölcsészkar világtörténeti tanszékén (Köznevelés, 1948, 19.).

Az 1948-ban kinevezettek között Várkonyi Hildebrand volt az egyetlen, aki nem a párt delegáltja volt, de ugyanakkor élvezte az MKP bizalmát. Valamennyi kinevezett „megfelelt” a népi demokrácia politikai elvárásainak, tényleges tudományos teljesítményük azonban sok kívánnivalót hagyott maga után. Egyesek, mint Fogarasi Béla, már a háború előtt és után is a marxista filozófia népszerűsítése mellett önálló tudományos munkát is folytattak. Andics Erzsébet tudományos teljesítményét felülmúlta ideológiai-politikai tevékenysége. Trencsényi Waldapfel Imre politikai szerepvállalása mindig nagyobb volt kutatási kvalitásánál.

A II. világháború utáni két év habilitációit elemezve megállapíthatjuk, hogy a bölcsész tudományok területén viszonylag kevesen szereztek magántanári fokozatot. 1946 őszétől viszont nagy számban habilitáltak a magyar egyetemeken. 1946 októberében szerzett fokozatot művészettörténetből Csapodi István, történelemből Borsodi István, egyháztörténetből Horváth Konstantin, Bakát István gazdaság- és társadalomtörténetből, Decsényi Béla pedig sajtótörténetből (Köznevelés, 1946, 20.). Novemberben habilitált irodalomból Vatai László (Köznevelés, 1946, 21.), decemberben szintén irodalomtudományból Gábor Rezső, nyelvészetből szerzett fokozatot Hasan Eren. Makkai László a történettudomány területén védett ugyancsak decemberben (Köznevelés, 1946, 23.).

1947 februárjában hadtörténelemből szerzett minősítést Szilágyi János (Köznevelés, 1947, 3.), Nagy Tibor pedig történettudományból (Köznevelés, 1947, 5.). Márciusban habilitált Telegdy Zsigmond nyelvtudományból, Dobrovics Aladár pedig történelemből (Köznevelés, 1947, 5.). 1947 májusában művelődéstörténetből védett Juhász Vilmos (Köznevelés, 1947, 9.), júliusban pedig irodalomból Füst Milán (Köznevelés, 1947, 12.). Augusztusban két védés volt: Kósa János történettudományból, Szendrey Ákos pedig néprajzból

(Köznevelés, 1947, 15.). Szeptemberben lett irodalomból habilitált doktor Kardos László, Kiss Tibor pszichológiából, Hegedűs Lajos pedig nyelvészetből (Köznevelés, 1947, 18.). Októberben szereztek fokozatot: Balogh István történettudományból, Kádár Zoltán klasszika-filológiából, Koczog Ákos pedig irodalomból (Köznevelés, 1947, 20.). Decemberben három védés történt: Harmatta János ókortudományból, Kovács Gyula pedagógiából, Gerevich László pedig régészetből (Köznevelés, 1947, 23–24.).

Meglepő módon az 1948-as évben a habilitációk száma erősen csökkent. Februárban Kardos Lajos pszichológiából, Péter Zoltán és Temesi Mihály filozófiából védett (Köznevelés, 1948, 3.), Jajczay János pedig egyháztörténetből (Köznevelés, 1948, 5.). 1948 áprilisában Tóth Zoltán történettudományi védésére került sor (Köznevelés, 1948, 8.). Ezután hosszú időn keresztül nem került sor habilitációkra, majd augusztusban volt még négy védés, Beke Ödön és Lutter Tibor nyelvészetből, Körmeny Dénes filozófiából és Kádár Jolán színháztörténetből szerzett fokozatot (Köznevelés, 1948, 16.).

BÖLCSÉS Z VEZETŐK AZ EGYETEMEKEN

Az 1946 őszi „fordulat” után a bölcsészettudományok súlya a tudományos életben csökkent, ennek ellenére három egyetemi rektor bölcsészek közül került ki. Az 1946/47-es tanévben Tóth László történész lett a szegedi egyetem rektora (Köznevelés, 1946, 19.). 1946–1948 között Németh Gyula akadémikus, nyelvész a Pázmány Péter Tudományegyetem, Pukánszky Béla akadémikus, irodalomtörténész nyelvész pedig a debreceni egyetem rektora lett (Köznevelés, 1947, 16.). A bölcsészkarok vezetése úgy alakult, hogy 1946 őszétől Mészáros Ede klasszika-filológus lett a debreceni egyetem dékánja, a szegedi dékán pedig a filozófus Mester János (Köznevelés, 1946, 19.). 1947 őszétől Zsirai Miklós nyelvész akadémikus irányította a Pázmány Péter Tudományegyetem Bölcsészettudományi Karát, Pukánszky Béla a debreceni bölcsészkart, Birkás Géza pedig a szegedi BTK-t (Köznevelés, 1947, 16.). A rektorok terén az 1948–49-es tanévben annyi „váltás” történt, hogy továbbra is rektor maradt Németh Gyula a budapesti, Pukánszky Béla pedig a debreceni egyetemen (Köznevelés, 1948, 16.).

A bölcsészkar rektorok kinevezésével kapcsolatosan megállapíthatjuk, hogy szakmailag valamennyien kiváló tudósok voltak, politikai kötődésüket nézve viszont igen eltérő talajon „álltak”. Németh Gyula, a nemzetközileg is jegyzett orientalista 1947-től egyetértő „támogatója” volt a Magyar Kommunista Párt tudománypolitikájának. Tóth László egyetemes történész is elismert kutató volt, aki azonban pártpolitikai szerepet is vállalt. 1947-ben a kisgazdapárt parlamenti képviselője lett. Pukánszky Béla a német kultúra és a magyarországi német irodalom elismert művelője viszont nem kötődött egyetlen politikai irányzathoz és párthoz sem.

A dékánok közül Mészáros Ede a klasszika-filológia területén széles körű és sikeres kutatásokat folytatott, politikai aktivitása nem volt. Mester János filozófus-pedagógus is széles körű tudományos munkásságot fejtett ki, politikával nem foglalkozott, ugyanakkor erős kötelekék fűzték a katolikus egyházhoz. Zsirai Miklós kiváló finnugor nyelvész a magyar nyelvtudomány számos problémáját kutatta, és a szovjet nyelvészeti irodalom eredményeit elsőként „ismerte fel” a magyar nyelvészek között, és a Magyar Kommunista Párt tudománypolitikájának támogatójává vált. Birkás Géza a francia irodalom és kultúra kutatója is színvonalasan művelte tudományterületét, de politikai kötődése nem volt. Bárcei Géza akadémikus nyelvész, kiváló tudományos tevékenysége mellett pártpolitikával nem foglalkozott.

Látható, hogy a nyolc bölcész vezető közül egy professzor vállalt aktív politikai szerepet a polgári erők oldalán mint országgyűlési képviselő, amiről 1948 novemberében azonban lemondott. Két bölcész vezetőnek baloldali pártkötődése volt, egy dékánál kimutatható korábbi katolikus egyházi kapcsolata.

A bölcészrektorok és -dékánok fele arra törekedett, hogy elsődlegesen a szakmáját művelje.

BÖLCÉSZEK A MAGYAR TUDOMÁNYOS AKADÉMIÁN

1946-ban két olyan nagygyűlés volt az Akadémián, amelyen új tagokat választottak. Az első július 24-én, a második már a fordulat után, december 19-én zajlott le. A decemberi MTA gyűlésen a bölcészek közül a testületbe nem került be egyetlen levelező és rendes tag sem, ugyanakkor a tiszteleti tagok közé választották Kodály Zoltán zenetudóst (Feketéné, 1975, 548–550.).

Az 1947. június 8-i tagválasztások alkalmából mindösszesen egyetlen bölcész levelező tagot választottak a testület tagjának: Szemerényi Oszvald nyelvészt. Viszont heten rendes tagok lettek a humán tudományok képviselői közül. Ekkor lett rendes tag Bárcei Géza nyelvész, Alszeghy Zsolt irodalomtörténész, Eckhardt Sándor irodalomtörténész, Kniezsa István nyelvész, Ligeti Lajos nyelvész, Rédei Tivadar irodalomtörténész és Turóczi-Trostler József irodalmár (Feketéné, 1975, 550.).

Az akadémiai tagválasztások jellemzője volt, hogy különösebb társadalmi és politikai véleménynyilvánítás nélkül zajlottak, és hogy a megválasztott új akadémikusok szakmai tudományos teljesítményük alapján lettek az MTA tagjai. Konkrét pártkötődése, illetve politikai aktivitása csak Turóczi-Trostler József irodalmárnak volt, aki azonban már az 1918-as, 1919-es forradalmak idején is szerepet vállalt a baloldalon, a Horthy-rendszer idején a Szociáldemokrata Párt munkásakadémiájának előadója volt, eközben sokoldalú tudományos munkát is folytatott, folyóiratot szerkesztett, 1945 után pedig az SZDP országgyűlési képviselője lett, de szakmai tevékenységével nem szakított politizálása idején.

Az MDP létrejötte után, a formálisan is már létező diktatúra idején 1948. július 2-án ülésezett az MTA következő tagválasztó nagygyűlése, ahol levelező tag lett Fülep Lajos művészettörténész, Gyóni Mátyás bizantológus, Hadrovics László nyelvész, Keresztury Dezső irodalomtörténész, Lakó György nyelvész, Mátrai László filozófus, Rubinyi Mózes nyelvész, Szalai Sándor szociológus, Tolnai Gábor irodalomtörténész, Waldapfel József irodalomtörténész. Ekkor lett rendes tag Zsolnai Béla irodalomtörténész, tiszteleti tag pedig Fogarasi Béla, Lukács György filozófusok, Molnár Erik filozófus-történész, valamint Schöflin Aladár irodalomtörténész (Feketéné, 1975, 550.).

Annak ellenére, hogy a totalitárius diktatúra már kiépült, az új akadémiai tagok összetételében ez még egyértelműen nem képződött le. Voltak olyanok, akik a kommunista tudománypolitika delegáltjai voltak, de közöttük is differenciálni kell. Például Lukács György világszínvonalat képviselt, de viszonylagos színvonalat jelentett Fogarasi Béla és Molnár Erik is. Az 1950-es évek törvénysértésének áldozatává vált, majd a 60-as évektől tudományosan kibontakozó Szalai Sándornak a tudománypolitikai és tudományos szervező tevékenysége ekkor még meghatározóbb volt, mint a tudományos teljesítménye. Az irodalmárok és a nyelvészek eddigi tudományos teljesítményeik alapján lettek akadémikusok, illetve a későbbiek során bizonyított szakmaiságuk alapján megérdemelten kerültek a tagok közé. A polgári irracionalizmust bíráló Mátrai László is a tudományt képviselte az akadémikusok között. Az 1948-ban beválasztott MTA tagok közül hárman az 1948-es „reform” után elsődlegesen politikai okok miatt már nem lehettek a testület tagjai.

A „reform” után az Akadémián 1949. október 31-én megválasztott akadémikusok között több szakmai-tudományos kiválóság is volt, de a taktikát kiszolgáló tudománypolitika koncepciója az 1949-es, majd az azt követő tagválasztásoknál már egyértelműen kimutatható. Ezekben az években a Révai-féle kultúrpolitika következményeként több olyan „tudóst” is bejuttattak az MTA-ra, akik nem tudományos kvalitásuk alapján, hanem politikai érdemeik révén lettek az Akadémia tagjai. Az átszervezés következménye volt az is, hogy az MTA tagjainak nagy részét, százhatvan főt tanácskozó taggá minősítettek, ami által gyakorlatilag megfosztották őket akadémiai tagságuktól. A jogfosztottak 50%-a a humán tudományokhoz tartozott.

IRODALOM

- Beck M. – Peschka V. (főszerk.) (1989): *Akadémiai Kislexikon*. II. kötet. Budapest: Akadémiai Kiadó
- Fekete G.-né (összeáll.) (1975): *A Magyar Tudományos Akadémia Tagjai, 1825–1975*. Budapest: MTA Könyvtára
- Gergely J. – Izsák L. (2000): *A huszadik század története*. Budapest: Pannonica Kiadó

- Kónya S. – Pach Zs. P. (főszerk.) (1975): *A Magyar Tudományos Akadémia másfél évszázada 1825–1975*. Budapest: Akadémiai Kiadó
- Ladányi A. (1989): *Mennyiségi fejlődés és strukturális változások: a felsőoktatás útja a felszabadulás után*. Budapest: Tankönyvkiadó–Oktatáskutató Intézet
- Rákosi S. – Szabó B. (szerk.): (1967): *A Magyar Kommunista Párt és a Szociáldemokrata Párt határozatai 1944–1948*. Budapest: Kossuth Könyvkiadó
- Urbán K. (1985): *Lukács György és a magyar munkásmozgalom*. Budapest: Kossuth Könyvkiadó

Források

Köznevelés (1946–1948)

Magyar Közlöny 1946

Népszava 1947

Nemzetgyűlés Naplója (1952) VI. kötet, Budapest. Hiteles kiadás

Országgyűlés Naplója (1948) III. kötet, Budapest. Hiteles kiadás

Párttörténeti Intézet Archívuma (PTI. Arch.)

Szabad Nép (1946–1948)

RENDSZERSZEMLELETŰ GAZDASÁGITELJESÍTMÉNY-MÉRÉS

MEASURING ECONOMIC PERFORMANCE IN SYSTEM APPROACH

Csath Magdolna

az MTA doktora, professor emerita, Szent István Egyetem, Gödöllő,
magántanár, kutatóprofesszor, Nemzeti Közszolgálati Egyetem
csath.magdolna@gtk.szie.hu

ÖSSZEFOGLALÓ

Fontos, hogy a gazdasági eredményekkel ne csak a döntéshozók, de a társadalom is tisztában lehessen. A gazdasági teljesítmény objektív mérése azonban a globalizált világ gazdaságban egyre nehezebb feladat. Ennek egyik oka, hogy a globális cégek értékláncai mentén keletkező új, hozzáadott érték két fő eleme, a nyereség és a bér, gyakran nagyon eltérően oszlik meg a fejlett és a kevésbé fejlett gazdaságokban. A tőke szabad áramlása lehetővé teszi a profit repatriálását, így a helyben megtermelt érték jelentős arányban vándorolhat vissza a cégek központokba. Ez azt jelenti, hogy a megtermelt érték nem járul hozzá a helyi fejlődéshez. Ezért célszerű górcső alá venni, hogy a gazdasági eredmények mérésére és a gazdaságpolitikai döntéshozatal megalapozására széleskörűen használt makromutatók, közöttük is különösen a GDP – amelyet egyébként nyolcvan éve dolgoztak ki –, alkalmasak-e megalapozott információval szolgálni a valós eredményekről, és alapját képezni a jövőről szóló döntések meghozatalának korunk bonyolult viszonyai közepette.

A cikk bemutatja a GDP-mutató tipikus gyenge pontjait, és felhívja a figyelmet azokra a veszélyekre, amelyeket ezen gyenge pontoknak a döntések során történő figyelmen kívül hagyása előidézhet. Részletesebben is elemzi, hogy a múlt eredményeit pénzügyi szemléletben mérő GDP-re mint legfontosabb teljesítménymérő mutatóra való támaszkodás a gazdaságpolitikai döntéshozatalban miért nehezíti a rendszerszemléletben való gondolkodást, a gazdasági tevékenységek okozta társadalmi és környezeti hatások, az externáliák jövőorientált számbavételét. Figyelmeztet arra is, hogy a GDP káros következményekkel járó gazdasági műveleteket eredményként képes feltüntetni, miközben olyan fontos emberi tevékenységeket, amelyek pénzben nem mérhetők, nem tud értékként figyelembe venni. Kitér továbbá az értéklánccok gazdasági teljesítményt és társadalmi lehetőségeket befolyásoló szerepére, az alacsony szintű gazdasági diverzifikáltság és a gazdasági-társadalmi fejlettség regionális egyenszilárdsága hiányának veszélyeire. Bizonyítja, hogy a humán tőke hatékony hasznosítása a legfontosabb teljesítménynövelő tényező nemcsak a gazdaság, de a társadalom szempontjából is. Következtetésként megfogalmazza, hogy olyan döntés-előkészítő módszertant célszerű alkalmazni, amely képes figyelembe venni, hogy a gazdaság további rendszerek része, ezért növekedésének elsőrendű célként való meghatározása egyrészt fenntarthatósági korlátokba ütközhet, másrészt hosszabb távon előnytelen fejlődési út választására vezethet.

ABSTRACT

It is important to measure economic performance in order to guarantee that not only the decision makers, but also the members of the society shall be able to judge economic results. However measuring economic performance objectively in a globalized world economy is an increasingly complicated task. One reason is that the new added value created along the value chains of global businesses is often distributed very differently between profit and wages in the developed and less developed economies.

The free movement of capital facilitates the repatriation of profit, therefore a significant proportion of the locally created new value can migrate back to the corporate headquarters, this way not contributing to local development. Therefore it is justified to think over whether those macro indicators, among them especially the GDP – developed about 80 years ago – which are applied extensively to measuring economic performance and also for making economic policy decisions are still suitable for providing reliable information on real results and for future oriented decision making in the present fast changing and very complex environment.

The article presents the typical weaknesses of the GDP indicator, and directs attention to those potential problems which may be caused by not taking into consideration these deficiencies in the decision making process. It details why relying on GDP – as the best indicator of economic performance for decision making when it measures basically past results from a financial perspective – may impede system thinking and future oriented consideration of social and environmental impacts like externalities of economic activities. The article also warns that GDP may take into account environmentally or socially harmful economic activities as economic achievements, while it is unable to count in socially very useful but financially unmeasurable human activities.

The article also discusses the impacts of global value chains on economic performance and social opportunities, and reminds to the potential dangers of low level economic diversification and weak homogeneity in regional economic and social development. It argues that efficient utilization of human capital is the most important factor for improving performance not only from the economic but also from societal point of view.

The key conclusion of the paper is that makers of economic policy have to turn to applying more complex measurement systems, which are capable of treating the economy as a subsystem of other systems. By doing so they would be able to avoid collision between aiming at growing the economy and considering sustainability requirements. The changing of the measurement system could also increase the probability of governing the economy and the society to the most favorable development path.

Kulcsszavak: gazdasági teljesítmény, GDP, externáliák, rendszerszemlélet, fenntarthatóság, értéklánc

Keywords: economic performance, GDP, externalities, system view, sustainability, value chain

„Jogot tehát az árnyalatnak,
melyben a holnap rajza áll
s a kivételnek,
mely holnapra talán szabály;”
(Illyés Gyula: *Óda a törvényhozóhoz*, 1965)

BEVEZETÉS

A gazdasági teljesítmény objektív mérésének igénye már hosszú ideje foglalkoztatja a közgazdászokat. Az objektivitáson általában azt értették, hogy jól mérhető mutatókat kell kialakítani és vizsgálni. Ehhez a gondolathoz járult az a meggyőződés is, hogy az emberek és a cégek gazdasági kérdésekben egyaránt objektív és racionális döntéseket hoznak, tehát a gazdasági teljesítményt is objektív és mérhető tények befolyásolják. Ennek következtében alakult ki konszenzus arról, hogy a gazdasági teljesítmény mérésére leginkább néhány makromutató, közöttük is nagy hangsúllyal a bruttó hazai termék (GDP) mutató a legalkalmasabb. Továbbá az objektív mérhetőség érdekében zárt rendszerként be is határolták a gazdaság területét, nem véve figyelembe, hogy valójában a gazdaság további nagyobb rendszerek része. A körünkben zajló technológiai, társadalmi és környezeti változások azonban egyaránt arra figyelmeztetnek, hogy a gazdasági eredményeket csak ezek összefüggésében szabad értékelni. A pénzmozgások mérésén alapuló makromutatószám-rendszer, bár kétségtelenül fontos információt közvetít a döntéshozók számára, mégsem képezheti a jövőről való mai döntések kizárólagos alapját. A döntés-előkészítési folyamatban helyük van a számszerűen nem pontosítható és mérhető, úgynevezett puha mutatóknak is, és nem engedhető meg az sem, hogy gazdasági döntéseket azok társadalmi és környezeti hatásainak alapos elemzése nélkül hozzuk meg.

MI A BAJ A GAZDASÁGI TELJESÍTMÉNY MÉRÉSÉNEK ELFOGADOTT MÓDSZERÉVEL, A GDP-VEL?

A nagy gazdasági válság kapcsán merült fel az az igény, hogy az országok gazdasági teljesítményét objektív statisztikai adatokkal kellene mérni.

A feladatot Simon Kuznets, aki az USA-ban a Gazdaságkutatás Nemzeti Irodájának (National Bureau of Economic Research, NBER) közgazdásza volt, oldotta meg 1937-ben. Abból indult ki, hogy ehhez az egyének, cégek és a kormány gazdasági teljesítményét kell megmérni, és egyetlen mutatóba foglalni, amelyet „gross domestic product”-nak, azaz bruttó hazai terméknek (GDP) nevezett el. A GDP-mutató tehát éppen nyolcvan éves. Ez alatt az idő alatt hatalmas változások zajlottak le a világgazdaságban. A GDP hitelességét érintő talán legnagyobb változás a világgazdaság globalizálódása, a globális cégbirodalmak létrejötte, és tevékenységük gyors nemzetközi elterjedése volt. Ezek a vállalkozások, a tőke

szabad áramlása következtében cégeket telepítenek a világ több országába is, az ott megtermelt új értéket pedig tetszés szerint ruházzák be helyben, utalják másik telephelyükre vagy esetleg oda, ahol a központjuk van. Miért jelent ez gondot? Azért, mert a létrehozott profit a GDP része. Ily módon szerepel a nemzeti számlákban mint helyben létrehozott érték. Viszont, ha elviszik az országból, akkor helyben megtermelt, de helyben nem hasznosítható érték keletkezik, vagyis a megtermelt érték nem járul hozzá a helyi fejlődéshez.

További, a globális cégek stratégiájával összefüggő jelenség, hogy általában költségminimalizáló céllal hoznak létre külföldön telephelyeket. A költségek között fontos tétel a bér, amely szintén a GDP összetevője. Ezért a helyi fejlődési lehetőségek megítélése szempontjából fontos jellemző az is, hogy a létrehozott új értéken belül mekkora a helyben maradó bér és a helyben megtermelt, de nem feltétlenül helyben hasznosított profit egymáshoz viszonyított aránya. Galgóczy Béla (2017) szerint 2015-ben az EU-ban átlagosan a megtermelt összes új érték közel 70%-a volt a bérek aránya. Ezzel szemben ez az érték Kelet-Közép-Európában átlagosan 10-15%-kal alacsonyabb, például Magyarországon csak 55,7%. A Központi Statisztikai Hivatal (KSH) legfrissebb adatai szerint a 2017. harmadik negyedévben Magyarországon összesen létrehozott bruttó hozzáadott érték – 8158,8 Mrd Ft – átlagosan 41,4%-a volt bér és kereset, és 58,6%-a működési eredmény, vagyis profit. Az arány valamivel kedvezőbb a szolgáltatások (49,8–50,2%) és kissé kedvezőtlenebb az ipar (44,4–55,6%) esetén, amelynek oka, hogy a külföldi nagyvállalkozások telephelyeinek jelenléte jellemzőbb az iparra, mint a szolgáltatásokra.

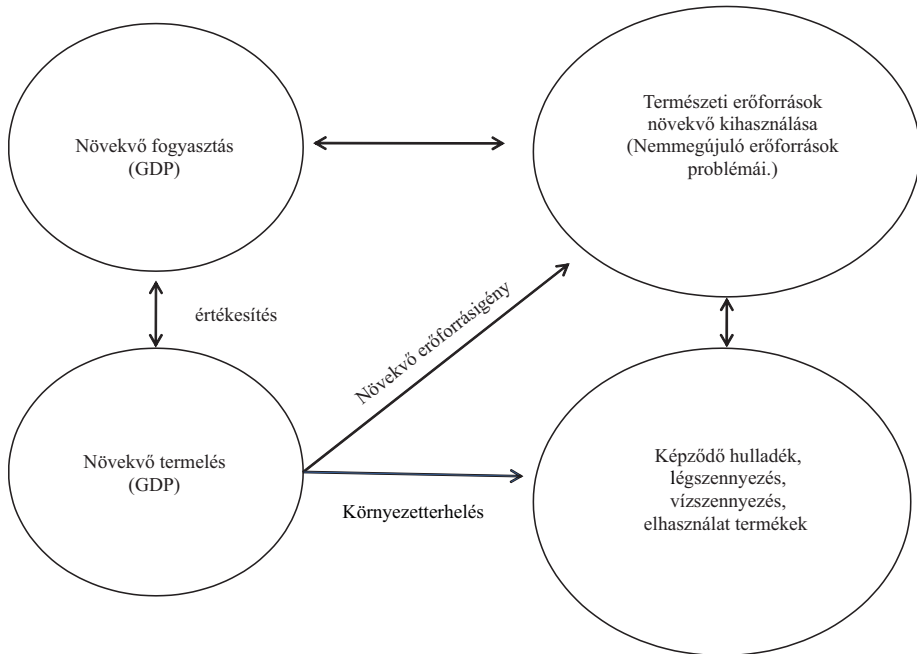
Megjegyzendő, hogy az Eurostat adatai szerint 2015 óta Nyugat-Európa országaiban is – eltérő mértékben, de – csökken a megtermelt új értéken belül a bérek és keresetek aránya a profithoz képest. A GDP-mutató tehát jelentősen javulhat úgy is, hogy a profit nő meg, de az nem járul hozzá a helyi fejlődéshez. Minél nagyobb egy ország gazdaságában a külföldi vállalkozások aránya, és minél inkább az alacsonyabb bérek jelentenek vonzerőt a betelepülésre, annál nagyobb az esélye annak, hogy a helyben megtermelt profit aránya nagyobb lesz, mint a bérearány. Ugyanakkor csak a bérekről feltételezhetjük, hogy azokat a munkavállalók döntő többségükben helyben költik el. A profitot viszont, az adózás után, a cégek saját döntésük alapján el is vihetik a megtermelés helyéről. Ezért a GDP-mutató az ilyen országokban alkalmatlan a valós fejlődés mérésére.

TOVÁBBI PROBLÉMÁK: A GAZDASÁG EGY NAGYOBB RENDSZER RÉSZÉ

A GDP, mint láttuk, a gazdasági teljesítményt sem tudja korunkban már jól mérni. De az is hiányossága, hogy csak azt tudja mérni, ami pénzben mérhető. Így nem növeli a GDP-t például a közösségért térítésmentesen végzett munka, a

gyermekgondozás vagy az idősokkal való törődés. Pedig ezek nemcsak egyszerű munkavégzést, hanem a társadalom szempontjából komoly értéket teremtő tevékenységeket jelentenek. Érzéketlen továbbá a társadalmi és környezeti hatásokra is. Például, ha egy parkban kivágnak minden fát azért, mert oda egy új épületet akarnak felhúzni, akkor a fák kivágása – akkor is, ha az indokolatlan, mert az épületet máshol is fel lehetne építeni – növeli a GDP-t. De szintén növeli a GDP-t a túlduzzasztott turizmus, amely esetleg az éjszakai mulatozások miatt élhetlenné teszi a környezetben lakók életét.

Mindezen problémák ellenére a gazdaságpolitikusok előszeretettel használják a GDP-t a gazdasági teljesítmény, eredményesség mérésére, sőt legfontosabb célként annak növelését jelölik meg. Mire ösztönzik ezzel a gazdaságot és a társadalmat? Természetesen a pénzben mérhető tevékenységek állandó bővítésére. Ennek a körforgásnak egyszerűsített modelljét látjuk az *1. ábrán*.



1. ábra. A GDP-vel mért gazdasági körforgás és externáliái

Az *1. ábrából* kitűnik, hogy az egyre bővülő termelés és fogyasztás a környezetet terhelő kibocsátással jár, amelynek költségeit általában a cégek nem térítik meg, hanem azok fedezését a társadalomra hárítják. A gazdaság eredményeit pedig a GDP-mutató a gazdaságon kívüli hatások nélkül, vagyis úgy méri, mintha a gazdaság elszigetelt, önmagában elemezhető rendszer lenne.

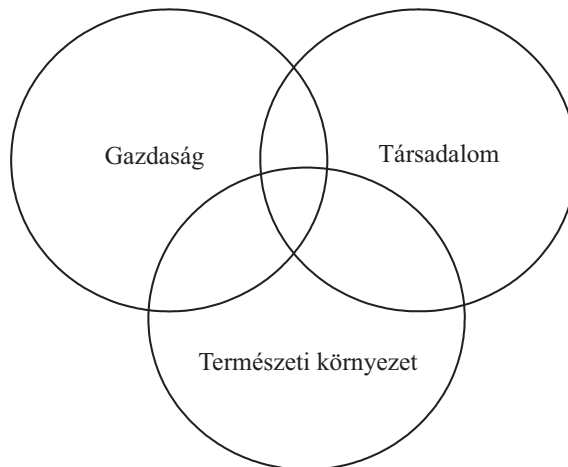
A GAZDASÁG TOVÁBBI NAGYOBB RENDSZEREK EGYIK ELEME

A döntéshozók is hajlamosak rá, hogy a gazdaságot zárt rendszerként szemlélve, a működéséhez szükséges természeti erőforrásokat egyszerűen megvásárolható *input*ként kezeljék, a keletkező hulladék tárolásának, feldolgozásának vagy megsemmisítésének költségeit pedig a társadalom által viselendő költségeknek tekintsék.

Az ipari és kereskedelmi tevékenység másokra, köztük a bioszférára gyakorolt olyan káros hatásait, amelyek költségei nem jelennek meg az árakban megtérítendő kiadásként, externáliáknak nevezzük. A levegőbe, a termőföldre vagy a természetes vizekbe kerülő szennyeződés ilyen közvetlen externália. De annak tekinthető a gyártás során keletkező szemét vagy az elhasználódás után hulladékká váló termék. Újabban az egészségtelen munkakörülmények miatt megbetegedő munkavállaló kezelésének az egészségügyben megjelenő költségeit is externáliának tekintik. Az externáliák tehát az üzleti világ által másokra áthárított költségek.

Az externáliák figyelmen kívül hagyása pedig a fenntartható fejlődést veszélyezteti, amelyet a Brundtland-bizottság tanulmánya (Brundtland, 1987) a következőképpen definiált: „A fenntartható fejlődés olyan fejlődés, amely anélkül elégíti ki a jelen generációk igényeit, hogy azzal veszélyeztetné a jövő generációk igényeinek kielégíthetőségét.”

Erkölcsi és tisztánlátási szempontból jogos, illetve fontos tehát, hogy a gazdaságot annak szereplőivel együtt az azt körülvevő nagyobb rendszerekkel való összefüggésében vizsgáljuk.

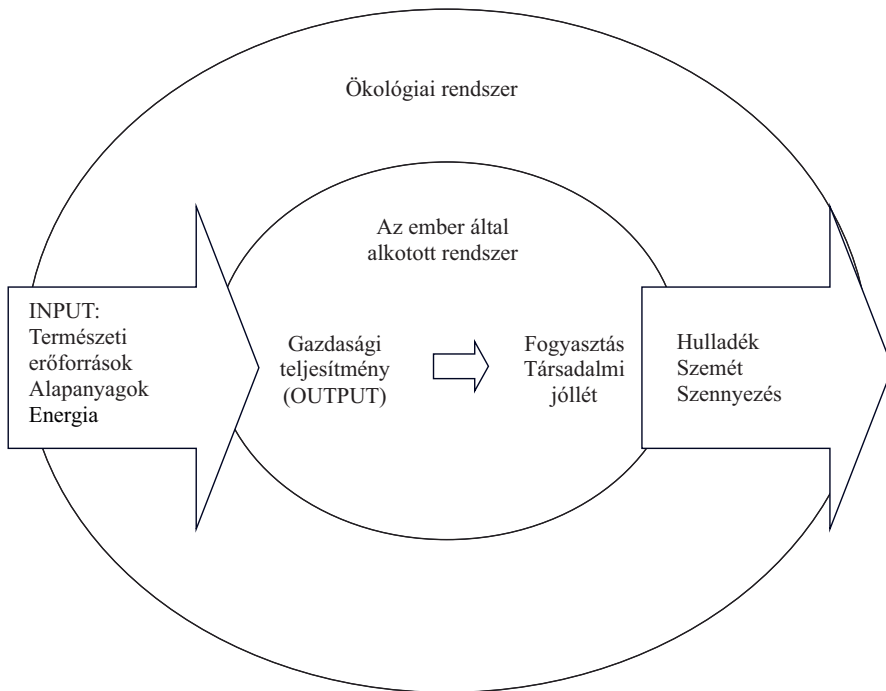


2. ábra. A fenntarthatóság átfedő körei

Forrás: Brundtland Report, 1987

A gazdaság, ahogyan a 2. ábra szemlélteti, a társadalmi és természeti környezet része, ezért hatásai miatt a fenntarthatóság szempontjából nem vizsgálható ezektől a rendszerektől elszigetelten.

Az ökológiai rendszerrel való szoros kapcsolatát, vagyis azt, hogy hol jelennek meg a gazdaság működésének input igényei és annak externáliái, a 3. ábrán látjuk.



3. ábra. A gazdasági működés az ökológiai és társadalmi rendszerben

Forrás: Clugston alapján (2012)

A gazdasági teljesítmény GDP-vel való mérése tehát egyrészt kiszakítja a gazdasági tevékenységeket azokból a rendszerekből, amelyeknek nyilvánvaló része. A növekedést és a fogyasztást a lehető legfontosabb eredménynek tünteti fel úgy, hogy közben azok káros hatásait nem méri. Ezért a GDP-növelés mint legfontosabb cél alapján történő döntéshozatal már rövid távon is okozhat gondokat, de hosszabb távon biztosan rossz irányba mozdítja el a gazdaság és a társadalom fejlődését. Hiszen jelentősen nőhet a GDP értéke úgy, hogy közben megnő a környezetkárosodás szintje, vagy a növekedés eröltetése miatt romlik a munkavállalók egészségi állapota, károsodik a fenntartható fejlődés fontos elve. De tovább-

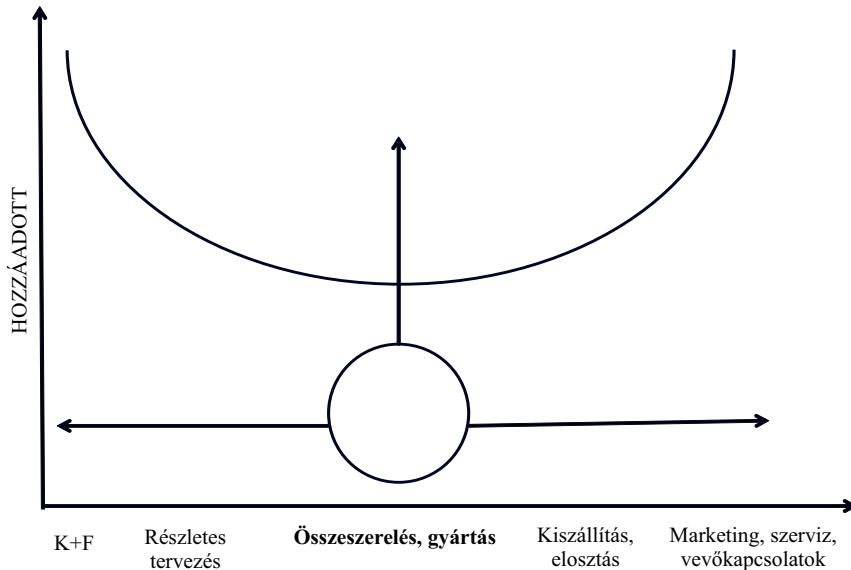
bi károk, a jövő szempontjából jelentős torzulások is bekövetkezhetnek a GDP növelésének velejárójaként. Például komoly aránytalanságok, egyensúlytalanságok alakulhatnak ki gazdasági ágazatok, régiók és társadalmi csoportok között. A gazdaságban létrejöhet egy monokultúras szerkezet, amelyben az eredmények kevés, nagyra nőtt ágazattól függenek. Ez azért veszélyes, mert éppen úgy, mint a természetben a biodiverzitás, a gazdaságban is a sokszínűség, a diverzifikáltság a kiegyensúlyozott fejlődés feltétele.

Másrészről a nagyon leszakadó gazdasági területek és elszegényedő embercsoportok rontják az egész rendszer működésének hatékonyságát, hiszen ahogy a fizikában is a stabilitás feltétele az egyenszilárdság, a gazdaságban is veszélyt okoznak a fejlődésben részt venni nem tudó, gyenge láncszemeknek tekinthető lemaradó országrészek. De torzító hatású lehet a gazdaság szerkezetére nézve az is, ha például a GDP-vel mért gazdasági növekedés maximalizálása érdekében a gazdaságpolitika a külföldi tőkebefektetőket, attól függetlenül, hogy azok milyen tevékenységek végeztetését tervezik, vagyis milyen minőségű munkahelyeket hoznak létre, jelentős támogatásokkal ösztönzi betelepülésre. Ha ugyanis olyan nemzetközi cégek települnek az országba, amelyek a teljes tevékenységi láncolat, vagyis az értéklánc csupán rövid szakaszát valósítják meg, az, még ha hozzájárul is a gazdasági növekedéshez, egyrészt kockázatos függőséget teremthet, másrészt a gazdasági szerkezet túlzott leegyszerűsödését és ezzel a kínált munkalehetőségek szűkülését, ezek következtében pedig a lehetségesnél alacsonyabb szintű hozzáadott érték teremtését eredményezheti. További káros hatás lehet a hazai kisvállalkozások háttérbe szorulása.

AZ ÉRTÉKLÁNCOK PROBLÉMÁI

Az értéklánc egy teljes termelési vagy szolgáltatási ciklus a termék vagy szolgáltatás ötletének felmerülésétől a tervezésen, a létrehozáson át a piacra vitelig. Az értéklánc különböző szakaszaiban eltérő nagyságú érték keletkezik. A legnagyobb hozzáadott érték a folyamat elején, az ötlet megszületésekor és tervbe foglálásakor, valamint az eredmény hasznosulásakor, azaz a piacra kerülésekor jön létre. A legalacsonyabb érték pedig az előállítási szakaszban, az összeszerelésnél keletkezik. Ezt szemlélteti a 4. ábra, egy termelési folyamat példáján.

A hozzáadott érték nagysága pedig összhangban van a bérszínvonallal. Ez egyben arra is rámutat, hogy minél hosszabb értéklánc van jelen egy országban, annál többféle lehetséges munkahelyet és eltérő fizetési szintet kínálnak a munkaadók a munkavállalóknak. Az értékláncok hossza az adott országban azonban más oldalról is fontos jellemző. Azt állíthatjuk, hogy a gazdaság sikeressége nem önmagáért való, hanem annak hatásait a társadalomnak is érzékelnie kell. A társadalom sikerességének egyik jele, ha a munkavállalóknak lehetőségük van arra,



4. ábra. Hozzáadott érték az értéklánc mentén

hogyan képességeiket egyrészt maximálisan hasznosíthatják, másrészt állandóan továbbfejleszthessék. Ez egyébként nemcsak az egyénnek, hanem a társadalomnak is érdeke. Ha a munkavállalóktól a munkahelyek kevesebbet igényelnek, mint amennyire képesek, akkor kihasználatlan képességekről, másképpen alulfoglalkoztatottságról beszélünk. Viszont az összeszerelési munka tipikusan nem igényel mást, mint fizikai és idegi tűrőképességet, vagyis kreativitásra, új ötletek, megoldások keresésére általában nem kínál lehetőséget.

Ebből az következik, hogy ha az összeszerelő munkahelyek aránya túl magas, akkor felléphet a mérsékelt bérű alulfoglalkoztatás. Ha pedig ezt a helyzetet együtt elemezzük az egyéni elégedettség mérő Abraham Maslow-i (1943) hierarchia társadalmi üzenetével, akkor arra a következtetésre kell jutnunk, hogy ha egy országban sok az összeszerelő munkahely, akkor azok a munkavállalók többségének szükségleteit csak a maslowi hierarchia alsó szintjein képesek kielégíteni akkor is, ha igényük és képességeik lennének a felsőbb szintekre lépésre is. Ez alulfoglalkoztatottságot, a humán vagyont nem hatékony hasznosítást, tudásvesztést okozhat, ami a társadalom dinamikus fejlődését is gátolhatja.

Érdekes idézni ezzel kapcsolatban a statisztikai adatokat. A *Jó állam jelentés* (Kaiser, 2017) *Pénzügyi stabilitás és gazdasági versenyképesség* fejezete bemutatja, hogy mi jellemzi a magyar gazdaságban működő értéklánc hosszát. Az Eurostat adataira támaszkodó elemzés szerint a szolgáltatások esetén hosszabbak, az ipar esetén rövidebbek az értékláncok. A 2004-es évtől 2015-ig terjedő



5. ábra. Maslowi szükséglethierarchia

Forrás: Maslow, 1943

vizsgálat szerint az összes új értékből a helyben előállított, bruttó hozzáadott értékkel mért új érték, ami egyben az értéklánc hosszát is jellemzi, ezen időtávon a szolgáltatások esetén átlagosan 57%, az ipar esetén pedig 26-27%. Ez azt mutatja, hogy az ipar esetében sokkal jelentősebb a technológiai tudás, részegység stb. formájában kívülről behozott és nem helyben előállított új érték.

A KSH legfrissebb, 2017. harmadik negyedévi adatai is megerősítik a kutatás eredményeit. Ebben az időszakban a szolgáltató szektor a teljes új, hozzáadott érték 65, az ipar pedig 25%-át hozta létre. Vagyis a szolgáltató szektorban a teljes értéklánc nagyobb szakasza van jelen, ami lehetővé teszi azt is, hogy színesebb és jobban fizető munkahelykinálat álljon rendelkezésre a munkavállalók számára. Az iparban pedig az értékláncok rövidségét az összeszerelő tevékenység elterjedtsége okozza.

Az értékláncok hossza tehát jelentősen határozza meg a létrehozható új érték nagyságát és egyben azt is, hogy a társadalom számára milyen munkalehetőségek jönnek létre. A környezeti hatások szempontjából pedig fontos szempont az is, hogy az összeszerelő tevékenységnek jelentős az energiaigénye. Ezeknek a tényezőknek a finomhangolású elemzésére azonban a gazdasági eredményeket mérő tipikus makromutatók nem alkalmasak.

ÖSSZEZÉS

Vannak olyan gazdasági mutatók, amelyek részben kiküszöbölik a GDP használatának gondjait. Ilyen mutató például a bruttó nemzeti jövedelem (Gross National Income, GNI). Azonban ennek a mutatónak a használata nem terjedt igazán el. Az elemzők és a gazdaságpolitikuskok egyaránt a GDP-t részesítik előnyben a gazdasági teljesítmény mérésekor.

Nyilvánvaló, hogy a gazdasági teljesítményt mérnünk kell, és ehhez használhatjuk a ma is kiterjedten alkalmazott hagyományos makromutatókat, közöttük a GDP-t is. Azonban, elsősorban a jövőre vonatkozó döntéshozatalnál, figyelemmel kell lenni ezen mutatók korlátaira. A makromutatók szigorúan a mérhető, pénzben kifejezhető értékekkel számolnak, ráadásul csak a múlt és a jelen eredményeit tükrözik. A gyors változások közepette viszont, amikor mai döntéseinkkel az egyre nehezebben előrebecsülhető és szerteágazó hatásoknak kitett jövőre akarunk felkészülni, több lábón álló mérőrendszert kell alkalmaznunk. Olyan mutatószám-rendszerre van szükség, amely az aktuális gazdasági eredmények mérése mellett a hosszú távon fontos gazdasági szerkezetváltozást, az értékláncok hosszának alakulását és a gazdasági teljesítmény externáliáit rendszerbe foglalva képes elemezni, és ezzel megbízhatóbb döntés-előkészítő segédeszközként tud szolgálni.

Az ilyen mérés lehetővé teszi a hatásvizsgálatokat: nevezetesen annak feltérképezését, hogy a mai gazdasági döntések és azok velejárója, az erőforrás-allokáció milyen hatással lesznek a jövő gazdasága mellett a jövő társadalmára és természeti környezetére. Fontos, hogy amint a természetben a biodiverzitás a válságokkal szembeni ellenálló képesség egyik fontos feltétele, éppen úgy a gazdaságban is elemezzük és tudatosan alakítsuk a gazdasági szerkezetet, biztosítva a gazdasági sokszínűséget, egészséges ellenálló képességet.

Másrészről vizsgáljuk a gazdasági és társadalmi fejlődés kiegyensúlyozottságának alakulását, a gazdasági és társadalmi fejlettség egyenszilárdságát azért, hogy a gyenge láncszemek ne rontsák le az egész rendszer teljesítményét. Ezekre a mérésekre a hagyományos mutatók önmagukban nem alkalmasak, ezért a rájuk alapuló döntéshozatal akár sikertelen fejlődési pályára is irányíthatja az országot. Szükség van tehát rendszerszemléletre és hosszabb távon történő gondolkodásra, és ezzel összhangban arra, hogy a gazdasági, társadalmi és környezeti rendszerek kapcsolódó területein zajló változásokat is mérni tudjuk.

Biztosítani szükséges, hogy elsősorban az erőforrások szétosztásánál az összefüggő rendszerek elemeit ne egymástól függetlenül kezelje a döntéshozás, mert az párhuzamosságokat hoz létre, és így lehetetlenné teszi a hatékony erőforrás-allokációt, sőt egymásnak ellentmondó, egymás hatását lerontó döntésekhez is vezet. Az összhangot pedig elsősorban a gazdasági, társadalmi és környezeti rendszerek között kell megteremteni a 6. ábrán látható egyszerűsített gondolatmenet segítsé-

gével. Az ebben a szemléletben megvalósuló döntéshozatal lehetővé tenné, hogy a gazdasági növekedés mellett a társadalmi és környezeti fenntarthatóság szempontjai is a döntéshozók asztalára kerüljenek.

Gazdaság	Társadalom	Környezet
Hatékonyság	Emberi szükségletek	A környezet megőrzése a jövő generációk számára
Termelékenység Gazdasági szerkezet	„Jól-lét”, életminőség	
Gazdasági növekedés	Tanultság, egészség	Energia- és anyaghatékonyság
Versenyképesség	Egyenlőtlenségi szint	Biodiverzitás
Gazdaságpolitika ← → Társadalompolitika ← → Környezeti politika		

6. ábra. Az elszakított rendszerek összekapcsolása

Forrás: saját szerkesztés

IRODALOM

- Brundtland, G. H. (1987): *Our Common Future*. United Nations, <http://www.un-documents.net/our-common-future.pdf>
- Clugston, Ch. O. (2012): *Scarcity: Humanity's Final Chapter?* Port Cahrlotte, Florida: Booklocker. Com. Inc.
- Galgóczi B. (2017): *Why Central and Eastern Europe Needs a Pay Rise?* Working Paper 2017. 01. Brussels: Etui., European Trade Union Institute, <https://www.etui.org/content/download/31470/290283/file/WP+2017-01-pay+rise-HD-web+version+finale.pdf>
- IMD (2017a): *IMD World Talent Ranking 2017*. Lausanne, Switzerland: IMD, Institute for Management Development, https://www.imd.org/globalassets/wcc/docs/talent-ranking/Talent_Ranking_2017_web.pdf
- IMD (2017b): *Digital Competitiveness Ranking 2017*. Lausanne, Switzerland: IMD, Institute for Management Development, https://www.imd.org/globalassets/wcc/docs/release-2017/world_digital_competitiveness_yearbook_2017.pdf
- Kaiser T. (szerk.) (2017): *Jó állam jelentés 2017*. Budapest: Nordex Nonprofit Kft.–Dialóg Campus Kiadó, http://akfi-dl.uni-nke.hu/jo_allam_jelentes_2017/JoAllamJelentes_2017_web.pdf
- Schmuck E. (szerk.) (2010): *Jövőkereső: A Nemzeti Fenntartható Fejlődési Tanács jelentése a magyar társadalomnak*. Budapest: NFFT, http://kornyezetineveles.hulladekboltermek.hu/files/pdf/jovokereso_az_NFFT_jelentese_a_magyar_tarsadalomnak.pdf
- Maslow, A. H. (1943): A Theory of Human Motivation. *Psychological Review* (Washington DC), 50, 4, 370–396. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.318.2317&rep=rep1&type=pdf>

BETHLEN-KÉZÍRÁS A KOLOZSVÁRI TÉRKÉPTÁRBAN

THE AUTOGRAPHY OF GABRIEL BETHLEN IN A MAP COLLECTION OF CLUJ

Bartos-Elekes Zsombor

PhD, egyetemi docens, Babeş–Bolyai Tudományegyetem (Kolozsvár), Magyar Földrajzi Intézet
bezsombor@yahoo.com

ÖSSZEFOGLALÁS

Bethlen Gábor erdélyi fejedelem könyvtárából eddig öt kötetet sikerült azonosítani. A kötetek egyike egy 1592-es latin nyelvű Ortelius-világatlasz, amelybe Bethlen 1608 júliusában írt bele: latin nyelvű keltezés és aláírás mellett egy magyar nyelvű fohászt. Később az atlasz Csáky Katalin tulajdona lett, majd Csáky Györgytől jutott a kolozsvári Erdélyi Múzeum könyvtárába, amiről a könyvtáros Szabó Károly 1882-ben adott hírt; jelenleg az atlaszt a kolozsvári Központi Egyetemi Könyvtár tárolja. A tanulmány áttekinti a példányra vonatkozó irodalmat, bemutatja a példányt, először közöl fényképet a beírásról, rekonstruálja a példány történetét.

ABSTRACT

From the library of Gabriel Bethlen, Prince of Transylvania only five books has been identified. One of them is the Latin edition of the world atlas of Abraham Ortelius published in 1592, with an autography of Bethlen from July 1608: he had entered the date and signature in Latin, and a petition to God in Hungarian. Later the atlas had become the property of Katalin Csáky, and then arrived from György Csáky to the library of Transylvanian Museum, which was reported by the librarian Károly Szabó in 1882; the volume is now kept by the Central University Library of Cluj. This reviews the literature of the topic, presents the volume, publishes a copy of the note the first time, and reconstitutes the volume's history.

Kulcsszavak: Abraham Ortelius, atlasz, Bethlen Gábor, 1608, Csáky Katalin, Csáky György, Szabó Károly, Erdélyi Múzeum könyvtára, kolozsvári Központi Egyetemi Könyvtár

Keywords: Abraham Ortelius, atlas, Gabriel Bethlen, 1608, Katalin Csáky, György Csáky, Károly Szabó, the library of Transylvanian Museum, Central University Library of Cluj

A TÉMA IRODALMA

Szakirodalomban több helyen találkozhatunk olyan adatokkal, hogy Bethlen Gábor az általa forgatott atlaszba beleírt, és ez megtalálható Kolozsváron.

Az egyik ilyen információt Plihál Katalin írta le (2009, 10.), aki szerint Borbély Andornak van olyan feljegyzése, hogy Bethlen Mercator-atlaszt használt utazásai során, a térképeket széljegyzetekkel egészítette ki, és az atlasz 1942–44-ben megvolt a kolozsvári Egyetemi Könyvtárban. A szerző hozzátette, hogy az atlasz mai sorsáról semmit sem tudunk.

2013 márciusában a fenti információ alapján kerestem a kolozsvári Központi Egyetemi Könyvtár Térképtárban¹ a szóban forgó atlaszt, de nem jártam eredménnyel. A könyvtárosok nem tudtak Bethlen-kézírásos atlaszról. A térképtárnak nincs Mercator-atlasza. Ugyan van pár Mercator-térképe (*Sclavonia, Croatia, Bosnia cum Dalmatiae parte; Walachia, Servia, Bulgaria, Romania; Transylvania*), de azokon nem találtam kézírást. Plihál Katalin nem adott meg hivatkozást Borbély Andor feljegyzéséről, így keresnem kellett az eredeti Borbély-feljegyzést. A Bende-fy László (1972) által felsorolt Borbély-tanulmányok közül az 1942 utáni lehetőségekben szintén nem találtam meg a Plihál Katalin által említett feljegyzés eredetijét; aki később megkeresésemre sem tudta pontosítani az adatot. Így nem sikerült azt sem azonosítanom, hogy Borbély Andor milyen munkájában írhatott erről.

Később, egy másik nyom viszont eredményre vitt. Plihál Katalin hívta fel a figyelmemet Szabó Károly kolozsvári történész, könyvtáros rövid közlésére (Szabó, 1882), amely szerint gróf Csáky György váralmási könyvtárával jutott az Erdélyi Múzeum könyvtárába egy 1592-es Ortelius-atlasz, benne Bethlen Gábor saját kezű feljegyzése. A rövid híren kívül Szabó Károly idézte a Bethlen-kézírást szövegét.

A kolozsvári, egykori Bethlen-féle Ortelius- (és nem Mercator!) atlasz többször jelenik meg a szakirodalomban, leginkább Szabó Károly írására hivatkozva, más megvilágításokkal kiegészítve azt. Az egyik korai ilyen hivatkozás a szintén a kolozsvári egyetemi könyvtárban dolgozó Kelemen Lajosé (1907, 59.) volt, aki – bár Fröhlich Dávidról értekezett, mégis – fontosnak tartotta Bethlen Orteliusával indítani írását. A hetvenes–nyolcvanas években Jakó Zsigmond és Vita Zsigmond is Bethlen könyvtáráról írt tanulmányt. Az előbbi (Jakó, 1976, 207.) felsorolta azt az öt kötetet, amely fennmaradt, és biztosan része volt Bethlen könyvtárának, közte az Ortelius-atlasz. Az utóbbi (Vita, 1983, 9–10.) idézte Milotai Nyilas Istvánt, hogy a fejedelem táborozás alkalmával könyveket hordozott magával, és azokat olvasta, a szerző szerint ilyen volt a Kolozsváron fennmaradt Ortelius-atlasz. Csetri Elek Bethlen-életrajzában (1992, 36., 57., 145.) – amellet, hogy elírva, 1552-esnek jelölte az atlaszt – hozzátette, hogy Bethlen Báthori Gábor fejedelem tanácsosaként,

¹ A kolozsvári Központi Egyetemi Könyvtár Térképtárának katalógusa szkennelve elérhető az alábbi címen: URL1.

1608-as konstantinápolyi útja elején, Déván írt bele az atlaszba. Közölte továbbá – Szabó Károly után ismét – a bejegyzés szövegét. Csetri Elek szerint Bethlen utazásai során vitte magával azt. Az újabb szakirodalomból Monok István (2010, 55–56.) megadta, hogy az öt megmaradt kötetből három a fejedelem számára készült kötésben *supralibrosával* maradt ránk. R. Várkonyi Ágnes (2013, 28–29.) Bethlen Ortelius-atlaszát a korszerű kormányzás nélkülözhetetlen segédeszközének nevezte, a szerző a fejedelem bejegyzését – hibásan – az atlasz címlapjára tette.

Az Ortelius-atlaszt 2017 októberében vizsgáltam meg a kolozsvári Központi Egyetemi Könyvtár Térképtárában. A térképtárnak két Ortelius-atlasza is van: az 1608-as olasz nyelvű (jelzete Atlas 64), és az 1592-es latin nyelvű (jelzete: Atlas 62) kiadás. Az utóbbi eredeti katalóguscéduláján megjegyzések szerepelnek Bethlen saját kezű bejegyzéséről, az egyik Magyarország-térképlap hiányáról, továbbá egy 1942. IV. 18-i jegyzet, miszerint a kötetet visszahelyezték a díszművek közé. Vajon Borbély Andor egy 1942-es kiállítás során figyelhetett fel az atlaszra?

A PÉLDÁNY FELÉPÍTÉSE, BEJEGYZÉSEI

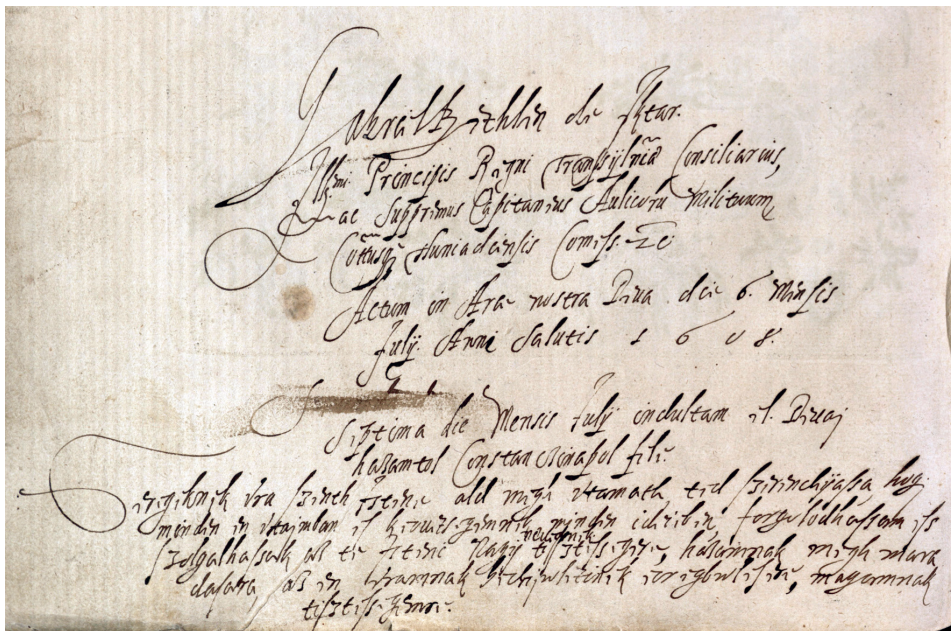
A díszes, de nem Bethlen-címeres 43×28 cm-es kötéstábla az atlasz három részét köti össze: az első és a harmadik szöveges rész, a második a térképeket tartalmazza.

Az első részben a kontinenseket jelképező nőalakokkal díszített címlapon a példány egykori hovatarozását jelzi a Csáky-család pecsétje (lásd 1. ábra), e címlap hátoldalán pedig az Erdélyi Múzeum könyvtárának pecsétje található, amely a példány későbbi sorsát bizonyítja.

Az ajánlásból és versből is álló bevezetőt a *Catalogus auctorum* (bibliográfia: a közreműködők felsorolása, akikről térképet vett át az atlaszába, továbbá a kor neves európai térképei, amiket felhasználott) követi, majd utána az *Index tabularum* (a térképlapok tartalomjegyzéke) következik. Ez utóbbinál „+” jel beírásával jelölte valaki az akkor hiányzó, mintegy kilenc térképlapot. Valamivel később Philip Galle Ortelius-portréja következik, amelynek hátoldalán Johannes Posthius hatsoros verse van, ez alatt található Bethlen Gábor saját kezű bejegyzése (lásd 2. ábra).



1. ábra. A Csáky-család pecsétje az Ortelius-atlasz címlapjának szélén



2. ábra. Bethlen Gábor saját kezű bejegyzése Posthius verse alatt

Bethlen bejegyzése latinul kezdődik nevének, címének megadásával és a keltezéssel, majd magyarul folytatódik egy fohással.

„Gabriel Bethlen de Iktar.

Illustrissimi Principis Regni Transsylvaniae Consiliarius, ac Supremus Capitaneus Aulicorum Milituum, Comitatusque Huniadiensis Comes etc.

Actum in Arce nostra Deua die 6. Mensis Julij Anni salutis 1608.

Septima die Mensis Julij² indultam el Deuaj hazamtol Constanczinapol fele.

Seregeknek Ura Szenth Istene ald megh utamath, ted szerenchyassa, hogj minden en utaimban es keovetsegemnek minden ideiben fogolodhassam ess Szolgalhassak az te Isteni nagy nevednek tisztessgere, hazamnak megh maradasara az en uramnak bechywletinek eoregbwlesere, magamnak tisztessgemre.”

² A latin nyelvű rész magyar fordítása (fordította Gálfi Emőke): *Iktári Bethlen Gábor. A fenséges erdélyi fejedelem tanácsura, az udvari hadak főkapitánya, Hunyad vármegye ispánja. Kelt dévai várunkban az üdvözítés 1608. évében, július hó 6-án. Július hó hetedikén...* A magyar nyelvű részt szövegűen közlöm (Gálfi Emőke átírásában). A részben modernizált változat Szabó Károly és Csetri Elek átírásában érhető el (Szabó, 1882; Csetri, 1992, 183.).

A második rész 108 darab kétoldalas térkép, a térkép előtt leírással az ábrázolt területről. Ez a rész hiányos, a térképek közel fele (kb. ötven) ki van vágva: összesen mintegy ötvennyolc térképlap maradt meg a példányban. Jóval több hiányzik tehát, mint amennyit a tartalomjegyzékben jelöltek valamikor, az még a hiányok egy korábbi állapotát jelölte. Az is elképzelhető, hogy a hiányzó térképlapok között találhatóak a korábban még megfigyelt széljegyzetek, ma már csak pár beírás található.

A 26., Pikárdiát (*Picardia*) ábrázoló térképlap alján egy „G. Csaki Katalin” beírás található (lásd 3. ábra).

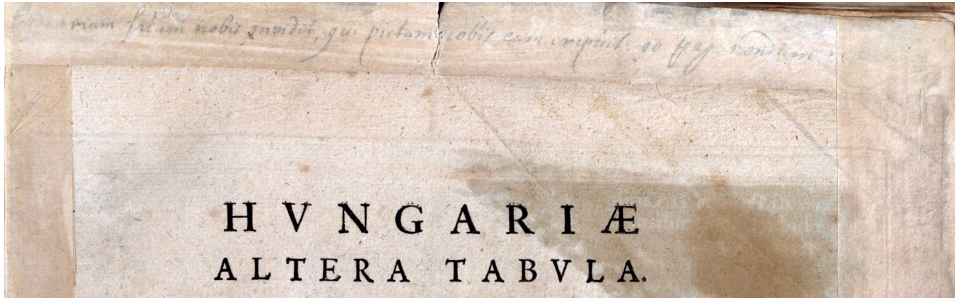


3. ábra. Csáki Katalin bejegyzése a Pikárdia-térkép alján

A példányból hiányzó, 90-es Lazius-féle Magyarország térképlap³ után következő 91-es *Hungariae altera tabula* Zsámboky-féle Magyarország-térkép szöveges előoldalának fejlécében egy latin nyelvű bejegyzés található, de papírcsíkkal le-

³ Az Ortelius-atlasz első, 1570-es kiadásától kezdve jóformán az összes kiadásban szerepelt Wolfgang Lazius *Hungariae Descriptio* térképe, így ebben az 1592-es latin kiadásban is, itt a 90. térképként. Bár általában Ortelius egy országról egy térképet jelentetett meg, mégis 1579-től Zsámboky János Magyarország-térképét is szerepeltette jóformán az azt követő összes kiadásban, itt a 91. térképként. Ortelius a lap szövegrészában pontosabbnak és valóbbnak tekintette e második térképet, de a szorgalmas olvasóra bízta, hogy melyiket használja inkább (Gróf, 1992).

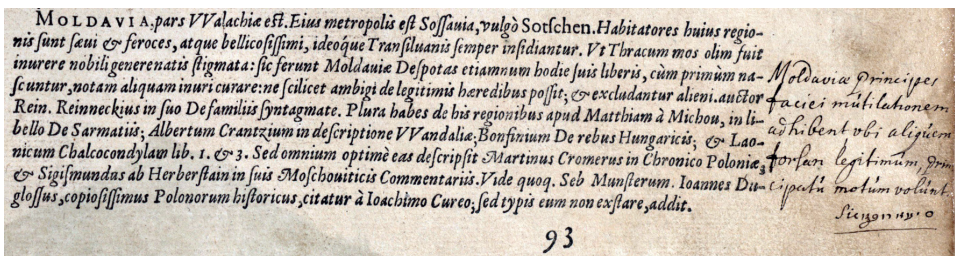
ragasztva. A takarás miatt nehezen kibetűzhető írás a következő: „*Patriam solam nobis invidit, qui pictam nobis eam eripuit 90 pag nondum*”⁴ (lásd 4. ábra).



4. ábra. A Magyarország-szöveg feletti leragasztott beírás

Elgondolkodtató a három tény együttese: a Magyarország-térkép kivágása, az erre vonatkozó szentencia leírása (hogya a tolvaj tettének az oka nem a térképlap megszerzése volt, hanem az irigyelt ország ábrázolásának eltüntetése), és a vétség ellen emelt szó elnémitása a letakarással.

A 92-es Erdély térképlapon nincs semmiféle bejegyzés. A 93-as, *Poloniae regnum* Lengyelország-térkép szöveges előoldala Lengyelország után, többek között Moldvát is tárgyalja. Ez utóbbi bekezdés szélén ismét egy latin bejegyzés található. A szöveg a következő: „*Moldaviae principes faciei mutilationem adhibent, ubi aliquem forsitan legit[ti]mum principatu motum volunt. Sic...*”⁵ (a szöveg egy kiolvashatatlan görög vagy cirill betűs szóval zárul, lásd 5. ábra).

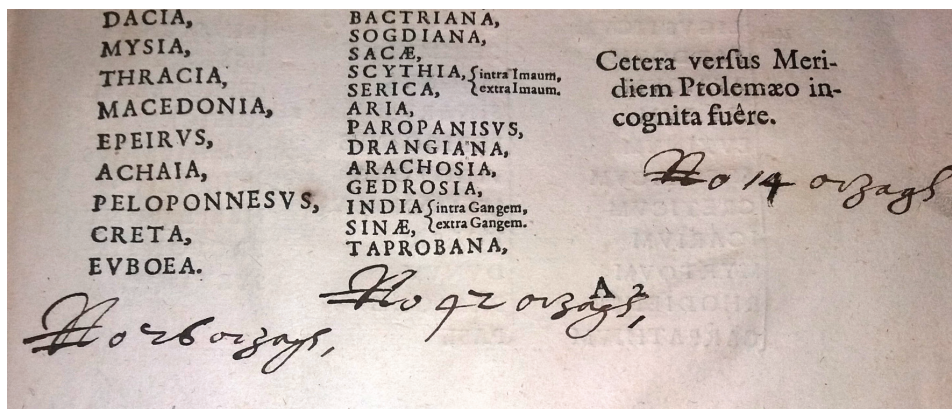


5. ábra. A Moldva-leírás szélén található bejegyzés

⁴ Csak a mi hazánkra volt irigy, aki ezt a rajzot a 90. oldalról tőlünk elvette. Fordította Gálfi Emőke.

⁵ A moldvai uralkodók/fejedelemek az arc megcsonkításával torolják meg, mihelyt valaki esetleg a fejedelemség/ország törvényes rendjére akar törni. Fordította Gálfi Emőke.

A harmadik rész legnagyobb részét az 1591-es *Nomenclator Ptolemaicus* (az ókori helynevek névmutatója országokénti bontásban), amelynek kezdetén az akkor ismert három földrész országai vannak felsorolva. Ennek alján három rövid bejegyzés van, amely azt adja meg, hogy hány ország van felsorolva Európában („No. 26 orszagh”), Ázsiában („No. 42 orszagh”), és Afrikában („No. 14 orszagh”) (lásd 6. ábra).



6. ábra. Az országok számára vonatkozó bejegyzések

A munka utolsó lapján ismét az Erdélyi Múzeum könyvtárának pecsétje, ennek hátlapja, a kolofon oldal zárja az atlaszt, ahol az atlasz kiadásának 1592-es voltát jelölték.

A kötet bejegyzései mind különböző személyek kézírásainak tűnnek.

A PÉLDÁNY TÖRTÉNETE

Abraham Ortelius (1527–1598) *Theatrum Orbis Terrarum* című világtalasa az egyetemes térképtörténet egyik legjelentősebb munkája: az első modern világtalasznak tekinthető könyv. Az atlasz első kiadása 1570-ben jelent meg, a könyv értékét bizonyítja, hogy – a rendkívül magas ára ellenére – egyre javított és bővített tartalommal jelent meg rendszeresen Antwerpenben 1612-ig. A kutatók negyvenhat kiadást különítenek el: harminckét rendes atlaszt, tizenkét kiegészítést (*Additamentum*) és két antik világot bemutató munkát (*Parergon*). A rendes kiadások közel fele latin nyelvű, de másik részük további hat nyelv (holland, francia, német, spanyol, angol és olasz) valamelyikén jelent meg. A szóban forgó kolozsvári példány 1592-ben jelent meg, latin nyelvű (Ortelius, 1592). Ezt Ort27AB-kiadás-ként jelölik, vagyis a negyvenhat változatból a huszonhetedik térkép-összeállítás,

amely két szövegváltozatban jelent meg. Az összes rendes kiadást 7300 példányban nyomtatták ki, amelyből az 1980-as években közel kilencszázat azonosítottak, ennyi legalább fennmaradt. Az Ort27AB-változatok együttes eredeti példányszámát háromszázra becsülték, a nyolcvanas években ötvenegy fennmaradt példány volt ismeretes, fele könyvtárakban, fele magángyűjteményekben vagy kereskedőknél (Broecke, 1986; Gróf, 1992; Koeman et al., 2007).

Az atlasz megjelenésekor Bethlen tizenkét éves volt, tehát szinte biztos, hogy volt előző tulajdonosa is az atlasznak, Bethlen nem új példányként tett rá szert. Arról nincs adatom, hogy Bethlen mikor, hogyan szerezte meg, illetve mikor használta az atlaszt. Atlaszról lévén szó, érdemes áttekinteni az utazásait, az általa bejárt területet. Az 1580-as születésű Bethlen fiatalkorában – Erdélyen kívül – már megfordult sereggel Havasalföldön, fejedelmi kíséret részeként a Német-római Birodalomban (annak székhelyén, Prágában 1598-ban biztosan, de akár más területein is: Németalföldön, Flandriában és Észak-Itáliában), majd az Oszmán Birodalomhoz tartozó Nándorfehérváron bujdosóként (Csetri, 1992; Romsics, 2015, 194–217).

A huszonhét éves Hunyad vármegyei főispán, Bethlen Gábor 1608. július 7-én a dévai uradalomból főkövetként, ötvenfőnyi kísérettel indult Konstantinápolyba, ott augusztus 15-én elnyerte Báthori Gábor részére az *athnamét* (vagyis a fejedelmet elismerő szultáni okmányt). Ezt a Bethlennel érkező kapucsi basa (vagyis a Portától a tartományokhoz leveleket hordó személy) november végén adta át a fejedelemnek. Ez út előestéjén írt bele az atlaszba Bethlen, Déván – tehát ekkor már biztos nála volt az atlasz. Mivel az atlaszt – árához és ritkaságához képest – igen fiatalon szerezte be, ez számomra jelzi, hogy azt mennyire értékelte, fontosnak tartotta, könyvszeretetének bizonyítéka (Csetri, 1992, 36.) (lásd 7. ábra).

Bethlen az atlaszt az 1608-as beírás után a következő helyszíneken használhatta: 1608 után újabb hadjárat következett Havasalföldön, diplomáciai utak az Oszmán Birodalomban (Kanizsa, Drinápoly), majd 1613-as fejedelemmé választása után több hadjárata volt Besztercebányán át Bécsig, Morvaországig (Csetri, 1992). Nem tudom megerősíteni azt a kézenfekvő, valószínű feltevést, hogy utazásai során használta ezt az atlaszt. Bár külön figyeltem azokra a térképlapokra, amelyek olyan területeket mutattak be, amelyeket beutazott, nem találtam ilyen jellegű széljegyzetet. Természetesen az atlasz nemcsak a helyszínen használható közvetlen tájékozódásra, hanem otthon is az ismeretek bővítésére, így valóban fontos kézikönyve, segédeszköze lehetett Bethlennek.

Ahogy nincs adat, hogy hogyan lett Bethlené a példány, ugyanúgy az is homályban marad, hogy hogyan került a kivételek közé az atlasz, vagyis miképpen maradt fenn: egyike annak az öt kötetnek, amit Bethlen könyvtárából ismerünk. A könyvtár már Bethlen 1629-es halála után felbomlott, az 1630 és 1648 között uralkodó I. Rákóczi György idején a gyulafehérvári és a sárospataki kollégium



7. ábra. Az Ortelius-atlasz Rumélia (96. Romania) térképe, e térséget járta be Bethlen a bejegyzése évében

között megosztották, de a pusztulás a krími tatárok gyulafehérvári, 1658-as rombolásakor következett be: a könyvtárban tűz pusztított. A maradék könyvekből a Bethlen Kollégium könyvtára már Nagyenyeden szerveződött újjá (Jakó, 1976, 199–208.; Vita, 1983, 8–15.).

A fennmaradt, eddig azonosított öt kötetnek, köztük az atlasznak ekkor akár már más útja lehetett. Kapocs az eredeti és jelenlegi elhelyezés között a „*G. Csáki Katalin*” beírás. Feltételezéseim szerint gróf Csáky Katalin (1726–1794) íróőné a kézírás, akinek a férje gróf Bethlen (VIII.) Miklós (1720–1781) erdélyi kincstárnok volt. A beírás annak is jele lehet, hogy ekkor került át az atlasz a Bethlenektől a Csákyakhoz.

Csáky Katalinig több lehetséges örökítési vonal képzelhető el. Egyrészt iktári Bethlen Gábor fejedelem tanácsosa volt bethleni Bethlen (II.) Farkas, az ő fia Bethlen (IX.) János (1613–1678) történetíró–kancellár, akinek javaslatára az 1658-ban felégetett gyulafehérvári főiskolát Nagyenyedre költöztették, ő a nagyenyedi főiskola első gondnoka volt, emellett az udvarhelyi református kollégium könyvtárának alapítója (Bicsok–Orbán, 2012, 170–172.). Akár az ő kezébe is kerülhetett az atlasz, az ő dédunokája volt Bethlen Miklós, Csáky Katalin férje. Másrészt a fejedelem öccse, Bethlen István fejedelem alakította át a küküllővári kastélyt, amelyet később Csáky Katalin férje, Bethlen Miklós örökölt, és szintén átépített (Bicsok–Orbán, 2012, 180–181.). E várkastély is lehetett az atlasz tárolási helye Váralmás előtt.

Csáky Katalintól nagyobb bizonyossággal rajzolható meg a példány története: ő 1737-ben örökölte meg Váralmást, amely innentől kezdve az ő leszármazottaié volt. Bethlen Miklóssal leányuk a nagy műveltségű Bethlen Rozália (1752–1826), unokájuk Csáky (II.) József királyi kamarás, szepesi főispán (1782–1848), dédunokájuk Csáky (VII.) György (1817–1890) volt (Papp, 2009, 64–80.; Bicsok–Orbán, 2012, 218.). Szabó Károly tudósításából tudjuk, hogy Csáky György váralmási könyvtárával került át az Erdélyi Múzeum könyvtárába az atlasz, tehát feltehetőleg közvetlen 1882 előtt (Szabó, 1882). Azóta ez a könyvtár (vagyis a kolozsvári egyetemi könyvtár) őrzi azt, korábban az atlasz fontosságát számon tartották, a díszművek között tartották, ma a térképtár több ezer tételének egyike.

IRODALOM

- Bendefy László (1972): Borbély Andor emlékezete. *Földrajzi Közlemények*, 4, 353–356.
- Bicsok Zoltán – Orbán Zsolt (2012): „*Isten segedelmével udvaromat megépítettem ...*”. Csíkszereda: Gutenberg Könyvkiadó
- van de Broecke, Marcel (1986): How Rare Is a Map and the Atlas It Comes from? Facts and Speculations on Production and Survival of Ortelius' *Theatrum Orbis Terrarum* and Its Maps. *The Map Collector*, 36, 2–15. <http://www.orteliushmaps.com/essays/mapcollector1986.htm>
- Csetri Elek (1992): *Bethlen Gábor életútja*. Bukarest: Kriterion Könyvkiadó

- Gróf László (1992): Magyarország térképei az Ortelius-atlaszokban. *Cartographica Hungarica*, 1, 26–36.
- Jakó Zsigmond (1976): *Írás, könyv, értelmiség*. Bukarest: Kriterion Könyvkiadó, <http://mek.oszk.hu/03200/03227/html/>
- Kelemen Lajos (1907): *Erdély leírása 1639-ből. Földrajzi Közlemények*, 2, 59–71.
- Koeman, Cornelis – Schilder Günter – van Egmond, Marco – van der Krogt, Peter (2007): Commercial Cartography and Map Production in the Low Countries, 1500 – ca. 1672. In: Woodward, David (ed.): *The History of Cartography*. Volume Three (Part 2). *Cartography in the European Renaissance*. Chicago: The University of Chicago Press, http://www.press.uchicago.edu/books/HOC/HOC_V3_Pt2/HOC_VOLUME3_Part2_chapter44.pdf
- Monok István (2010): *A magyarországi főnemesség könyves műveltsége a XVI–XVII. században*. Értekezés az MTA doktora cím elnyeréséért. Szeged–Eger–Chatou: Eszterházy Károly Főiskola, <http://real-d.mtak.hu/368/>
- Ortelius, Abrahamus (1592): *Theatrum Orbis Terrarum*. Antverpiae: In officina Plantiniana, <https://archive.org/details/theatrumorbister00orte>
- Papp Klára (2009): Az erdélyi Csákyak birtokviszonyai és családi stratégiája. In: Papp Klára – Püski Levente (szerk.): *Arisztokrata életpályák és életviszonyok*. Debrecen: Debreceni Egyetem Történeti Intézete, 57–80.
- Plihál Katalin (2009): *Magyarország legszebb térképei 1528–1895*. Budapest: Kossuth Kiadó–Országos Széchényi Könyvtár
- Romsics Ignác (2015): *A múlt arcai*. Budapest: Osiris Kiadó
- R. Várkonyi Ágnes (2013): Bethlen Gábor jelenléte Európában. In: Kármán Gábor – Teszelszky, Kees (szerk.): *Bethlen Gábor és Európa*. Budapest: ELTE BTK–Transylvania Emlékeiért Tudományos Egyesület, <http://mek.oszk.hu/13200/13299/13299.pdf>
- Szabó Károly (1882): Bethlen Gábor saját kezű följegyzése. *Történelmi Tár*, 1, 207.
- Vita Zsigmond (1983): *Művelődés és népszolgálat*. Bukarest: Kriterion Könyvkiadó, <http://mek.oszk.hu/03200/03250/html/index.htm>

URL1: <https://www.bcucuj.ro/ro/resursele-bibliotecii/cat/csh>

Vélemény, vita

TAXONÓMIA ÉS SZISZTEMATIKA – FOGALMAK, FORRÁSOK, FOLYAMATOK

TAXONOMY AND SYSTEMATICS: CONCEPTS, SOURCES, PROCESSES

Varga Zoltán

professor emeritus, a biológiai tudomány doktora
Debreceni Egyetem Természettudományi Kar, Evolúciós Állattani Tanszék
varga.zoltan@science.unideb.hu

ÖSSZEFOGLALÁS

Az egyedfözlötti rendszerekkel foglalkozó biológiai tudományterületeken az új, főképpen a molekuláris módszerek megjelenésével a korábban leíró-regisztráló (idiografikus) kutatások egyre inkább a hipotézisvizsgáló, prediktív (nomothetikus) tudomány irányába mozdultak el. Így van ez a taxonómia és szisztematika vonatkozásában is. Bár az evolúciós paradigma a múlt század derekától általánossá vált, a bioszisztematika ún. „darwini” forradalma csak fáziskéséssel jött létre: a morfológiai alapú jellegelmezés hennigi fogalomrendszerére alapozott filogenetikai módszer – módosulásokkal – meghatározóvá vált a molekuláris alapú rendszerezésben is. Ezáltal a taxonómiában is egyre általánosabbá vált a filogenetikai-hipotetikus kérdésfeltevés, és a taxonómiai fogalmak filogenetikai értelmű használata. Mindez a taxonómia és szisztematika szemléleti és módszertani közeledéséhez vezetett, olyannyira, hogy a taxonómia egyre inkább a szisztematika praktikus részévé, „nyelvévé” vált, a céljaikban és módszereikben fennálló különbségek ellenére. A biodiverzitás rohamos csökkenése miatt a taxonómiai és szisztematikai kutatások egyre sürgetőbbé váltak. Feladataik megoldásához nélkülözhetetlenek a nemzeti természettudományi gyűjtemények, anyagaik a biodiverzitási adatok felbecsülhetetlen forrásai. Ennek tulajdonítható, hogy Európa legtöbb országában ezek a nemzeti gyűjtemények egyben biodiverzitás-kutató központokká váltak. Sajnos, ebből a szempontból a Magyar Természettudományi Múzeum helyzete a legkevésbé sem kielégítő.

ABSTRACT

Due to the appearance of new, mostly molecular methods, the formerly descriptive, registering (idiographic) branches of supra-individual biology became increasingly hypothesis-testing, predictive (nomothetic). Such changes also have proceeded in the relation of the taxonomy and systematics. Despite the general acceptance of the evolutionary thinking from about the middle of the last century, the „Darwinian” revolution of the systematic biology have only proceeded with some time-lag. Although the Hennigian conceptual categories were originally elaborated

on morphological characters, they were also introduced – *mutatis mutandis* – into the molecular phylogenetics. These changes resulted in the advancing of the taxonomy and systematics. The taxonomy became more and more the practice, the „language” of the systematics, independently of the differences in their aims and methods. Considering the rapid decrease of the biodiversity the tasks of taxonomy and systematics are pressing. To these, the natural history collections provide inestimable record of biodiversity. Thus, several major European national collections are increasingly functioning as research centres of biodiversity. Concerning this the situation of the Hungarian Natural History Museum is rather insufficient.

Kulcsszavak: taxonómia, szisztematika, filogenetikus, biodiverzitás, populáció, idiografikus, nomothetikus, diverzitás

Keywords: taxonomy, systematics, phylogenetic, biodiversity, population, idiographic, nomothetic, diversity

BEVEZETÉS

Napjainkban, amikor mindenfelől a biológiai sokféleség olyan mértékű csökkenéséről kapjuk a jelzéseket, amely a nagy földtörténeti kipusztulásokhoz hasonlítható, ám összemérhetetlenül rövidebb idő alatt zajlik, sürgősen át kell gondolnunk, milyen eszközeink vannak annak feltárására és megértésére, talán a megmentésére is, valójában mi is ez az „*eltűnő sokféleség*” (vö. Juhász-Nagy, 1993). Thomas Kuhn (2000) szerint a természettudományokban paradigmaváltások zajlanak. Ezeket nagy felfedezésekhez köthetjük, ahogy ez például a fizika történetében világosan kirajzolódik, és ahogy – sejtetően – ma éppen ilyenek a küszöbén állunk. Az előbbiekből következik, hogy azokban a tudományokban is mély változások zajlanak, amelyek a biológiai sokféleség különböző szintjeinek a feltárásával foglalkoznak. Ám itt már nem az ismeretek bővülésével együtt járó specializációról van szó, hanem arról, hogy az új módszerek és szemléletmódok megjelenésével korábban távolinak tűnő kutatási területek kapcsolódnak össze. Vegyük példának a biológiai sokféleség földrajzi vetületével foglalkozó tudományt, az életföldrajzot, a *biogeográfiát*. Ezt korábban leíró-rendszerező, Juhász-Nagy Pál (a kantianus német filozófus, Wilhelm Windelbandtól kölcsönzött) szóhasználatával *idiografikus* tudománynak tekintették, amelynek a feladata hierarchikus felépítésű regionális rendszerek (flóra- és faunabirodalmak és alárendelt egységeik) létrehozása volt. Míg nem a századforduló előtti években megérkeztek a molekuláris módszerekkel dolgozó evolúciógenetikusok, akik feltárták, hogy a különböző fajok populációinak tér-idő dinamikájának hátterében milyen evolúciós változások, akár fajképződési folyamatok zajlanak. Mivel ezeknek a folyamatoknak a földtörténeti múlt és a jelen éghajlatváltozásaival való kapcsolatát is megértették, világossá vált, hogy nemcsak a múltbeli folyamatok közös min-

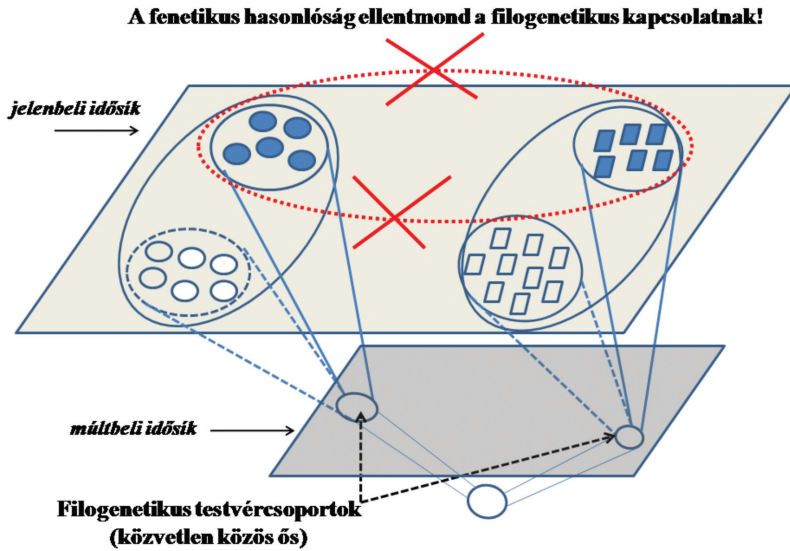
tázatait érthetjük meg, hanem a klímamodellekkel együtt a fajok elterjedésének változásait is megjósolhatjuk. Tehát tudományunk oknyomozóvá, egyúttal *prediktív*vé, az előbbi terminológiát folytatva *nomothetikussá* („törvényalkotóvá”) vált, ahogy ezt egy rendes természettudománytól el is várhatjuk.

BIOSZISZTEMATIKA: EVOLÚCIÓS SZEMLÉLET, FILOGENETIKUS MÓDSZER

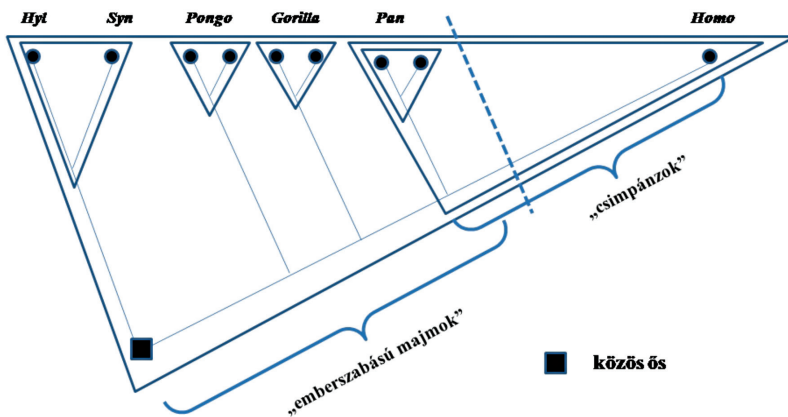
Miért lenne ez másként az élőlények sokféleségét leíró, megnevező, rendszerező és értelmező tudományban, a *bioszisztematikában*? A linnéi korszak paradigmája még az *esszenciális* jellegek alapján jellemzett és elkülönített (*diagnosis!*) fajok állandósága volt, ennek arisztotelészi logikáját örököltük meg a fajleírás és -megnevezés szabályrendszerében (lásd: *genus proximum* és *differentia specifica*). Jórészt annak köszönhetően, hogy még ma is szinte áttekinthetetlenül nagy az ismeretlen, illetve tudományosan leíratlan fajok száma, ez az egyszerű nevezék-tani apparátus a szemléletmód gyökeres változása és a kutatási módszerek rohamos fejlődése ellenére megmaradt: továbbra is „*Homo sapiens*” vagyunk, a barna medve pedig *Ursus arctos* (vagyis medve-medve latin-görögül). A helyzetet bonyolítja azonban az, hogy az evolúciós paradigma általánossá válása önmagában még nem tette nyilvánvalóvá, hogyan történjék a következő folyamat: a leírt fajok rendszerezése – a *szisztematizálás*, bár ennek első lépése az arisztotelészi logikából következően adott volt: az adott fajt egyben valamelyik *genusba* (nemzetségbe) is be kellett sorolni, és így tovább. Amivel egyúttal azt is állítottuk, pontosabban *feltételeztük*, hogy a szóban forgó faj mely további fajokkal hozható közelebbi – itt még ne mondjunk többet! – rokonságba. Nem nehéz belátni, hogy itt kapcsolódik be a folyamatba az *evolúciós szemlélet* és a *filogenetikus módszer*.

Nyilván indokolnom kell, miért így különböztetem meg ezeket. Az első körülmény az időbeli sorrend: a fentiek *nem egyszerre* és *nem együtt* jelennek meg. Először a *New Systematics* (Huxley, 1940) bontott zászlót, amely szemléletet adott, a *populációs szemléletet*. Eszerint a fajok meghatározott kritériumok alapján összetartozó („specifikált”) *populációk*, és amelyek adott kritériumok szerint el is különülnek más fajoktól mint populációktól. Ebből viszont az is következik, hogy mivel az elkülönülés *folyamat*, az így szétváló csoportokra létrehozott és megnevezett egységek, a *taxonok*, *dinamikus folyamatok* adott időbeli keresztmetszetben létező stádiumainak *statikus keretekben* való rögzítését jelentik (*1. ábra*). Látható tehát, hogy bár szemléletünk *evolúciós* és rendszerezésünk *hierarchikus*, eddig mindez leképezhető két dimenzióban. Ekkor az evolúciós szemléletben még együttesen érvényesül az ősoktól való leszármazás gondolata, és a tőlük való eltérések létrejöttének felismerése, a fosszilis leletek, illetve a ma élő fajok sajátosságai alapján. Evidenciává válik, hogy az embernek és az emberszabású majmoknak közös őseik voltak, de ezzel együtt él az a felfogás is, hogy az egyik,

bizonyos fajokat összefoglaló kategória az *emberszabású majmok*, és ettől elkülönül, mint önálló faji egység, az *ember*. És ez mindaddig nem is okoz problémát, amíg a folyamat időbeli dimenzióját nem kezdjük el vizsgálni.



1. ábra. A fajok dinamikus folyamatok adott időbeli keresztmetszetben létező stádiumai



Hasonlóság szerint: [(1 + 2) + (3 + 4) + (5 + 6) + (7 + 8)] + 9

Eredet szerint: {(1 + 2) + (3 + 4) + (5 + 6) + [(7 + 8) + 9]}

2. ábra. Beágyazottan hierarchikus monofiletikus csoportok a hominoid filogenetikai ág (kládusz) fejlődéstörténetében

Ehhez pedig nélkülözhetetlen a folyamat harmadik, *időbeli dimenziójának* a feltárásához szükséges módszer. Ennek az algoritmus a *filogenetika* fogalomrendszeréből következik, ahogy ezt a korábbi egyszerű ábrán szemléltethetjük (2. ábra). Döntésünk eltérő lesz aszerint, bár mindkét esetben feltételezzük a (valamikor) közös ősből való leszármazást, hogy a két lehetőség közül melyiket választjuk: a *változás mértékét* tekintjük mérvadónak, vagy pedig a filogenetikai szétválások időrendjét, a *közvetlen közös őstől* való leszármazást. Előbbi esetben a két csimpánzfaj a gorillákkal (két faj), orangutánokkal (három faj) és a gibbonokkal (tizenkilenc faj) együtt „emberszabású majom”, ha viszont a közvetlen közös őst vesszük alapul, akkor (a kihalt fajokat nem számítva) a legközelebbi rokonaink a csimpánz(ok). Filogenetikai értelemben tehát az ember a „harmadik csimpánz” (Diamond, 2002). A többlépcsős leszármazási viszonyokat legegyszerűbben decimális számozással fejezhetjük ki (*kládusz* = filogenetikai ág):

- | | | | |
|----------|-----------------------------|--------------|-------------------------|
| 1. | Hominoidea | | |
| 1.1. | Hylobatidae | 1.2.2. | Homininae |
| 1.1.1. | <i>Hylobates</i> kládusz | 1.2.2.1. | <i>Gorilla</i> |
| 1.1.1.1. | <i>Hoolock</i> (2 faj) | 1.2.2.1.1. | <i>Gorilla gorilla</i> |
| 1.1.1.2. | <i>Hylobates</i> (9 faj) | 1.2.2.1.2. | <i>Gorilla beringei</i> |
| 1.1.2. | <i>Synphalangus</i> kládusz | 1.2.2.2. | <i>Homo</i> |
| 1.1.2.1. | <i>Synphalangus</i> (1 faj) | 1.2.2.2.1 | <i>Pan</i> kládusz |
| 1.1.2.2. | <i>Nomascus</i> (7 faj) | 1.2.2.2.1.1. | <i>Pan satyrus</i> |
| 1.2. | Hominidae | 1.2.2.2.1.2 | <i>Pan paniscus</i> |
| 1.2.1. | Ponginae | 1.2.2.2.2. | <i>Homo</i> kládusz |
| 1.2.1.1. | <i>Pongo abelii</i> | 1.2.2.2.2.1. | <i>Homo sapiens</i> |
| 1.2.1.2. | <i>Pongo pygmaeus</i> | | |
| 1.2.1.3. | <i>Pongo tapanuliensis</i> | | |

Teljesen hasonló a helyzet a „hüllők” és a madarak vonatkozásában. A fosszilis bizonyítékok alapján egyértelmű, hogy a madarak a két lábon járó, tollas, fészekrakó, ún. *theropod* (a madarakhoz hasonlóan négyujjú hátsó lábú) dinoszauruszok leszármazottai. A madarak bármennyire is változatosak, és számos csoportra (hagyományosan: rendekre és családokra) tagolódnak, filogenetikai értelemben a *theropod* dinoszauruszokba „beágyazottak”. Lényegében megdőlt tehát a megszokott „hüllők-madarak” megkülönböztetés. Más szóval a *theropod* dinoszauruszok nagy filogenetikai ága, *kládusza* a madarak nélkül „csonka” (*parafiletikus*), és csak a madarakkal együtt válik filogenetikailag teljessé (*holofiletikussá*). A különbség csupán az, hogy míg a hominid főemlősök fenti filogenetikai tagolódását a morfológiai (csonttani!) jellegek mellett a kromoszomális és molekuláris (a teljes, lásd: *Homo*, vagy csaknem teljes genomra kiterjedő) bizonyítékok is támogatják, addig a *theropod* dinoszauruszok esetében ez utóbbi jellegek nem állnak rendelkezésünkre, így fejlődéstörténetüket a recens anyag mellett a fosszilis, főleg csonttani leletekre alapozva rekonstruálhatjuk.

TAXONÓMIA ÉS SZISZTEMATIKA

Ezek után jogos a kérdés: mi köze mindennek a taxonómia és a szisztematika viszonyához? Nem is kevés! A *New Systematics* időszakában, a múlt század 40-es éveitől, a leíró-megnevező-rendszerező taxonómia mögött már jelen volt az evolúciós háttér, de eleinte még egyértelmű fogalomrendszer és módszer nélkül. Ezáltal a kodifikált eszközrendszerű taxonómia valóban elég élesen elkülönült a szisztematikától, hiszen a szisztematika, egyrészt az axiomatizált, „megismételhető” metodika nélkül, másrészt a molekuláris „kemény” bizonyítékok híján továbbra is erősen hipotetikus maradt. Mivel emiatt a legkülönbözőbb részterületekről gyűjtött tényekre mint „az evolúció bizonyítékaira” támaszkodhatott, ahogy ezt sok tankönyvben ma is látjuk, ezért egyrészt induktív alapon általánosító volt, másrészt pedig eléggé „parttalanak” is bizonyult, ahogy az ún. „evolúciós szisztematika” több vezető alakjának gyakran idézett megnyilvánulásából leszűrhetjük (például a szisztematika mint a biológia „fókuszpontja”, vö. Huxley, 1940; illetve Simpson, 1961: „Systematics is the science that deals with the organization, history, and evolution of life”).

Mivel a taxonómia újabb módszereinek (például: molekuláris genetikai módszerek, pásztázó elektronmikroszkópia, geometriai morfometria) egyre általánosabbá válása az egyes taxonok fejlődéstörténetének pontosabb feltárását tette lehetővé, ezért a taxonómiai vizsgálatokban is egyre általánosabbá vált a filogenetikai-hipotetikus kérdésfeltevés, sőt a taxonómiai fogalmak (például faj- és fajfeletti kategóriák) filogenetikai értelmű használata is. Gyakran beszélnek például molekuláris taxonómiáról, amikor egy-egy *standardizált molekuláris jelleg* (például a bárkód, a COI mitokondriális gén adott szakasza) mint *diagnosztikus bélyeg* szerepel. Ami csak akkor válik naiv tévedéssé, amikor a bárkód-nukleotidszekvencia alapján szerkesztett hasonlósági fát egyúttal törzsfának is tekintik. Mi sem természetesebb, mint hogy – a Hennig és követői által – a jellegek elemzésére kidolgozott filogenetikai fogalomrendszernek, illetve a molekuláris filogenetika erre alapozott fogalomrendszerének (például *ortológ* és *paralóg* gének) az általános használata viszont a szisztematika fogalmát is pontosabban körülhatároltá tette.

Ettől persze a szisztematika mint kutatási terület nem vált *azonossá* a taxonómiával, hanem egyre inkább *a taxonómia vált a szisztematika részévé*, ahogy ezt Michael J. Novacek (1998) megfogalmazta: „A taxonómia az élőlények leírásának és megnevezésének a gyakorlata és olyan osztályozása, amely a rokonsági viszonyokat tükrözi. *Ez a szisztematika nyelve*; a szisztematika része, nem a szisztematika egésze”. A szisztematika pedig, ahogy már a 90-es években megállapították (De Queiroz, 1998), szinte „darwini” forradalmon ment keresztül. Azzal, hogy a genom egyre nagyobb szakaszai, s nemcsak a mitokondriális genom, hanem egyre teljesebben a nukleáris gének is ismertté válnak, egyre megbízhatóbban

válik megrajzolhatóvá az „élet fája”, és a törzsfajlódás kutatását egyre erősödő bioinformatikai háttér is segíti.

Ebből a tendenciából adódnak kedvező és kedvezőtlen fejlemények egyaránt. A taxonómiát és a szisztematikát nyilvánvalóan megkülönböztetik egymástól az eltérő célok és a jelentősen eltérő módszerek is. A taxonómia alapfeladata a fajok minél pontosabb diagnosztizálása, leírása és a nevezéktani szabályok szerinti megnevezése. Ennek lehetőleg *revíziós munka* keretében, a már leírt fajok lehető teljességének ismeretében kell történnie, a „spanyolviasz”-taxonok újraleírásának és a szinonim nevek létrehozásának elkerülése érdekében. Ehhez pedig a taxonómusnak részletekbe menően ismernie kell a vizsgált élőlénycsoportot, ismernie kell a jelentős gyűjtemények anyagát, ezek közül is kiemelten azokat, amelyek az általa vizsgált csoport fajainak elsődleges típuspéldányait, „etalonjait” őrzik. Az alapos fajismeret mellett korszerű módszertani felkészültséggel is rendelkeznie kell. Ezért is van az, hogy egy kompetens taxonómus kiképzése rendszerint jóval meghaladja egy PhD vagy posztdoktori ösztöndíj időkeretét.

Némileg más a helyzet a szisztematikai munka esetében, amelynek az alapvető célja az adott élőlénycsoport fejlődéstörténeti viszonyainak tisztázása és filogenetikus rendszerének létrehozása. Bár az lenne az alapvető kíváncsi, hogy aki filogenetikus szisztematikai munkát végez, a széles alapozású evolúciobiológiai tudása mellett tisztában legyen legalábbis az általa intenzíven vizsgált élőlénycsoport taxonómiájával is. Ez garantálhatja ugyanis azt, hogy az egyre duzzadó molekuláris adatbázisokba taxonómiailag helyesen diagnosztizált fajok adatai kerüljenek. Ez azonban – csoportja válogatja – nem mindig van így, hiszen a ma már zömmel molekuláris adatokkal és számítógépes programokkal dolgozó filogenetikusok nem szükségképp rendelkeznek egyúttal taxonómiai tapasztalati ismeretekkel is (Wägele, 2005). Ezért is volna szükséges, hogy a taxonómiai-szisztematikai munka is egyre inkább csapatmunkává váljék, és ezáltal a taxonómia eszköztára hatékonyabban szolgálja a modern filogenetikai-szisztematikai kutatásokat, *feleslegessé téve a taxonómia és szisztematika éles megkülönböztetésével kapcsolatos korábbi, zömmel formális vitákat.*

ZÁRÓ GONDOLATOK

A biológiai sokféleség rohamos csökkenése nem új keletű felismerés. Legfeljebb az a vita tárgya, hogy mikor kezdődött. Nagyobb figyelem viszont inkább csak újabban terelődött rá azáltal, hogy az „iparosodott” mezőgazdaságú országokból egyre inkább eltűnnek a vetőmag-, a méz- és gyümölcsstermelés számára nélkülözhetetlen virágmegporzó, az ún. ökoszisztéma-szolgáltatásokat végző rovarok (Potts et al., 2010; Hallmann et al., 2017). Önkéntelenül eszünkbe juthat a filozófus Karl Poppernek (1972!) az evolúciós ismeretelmélet kapcsán megfogalmazott

intelme: „Az a filozófia legnagyobb botránya, hogy míg körülöttünk a természet – és nemcsak az – tönkremegy, a filozófusok továbbra is arról beszélnek – hol értelmesen, hol kevésbé –, hogy ez a világ egyáltalán létezik-e.”

A taxonómiai-szisztematikai kutatómunka jelenleg egyre inkább múzeumi tevékenységgé válik, és kiszorul a felsőoktatási intézményekből, holott az egyre rohamosabban „eltűnő” biodiverzitás feltárása és megvédése megfelelően képzett tudományos kutatói utánpótlást igényel. Ennek a feladatnak a fontossága különösen azóta került a nemzetközi figyelem középpontjába, mióta a tudományos gyűjteményekben őrzött anyagokon végzett molekuláris filogenetikai vizsgálatok alapján rájöttek arra, hogy ezek a természettudományos (botanikai, zoológiai és őslénytani) gyűjtemények milyen hatalmas és pótolhatatlan értéket jelentenek az egyre sürgősebbé váló kutatások számára (lásd Novacek–Goldberg, 2013; Holmes et al., 2016). Kitűnt, hogy a különböző időszakokban gyűjtött és megőrzött anyagokon végzett morfológiai és molekuláris vizsgálatok a különböző fajok népségeiben végbement fenotipikus és genotipikus változások feltárásával mintegy „ablakot nyitnak” az evolúciós változásokra. Ezek a kutatások, amelyek gyakran mára már kihalt népségek egyedein folynak, nemcsak azért különösen jelentősek, mert olyan folyamatokat tesznek megismerhetőkké, ahol „nem voltunk jelen”, hanem azért is, mert jelezhetik, sőt előre is jelzik bizonyos fajok népségeinek sebezhetőségét, kipusztulástól veszélyeztetettségét, ezáltal közvetlenül is szolgálhatnak gyakorlati védelmi programokat.

Ha körülnézünk Európában, azt látjuk, hogy az idézett felismeréseknek már sokféle vannak tudománytámogatási és -szervezési következményei is. Ma már egyben *biodiverzitás-kutató központ* is a Museum of Natural History, London, a Muséum National d’Histoire Naturelle, Paris, a Naturhistorisches Museum, Berlin, a Naturmuseum Senckenberg, Frankfurt am Main (logója: *the World of Biodiversity*), a leideni Naturalis, és ezzel együtt vagy nagy kutatóintézetekhez (Frankfurt) vagy kutatóegyetemekhez (például: Stockholm, Berlin, Bonn, München, Leiden stb.) tartoznak. Ezekhez képest a hazai helyzet kirívóan mostoha, de ez egy következő írás tárgya lehet.

IRODALOM

- De Queiroz, K. (1988): Systematics and the Darwinian Revolution. *Philosophy of Science*, 55, 238–259.
- Diamond, J. (2002): *A harmadik csimpánz felemelkedése és bukása*. Budapest: Typotex Könyvkiadó
- Hallmann, C. A. – Sorg, M. – Jongejans, E. et al. (2017): More than 75 Percent Decline over 27 Years in Total Flying Insect Biomass in Protected Areas. *PLoS ONE*, 12, 10, 1–21. DOI: 10.1371/journal.pone.0185809, <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0185809>

- Holmes, M. W. – Hammond, T. T. – Wogan, G. O. U. et al. (2016): Natural History Collections as Windows on Evolutionary Processes. *Molecular Ecology*, 25, 864–881. DOI: 10.1111/mec.13529, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4755843/>
- Huxley, J. A. (ed.) (1940): *The New Systematics*. Oxford: Clarendon Press
- Juhász-Nagy P. (1993): *Az eltűnő sokféleség*. Budapest: Scientia Kiadó
- Kuhn, T. S. (2000): *A tudományos forradalmak szerkezete*. Budapest: Osiris Kiadó
- Novacek, M. J. (1998): ‘*Mapping the Ebb and Flow of Life*’. Presentation, The American Museum of Natural History. Előadás
- Novacek, M. J. – Goldberg, S. L. (2013): The Role of Museums and Institutions in Biodiversity Science and Education. In: Levin, S. A. (ed.): *Encyclopedia of Biodiversity* (2nd ed.) Vol. 5, 404–420. Waltham, MA: Academic Press, <https://www.researchgate.net/publication/255982249>
- Popper, K. (1972): *Objective Knowledge. An Evolutionary Approach*. Oxford: Clarendon Press
- Potts, S. G. – Biesmeyer, J. C. – Kremen, C. et al. (2010): Global Pollinator Declines: Trends, Impacts and Drivers. *Trends in Ecology and Evolution*, 25, 6, 345–353. DOI: 10.1016/j.tree.2010.01.007, https://www.researchgate.net/publication/41621021_Global_pollinator_declines_Trends_impacts_and_drivers
- Simpson, G. G. (1961): *Principles of Animal Taxonomy*. New York–London: Columbia University Press
- Wägele, H.-J. (2005): *Foundations of Phylogenetic Systematics*. (2nd ed.) München: Verl. F. Pfeil

SZEMLEÉLVÁLTÁS SZÜKSÉGES A TAXONÓMIA ÉS SZISZTEMATIKA VISZONYÁRÓL ÉS MEGHATÁROZÁSÁRÓL

PARADIGM CHANGE IS NECESSARY ON RELATIONSHIP AND DEFINITION OF TAXONOMY AND SYSTEMATICS

Páll-Gergely Barna

PhD, MTA Agrártudományi Kutatóközpont Növényvédelmi Intézet Állattani Osztály
pall-gergely.barna@agr.ar.mta.hu

ÖSSZEFOGLALÓ

A jelen írás fő megállapításai a következők:

(1) Rá szeretnék mutatni, hogy a magyar szakirodalomban jelen lévő, szélesen értelmezett szisztematika meghatározáson kívül van egy szűkebben értelmezett változat is, amelyet a nemzetközi szakirodalomban gyakrabban használnak. A szisztematika tág értelmezése vezetett oda, hogy a taxonómia és a szisztematika elkülönülése élesen jelenik meg a hazai szakirodalomban.

(2) A molekuláris filogenetika (molecular phylogeny) mindennapos és rutinszerűvé, és manapság sokak szemében nagyrészt a *systematics* szinonimájává vált. Míg a fajok leírása és meghatározhatóvá tétele (szűk értelemben vett taxonómia) elsősorban morfológiai alapon működik, addig a csoportok közötti rokonsági viszonyok felderítése a morfológián túl egyre gyakrabban molekuláris markerek segítségével történik.

(3) Az a diszciplína, amely az élőlényeket leírja, elnevezi, és a köztük lévő rokonsági viszonyokat kutatja, többszintű hipotézisvizsgálattal működik. Ez a tudományág céljait és tudományfilozófiai módszereit tekintve egységes a faj alatti szintektől a magasabb rendszertani kategóriákig, és az elnevezése szerzőnként eltérő. Az egységes volta miatt hajlanék a taxonómia elnevezés felé, viszont tudománytörténeti és módszertani különbségekből kifolyólag javaslom a taxonómia és a szisztematika elnevezések használatát.

ABSTRACT

The aims of the present paper are the following:

(1) I would like to show that besides the definition of systematics as a broad discipline, as appears in the Hungarian literature, there is another, narrower approach, which is more often used in the international literature.

(2) Molecular phylogenetics has been used routinely, and for many researchers became the synonym of systematics. While describing and making species identifiable (taxonomy in a stricter sense) is mostly done by examining morphology, molecular methods take increasingly large share in revealing their phylogenetic relationships.

(3) The discipline that describes, names organisms, and examines their phylogenetic relationships operates as hypothesis testing on multiple levels. Considering its aims and philo-

sophical methods, this discipline is unified from the level of species and higher taxa, and its name differs between authors. Because it is unified, I considered using the word taxonomy for it, although due to the historical background and methodological differences, I suggest using the terms taxonomy and systematics in the Hungarian literature.

Kulcsszavak: rendszertan, taxonómia, szisztematika, homológ karakterek, tudományfilozófia, biodiverzitás

Keywords: systematics, taxonomy, sistematics, homologous characters, history of science, biodiversity

BEVEZETÉS

A taxonómiát sokan nem tartják 21. századi modern tudománynak, sőt, talán még önálló tudományágnak sem. Az okok között sokan sokfélélt említettek már, ezért a teljes összegzés csaknem lehetetlen, és nem is térnék ki erre. Remek magyar nyelvű összefoglaló található Werner E. Holzinger (2007) írásában, de az érdeklődők könnyedén találhatnak vonatkozó cikkeket a „taxonomic impediment” kulcsszóra keresve. Végeredményben a taxonómia egyre kevesebb támogatást kap, a taxonómiai munkahelyek száma világszinten egyre csökken, és egyre kevesebb diák találja vonzó karrierlehetőségnek a taxonómiát. Ezek a trendek nemcsak magyar, hanem globális problémát is jelentenek (Ebach et al., 2011). Elcsépelet gondolatnak tűnik, de igaz: jelenleg az emberi tevékenységgel összefüggésben hatalmas mértékű kihalási hullámot tapasztalunk, és ismeretlen kártevő fajok százai jelennek meg eredeti elterjedési területüktől távol. Ezek miatt minden eddiginél fontosabb lenne a taxonómus szakemberek munkája, akik képesek a bolygónkat benépesítő élőlények dokumentálására, a természetvédelmi prioritások meghatározására, valamint az invazív fajok azonosítására.

Magyarországon jelenleg húsz-huszonöt körüli az aktívan dolgozó, nemzetközi folyóiratokban folyamatosan publikáló zootaxonómus. Mahunka Sándor és Vásárhelyi Tamás (1990) a „revízióképes” zootaxonómusok számát negyvenhatra tette, amely Papp László és szerzőtársai szerint (2005) túlbecsült szám. Ennél természetesen több szakember az, aki pontosan tud határozni, de a taxonómusok és megbízható faunisták csökkenése együtt jár. Bár a vészharangot többen megkongatták, nincs egyértelmű recept a taxonómia vonzóbbá és elismertté tételére. Véleményem szerint a taxonómia problémáit a szakmán kívül és belül is keresnünk kell. A jelen írás fő célja, hogy bemutasson egy, a hazai szakirodalomban még ismeretlen szemléletmódot a taxonómia és szisztematika viszonyrendszeréről és a taxonómia tudományfilozófiai megítéléséről. Ez talán jobban integrálhatóvá teszi a taxonómiát a többi „modern” diszciplína közé, és – reményeim szerint – enyhít a taxonómia „külső” problémáin.

Bár nem írom minden alkalommal a taxonómia és a szisztematika szavak elé a „zoo-” előtagot, a jelen írás az állatrendszertanra korlátozódik.

A TAXONÓMIA ÉS SZISZTEMATIKA VISZONYA

A taxonómiát a magyar szakirodalomban (is) a szisztematikával szokás szembeállítani és meghatározni. Ennek az alapjait Dudich Endre és Loksa Imre (1969) fektették le. Az országosan jelenleg elfogadott szemléletet Papp László (1986) fogalmazta meg egy dolgozatában. Papp László egyéb publikációkban is érintette a témakört (Papp, 1983, 1994; Balogh–Papp, 1988; Papp et al., 2005), és mindig következetesen hangsúlyozta a taxonómia és a bioszisztematika éles szétválasztásának szükségességét. Nézetei az általa szerkesztett *Zootaxonómia* című könyvben (Papp, 1996) is megtalálhatóak, és ez a könyv adja a mai napig a hazai felsőoktatásban az állatrendszertan oktatásának gerincét.

Változtatás nélkül idézem a taxonómia és a bioszisztematika meghatározását Papp László (1996) szerint (ez az eredeti, 1986-os meghatározás valamelyest módosított változata).

„A (*zoo*)taxonómia az élőlények (állatok) sokféleségének az egyedek, populációk szintjén való megnyilvánulásaival foglalkozik: kutatásainak szintje ugyan nem egyértelműen idiobiológiai, viszont a vizsgálatok közvetlenül (állat)egyedeken folynak. A jelenségek szférájának tudománya, amely a fajok nevének referenciaalapjául szolgáló morfológiai, élettani, biokémiai stb. adatok gyűjtésén és feldolgozásán kívül a fajok életmódjának alapadatait is rögzíti. A taxonómia a törzsféjlődés »végeredményeivel«, a bioevolúció eredményeként létrejött, populációkat, fajokat képviselő egyedekkel foglalkozik, és bár az evolúciós szemlélet megtermékenyítőleg hatott rá, önmaga nem történeti tudomány. A taxonómia alapvetően összehasonlító és analitikus szemléletű, ún. idiografikus tudomány.

A *bioszisztematika* ezzel szemben az élővilág diverzitása kialakulásának okaiival és módjaival, az evolúció révén kialakult sokféleséget megtestesítő fajok, monofiletikus csoportok stb. egymáshoz való viszonyával foglalkozik. Nemcsak és nem elsősorban a taxonómia tudományának eredményeit dolgozza fel, hanem az evolúciobiológia, a populációgenetika, az őslénytan és több más infra- és szupraindividuális tudományág eredményeit is; a bioszisztematika általános információfeldolgozó, oknyomozó történeti tudomány.”

A taxonómia és szisztematika meghatározása után további magyarázatot találunk a két tudományág közötti különbségekről: „Míg a taxonómia ismeretanyaga döntően maguknak az élőlényeknek a közvetlen vizsgálatán, tehát empirikus, tapasztalati ismeretanyagon nyugszik, addig a bioszisztematika általánosító, absztraháló, szintetikus szemléletű biológiai diszciplína; Julian Huxley szavaival a biológia »fókuszpontja«. Ezért benne különösen érzékletesen nyilvánulnak meg a különböző »tudományfilozófiák« eltérő nézőpontjai. A bioszisztematika így a »biológiai tudományok tudománya«. A biológiai rendszerezés elméletével a bioszisztematika foglalkozik, amely hipotéziseket és szabályokat (invariancia-elvet) felállító, nomothetikus tudomány.”

A két fogalom meghatározásának és elkülönítésének okait Papp (1994) arra vezeti vissza, hogy Julian Huxley (1940) *New Systematics* című könyvének megjelenésével, valamint George G. Simpson (1961) és Ernst Mayr (1963) munkáival kezdve a taxonómia a „modernebb” tudományágakkal szemben egyre inkább háttérbe szorult, és művelőit, eredményeit egyre inkább lenézték. A taxonómusok addigi fő munkáját – az ismeretlen fajok leírását – értéktelennek titulálták, és más, „modern” típusú információk összegyűjtésére kényszerítették. Papp László magát Richard E. Blackwelder (1962) követőjének tartja, aki szerint Huxley céljait Mayr és Simpson elferdítette. Blackwelder a klasszikus taxonómiai vizsgálat eredményeit (fajok elnevezése, leírása, besorolása) továbbra is fontosnak tartotta, ellentétben a korszak nagy szószólóival, akik a populációgenetikai, evolúcióbio-lógiai stb. kutatásokat preferálták (Papp, 1994).

SZISZTEMATIKA „SENSU LATO” ÉS „SENSU STRICTO”

A szisztematika (systematics) és a taxonómia (taxonomy), valamint a rendszerezés (classification) meghatározása szerzőnként eltér a témával foglalkozó szakcikkekben és könyvekben, sőt, gyakran megemlítik, hogy a két fogalom egymás szinonimájaként is használatos. A taxonómia megítélése egységesebb, viszont a szisztematikát alapvetően két különböző módon értelmezik. Donald L. J. Quicke (1993) például a szisztematikát szélesebb értelemben használja („tartalmazza a klasszikus taxonómiát az evolúció, a genetika és a fajképződés elméleti és gyakorlati megközelítéseivel”), valamint a taxonómia részének tekinti a rendszerezést (classification). Ez a fajta elgondolás az, amelyet Simpson és Mayr fogalmaztak meg. Papp (1986, 1994) szisztematika-meghatározása is Simpson (1961) elgondolásán alapul, aki a *systematics* nevű diszciplínát a „biológiai tudományok tudományának” tekintette. Amit Simpson systematicsnak tartott, az ma az összehasonlító biológiának (comparative biology) felel meg (Minelli, 1993). Ezekkel szemben például Wägele (2005) szerint a taxonómusok a fajok leírásával és elnevezésével foglalkoznak, míg a szisztematikával foglalkozó kutatók (systematists) a törzsfajlódási rendszert kutatják. Randall T. Schuh és Andrew V. Z. Brower (2009) szerint a szisztematika (amelyet szerintük gyakran hívnak taxonómiának is) „a biológiai rendszerezés tudománya”. Ez utóbbi, szűkebb értelemben vett szisztematika az, amelynek szellemében a legtöbb szakkikk születik manapság. Vagyis, bár a szisztematika fogalmát érintő szakkönyvekben és cikkekben nincs egyetértés, egy trend figyelhető meg a fogalom használatában az utóbbi verzió javára. A systematics tehát jelenleg az élőlények egymáshoz való leszármazási viszonyait kutatja, vagyis például a diverzitás „okait” és „módjait” (Papp, 1986) nem. Természetes, hogy nemcsak morfológiai, hanem például ökológiai, viselkedési és egyéb tudományágak által generált adatokat is felhasznál, de azokat

karaktereknek tekinti, és a célja nem más, mint az egyetlen, leszármazási viszonyokat ábrázoló törzsfát a lehető legpontosabban rekonstruálni. Balogh János és Papp László (1988) meghatározása a rendszerezésről (classification), amelyet a szisztematika ágának tekintenek, megegyezik azzal, amit ma a leggyakrabban a „systematics” nevű tudományágon értenek.

Ha megnézünk egy random publikációt, melynek *Taxonomy and systematics of 'ez és ez' a genus* a címe, fogunk találni egy morfológiai alapon álló taxonómiai revíziót és egy molekuláris filogramot. Bár az ilyen cikkek jellemzően nem térnek ki meghatározásbeli kérdésekre, sejthető, hogy a revízió részt értik „taxonomy”, a filogram(ka)t pedig a „systematics” alatt. Nem látom tehát értelmét annak, hogy a világosan kirajzolódó nemzetközi szóhasználattól eltérő meghatározásokat használjunk a magyar szakirodalomban. A „taxonomy” jelentsen taxonómiát, a „systematics” pedig szisztematikát.

A TAXONÓMIA MINT HIPOTÉZISEKEN ALAPULÓ TUDOMÁNYÁG

A taxonómia meghatározását Dudich Endre a formális logikából vezette le. Eszerint a taxonómia önmagában nem természettudomány, hanem a logikának az ismeretek rendszerezésével foglalkozó ága. Bár ezt az álláspontot a hazai zoológusok széles körben elfogadták (Papp, 1986), szerencsére már egyre kevésbé kerül elő a taxonómia meghatározásánál. Ha a taxonómiát, mint a rendszerezés formális logikáját tekintjük, akkor Dudichnak igaza van. Azonban a taxonómiát sokkal inkább egy önálló tudományágnak kell felfogni, nem pusztán egy logikai keretrendszernek. Számos közelmúltban született cikk (például Nixon–Carpenter, 2011) fejti ki, hogy a taxonómia nem ugyanúgy működik, mint az élettelen dolgok rendszerezése, hiszen a leszármazási alapú rendszerezésnél közös ősökben és homológ karakterekben (mind egytől egyik hipotézisek) gondolkozunk.

A taxonómiát sokan „leíró” tudományágnak tekintik/tekintették kül- és belöldön egyaránt. Megjegyzem, nem azzal van bajom, ha valami „leíró jellegű”. A határmezsgye a tudomány és a tudománytalanság között nem a „leíró” és „hipotézistesztelő” tudományágak között van. A „leírás”, tehát a morfológiai struktúrák milyenségének megfigyelése nagyon fontos, de nem adja vissza a taxonómia mögött létező elméleti hátteret. A tudományfilozófiai írások közül a legjobban Quentin D. Wheeler (2004) fogalmazta meg a taxonómia hipotézistesztelő jellegét: „Az úgynevezett leíró taxonómia hipotéziseket tesztl több szinten. A kiinduló megállapítás, hogy két struktúra homológ, nem más, mint egy hipotézis. A konklúzió, hogy a homológ tulajdonságok eloszlási mintázata egy fajra jellemző karakterek összességét határozza meg (vagy egy magasabb rendszertani kategória szünapomorfiáit), ugyancsak hipotézis. Minden faj hipotézis, és minden klád bármely linnéi rendszertani kategóriában hipotézis. Ezen hipotézisek mind-

egyike általánosítás, amelyből jövőbeni megfigyelések vetíthetők előre. Amint a megfigyelések megszületnek, megerősítik vagy megcáfolják a hipotézist.”

Feltűnő, hogy a fenti idézetben Wheeler nem választja szét a taxonómiát és a szisztematikát. Az ok egyrészt valószínűleg az, hogy a szisztematika túlságosan a molekuláris filogenetika (Ebach et al., 2011, szemléletes kifejezésével „taxonomy light”) szinonimájává vált. Másrészt pedig a felismerés, hogy a taxonómia a rendszer alacsony szintjein (faj és alfaj) ugyanúgy hipotéziseket tesz fel, mint a felsőbb szinteken. A taxonómia és a szisztematika meghatározásánál valóban gyakran előkerül, hogy a taxonómia közvetlenül populációkból származó élőlényegyedeket vizsgál, míg a szisztematika magasabb rendszertani kategóriák közötti leszármazási viszonyokat (például Wägele, 2005). Úgy tűnik, mintha a populációk fajokba való csoportosítása alapvetően másfajta folyamat lenne, mint a fajok génszba, a génuszok családokba stb. csoportosítása, holott ez egyáltalán nincs így.

A SZISZTEMATIKA ÉS A MOLEKULÁRIS FILOGENETIKA

Nem mehetünk el amellett, hogy az elmúlt három évtizedben gyökeres változások történtek az élőlények evolúciós viszonyainak feltárásában. Elsősorban azért, hogy a DNS-szekvencián alapuló dendrogramok készítése mindennapos eljárássá vált. Papp (1994) megállapítása arról, hogy rengeteg cikk címe „taxonomy and phylogeny”-vel kezdődik, tökéletesen igaz, és a hasonló cikkek száma azóta jelentősen (valószínűleg exponenciális mértékben) nő. A *Web of Science* adatbázisban az 1987 és 2017 közötti cikkek között keresve (2017. júniusi lekérdezés alapján) a „taxonomy” + „systematics” kulcsszavakra 226, a „taxonomy” + „phylogeny” kulcsszavakra 601, valamint a „phylogeny” + „systematics” szavakra 414 találatot adott a publikációk címeire fókuszálva. A közlemények döntő része a „taxonomy and systematics of ... (taxon neve)” vagy „molecular phylogeny and taxonomy of ... (taxon neve)” típusú címmel rendelkezett. Természetesen előfordul, hogy nem lehetséges molekuláris dendrogramok összeállítása. Ennek több oka lehet, például a vizsgált élőlényeket oly módon kezelték begyűjtés után, hogy DNS kivonása nem lehetséges, esetleg kizárólag a vizsgált fajok DNS-t nem tartalmazó vázai álltak rendelkezésre, nem is beszélve a rég kihalt, csak fosszilis formában ismert fajokról (Páll-Gergely, 2017). Ettől eltekintve, bár évtizedekkel ezelőtt a systematics szó a biológiai tudományok tudományaként sok mindent magába foglalt (Simpson, 1961; Mayr, 1969), a mai szóhasználatban nagyrészt a phylogeny (= molecular phylogeny) szó szinonimája lett a nemzetközi szakirodalomban.

Mivel senki nem „volt ott” a ma élő kládok múltbéli elválásánál, a csoportok leszármazási viszonyairól csak hipotéziseink lehetnek. Még a legmodernebb DNS-mátrixok alapján készült törzsfák is csak hipotézisek, és csak közelíthetik a valóban megtörtént eseményeket. A leszármazási hipotézist felállíthatjuk ho-

mológ struktúrák összehasonlításával morfológiai alapon, és tesztelhetjük újabb, korábban nem vizsgált morfológiai bélyegek vagy például DNS-szekvenciák felhasználásával. Az, hogy egy kutya bal mellső lába és egy denevér bal szárnya homológ szervek, ugyanúgy hipotézis, mint hogy a két nevezett állat valamely génjének háromszázadik pozíciójában lévő bázispár homológ. Morfológiai alapon legfeljebb néhány tíz tulajdonságot hasonlítunk össze, és gyakran verbális értékelést adunk (nem használunk szoftvereket), a molekuláris filogenetikában (jó esetben) több ezer, de sokszor csupán néhány száz bázispár adatait futtatjuk le egy-egy elemzésben. A különbség könnyen látható, de nem kvalitatív, csak kvantitatív.

Egyértelmű, hogy a molekuláris filogenetika teljesen új távlatokat jelent az élőlények közötti leszármazási viszonyok felderítésében, és a filogenetikai eredmények számtalan meglepő információt szolgáltatottak. Nem szabad azonban elfelejtenünk, hogy a morfológiai bélyegek leggyakrabban elegendőek ahhoz, hogy a főbb monofiletikus csoportokat meghatározzuk. Továbbá, a fajok rokonsági viszonyait szemléltető molekuláris kladogramoknak semmi értelmük (legalábbis nem izgalmasak) anélkül, hogy tudnánk, mi is van a törzsfa ágai végén (Wheeler, 2004). A törésvonal „jó” és „rossz” tudomány között nem a morfológiai és molekuláris módszerek alkalmazása mentén húzódik. A magas minőség a pontosan azonosított példányokból, helyesen végrehajtott elemzésekből és magyarázatukból, valamint a jól feltett kérdésekből következik. A filogenetika egyik legnagyobb hibája és kihívása éppen az, hogy a molekuláris analízishez felhasznált minták nagy hányadát pontatlanul azonosították. A gombák esetén például 20%-ra tehető a rosszul meghatározott adatbanki (GenBank) szekvenciák aránya (Nilsson et al., 2006). A rendszertan jövője mindenképpen a masszív morfológiai alapon álló, molekuláris módszereket is alkalmazó rendszerezés, vagyis divatos kifejezéssel élve az integratív taxonómia.

TAXONÓMUS, REVÍZIÓ, NÉVADÁS

A taxonómiával kapcsolatban több fogalom van, amelyet a taxonómusok és a nem taxonómusok is használnak, de talán mindenki mást és mást ért rajtuk. Először is, kik a taxonómusok? Papp és szerzőtársai (2005) a következők írájk: „számunkra a zootaxonómus olyan fajta tudományos munkás, aki eddig nem ismert fajok leírásait és megbízható faunalistákat publikál pontosan azonosított bizonyítópéldányok alapján, és aki állatok referenciagyűjteményével dolgozik stb.”. Ezzel szemben Marcelo de Carvalho és szerzőtársai (2014) visszautasítják azt a leegyszerűsítést, amely a taxonómust a fajok leírását elvégző személlyel azonosítja. Szerintük „messze a fajok felfedezésén és elnevezésén túl a taxonómia tudománya evolúciós hipotézisek által vezérelt, melyek prediktív módon rendszerezik az élőlényeket,

és segítik az élővilág sokféleségének megértését aprólékos rendszertani revíziók és a homológ struktúrák értékelésén keresztül”. Wheeler (2014) a következőképpen vélekedik: „Ha arra összpontosítasz, hogy annyi fajt ismerj meg egy adott kládon belül, amennyit csak lehetséges, gondosan értelmezd és analizáld annyi homológ karakter átalakulásának történetét, amennyit csak lehet; ismerd meg az összes vonatkozó szakirodalmat 1753 óta; informatív neveket használj; és végül egy filogenetikai rendszerben összegezd mindazt, amit az adott taxonról tudni lehet, akkor egyértelműen taxonómus vagy.” Személy szerint az utóbbi két meghatározást tartom elfogadhatónak. Egy-egy faj leírása, sőt, számos új faj elhelyezése a rendszerben revízió nélkül sem tesz senkit taxonómu ssá. A taxonómusi munka elengedhetetlen feltétele a revíziók készítése, melyek legfontosabb eleme addig még nem vizsgált vagy félreértelmezett homológ struktúrák sokaságának összehasonlítása és értelmezése. Ezért a gyakran említett „revízióképes taxonómus” kifejezésből a jelzõt el is lehet hagyni: aki nem revízióképes, az nem taxonómus.

Itt ki kell térnem még egy gyakori félreértelmezésre. A szakirodalomban és a kutatók közötti társalgásokban is gyakran fordul elő, hogy „elnevezett” vagy még „el nem nevezett” fajokról beszélünk, illetve olyasmirõl, hogy a fajok kipusztulnak még azelőtt, hogy „elneveztük volna õket”. Hangsúlyoznom kell, hogy a tudományos érték nem önmagában a név, hanem a mögötte lévõ tartalom. A névadás természetesen a taxonómia *velejárója*, de nem *célja*; a nómenklatúra eszköz, de nem tudomány (Papp et al., 2005). A taxonómiai revízió célja az élőlények meghatározhatóvá tétele és a köztük lévõ leszármazási kapcsolatokról alkotott hipotézisek felállítása. Egy taxon tudományos neve nem más, mint egy, a jellemzõk eloszlásáról alkotott hipotézis rövid jelölése, amely lehetővé teszi, hogy a kutatók a lehető legprecízebben és a lehető legkevesebb szó felhasználásával értekezzenek (Wheeler, 2004).

JAVASLAT A TAXONÓMIA ÉS A SZISZTEMATIKA MEGHATÁROZÁSÁRA

A következõ javaslatot szeretném tenni a taxonómia és a szisztematika meghatározásáról és a kettõ egymáshoz füzõdõ viszonyáról. Javaslom, hogy a magyar szakirodalomban a taxonómiát és a szisztematikát együtt rendszertan néven nevezzük. A szisztematika fogalmát tudománytörténeti okokból, valamint azért mert a taxonómiától részben eltér módszereiben, fontosnak tartom fenntartani.

A taxonómia feladata elsõsorban az, hogy (fõleg morfológiai) karakterek alapján pontosan azonosíthatóvá tegye az élőlényeket, és elkülönítse a fajokat. A szisztematika homológ karakterek (morfológiai, molekuláris vagy egyéb) segítségével hipotéziseket állít fel és tesz tel az élõ és kipusztult fajok, valamint magasabb rendszertani kategóriák egymáshoz füzõdõ leszármazási viszonyairól.

A taxonómiát nem lehet elképzelni anélkül, hogy a vizsgált organizmusok rokonsági viszonyairól legalább érintőlegesen ne értekezni. Egy fajt lehetetlen leírni anélkül, hogy ne helyeznénk bele a rendszerbe, és ez alkalmanként a rendszer megváltozását vonja maga után. Megjegyzem, hogy itt nem csupán arról van szó, hogy az új fajoknak egyből génusznevet is kell adni (amit Podani János [2010] Linné „óriási tévedésének” hív). Leszármazási viszonyokról pedig nincs értelme anélkül értekezni, hogy az élőlények tulajdonságaival ne lennénk tisztában. Utóbbira kizárólag a taxonómia művelőinek aprólékos munkája eredményeként látunk rá. Ezért a taxonómia a szisztematikától nem választható el szorosan; a revíziókat végző taxonómusok szükségszerűen taxonómiával és szisztematikával is foglalkoznak.

ÖSSZEFOGLALÁS

Az írásom fő megállapításai a következők:

(1) Rá szeretnék mutatni, hogy a magyar szakirodalomban jelen lévő, szélesen értelmezett szisztematika meghatározáson (Papp, 1986, 1994, 1996; Balogh–Papp, 1988; Papp et al., 2005) kívül van egy szűkebben értelmezett változat is, amelyet a nemzetközi szakirodalomban gyakrabban használnak. A szisztematika tág értelmezése vezetett oda, hogy a taxonómia és a szisztematika elkülönülése élesen jelenik meg a hazai szakirodalomban.

(2) A molekuláris filogenetika (molecular phylogeny) mindennapossá és rutinszerűvé, és manapság sokak szemében nagyrészt a systematics szinonimájává vált. Nem arról van szó, hogy a kutatók nem értenék a fogalmak jelentését, hanem arról, hogy – kissé sarkítva – ma már kinek jutna eszébe nem DNS-alapú dendrogramokkal vizsgálni a leszármazási viszonyokat? Míg a fajok leírása és meghatározhatóvá tétele (szűk értelemben vett taxonómia) elsősorban morfológiai alapon működik, addig a csoportok közötti rokonsági viszonyok felderítése a morfológián túl egyre gyakrabban molekuláris markerek segítségével történik.

(3) Az a diszciplína, amely az élőlényeket leírja, elnevezi, és a köztük lévő rokonsági viszonyokat kutatja, többszintű hipotézisteszteléssel működik. Ez a tudományág céljait és tudományfilozófiai módszereit tekintve egységes a faj alatti szintektől a magasabb rendszertani kategóriáig. Tehát a hipotézistesztelő tulajdonság már a szűkebb értelemben vett taxonómia sajátja, nem csak a szűkebb értelemben vett szisztematikáé. Az elnevezése Wheelernél (2004) „taxonomy”, másoknál „taxonómia és szisztematika” vagy „taxonómia és rendszerezés”. Egységes volta miatt hajlanék a taxonómia elnevezés felé, viszont egyrészt tudománytörténeti okokból, másrészt módszertani (morfológiai vs. molekuláris) különbségekből kifolyólag javaslom a taxonómia és a szisztematika elnevezések használatát.

Balogh János és Papp László (1988) felhívta a figyelmet arra, hogy a taxonómusok ne kezdjenek parttalan vitákba, amelyek hátráltatják a taxonómiai kutatásukat. Ezzel egyetértve a fenti megállapításaim papírra vetését mégis fontosnak tartottam, főleg azért, hogy felhívjam a figyelmet a taxonómia és a szűk értelemben vett szisztematika szoros kapcsolatára és hipotézistesztelő természetére.

KÖSZÖNETNYILVÁNÍTÁS

Elsősorban szeretném kifejezni köszönetemet Varga Zoltánnak, aki a kéziratot bírálta, és számos hiányosságára rámutatott. Köszönetemet fejezem ki Dorian Dörge-nek és Adrienne Jochumnak a szakirodalom beszerzésében nyújtott segítségükért, valamint Fehér Zoltánnak, Kontschán Jenőnek, Murányi Dávidnak és Szekeres Miklósnak hasznos tanácsaikért.

IRODALOM

- Balogh J. – Papp L. (1988): How to Avoid Unprofitable Debate on Systematics among Taxonomists. *Opuscula Zoologica (Budapest)*, 23, 83–92. http://epa.oszk.hu/02300/02340/00032/pdf/EPA02340_opuscula_zoologica_tom23_1988_04.pdf
- Blackwelder, R. E. (1962): Animal Taxonomy and the New Systematics. *Surveys of Biological Progress*, 4, 1–57.
- de Carvalho, M. R. – Ebach, M. C. – Williams, D. M. et al. (2014): Does Counting Species Count as Taxonomy? On Misrepresenting Systematics, Yet Again. *Cladistics*, 30, 322–329. DOI: 10.1111/cla.12045, http://www.ib.usp.br/~lfsilveira/pdf/a_2013_counting.pdf
- Dudich E. – Loksa I. (1969): *Állatrendszertan*. Budapest: Tankönyvkiadó
- Ebach, M. C. – Valdecasas, A. G. – Wheeler, Q. D. (2011): Impediments to Taxonomy and Users of Taxonomy: Accessibility and Impact Evaluation. *Cladistics*, 27, 1–8. DOI: 10.1111/j.1096-0031.2011.00348.x, <https://bit.ly/2ke3G1B>
- Holzinger, W. E. (2007): A zootaxonómia nemzetközi irányvonalai. *Magyar Tudomány*, 168, 11, 1387–1393. <http://www.matud.iif.hu/07nov/03.html>
- Huxley, J. A. (ed.) (1940): *The New Systematics*. Oxford: Clarendon Press
- Mahunka S. – Vásárhelyi T. (1990): A zoológia Magyarországon. Fontos-e kutatnunk hazánk állatvilágát? *Magyar Tudomány*, 90, 9, 1055–1060.
- Mayr, E. (1963): *Animal Species and Evolution*. Cambridge, Massachusetts: Harvard University Press
- Mayr, E. (1969): *Principles of Systematic Zoology*. New York: McGraw-Hill
- Minelli, A. (1993): *Biological Systematics. The State of the Art*. London: Chapman & Hall
- Nilsson, R. H. – Ryberg, M. – Kristiansson, E. et al. (2006): Taxonomic Reliability of DNA Sequences in Public Sequence Databases: A Fungal Perspective. *PLoS ONE*, 1, 1, e59. DOI:10.1371/journal.pone.0000059, <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000059>
- Nixon, K. C. – Carpenter, J. M. (2011): On Homology. *Cladistics*, 28, 2, 160–169. DOI: 10.1111/j.1096-0031.2011.00371.x, http://lhc.ib.usp.br/BIZ0433/lib/exe/fetch.php?media=literature:nixon_and_carpenter_2012.pdf

- Páll-Gergely B. (2017): Should We Describe Genera without Molecular Phylogenies? *Zootaxa*, 4232, 4, 593–596. <https://bit.ly/2IC6STq>
- Papp L. (1983): A zootaxonómia hatékonyságának egyes kérdései. *Állattani Közlemények*, 70, 63–67.
- Papp L. (1986): A taxonómia és a szisztematika viszonyáról. *Állattani Közlemények*, 73, 61–68.
- Papp L. (1994): A zootaxonómia („zoológia”) száz éve és bizonytalan jövője. *Magyar Tudomány*, 12, 1428–1436.
- Papp L. (szerk.) (1996): *Zootaxonómia egységes jegyzet*. Budapest: Dabas-Jegyzet Kft.
- Papp L. – Peregovits L. – Ronkay L. (2005): *Zootaxonomy from a Hungarian Perspective. A Status Report and/or A Pamphlet*. Vácrátót: Hungarian Biodiversity Platform, <https://bit.ly/2IZ8NRd>
- Podani J. (2010): Evolúció, törzsfa, osztályozás. *Magyar Tudomány*, 171, 10, 1179–1192. <http://www.matud.iif.hu/2010/10/03.htm>
- Quicke, D. L. J. (1993): *Principles and Techniques of Contemporary Taxonomy*. Dordrecht, The Netherlands: Springer
- Schuh, R. T. – Brower, A. V. Z. (2009): *Biological Systematics. Principles and Application*. Second Edition. Ithaca–London: Cornell University Press
- Simpson, G. G. (1961): *Principles of Animal Taxonomy*. New York–London: Columbia University Press
- Wägele, J-W. (2005): *Foundations of Phylogenetic Systematics*. München: Verlag Dr. Friedrich Pfeil
- Wheeler, Q. (2004): Taxonomic Triage and the Poverty of Phylogeny. *Philosophical Transactions of The Royal Society, London B, Biological Sciences*, 359, 571–583. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1693342/>

Könyvszemle

SIPOS JÚLIA GONDOZÁSÁBAN

„MEGROSTÁLÁS” HITVITA KRITIKAI KIADÁSBAN

Az 1609-es év a magyarországi hitvitázó irodalom történetében különleges jelentőségűnek mondható. Gyors egymásutánban jelentek meg Pázmány Péternek és vitapartnerének, Alvinczi Péternek az írásai, összesen négy kiadvány, ami az akkori nyomdai lehetőségek mellett korántsem gyakori teljesítmény. A sort a méltán elhíresült ironikus polémia, a szerző nevének feltüntetése nélkül megjelent *Őt szép levél* (Grác, 1609) nyitotta meg az év tavaszán, erre felelt meg a kassai református prédikátor az azóta sajnálatos módon elveszett írásában (*Egy tetetes, neve vesztett pápista embertől S.T.D.P.P.től küldetett színes öt levelekre rend szerint való felelet Alvinczi Péterőtől a kassai m. ekklesia lelki pásztorától*, Debrecen, 1609). Alig néhány hét múlva megszületett a viszontválasz a jezsuita szerző most közreadott, *Megrostálás* című vitairata révén, és úgy, hogy közben még a hitviták korának legfullánkosabb, a protestáns országgyűlési rendeket minden korábbinál erősebben felháborító írását is kibocsátotta (*Az nagy Calvinus Jánosnak Hiszekegy Istene*). A Pázmány-szövegek a közismert millenniumi kiadásban ugyan eddig is a kutatás rendelkezésére álltak, művelődéstörténeti értékük és irodalmi-nyelvi jelentőségük vitán felül állt, azonban a korszerű textológiai és forráskritikai követelményeknek ezek a 19. század végi edíciók mostanra már nem felelhetnek meg. Indokoltan merülhet fel a kérdés, hogy a több mint négy évszázaddal ezelőtti hitvitázó írások kiadásának van-e ma is hasznosítható, a történeti értékeken túlmutató jelentősége. Úgy véljük, e kérdésnek a megválaszolásához jelen kiadás több vonatkozásban is hozzájárulhat, és támpontokat adhat.

Mint köztudott, ezzel a kötettel már kilencedik darabjához érkezett a Pázmány-életmű kritikai kiadása, a nagyívű vállalkozásnak így most újabb darabja csatlakozhat az eddig megjelentekhez. Nem kétséges, hogy a rendkívül terjedelmes szövegkorpusz teljességre törekvő kiadásának történetében mérföldkönek tekinthető az összes művek kritikai edíciójának programja, amelyet a Pázmány Péter Katolikus Egyetem kutatócsoportja Hargittay Emil vezetésével folytat. A mostani kötetnek már hosszú címe (*Alvinczi Péternek sok tétovázó kerengésekkel és cégéres gyalázatokkal felhalmozott feleletinek rövid és keresztyéni szelíd-séggel való megrostálása*, Pozsony, 1609) is jelzi, hogy stilisztikai leleményekben

bővelkedő, a régi magyar nyelvet, retorikát reprezentáló szövegről van szó ebben az esetben, ami a kommunikációtörténet oldaláról ugyancsak érdeklődésre tarthat számot. Főként talán azért, mert argumentumok feszülnek itt egymásnak, teológiai érvek és ellenérvek csapnak össze a retorikai és logikai szövegszerkesztési elvek érvényesülése mellett, s noha a személyeskedő megjegyzések sem hiányoznak egyes mondatokból, mégsem azok uralják az eszmecserét, az *ad personam* alakzatok elsősorban dekoratív (többnyire nyelvteremtő erejű) élénkítő alakzatai a fejtegetéseknek. Csupán ízelítőül néhány ezek közül: Pázmány vitapartnerének „forgószelehez hasonló gyors elméjéről”, „elálmétkodtató eszeskedéséről”, „buborék módgyára felfútt szókkal való nyelveskedéséről”, „foltos nyálaskodásáról”, „istállóhoz illendő poshadt szitkozódásáról” ejt szót, de azt is elismeri, hogy annak szövegében van „ennéhány jó módgyával ejtett magyarismus”. Látható, hogy az anyanyelvi leleményeket még az oly élesen szembenálló vitapartnernek is elismeréssel nyugtázták egymás írásaiban.

Mivel Alvinczi vitairata az *Öt szép levélnek* csak az első kettőjére adott választ, ezért a *Megrostálás* is két egységre tagolja a választ. Az elsőnek fő témája a szentek és képek tisztelete, a második pedig „az hittelt kötött fogadásnak erősségéről”, azaz a szószegésnek a különböző konfesszionális felfogásáról és történeti eseteiről (Husz János sorsa a konstanzi zsinaton, békekötés és békeszegés a törökkel, várnai vereség) értekezik.

A főszöveg után Báthory Orsolya tanulmánya részletesen ismerteti a mű keletkezési körülményeit és utóéletét. Megtudjuk innen azt is, hogy a *Megrostálás* 1609-es eredeti pozsonyi kiadása mindössze két példányban maradt fenn, ebből egyet Debrecenben, egy másikat Prágában őriznek, mindkettő kolligátumba kötve maradt fenn, s mindkettőt hasznosította a jelen kiadás.

A *Megrostálás* már csak azért is különleges helyet foglal el a polemikus műfaj történetében, mivel annak szövegéből részben rekonstruálni lehet az elveszett Alvinczi-vitairat tartalmát. Pázmány mintegy nyolcvan esetben szó szerint idézi az ellene felhozott vádakát, összegzi vitapartnerere érveit, szembesíti egymással a kétféle argumentációt. Természetesen ez nem helyettesítheti ismeretét az eredeti Alvinczi-szövegnek, amelynek előkerülésére már aligha lehet reményünk, de mégis ad róla egyfajta információt, a vitapartner tükréből is felsejlenek a kontroverz fejtegetés kontúrjai. Pázmány szigorú logikával dolgozott, mindig pontról pontra szó szerint citálta vitapartnerere szavait, s azokhoz fűzte cáfolatát, így lehetőség kínálkozott az elveszett Alvinczi-szöveg viszonylagos rekonstruálására. Báthory Orsolya ezt a nem gyakran adódó lehetőséget ki is aknáztta, s a szövegrészletek összeállításából (Heltai János korábbi stilisztikai megállapításait is kamatoztatva) körvonalazni tudta a ma már példányból nem ismert kiadványt.

A mű keletkezési körülményeit és későbbi sorsát bemutató tanulmányt követik a tárgyi jegyzetek mintegy hatvan lapon át, ezek minden szempontból megvilágítják és értelmezik a szövegben előforduló filológiai utalásokat. Vannak köztük

szómagyarázatok, a régies fogalmak értelmezései, latin citátumok fordításai, a szövegben említett történelmi személyek fontosabb életrajzi adatai. Közismert, hogy az ilyen munkában nem könnyű megtalálni a középutat egyfelől a túl szűkmarkú, másfelől a túlzottan bőbeszédű magyarázatok közt, itt a sajtó alá rendezőnek sikerült e téren a megfelelő arányt megtartania.

A tárgyi magyarázatokat a mutatók egész sora követi. Előbb a Pázmány által használt művek listáját kapjuk meg, majd az általa citált művek szerzők szerinti mutatója következik a rövidítések feloldásával. Mindennek alig túlbecsülhető jelentősége, hogy bepillantást enged az írói műhelybe, a kora újkori szövegszerkesztői gyakorlatba, mégpedig a korszak egyik legtöbbit idéző, legszélesebb olvasottságú szerzőjének révén. A marginális rövidítések feloldása olykor rejtvényfejtéshez hasonlítható művelet, amelyet természetesen ma már az elektronikus könyvtárak használata tesz eredményessé. Természetesen a személy- és helynévmutató, valamint a bibliai könyvek regisztrációja sem hiányzik a kiadványból, amely így a további kutatások lehetőségét is felkínálja a kora újkort kutatók számára.

Az eddig megjelent kötetek visszhangja igazolja, hogy a szövegközlés és a jegyzetelés során kialakított módszer eredményes és nagyon tanulságos a többi régi magyar irodalmi szöveg kiadása számára is. A most elkészült kötet viszonylag kisebb terjedelme tette lehetővé, hogy tárgyi jegyzetei is egyidejűleg elkészüljenek a főszöveghez csatoltan, így azok azonos kötetben kaphattak helyet, ellentétben a korábbi terjedelmesebb művek esetében alkalmazott eljárással. Pázmány hallatlanul bőséges utalásrendszere, szövegeinek hivatkozásokban és marginálisokban való gazdagsága rendkívüli szakmai felkészültséget kívánó, időigényes munka. Ennek elvégzéséhez a korszak történetének elmélyült ismerete, irodalom- és nyelvtörténeti, teológiai, patrisztikai, klasszika-filológiai tájékozottság egyaránt szükségesnek mutatkozott. A sajtó alá rendező Báthory Orsolya elismerésre méltó pontossággal végezte el a textológiai és a jegyzetelési munkát egyaránt. Mind a magyar, mind a latin részek betűhív átírásban készültek, a szöveg valamennyi sajátosságát megőrizve, csupán a mai magyar ábécében nem használt betűalakok és jelek átírása történt meg. A gondos textológiai lektorálás – Vadai István munkája – ugyancsak biztosítékot jelentett afelől, hogy ez a kötet is a korábbiakhoz hasonló módszerrel és szakszerűséggel készüljön el. Nem véletlen a belső címlapon olvasható információ, mely szerint a sorozat egésze „a Magyar Tudományos Akadémia Textológiai Munkabizottsága által elismert kritikai kiadás”.

Jelen kötet nem csupán a mai kornak megfelelő textológiai és filológiai elveket érvényesítette maradéktalanul, hanem a modern technika segítségét is igénybe vette, a kiadványhoz csatolt CD-n is közreadta a szöveget. S ami szintén nagy nyereség: a *Megrostálás* mellett az 1609-es év vitaindítója, az *Őt szép levél* textusa is szerepel a lemezen, mintegy előkészítve ennek a remélhetőleg hamarosan bekövetkező közreadását.

A sorozat kilencedik darabjaként most közreadott kötet mindenképpen erősíti azt a reményt, hogy „a magyar bíboros Ciceró” összes munkái belátható időn belül korszerű kritikai kiadásban állnak majd rendelkezésre, amelynek révén a korábbiaknál jóval árnyaltabban ismerhetjük meg a kora újkori vitakultúrát, ezen belül azokat a kommunikációs stratégiákat, amelyek a hitviták kontextusában nyelvteremtő erőt mutattak. A rangos vállalkozásnak mind eddigi munkáját, mind jövőre folytatását számos intézmény és pályázat (MTA, NKA, OTKA, Bolyai-ösztöndíj) által biztosított támogatás tette és teszi lehetővé.

(Pázmány Péter: Alvinczi Péternek sok tétovázó kerengésekkel és cégéres gyalázzatokkal felhalmozott feleletinek rövid és keresztyéni szelidséggel való megrostálása (1609). Sajtó alá rendezte és a jegyzeteket írta Báthory Orsolya. [Pázmány Péter Művei, kritikai kiadás, sorozatszerkesztő Hargittay Emil] Budapest: Universitas Kiadó, 2017, 222 o.)

Bitskey István

az MTA rendes tagja, professor emeritus, Debreceni Egyetem

HELYI ÉRTÉK A MŰVELŐDÉSTÖRTÉNETBEN

A 2017-ben megjelent egri *Kulturális örökség tanulmányok* 3. kötetének jelentőségét mindenekelőtt az a széles spektrum jelenti, amelyben a többnyire fiatal kutatók színvonalas tanulmányaikban beszámolnak kutatásaik eredményeiről. A könyvtörténeti és irodalomtudományi tematika mellett jelen van például az egyháztörténet is, noha szigorúan helytörténeti írásnak a könyvben csak Veres Gábor munkáját nevezném (*A kulturális örökség népi műemléki értékei Heves megyében*). A Czeglédi László, Verók Attila és Mizera Tamás szerkesztésében kiadott mű előszavában Monok István reflektált a kutatói társadalmat érintő, nyomasztó tudománymetriai jelenségre, mely ahhoz kapcsolódik, hogy a kutatókat publikációik mennyiségének mérőszámával jellemzik: így „[a] kutató ember közlési kényszerben él” (7). Úriemberhez illő eleganciával nem fedte fel ugyanakkor az ebből származó problémát: *A helyi értékhez* hasonló tanulmánykötetekben bizony gyakran megjelennek nem kellően kidolgozott munkák is. Szerencsére *A helyi érték* mind a tíz művelődéstörténeti tanulmányára igaz, hogy „történetileg hiteles, megközelítésében modern” (7), mondhatnánk, ez egy olyan összeállítás, amelyhez viszonyítani lehet.

A tanulmánykötet egészéből fiatalos lendület és módszertani változatosság sugárzik. A helytörténeti munkákhoz képest újdonság mutatkozik a témakezelésben is. A kötetben például helyet kap két nőtörténeti tanulmány is, melyek vallásos kontextusban tárgyalják a női életutakat. Szűcs Kata *Portugáliai Szent Erzsébet kultuszának újabb emlékeivel* foglalkozott munkájában. Portugáliai és Árpád-házi Szent Erzsébet hagiográfiájában egyaránt megtalálható a rózsacsoda, és a szerző tanulmányában egyrészt a két szent élete és csodái párhuzamosságainak feltárását végezte el, másrészt azt igazolta, hogy a portugáliai szent hagiográfiájába a magyar szent életéről szóló leírásokból került át ez a motívum. Számot tarthat az érdeklődésre az is, hogy Izabel, aki mélyen vallásos királynéként még a női nem iránt könnyen elgyengülő férje, I. Dénes törvénytelen gyermekeit is befogadta és felnevelte, atyai nagyanyjának testvére után kapta nevét, aki éppen Árpád-házi Szent Erzsébet volt. Utóbbi tiszteletére szenteltette fel a ma Santa-Clara-a-Velhaként ismert monostort, ahova férje halálát követő zárandoklata után klariszsa apácaként visszavonult, és ahol 1336-ban megtért az Úrhoz. (Némiképpen megkönnyebbülhetünk, olybá tűnik, mintha nem csak a magyar arisztokrácia bánt volna rútol az idegenből származott megözvegyült királynéival.) A Szent Királyné több kórházat, szegény- és lepraházat alapított. Szűcs Kata felmutatja a két női szent életrajzának és legendáriumának sok közös vonását, melyeket körültekintően végigvesz, csakúgy, mint a rózsacsoda (egyébként férfi mártírok-

hoz is kapcsolódó) hagiográfiai hagyományának szakirodalmát. A két asszony tiszteletét európai kontextusban is értelmezi: „A portugál szent elsőként képviselte az Ibériai-félszigeten azt az új típusú női szentséget, amely a 13. század első harmadától kezdve oly népszerűvé vált Európában, és melynek első kanonizált képviselője Árpád-házi Szent Erzsébet volt. A karitatív, ferences eszmeiségből eredeztethető modell képes volt másfajta szerepet biztosítani a nők számára a társadalomban – elsősorban az uralkodó- és a nemesi családok női sarjainak – azáltal, hogy egyre inkább a javaikról lemondó, fogadott szegénységben élő, életüket a rászorulóknak megsegítésének és a betegek ápolásának szentelő királynők és nemes asszonyok erősítették a dinasztia politikai befolyását” (16.). Szűcs Kata kutatása azt mutatja, hogy szükséges és érdemes elővinnünk régi filológiai kérdéseket is, hiszen a szerző a Szent Királyné életrajzainak, a portugál viták (életrajzok) és a latin fordítások filológiai elemzésével remek, logikailag is kiválóan felépített tanulmányában így meggyőző érvelésben állíthatja, hogy a rózsacsoda mindkét női szent hagiográfiájában utólagos betoldásnak bizonyul. Itt szükséges kitérni arra, hogy egy, a szentekétől módszertanában is eltérő kultuszvizsgálatról is olvashatunk a kötetben, hiszen Bényei Miklós *A debreceni Kossuth-kultusz tetőpontja: a kilencvenedik születésnap* című írásában a cívisváros 1892-es ünnepségét elemezte. A tanulmány nemcsak azért izgalmas, mert figyelemmel van Kossuth születésének napja körüli bizonytalanságra, hanem mert a debreceni kultuszt egy olyan politikus esetében vizsgálta a szórólapok, a publicisztika és a városiak körében a rendezvényre fókuszálva, aki ekkor már kiszorult a politikai életből, „élő szoborrá” (98.) vált. A dolgozat olyan groteszk lépésekre is kitér, mint például, hogy miután az ünnepség bevételeiből származó 1170 Ft-ból Kossuth nevére elnevezett lelenházat alapítottak, a csizmadia iparosok vendéglőjüket is szerették volna Kossuthról elnevezni, s erről a *Debreceni Ellenőr* megbotrányozását kifejező cikket közölt: „a nagy Cézár sincs valami nagyon megtisztelve, amikor nevére most itt-ott megtermett bulldog hallgat” (id. Bényei, 118.). Mivel a polgári átalakulásban jeles szerepet játszó Kossuth a névadáshoz nem járult hozzá, a csizmadiaik kénytelenek voltak elállni a tervtől.

A polgári átalakuláshoz is kapcsolódik a kötet második nőtörténeti tanulmánya. A királyi származású női életutak mellett ebből az írásból az arisztokrata és polgári női életpályákról is tájékozódhatunk. Rácsai Rita *Nőkép a 18. századi református halotti prédikációkban* című dolgozatában kötődik a feminista történetírás azon megállapításához, hogy a „történészek által feltárt múltból hiányoznak a nők” (55.). Kutatásában a Debreceni Református Kollégium Nagykönyvtárának 1720 és 1734 között keletkezett halotti prédikációkat közlő nyomtatványait (az ún. halottas könyveket) vizsgálta. A prédikációk „ismertetik a kor általános elvárásait a nőkkal szemben, másrészt konkrét élettörténeteket beszélnek el” (56.). A szerző által vizsgált elhunyt és megbecsülést érdemlő asszonyok férjei elismert református államférfiak vagy tudósok voltak. A vizsgált személyek Bethlen Ju-

dit, Kun Borbála, Köleséri Sára, Ubrisi Mária, Rhédey Krisztina és Bánffy Kata asszonyok. Rácsai Rita kritikus összegzésében nem kis bátorságról árulkodó és a posztmodern történetíráshoz közelítő soraiban leírta, hogy „a halotti prédikációk sosem mutatják be a teljes igazságot, inkább csak az elvárásokat” (74.). Meglepő megfigyeléseinek egyike, hogy a nőkkel szemben támasztott expektanciák századok óta ugyanazok (jó és csinos feleségnek, szerető anyának és okos gazdasszornynak egyszerre kellett lenni), s csak az változott, hogy ezeknek a várakozásoknak miként teszünk eleget. A kiváló tanulmány a női öltözködéstől az élet minden területéig (többek között a betegápolás, a szerelem, az anyaság, a szakmai és tudományos tevékenység iránti érdeklődés, a házasságban való együttélés) vizsgálja a 18. századi asszonyok életét, hogy a jó feleség képéről a lehető legteljesebb ábrázolást kaphassuk. Monok István tanulmányából ide kapcsolódik, hogy az asszonyhagyatékokban is lejegyezték a könyvállományt, ami – bármily meglepő is – azt jelenti, hogy a kassai polgárok számára is fontos volt a nők műveltsége már a 17. században is (29.). Rácsai Rita tanulmányának zárásában talán túlságosan is szerény: noha valóban csak egyféle vallású értelmiség feleségeinek életével foglalkozott, de tanulmánya az újkori nők életének kutatását gondos adatokkal gazdagította, s példákkal igazolta, hogy a nő műveltségére szükség lehetett, hiszen férje távollétében, betegeskedésekor még gazdasági kérdésekben is dönthetett (71.). Hasonlóan innovatív Faa-Lendvai Erzsébet precíz, hallgatói névsorokkal gazdagított nagyszombati matrikulák vizsgálatán alapuló tanulmánya, mely szerint már a bencés rend újrászervezését (1637) követő tanévben tanult Nagyszombatban bencés diák. Ugyan a pannonhalmi rendtörténet szerint a 18. században negyven bencés szerzetes tanult Nagyszombatban, a szerző kutatásai alapján csupán tizenötörről tudjuk ezt biztosan, mert csak ennyi bencés hallgató anyakönyve áll rendelkezésünkre ebben az időszakban.

A kötet írásai közül a szűkebb értelemben vett vallásos tematikához tartozik még az előbbieken kívül az 1848-as egri egyházmegyei zsinat előkészületeivel foglalkozó Tengely Adrienn magyar egyháztörténeti munkája. Írása a magyar történelem egyik legkutatottabb időszakához nyújt új adalékokat. A tanulmányból megtudhatjuk, hogy a cölibátus eltörlésének kérdése már ekkor is nagy vitákat generált, de más tartalmi elemeket, például egyházszerkezeti és vallási életre vonatkozókat is vizsgált az értekezés. Sajnálatos, hogy a forradalom miatt a nemzeti, illetve az azutánra ütemezett egyházmegyei zsinatra már nem került sor.

A tanulmánykötet módszertani sokszínűséget mutatja az is, hogy az írások közül a következő nagyobb csoportot az olvasástörténeti munkák adják, melyek közül két könyvtörténeti értekezés mindenképpen kiemelendő: Monok István *A kora újkori Kassa német polgárságának olvasmányairól* és Mizera Tamás *Erudíció és neveléstörténet: A kora újkori Felvidék tanítókönyvtárainak vizsgálatáról* szóló írásai. A Monok István által vizsgált téma a szerző saját bevallása szerint is kuriózumnak számít az egyébként divatosnak mondható kassai történetírás-

ban, ugyanis a polgárok olvasmányainak elemzése eddig elmaradt. Történt ez úgy, hogy Kassán a levéltári forrásokon túl a városban 1610-től kiadott művek és az országban működő könyvkereskedők által készített könyvösszeírások, a fennmaradt kötetek bejegyzései alapján képet tudunk alkotni a polgári olvasmányokról. A recenzióban először említett Szűcs Kata írásához hasonlóan, Monok István tanulmánya is európai keretekben gondolkodik a magyar művelődéstörténet vizsgálatakor. Ez az attitűd az egész kötetnek egy nem eléggé dicsérhető erénye. A kassai hagyatéki inventáriumok fennmaradása ugyanis Monok István elmondása szerint a nyugat-európai arányokkal azonos: „minden tizedik polgár halálakor készülhetett összeírás, és ennek 10-15%-a maradhatott fenn” (23.). A szerző izgalmas konkrét példaválasztásokkal jellemezte az elmagyarosodó kassai polgárság olvasmányait. Klisé, hogy az olvasmányok sokat elárulnak az emberekről, de ez jelen esetben is igaz. Venceslaus Machilles fegyverkovácsról például Monok István feltételezte, hogy „menekülni vagy állást változtatni kényszerült értelmiségi lehetett” (25.). Könyvgyűjteménye több mint harminc könyvet számlált, és az ókori szerzőkön kívül Martin Luther, Johann Brenz, több Philip Melancthon-kötet mellett példának okáért görög nyelvű Újtestamentum is gazdagította könyvespolcait. Miután Monok István a város vezetőinek olvasmányait is áttekintette, megállapította, hogy „[t]ermészetesnek vehető, hogy egy tanítónak, lelkésznek, patikusnak vagy éppen egy felcsernek voltak könyvei, de fontos figyelni arra is, hogy a 17. század második felétől egyre több mesterember, önöntő, rézműves, szíjgyártó, bodnár vagy éppen szűcs hagyatékában fordulnak elő könyvek, néha 5–15 kötet is” (29.). Nem tartja véletlennek, hogy a 17. század második felében a polgárok Johann Arndt kötetét és az *Imitatio Christi* szövegeit (a magyarok Pázmány Péter fordításában) olvasták.

Mizera Tamás a 16–18. századi felvidéki tanítói erudícióval foglalkozott az inventáriumokban, melyek jogi mivoltuknál fogva hiteles forrásnak számítanak. A könyvtörténettel foglalkozók számára rendkívül hasznosak, forrásaik 70%-a ezekhez köthető (37.). A tanítók magánkönyvtárai azon polgári gyűjtemények közé tartoznak, melyek az egyes foglalkozások mentén tipikusnak mondhatók (ilyenek többek között az orvosok, jogászok, gyógyszerészek könyvgyűjteményei is). A polgári könyvtárak állományairól már Monok István tanulmányából is kiderült, hogy jellemzően 5–15 kötetből álltak, Mizera Tamás megállapította, hogy a tanítói százpéldányos gyűjtemények jelentősnek minősíthetők. Innen és a korábbiakból is látható, hogy a kötet tanulmányainak egymás mellé kerülése nem véletlenszerű, ami oly sokszor szintén hibája a hasonló munkáknak, hanem egymással valóban élő diskurzust folytató, egymás állításait kiegészítő, pontosító és támogató írások. Mizera Tamás tanulmányában ismertette a tanítói gyűjteményeket. A sok izgalmas személy és életpálya közül is kitűnik Váradi György, akinek könyvállományára kettős élete is rányomta bélyegét: mivel Losoncon és Szabadszálláson iskolamester volt, Losoncon és Ráckeven ugyanakkor jegyző is. Mizera

Tamás egy lábjegyzetet szentel annak, hogy egyébként Váradi olyan sokszínű személyiséget tudhatott magának, hogy életének korábbi éveiben csizmadiaként és kocsmárosként is dolgozott. A jogi és számtankönyveken túl ugyanakkor érdekelhették a gyógynövények is, mert két herbárium-könyvet is vásárolt. A sok magyar nyelvű köteten túl az evangéliumokat magyarul, latinul és görögül olvasta, illetve a gazdasági, politikai könyveken kívül érdekelte a humanista irodalom is. Okos, hogy Mizera Tamás reflektál arra tanulmánya végén, hogy a könyvek birtoklásából nem következik, hogy a tulajdonosok biztosan olvasták is ezeket. Ehhez a könyvekbe tett bejegyzéseket szükséges vizsgálni.

A két irodalomtudományi tanulmány egyaránt műfaji megközelítésben vizsgált prózai szövegeket, s emiatt mindkét szerző innovatívan közelítette meg témáját. Bartók Béla és Kuser Judit az eddig többnyire történelemtudományi munkák sorát irodalomtörténeti módszertannal gazdagította, hiszen Bartók egy keveset kutatott szerző, Bródy Sándor *Az egri diákok* című ifjúsági regényét vizsgálta a robinzonád felől, míg Kuser Gárdonyi Géza kisregényének, az *Ida regényének* recepcióját elemezte. A két tanulmány ugyanazon kötetben való megjelenését az is indokolhatta, hogy a regények írói egyazon évben, 1863-ban születtek. Gárdonyi 1922-ben, Bródy pedig 1924-ben hunyt el.

Örömmel megjegyezhetem, hogy a szerkesztés és a szöveggondozás is kifogástalan *A helyi érték* című tanulmánykötetben. A jövőre vonatkozóan azonban hasznosnak gondolnám a tanulmányok szerzőiről akár csak egy-két soros bemutatkozást közölni, hogy a kutatókat is megismerhessük, ne csak kiváló munkáikat, ezen párját ritkító könyvsorozat által.

(Czeglédi László – Mizera Tamás – Verók Attila szerkesztők: *A helyi érték. Kulturális örökség tanulmányok 3. Művelődéstörténeti kalászatok*. Eger: Liceum Kiadó, 2017)

Szabó P. Katalin

irodalomtörténész, doktorandusz,
Eötvös Loránd Tudományegyetem Irodalomtudományi Doktori Iskola

ÉGHAJLAT, TUDOMÁNY, TÖRTÉNETEK

„A tudományról és a tudósokról számos, sokszor nem éppen hízelgő sztereotípiát él a köztudatban. E könyv megváltoztatja korábbi elképzeléseinket, sokszínűbbé, gazdagabbá téve azt a tudományról, annak történetéről, jövőjéről és művelőiről egyaránt” – olvasható Jankó Ferenc *Éghajlat Tudomány Történetek* című könyvének borítóján. Ha egy szóval kellene jellemezni a könyvet, valóban a sokszínűség lenne a legtalálób, mely egyaránt megmutatkozik az interjúalanyokban, személyiségükben, látásmódjukban és az általuk művelt szakterületekben, valamint az érintett éghajlattudományi, klímapolitikai, földrajzi, ökológiai, erdészeti, mezőgazdasági, történelmi, közgazdaságtani és fenntarthatósággal kapcsolatos témákban. Betekintést enged az éghajlat- és hatáskutatás kulisszatitkaiba, sokszínűbbé teszi, szélesíti az olvasó látókörét.

A kötet témájának aktualitására az előszóban Vida Gábor mutat rá. Kritikai értékelést fogalmaz meg a 2015-ös párizsi „klímacsúcs” eredményeiről, megítéléséről a médiában és a társadalomban, valamint a megfogalmazott célok megvalósíthatóságának realitásáról és a szükséges lépések elmaradásának következményeiről. Felhívja a figyelmet az ember szerepére, felelősségére és kihívásaira a felmelegedés már megfigyelt és várható hatásainak, veszélyeinek mérséklésében. Hangsúlyozza a globális együttműködés szükségességét az alkalmazkodásban, és a minden szintre kiterjedő ismeretterjesztés szerepét, mely képes „a közvélemény elaltatása és a pánikkeltés közötti felelősségteljes mértéket megtalálni”.

A téma komplexitásából adódóan a könyv struktúrájának összeállításánál az interdiszciplinaritás a fő vezérelv. A szerző geográfus diplomáját és PhD-fokozatát az Eötvös Loránd Tudományegyetemen szerezte. Az olvasó kíváncsiságát fokozhatja, hogy ebből adódóan más perspektívából, nem éghajlatkutatói háttérrel és szemlélettel közelíti meg a klímaváltozással és hatásaival kapcsolatos kérdéseket. A könyv egyediségét adja, hogy nem csupán a tudományos problémákat helyezi a középpontba, hanem a bemutatott történeteken keresztül megismerhetjük a beszélgetőpartnerek egyéni, szakmai életpályáját, témájához való kötődését és szerepét a tudományos élet alakításában. Az interjúalanyok megválasztásánál a szerző törekedett az eltérő szakterülettel, olykor eltérő véleménnyel rendelkező kutatók megszólaltatására. Az olvasóban azonban önkéntelenül felmerül a kérdés, hogy miért pont ők, mikor rajtuk kívül még vannak hasonló témákban jártas, több évtizedes nemzetközi tapasztalattal rendelkező szakemberek, de ez talán lehetőséget adhat a következő kötet összeállítására.

A három részre osztott kötet első részében Czelnai Rudolf és Probáld Ferenc történeteinek keresztül ismerhetjük meg a klímaváltozás egy-egy vetületét. A máso-

dik fejezetben három meteorológus, Bartholy Judit, Szépszó Gabriella és Lakatos Mónika tárják fel a klíma modellezésével, valamint az éghajlati adatbázisokkal kapcsolatos kérdéseket. A harmadik egységben Jolánkai Márton, Mátyás Csaba és Kerekes Sándor szól az ember és a természet kapcsolatáról.

Az egyes interjúk hasonló szerkezettel épülnek fel, az éghajlatváltozás problémája és a kutató közti egyéni kapcsolat mindegyikben kiemelt szerepet kap. A beszélgetőpartnerek beszámolnak a tudományos pályáik gyökereiről, pályaválasztásukról. Atfogó képet adnak az adott szakterület fejlődéséről, aktuális tudományos kérdéseiről, vitáiról, valamint saját helyükről, szerepükről, véleményükről, tudományos eredményeikről. Történeteiken keresztül betekintést nyerhetünk a nemzetközi kutatóintézetek, szervezetek, bizottságok működésébe, munkájába, főbb tudományos és klímapolitikai eredményeibe az elmúlt évtizedekben, melyekben az interjúalanyok is nagy szerepet vállaltak, és tevékenyek ma is. Feltárják az éghajlatváltozás kutatásának kezdeteit Magyarországon, az oktatás, ismeretterjesztés helyzetét, lehetőségeit, feladatait, valamint saját tapasztalataik alapján a magyar szakemberek nehézségeit, kihívásait a kutatói életpálya során. Megismerhetjük a beszélgetőpartnerek további terveit, vízióit, jövőképét, fejlesztési elképzeléseit is.

Minden interjút áthat a megszólított kutatók hivatástudata, kutatómunka iránti elkötelezettsége, szakterületük iránti lelkesedése. Ezért érzékelhető az éghajlati rendszerben és a társadalmi környezetben tapasztalható és jövőben várható változások miatt érzett aggodalmuk és tette sarkallásuk is. Az eddig tudományos cikkekből, szakmai bírálatokból, állásfoglalásokból, konferencia-előadásokról, katedráról ismert szakemberek most közvetlen hangvételben szólalnak meg, mely az interjúkészítő szerző felkészültségének is eredménye.

Az utolsó fejezet adja meg a könyv tudománytörténeti keretét. Ebben a szerző saját posztdoktori kutatási eredményeit összegzi a klímaváltozás, valamint a földhasználat-változás lehetséges éghajlatmódosító hatásai témákban folytatott vitákról és szerepükről a tudományban és társadalomban.

A könyv igazi értéke, hogy interdiszciplináris szemlélettel, mégis a tudományos mélységek kerülésével ad betekintést az éghajlatkutatás múltjába és aktuális kérdéseibe, melyet a kutatók személyes hangvételű háttértörténetei tesznek olvasmányossá. A bőséges hivatkozott szakirodalom lehetőséget ad az érdeklődőknek a szakmai háttér és részletek megismerésére. Ezért a könyv hasznos ismeretekkel szolgálhat a szakterületen és más tudományterületeken jártasok számára egyaránt, valamint minden olvasónak, aki nyitott a téma iránt.

(Jankó Ferenc: Éghajlat Tudomány Történetek. Budapest: Éghajlat Kiadó, 2017, 176 o.)

Gálos Borbála

egyetemi docens, Soproni Egyetem Erdőmérnöki Kar

KORSZERŰ MOTOR- ÉS SUGÁRHAJTÓMŰ ÜZEMANYAGOK I. MOTORBENZINEK ÉS II. DÍZELGÁZOLAJOK

A *Korszerű motor- és sugárhajtómű üzemanyagok* tankönyvsorozat új kiadásának első kötete az *I. Motorbenzinek* és második kötete a *II. Dízelgázolajok* címmel a közelmúltban jelentek meg (2014, 2016) második, bővített és teljesen átdolgozott kiadásban. Az új tankönyvsorozat tervezett további kötetei: *III. Alternatív motorhajtóanyagok*, *IV. Sugárhajtómű üzemanyagok*, *V. Belsőégésű motorok ke-
dőanyagai*.

A sorozat szerkesztője, Hancsók Jenő egyetemi tanár, okleveles vegyészmérnök, euromérnök (FEANI, Párizs), az MTA doktora, a Magyar Mérnöki Kamara és a Magyar Szabványügyi Testület szakértője. 2015-ig a Nemzeti Akkreditáló Testület külső szakértője. Öt tankönyv, négy szakkönyv, továbbá közel kilencszáz tudományos közlemény és előadás szerzője, illetőleg társszerzője; tizennégy, a vegyiparban részben megvalósított magyar, európai és világszabadalom egyik alkotója; 237 hallgató diploma-, szakdolgozat, diplomatervezés és harminckilenc tudományos diákköri dolgozat (nyolc fő országos I–III. helyezett), továbbá tizenöt sikeresen teljesített PhD-cselekmény témavezetője.

Az *I. Motorbenzinek* és *II. Dízelgázolajok* kötetek szerzői kollektívájának tagjai közül Hancsók Jenő professzor az egyes fejezetek szakmai elkészítését végezte, míg tanítványai és kollégái, Eller Zoltán, Kriván Eszter és Sági Dániel okleveles vegyészmérnökök feladata elsősorban a kéziratok gondozása, és azok megjelentetéshez való előkészítése volt.

A *I. Motorbenzinek* kötet első fejezete a mobilitást mint a fenntartható fejlődés egyik alappillért és kapcsolatrendszerét mutatja be. Ezen belül az egyes alfejezetek tartalmazzák a közlekedés és az áruszállítás tárgyi eszközeit és azok elterjedését, az energiaforrások korszerű rendszerezését, a közlekedés és a szállítás energiaigényét, a belsőégésű motorok környezetbarát és humánbiológiailag is megfelelő minőségű motorhajtóanyagaival szemben támasztott mennyiségi igények alakulását. Továbbá a hagyományos és újabb követelményeket, majd a motorhajtóanyagok osztályozási lehetőségeit. Ezt követően röviden áttekintik a benzinüzemű motorokat, azok csoportosítási lehetőségeit, károsanyag-kibocsátásuk szabályozási lehetőségeit, továbbá ezen belül a nemzetközi előírásokat (gépjármű- és motorgyártók előírásai, irányelvek, szabványok, rendeletek stb.).

A második fejezet a motorbenzinek előállításával foglalkozik. Ezen belül a keverőkomponensekkel és keverésükkel, illetőleg az adalékokkal. Ennek keretében részletesen tárgyalják a motorbenzinek keverőkomponenseinek célirányos előállítási lehetőségeit a kőolajiparban, a különböző szétválasztó és átalakító eljárásokat

(mint például a desztilláció, a kéntelenítés, a paraffinok alkilezése olefinekkel, az oxigéntartalmú vegyületek szintézise, a paraffinok izomerizálása, a benzinreformálás, a fluidkatalitikus krakkolás stb.), továbbá ezek kapcsolatrendszerét.

Az egyes eljárások tematikusan és logikusan felépített tárgyalásmódja (cél, szükségesség; nyers- és/vagy alapanyagok; kémia – reakciók és mechanizmusok, termodinamika stb. –; katalizátor(ok); műveleti paraméterek; általános és speciális elvi technológiai vázlatok; cél- és melléktermékek, valamint azok minőségi jellemzői és felhasználása stb.) a könnyű megértést szolgálja. A különböző eljárások során nyert benzináramok keverésének, valamint e művelet szabályozási és irányítási műveleteinek rövid bemutatását követi a motorbenzin-adalékok ismertetése, amelyek alapvetően a késztermékek végső minőségének beállítására szolgálnak.

A harmadik fejezet a motorbenzinek jellemző tulajdonságainak hatásait foglalja össze a motorra, az egész gépjárműre és a környezetre. Ehhez szorosan kapcsolódnak az utóátalakító katalizátorokat és hatásait (4. fejezet), a motorbenzinek és motorolajok kölcsönhatásait (5. fejezet), továbbá a motorbenzinek kezelését, tárolását és elosztását (6. fejezet) tárgyaló fejezetek. A szerzők külön fejezetben ismertetik a dugattyús motorral szerelt repülőgépek motorhajtóanyagainak működtetésére alkalmas, speciális igényeket kielégítő benzineket. Végül jövőképet festenek, bemutatva a motorbenzinek fejlesztési tendenciáit.

A II. Dízelgázolajok kötet első fejezete – az *I. Motorbenzinek* kötethez hasonlóan – a mobilitást mint a fenntartható fejlődés egyik alappilléret és kapcsolatrendszerét mutatja be. A második fejezet az energiaforrások korszerű rendszerezését, azok rendelkezésre állását, felhasználását, elsődleges források (kőolaj, földgáz, kőszén, nukleáris és alternatív energiák) szerint csoportosítva, a világ, az USA, EU és Magyarország tagolásban. A harmadik, főfejezet a szárazföldi mobilitás megvalósulását mutatja be, ezen belül a meghajtási módokat és elterjedésüket, a mobilitás tárgyi eszközeinek származtatott energiaforrásait, a közlekedés és a szállítás energiaigényét, a belsőégésű motorok környezetbarát és humánbiológiai is megfelelő minőségű motorhajtóanyagaival szemben támasztott mennyiségi igények alakulását, valamint a hagyományos és újabb minőségi követelményeket, majd a motorhajtóanyagok osztályozási lehetőségeit. A negyedik fejezetben röviden áttekintik a Diesel-üzemű motorokat, azok csoportosítási lehetőségeit, fejlesztésük főbb lépéseit és a további korszerűsítések lehetőségeit.

Az 5. fejezetben a Diesel-motorok működtetésére alkalmas kőolajalapú és alternatív forrásból származó hajtóanyagok összefoglaló bemutatása után a 6. fejezetben a dízelgázolajokat, ezen belül az azokkal szemben támasztott követelményeket, a minőség szabályozás közvetlen (emissziós előírások) és közvetett útjait (direktívák, szabványok, gépjármű- és motorgyártók előírásai stb.) ismertetik. A hetedik fejezet a dízelgázolajok kőolajalapú előállításával foglalkozik. Ennek keretében részletesen tárgyalják a dízelgázolajok keverőkomponenseinek

célirányos kőolajipari előállítási lehetőségeit, a különböző szétválasztó és átalakító eljárásokat (mint például a desztilláció, a kéntelenítés, az aromástelítés, az n-paraffinok szelektív hidrokrakkolása és/vagy izomerizálása, a hidrokrakkolás, a fluidkatalitikus krakkolás, a maradékok feldolgozása stb.), továbbá ezek kapcsolatrendszerét. Az egyes eljárások tárgyalásmódja (lásd a *Motorbenzinek* tankönyvnel) tematikusan és logikusan felépített; rendszerszemlélete segíti a könyvnyebb megértést.

Az ezt követő 8. fejezet a dízelgázolajok adalékolását mutatja be. Ennek keretében annak szükségességét és jelentőségét, a főbb adalékok hatásmechanizmusát és alkalmazási koncentrációit, továbbá az adalék-kompozíciókat ismerteti. A 9. fejezet röviden a dízelgázolajok keverésével, valamint e művelet szabályozási és irányítási műveleteivel foglalkozik. A 10. fejezet a dízelgázolajok jellemző tulajdonságainak hatásait foglalja össze a motorra, az egész gépjárműre és a környezetre. Ehhez szorosan kapcsolódnak a dízelgázolajok és motorolajok kölcsönhatásait (11. fejezet), utóátalakító katalitikus rendszereket és hatásait (12. fejezet), továbbá a dízelgázolajok kezelését, tárolását és elosztását (13. fejezet) tárgyaló fejezetek. A könyv ismeretanyagát a dízelgázolajok fejlesztési tendenciáinak felvázolása zárja.

A tudományos igényességgel készített tankönyvek – korunk korszerű műszaki követelményeit kielégítve – lehetővé teszik a tématerület átfogó ismereteinek magas színvonalú elsajátítását, mind a jelenleg tanulmányaikat végző egyetemi hallgatók, mind pedig a korábban diplomát szerzett gazdasági-, gépész-, közlekedési, vegyész- és agrármérnökök, energetikai mérnökök, továbbá motorhajtóanyagok kereskedelmével foglalkozó, valamint vám- és pénzügyőri teendőket ellátó szakemberek részére is.

(Hancsók Jenő – Eller Zoltán – Kriván Eszter: Korszerű motor- és sugárhajtómű üzemanyagok. I. Motorbenzinek, Veszprém: Pannon Egyetem Kiadó, 2014, 543 o. és Hancsók Jenő – Eller Zoltán – Sági Dániel: Korszerű motor- és sugárhajtómű üzemanyagok. II. Dízelgázolajok, Veszprém: Pannon Egyetem Kiadó, 2016, 654 o.)

Holló András

PhD, MBA, MOL Nyrt., DS Kutatás-Fejlesztés,
DS Innováció szenior szakértő

A FELSZÍNI VIZEK MINŐSÉGE. MODELLEZÉS ÉS SZABÁLYOZÁS

Sokan tudtuk a „vízszakmában”, hogy nagyon kéne egy jó, vízminőséggel foglalkozó szakkönyv, de nincs. Most megszületett egy kitűnő, igényes és vastag könyv.

Az első fejezet elejét, ami azt tárgyalja, mi is a víz, kötelező olvasmánnyá tenném minden tanult ember számára, már az óvodától. Ugyanide csatlakozik a fejezet másik lényege is, hogy az ember elszennyezte vizeinket, és a szennyezés most akár fokozódhat az éghajlatváltozás nyomán kialakuló egyre nagyobb csapadékok okozta szennyezőanyag-bemosódás miatt. A Duna vízgyűjtőről a Fekete-tengert elérő növényi tápanyag terheléseinek forrásait összesítő 1.11-es ábrából már kitűnik, hogy korunkban a diffúz, avagy nem pontszerű szennyezés a legnagyobb vízminőségi probléma, amit a szöveg is hangsúlyoz, és az 1.21. és 1.22. ábrák már a jövő megoldásai felé mutatnak.

A második fejezet eleje a globális mennyiségi vízhelyzetet ismerteti, vetíti ezt a földi népesség szaporodásának problémáira és az éghajlatváltozással, elsivatagosodással sújtott térségek számaival is érzékeltetve vázolja fel Földünk vízbajait. Ebből a hosszú gondolatsorból arra a tényre szeretném én is felhívni a figyelmet, hogy idézem: „...az ún. diffúz szennyezések váltak dominánsokká (például a Duna-vízgyűjtő esetében a tápanyag-emissziók 70–80%-a ilyen jellegű).” Sajnos Európában hajlamosak erről a meghatározó tényről megfeledkezni – még vízszakmai körökben is.

A vízminőséget tárgyaló fejezetből egy gondolatot emelhetek csak ki, ami megemlíti az USA-beli „zérus kibocsátásra” törekvést, ami egyúttal a diffúz szennyezések meghatározó fontosságának felismerésére is vezetett.

A *Vízminőség – Vízminősítés* című fejezetben fontos megállapítás, hogy „...a természetes víz minőségét a vízi ökoszisztémák állapotának jellemzésével kapcsolják össze”. Lényegi felismerés, amire kicsit későn érkezett meg a tudomány, vagy csak épp érkezőben van.

A vízminősítés módszereinek áttekintése során fontos a 4.5. ábra, ami a vizeink oxigénháztartását és így élővilágát legjobban meghatározó BOI-paraméter alakulásának fázisait mutatja be a szerzőtől megszokott precíz módon. Az egyik legfontosabb fejezet az Ökoszisztéma és anyagforgalom, biológiai jellemzők címet viseli. Fontos a 4.11. ábra, amely az ökológiai szempontból fontos paraméterek változását mutatja be. A vízminőség-változások történetéből a legérdekesebb az országba belépő és onnan kilépő vizek minőségének értékelése néhány fontos paraméter vonatkozásában (4.14. ábra).

A vízminőségi modellek következnek. Ez a főtárgy, mert mindmáig nincsenek rendszeresen karban tartott, hitelesített és igazolt, vízminőségi modellek vizeink-

re és vízgyűjtőinkre. Nem foglalkozhatunk a lehetséges modellek szépen bemutatott osztályozásával, sem magukkal a modellekkel. Annyit azonban érdemes megemlíteni, hogy roppant helyállóan először a tavaink és így a Balaton szempontjából oly fontos foszforforgalmi-modellek, majd a folyók oxigénháztartási alapmodellje, a Streeter & Phelps-féle „hagyományos oxigénvonal” kerül igényes taglalásra. Itt talán érdemes lett volna megemlíteni azt az UNESCO kiadványt (IHP CD-ROM series I, No.1), ami egy oktatóprogram, és ez a hagyományos modell is szerepel benne. A következő fejezetben a vízminőségi modellezés alap-egyenletei és néhány alkalmazás következnek. Az eutrofizálódás modellezési lehetőségek igen részletes, tudományos igényű taglalása következik ismét kitekin-téssel csodás tavunkra, a Balatonra.

A 7. fejezetnek már a címe (*Vízminőségi modellek és csapdák*) is meglepő lehet, pedig nem az, hiszen a világszinten is krónikus adathiány sokszor szinte kizárja bármely modell gyakorlati alkalmazhatóságát. A IV. rész címe *Tudomány, policy és alkalmazások. A Balaton vízminőségi stratégiai* fejezet már világossá teszi a címválasztást. Hosszan tárgyalja a Balatont érő foszforterhelést, ami meghatározó a tó eutrofizálódásában. Megdöbentő a 8.7-es ábra, amelyből kiderül, hogy régóta nem mériki rendszeresen a tavat érő foszforterhelést. A tó vízminőség-szabályozási döntés-előkészítés célú modellezésének leírása következik, amelynek közismert csúcspontja a Kis-Balaton létrehozásának várható, majd bekövetkezett hatásaira irányult. Talán az egész könyv legfontosabb fejezete a 10. fejezet, amely a Kis-Balaton foszfor-visszatartásával foglalkozik, hiszen jórészt ez határozza meg a Balaton vízminőségének mindenkori állapotát.

A 10.2. fejezet a tervek és üzemirányítás felülvizsgálata érdekében készített kétféle foszforforgalmi modellt mutat be. Mindkét modell egyszerűségében is lényegre tapintó, mely szerint az eutrofizálódási folyamatot egy belső foszfor-recirkuláció uralja, amelynek során az elhalt alगतөmegből ismét az algaszaporodást tápláló oldott foszfor lesz, amint ez már ismert volt egy 1979. évi MTA/IIASA Balaton-modellezési kiadvány kapcsán is. A következő fejezet a két modell sikeres kalibrációját és igazolását mutatja be a Kis-Balatonra. Ezt követően a foszforterhelési scenáriók előállításával foglalkozik a fejezet, majd ezek hatását vizsgálja a modellekkel, Monte Carlo-szimulációk alkalmazásával (10.9. ábra). Az optimista scenáriók a Langmuir-izotermával az 50%-os TP-visszatartást is valószínűsítették, ami így első hallásra kissé túl jónak tűnik. *A Hűsz évvel később, mit mutatnak az adatok* című fejezetben választ is kapunk a felmerült kérdéseinkre. Például a 10.10. ábrán jól látszik, mi történt: 30-35% körül látszik beállni a rendszer, ami a nemzetközi szakirodalom tükrében is jó eredménynek fogadható el. Nagyon fontos megállapítás, hogy: „Egyedül a visszatérő monitoring – mondjuk ötévenként – segítene megtanulni a rendszer viselkedését és a modell paramétereinek az illesztését.” A szerző azonban túl óvatosan fogalmaz, hiszen amíg nem mérünk eleget ahhoz, hogy modelljeink kalibrálva és verifikálva lehessenek, ad-

dig hiába dolgoztunk, mert modelljeink nem léteznek. *A jövőről – az üzemi irányítás felé* fejezet sokféle tanulságából én csak egyet emelhetek ki, hogy kívánatos a Hidvégi-tó vízszintjét úgy szabályozni, hogy abban minél többféle vízínövény jelenjen meg, hiszen mindig a tápanyag iránti versengés, a minél nagyobb növényi sokféleség kínálja a legjobb visszatartást, túl az üledék szorpciós kapacitásán.

A könyv végén egy fontos fejezet a Balaton csaknem bekövetkezett vízpótlásának esetét elemzi jelentős tanulságokkal.

A könyv utolsó mondata nagyon igaz és nagyon fontos, de a szerzőre jellemzően túlzottan elővigyázatos: „...a jövő gondjait egyszer már nem követnünk, hanem megelőznünk kellene. Azért is, hogy a Balaton ügye végre nyugvópontra kerüljön...”. Itt sajnos sokkal többről van szó, ugyanis tudomásom szerint ma nem mérik a legfontosabbat, az éghajlatváltozás miatt is egyre nagyobb és hevesebb esők okozta növényi tápanyag (és egyéb szennyezőanyag) terheléseket, és semmiféle modellezésről, azaz a beavatkozások megtervezéséről sem tudok. Ezen, a szerzőnek messzemenően igazat adva, sürgősen változtatni kellene.

(Somlyódy László: Felszíni vizek minősége. Modellezés és szabályozás. Budapest: Typotex Kiadó, 2018, 388 o.)

Jolánkai Géza

az MTA doktora, nyugalmazott egyetemi tanár,
DE, Műszaki Kar

Kitekintés

GIMES JÚLIA GONDOZÁSÁBAN

SÜTŐPOR-TERÁPIA

A sütőpor, kémiai nevén nátrium-bikarbonát (NaHCO_3), akár a gyulladáshoz autoimmun betegségek, például a reumatoid artritisz kezelésében is segíthet.

Az amerikai Augusta University kutatói a magas vérnyomás és a krónikus vesebetegség sajátosságait tanulmányozták állatmodelleken. A krónikus vesebetegség egyik következménye lehet, hogy a vér rendkívül szigorúan szabályozott pH-ja eltolódik a savas irányba, fokozva ezzel a szív- és érrendszeri betegségek, illetve a csontritkulás kialakulásának kockázatát.

Paul M. O'Connor és munkatársai kezdetben arra voltak kíváncsiak, vajon sütőporoldat rendszeres fogyasztásával csökkenthető-e a vér savassága. A válasz igen volt, ugyanakkor a kísérleti terápia a vesebetegség előrehaladását is lassította. Nos, a kutatók a továbbiakban erre a jelenségre kerestek magyarázatot.

Felfedezték, hogy a vesebeteg állatok lépében a gyulladáskeltő M1-makrofágok mennyisége csökken, míg a gyulladáscsökkentő M2-makrofágok száma nő. Ezt követően egészséges állatokon, majd egészséges egyetemi hallgatókon is megvizsgálták a nátrium-bikarbonát hatását, és mindkét teszt során a gyulladáshoz kapcsolódó folyamatok csökkenését találták.

A kutatók szerint ez a banális vegyület a lép ún. epitél sejteinek keresztül fejti ki hatását, ezek közvetítik az immunrendszer felé az üzenetet: nincs szükség az immunrendszer gyulladáshoz való reakcióira.

O'Connor szerint a „sütőpor-terápia” ígéretes lehet a gyulladáshoz autoimmun betegségekben is.

Ray, S. C. – Baban, B. – Tucker, M. A. et al.: Oral NaHCO_3 Activates a Splenic Anti-Inflammatory Pathway: Evidence That Cholinergic Signals Are Transmitted via Mesothelial Cells. *The Journal of Immunology*, 15 May 20018, 200, 10, 3568–3586.
DOI: 10.4049/jimmunol.1701605

DIAGNOSZTA BAKTÉRIUMOK

A neves MIT (Massachusetts Institute of Technology) kutatói olyan parányi megemészthető biológiai érzékelőt fejlesztettek ki, amely képes a gyomorvérzés detektálására. A szenzor lelke egy genetikailag módosított coli baktérium sejtjeiből áll, amelyek a hemoglobin hem csoportjával találkozáskor fényt bocsátanak ki. Ezt apró fototranzisztorok küldik tovább a mikroprocesszornak, amely vezeték nélküli jelet juttat egy adott számítógépre vagy mobiltelefonra. A kutatók az adatok elemzésére képes androidos alkalmazást is készítettek.

Az érzékelő egy kb. 3,8 cm hosszúságú henger, amelynek 17 mikrowatt teljesítményre van szüksége. Egy 2,7 voltos elemmel a rendszer kb. másfél hónapig folyamatosan működőképes.

A kutatók sertésekben sikeresen tesztelték a rendszert: mindig ki tudták mutatni a jelen lévő vért. Szerintük a szenzor alkalmas lesz egyszerű használatra is, és arra is, hogy néhány napig vagy hétig az emésztőrendszerben maradjon.

Mark Mimeo és munkatársai két másfajta szenzor kifejlesztésén is dolgoznak. Ezeket még nem tesztelték állatokon. Az egyik olyan baktériumot tartalmaz, amelynek fénykibocsátása egy gyulladáshoz vezető hatására indul meg, ami alkalmas lehet például a Crohn-betegség detektálására. A másik érzékelő olyan anyagot képes felismerni, amely bakteriális fertőzés jelenlétét valószínűsíti.

Szerintük a szenzorként működő coli baktérium megfelelő genetikai átalakításával nagyon sokféle anyag hatására tud a rendszer jelet produkálni. Már csak az a kérdés, hogy a gyomor-bélrendszer egyes betegségeinek diagnosztizálásához milyen anyag érzékelésére érdemes genetikai úton rávenni a colit.

Mimeo, M. – Nadeau, P. – Hayward, A. et al.: An Ingestible Bacterial-Electronic System to Monitor Gastrointestinal Health. *Science*, 25 May 2018. 360, 6391, 915–918. DOI: 10.1126/science.aas9315, <http://science.sciencemag.org/content/360/6391/915.full>

ODA FOLYIK, AHOL MÁR AMÚGY IS SOK VAN

A Föld édesvízkészletének területi eloszlása egyre egyenlőtlenebb, a száraz területek még szárazabbak, a nedvesek még nedvesebbek lesznek. Ez a fő megállapítása egy, a NASA irányításával végzett kutatásnak, melynek során műholdas megfigyelések adatait elemezték. A most publikált eredményekhez 2002 és 2016 között, tizennégy éven át gyűjtötték az adatokat.

A globális vízátcsoportosulás folyamatának több okát is azonosították. A legfontosabb tényezőnek az emberi vízkezelési gyakorlatot, valamint a természetes, illetve az emberi tevékenységekkel kapcsolatba hozható klímaváltozást találták.

A műholdas megfigyelések újdonságnak számítanak az édesvízkutatásban, a készletekkel kapcsolatos információk mostanáig zömükben földi megfigyelési pontokon végzett méréseken alapultak.

Rodell, M. – Famiglietti, J. – Wiesel, D.: Emerging Trends in Global Freshwater Availability. *Nature*, 17 May 2018. DOI: 10.1038/s41586-018-0123-1, <https://www.nature.com/articles/s41586-018-0123-1>

EURÓPA SZELESEBB LESZ

Közel háromszáz szélturbina tizenegy éven keresztül gyűjtött működési adatait egy klímamoddellel kombinálva brit kutatók kiszámolták, hogy egy kb. 1,5 °C-os felmelegedés esetén a szélerőenergia a jelenleginél sokkal fontosabb szerepet tölthet be.

A globális felmelegedés hatására Európában az Egyesült Királyság lesz a legszelesebb. A jelenleg kiépített kapacitásokkal a szél erőművekkel termelt elektromos energia kb. 10 százalékkal növekedhet. A kutatók szerint a Brit-szigeteken kívül a felmelegedés következtében Németország, Lengyelország és Litvánia is huzatosabb hely lesz.

Hosking, J. S. – MacLeod, D. – Phillips, T. et al.: Changes in European Wind Energy Generation Potential within a 1,5 °C Warmer World. *Environmental Research Letters*, 17 May 2018, 13, 054032

OK VAGY OKOZAT?

A kognitív képességek időskori hanyatlásának mértékében és ütemében jelentős eltérések tapasztalhatók. A különbségek egyebek között életmódbeli különbségekkel is kapcsolatba hozhatók. Összefüggést találtak már például az iskolázottsággal, a táplálkozással, az alkoholfogyasztással, a dohányzással vagy a testedzéssel is. Noha az időskori szexuális élet és közeli érzelmi viszony javítja a memóriatesztekben mért teljesítményt, a két évvel később megismételt vizsgálatok szerint a szexuális aktivitás a memória romlását nem lassítja.

A tesztek az Angliában évek óta zajló, idősek életkörülményeit, egészségi, fizikai és szellemi állapotát, anyagi helyzetét felmérő programok keretében történtek. A kutatás során felhasznált adatok az ötven év feletti angliai lakosság szempontjából reprezentatívak voltak. A program több mint tízezer résztvevővel kezdődött. A két évvel később megismételt felmérésen is részt vevő és a szexuális

életére vonatkozó kérdésekre is választ adó 2672 férfi és 3344 nő átlagéletkora hatvanhat év volt.

A memóriateszt egyszerű volt: tíz hétköznapi szót kellett hallás után megjegyezni, majd közvetlenül az elhangzás után, illetve némi szünetet és elterelő gyakorlatot követően visszamondani. Az értékelés is egyszerű volt. Minden találat 1 pontot ért, azaz összesen $2 \times 10 = 20$ pontot lehetett elérni. A szexuális kapcsolatokra és az aktivitásra vonatkozó kérdőívet az Egyesült Államokban és az Egyesült Királyságban korábban elvégzett hasonló témájú felmérésekkel összehangban állították össze.

Allen, M. S: Sexual Activity and Cognitive Decline in Older Adults. *Archives of Sexual Behavior*, 2018. First Online: 16 May 2018. DOI: 10.1007/s10508-018-1193-8, https://www.researchgate.net/publication/325187900_Sexual_Activity_and_Cognitive_Decline_in_Older_Adults