

VALÓS TÉRBEN – AZ ONLINE TÉRÉRT

Networkshop 31: országos konferencia

2022. április 20–22.
Debreceni Egyetem

Szerkesztette: Tick József, Kokas Károly, Holl András

HUNGARNET Egyesület
Budapest, 2022



A kötet megjelenését támogatta az
Energiaügyi Minisztérium

Szerkesztette: Tick József, Kokas Károly, Holl András

Tipográfia és tördelés: Vas Viktória

Workshop

2022. április 20–22. Debreceni Egyetem, konferencia előadásainak közleményei

ISBN 978-615-82243-0-7

DOI: [10.31915/NWS.2022](https://doi.org/10.31915/NWS.2022)

Kiadja a HUNGARNET Egyesület
az MTA Könyvtár és Információs Központ közreműködésével
Budapest
2022

Borítókép: [freepik.com](https://www.freepik.com)

TARTALOMJEGYZÉK

Előszó	5
Lencsés Ákos: A nyílt tudomány pénzügyi vonatkozásai	7
Farkas Katalin: Centenáriumi média-adattár és virtuális kiállítás létrehozásának tanulságai az SZTE Klebelsberg Könyvtárban	13
Bódog András: A nyílt archívumi információs rendszer (OAIS) szabványának honosítása.....	20
Perlaki Attila: Oktatást segítő gamifikációs alkalmazások, mint szakdolgozati témák	27
Csapó Noémi – Dani Erzsébet: APPropó fejlődés – A Bács-Kiskun Megyei Katona József Könyvtár mobilapplikációja.....	32
Simon András: Integrált könyvtári rendszerek tranzakciós rekordjainak vizsgálata, a könyvtári állomány digitalizálásának tervezésekor.....	41
Németh Márton: Az OSZK Webarchívum nemzetközi kapcsolatai.....	58
Antal Péter: A mesterséges intelligencia kihívásai a XXI. század társadalmára	70
Hajdu Csaba – Szilágyi Zoltán: Modern robotikai technológiai ismeretek oktatása „Teljes spektrumú” oktatási módszerrel	77
T. Nagy László – Boda István Károly – Tóth Erzsébet: E-tananyagfejlesztés virtuális 3D környezetben.....	84
Palencsárné Kasza Marianna: Digitális átállás – Minőség – lehetőségek az EQAVET terén.....	92
Nagy Gyula: Nemzetközi kitekintés a felsőoktatási könyvtárak világára: a EUGLOH könyvtári workshopja	99
Babocsay Gergely: Az európai természettudományi gyűjtemények digitális integrációja: határ a csillagos ég.....	108
Somorjai Noémi: Egyenlőtlenségek a tudományos kutatás területén. Az amatőr kutatók szerepe	114
Molnár Dániel – Dani Erzsébet: Robotok a könyvtárban: Hogyan válhat a robotika a könyvtári mindennapok részévé?	122
Horváthné Felföldi Helga: Digitalizáció a szakképzésben. A Szakmajegyzékben szereplő szakmák digitáliskompetencia jártassági szintjeinek felülvizsgálata	130
Kalcsó Gyula: Ne csak útra csomagoljunk! Miért fontos a csomagolás a digitális megőrzésben?	138
Karsa Zoltán István – Szeberényi Imre: A CIRCLE felhő elmúlt évtizede	146
Bobák Barbara – Kasza Péter: Az MI lehetőségei a kora újkori filológiában: Johannes Michael Brutus <i>Rerum Ungaricarum</i> libri kéziratának digitális kiadása (esettanulmány)	154
Egyed-Gergely Júlia – Vajda Róza, Gárdos Judit – Horváth Anna – Meiszterics Enikő – Micsik András – Martin Dániel – Marx Attila – Pataki Balázs – Siket Melinda: Szociológia, kutatási adatok, mesterséges intelligencia: lehetőségek és tapasztalatok	161
Szemes Botond – Bajzát Tímea – Fellegi Zsófia – Kundráth Péter – Horváth Péter – Indig Balázs – Dióssy Anna – Hegedüs Fanni – Pantyelejev Natali – Sziráki Sarolta – Vida Bence – Kalmár Balázs – Palkó Gábor: Az ELTE Drámakorpuszának létrehozása és lehetőségei.....	170



Sebestyén Ádám: Az ELTEdata szemantikus adatbázis legújabb fejlesztései.....	179
Szlamka Erzsébet: Új trendek a tanulási eredmények tanúsításában	185
Tóth Máté – Héjja Balázs: Webshop indítása közkönyvtári környezetben.....	192
Etlinger Mihály – Hernády Judit: A kiadás hagyatéka / a hagyatéka kiadása: A Régi Magyar Költők Tárának hálózati kiadásáról.....	199
Varga Emese – Makkai T. Csilla: „Ki a fenének kell collstok?” A digitális szöveg rejtett mértékegységei	204
Dobás Kata – Fazekas Júlia: ITIdata – Egy irodalmi adatbázis fejlesztése Wikibase alapon és ennek hasznosítása Kosztolányi Dezső forrásjegyzékénél	211
Sörény Edina: Kézai Simon Program – digitális családi fotóarchívum.....	219
Fülöp Tiffany – Molnár Tamás – Hoczopán Szabolcs: Open Monograph Press e-könyvplatform a Szegedi Tudományegyetemen	227
Palkó Gábor: Mesterséges intelligencia, digitális bölcsészet, kulturális örökség: trendek és eredmények.....	235
Pergéné Szabó Enikő – Bátfai Mária Erika: A tudományos publikálás támogatása a Debreceni Egyetemi és Nemzeti Könyvtárban	241
Csirmazné Rezi Éva: Nemzetközi kiadványazonosítók és kötelempéldányok kezelése az OSZK OKP (Országos Könyvtári Platform) rendszerében	250
Alföldi István – Dióssy Anna Laura: Digitálisan született kutatási anyagok megőrzése: a relációs adatbázis mint born-digital objektum	262
Fekete Norbert: HTR-modellépítés és kézírásfelismerés nagyméretű, többszerzős szövegtörzsen. A Transkribus alkalmazása az Arany János hivatali iratokon.....	271
Horváth Péter – Kundráth Péter – Palkó Gábor: ELTE Népdalkorpusz – magyar népdalok gépileg annotált adatbázisa	276
Nagy György: IKT eszközök alkalmazása az alsó tagozatos környezetismeret órákon.....	284
Köpösdí Zsuzsa – Molnár Tamás: Multimédiás, interaktív és adaptív tananyagok létrehozásának lehetőségei H5P keretrendszerrel	289
Jankó Tamás: Munka 4.0 – Ipar 4.0 – Szakképzés 4.0 – : A digitális kompetencia jövőbeni fejlesztési útjai	296
Békésiné Bognár Noémi Erika – Nagy Andor: Megújuló könyvtári statisztika: az egységes adatstruktúra és a korszerű megjelenítés kialakításának útján	304
Bolya Máttyás: Kézírtos dallamlejegyzések feldolgozása MI-vel támogatott digitális környezetben	310
Maróthy Szilvia – Seláf Levente – Vigyikán Villó: Régi magyar verskorpusz összeállítása stilometriai és számítógépes metrikai kutatásokhoz	324
Szűcs Kata Ágnes: Kézírtos források transzformációinak lehetőségei a közgyűjteményekben.....	330
Fellegi Zsófia: A digitális filológia infrastruktúrái. A DigiPhil megújulásáról.	338
Mihály Eszter: Mi az a dHUpla? A Digitális Bölcsészeti Platform bemutatása.....	345
Nemeskey Dávid Márk – Palkó Gábor: Szemantikus névelim-azonosítás magyar nyelvű szövegeken (a HuWikifier bemutatása)	359

„Ki a fenének kell collstok?”¹
A digitális szöveg rejtett mértékegységei

„Who the hell needs measuring stick?”
The hidden units of digital text

Varga Emese
Országos Széchényi Könyvtár, Digitális Bölcsészeti Központ
varga.emese@oszk.hu

Makkai T. Csilla
Országos Széchényi Könyvtár, Digitális Bölcsészeti Központ
makkai.csilla@oszk.hu

Abstract: How to reveal the results of the philological research behind a digital edition? These data are often hidden in the text of letters and critical notes, but when grouped and quantified they provide significant information.

In our presentation, we will present the results of machine-assisted research based on the digital critical edition of Móricz Zsigmond's correspondence and novel corpus, outlining how new contexts in the correspondence and novel corpus become visible with machine assistance. It allows us to look at our corpus and the author's oeuvre from a different perspective.

Keywords: data visualization, digital edition, graph visualization, stylometry, hungarian literature, 20th century literature, corpus analysis

Absztrakt: Hogyan lehet megmutatni egy digitális kiadás mögött álló filológiai kutatás eredményeit? Ezek olyan adatok, amelyek a levélszövegben, kritikai jegyzetekben, sokszor rejtve vannak jelen, csoportosítva és számszerűsítve viszont szignifikáns információval bírnak. Előadásunkban a Móricz Zsigmond-levelezés digitális kritikai kiadása, illetve a regénykorpusz digitális kiadása alapján gépi eszközökkel készült, további kutatásokra támaszkodó eredményeit prezentáljuk, felvázolva, hogy gépi támogatással milyen új összefüggések válnak láthatóvá a levelezésben. Ezek segítségével más perspektívából tudunk rátekinteni a korpuszunkra, innen tovább lépve pedig az írói életműre.

Kulcsszavak: adatvizualizáció, digitális kiadás, gráf alapú vizualizáció, stilometria, magyar irodalom, 20. századi irodalom, korpuszelemzés.

Egy digitális szövegkiadás készítése során szükségszerűen keletkeznek számszerűsíthető adatok, legyenek azok tulajdonnevek, évszámok, kritikai jegyzetek, műfaji kategóriák. Viszont az adatok közötti összefüggések felvázolása is hozzátartozik a kutatómunkához annak ellenére, hogy azok sokszor hosszú, nehezen átlátható Excel táblákba tömörülnek.

Tanulmányunk legfőbb kérdése, hogy milyen adatvizualizációk segítségével lehet megmutatni egy digitális szövegkiadás mögött álló filológiai kutatás eredményeit, hogyan lehet

1 Móricz Zsigmond, *Tükör IV.*, lásd: <https://dhupla.hu/text/o:mzs-tukor.tei.4>.

átláthatóbbá tenni egy strukturált adathalmazt, milyen eszközök léteznek az összefüggések megmutatására.

A továbbiakban három különböző adatvizualizációt szeretnénk prezentálni: két gráf alapút és egy stilometriai elemzés eredményeit. Ez esetben nem volt szükség *collstokra*, gépi támogatású méréseket végeztünk nagyméretű, digitális szövegtörzseken.

A Móricz-műhely keretein belül elkezdett közös munka fókuszában állt a PIM (Petőfi Irodalmi Múzeum) Kézirattárában található Móricz-különgyűjtemény kutathatóvá tétele. Ebből a sokéves munkából született meg a [Tükör](#) című világháborús jegyzetkötetek digitális kiadása, illetve a [Móricz-levelezés](#) digitális kritikai kiadása, melynek 2.0-s verziója olvasható a 2021-ben létrehozott [dHUpIa](#) nevű digitális bölcsészeti platformon. Többfajta technika többféle korpuszt igényel: a gráf alapú adatvizualizációk kiindulópontjaként a Móricz-levelezés digitális kritikai kiadása szolgál, a stilometriai elemzést pedig a Móricz-regénykorpuszon végeztük, melyet az Országos Széchényi Könyvtár [Magyar Elektronikus Könyvtárából](#) értünk el.

Munkánkat tehát a megszámlálható adatok elemzése jellemzi. A nyelv statisztikai alapú vizsgálatának különböző módzatai már a korai irodalomtudományi elemzésekben, például szógyakoriság mérésekben is felfedezhetőek. Hasonlóképpen a stilometriai mérések első kísérletei közé tartoznak a szerzőségi kérdések statisztikai alapú vizsgálatai, mellyel már 1851-ben foglalkoztak. Augustus de Morgan angol tudós és Thomas Mendenhall amerikai fizikus kutatásaikban a szavak hosszúságát és gyakoriságukat vizsgálták.² Ebből is látszik, hogy a számolás és az olvasás műveletei nem függetlenek egymástól, mivel a nyelv természetének statisztikai alapú rendeződése is szignifikáns információkkal bír.³

A digitális irodalomtudomány pedig erre a jellegre alapozva, a szöveg részeinek számszerűsíthetőségéből indul ki. A gépi támogatással végzett elemzések nemcsak felgyorsítják ezeket a statisztikai alapú vizsgálatokat, hanem a korábbiakhoz képest több és más jellegű adatra építve adnak lehetőséget irodalmi jelenségek értelmezésére. Természetesen a gépi, statisztikai alapú elemzések nem függetlenek a hagyományos filológiai elemzésektől, hanem új perspektívaként kiegészítik azokat.

A digitális kiadásoknak fontos szerepük van az adatok hatékony számszerűsítésében, hiszen azok önmagukban is adatbázisok – a metaadatolt, annotált szövegkiadások kereshető, lekérdezhető, számszerűsíthető adatokat tartalmaznak. A Móricz levelezés kiadás szűrési funkciója (ún. facettált kereső) listázott, számszerűsített adatokat jelenít meg, melyek segítségével az olvasó hamarabb eljut az információhoz, amit keres.

A tudományos kutatások egyre inkább adatközpontúvá válnak, a webes kultúra pedig egyre vizuálisabbá.⁴ A nagyméretű adathalmazok értelmezhetősége, befogadhatóvá tétele önmagában még nem valósul meg, ehhez különböző vizualizációs eszközök használata szükséges, melyek szemléletessége folytán a kutatási eredmények jobban megmutathatóak.

Gráf alapú adatvizualizációk

A genealógiai kutatás elkerülhetetlen egy teljességre törekvő levelezés kiadás készítésekor, ugyanígy az író kapcsolatrendszerének átláthatóbbá tétele is a kutatás lényeges részét képezi. Ennek érdekében indítottuk el Móricz Zsigmond leveleinek publikálása mellett

2 David HOLMES – Judit KARDOS, *Who Was the Author? An Introduction to Stylometry*, Chance 16. (2003/2.), 6. lásd: <https://id.art1lib.org/ireader/42465305>.

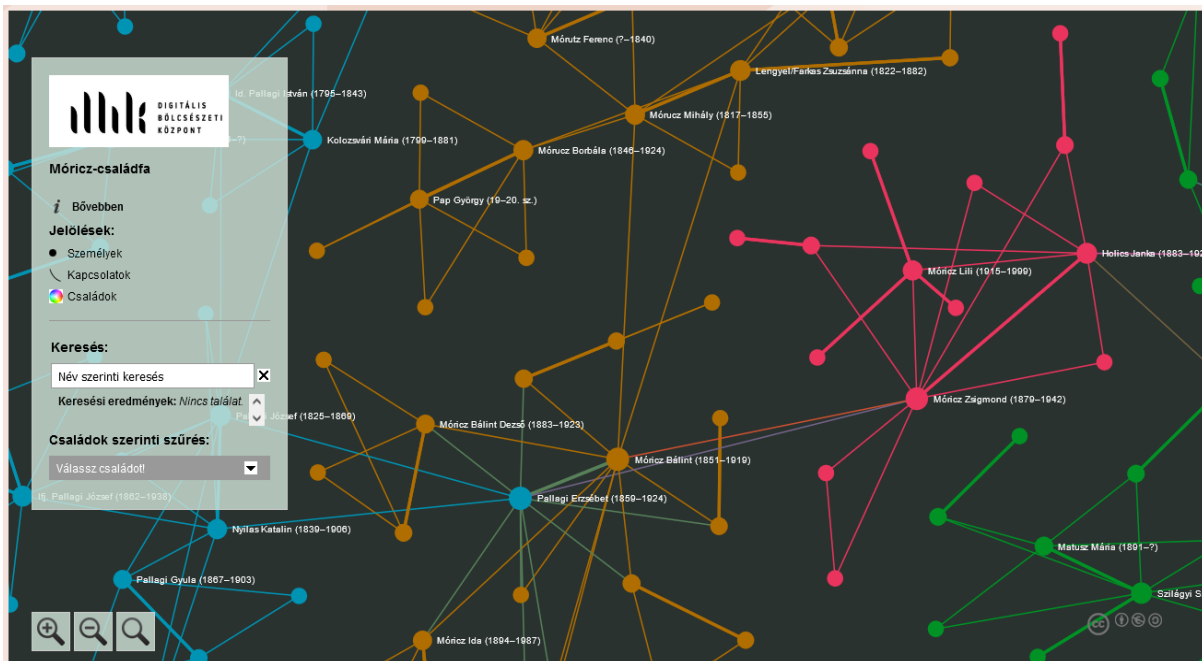
3 SZEMES Botond, *A digitális bölcsészet mint kódfejtés. A nyelv statisztikai vizsgálatának hagyományáról*, Műút 2021/079. Lásd: <https://www.muut.hu/archivum/37522>.

4 MARÓTHY Szilvia. 2020. *Digitális bölcsészet. A szövegtől az adatig*, Akadémiai, Budapest, 2020. Lásd: <https://doi.org/10.1556/9789634545781>.

Móricz családfájának és szakmai kapcsolati hálójának kutatását és interaktív vizualizációs megjelenítését.

Hogy minél pontosabb legyen a családfához készülő adatbázisunk, továbbléptünk a levelezésen és meglévő adatainkat összevetettük a PIM Kézirattárban lévő Móricz-hagyatékkal és a Móricz szakirodalom adataival, valamint különböző internetes, digitalizált adatbázisokban kutattunk. (Huntéka, [PIM Névtér](#), [Arcanum](#), [Hungaricana](#), [FamilySearch](#), [OSZK-Gyászjelentések](#))

A kutatómunka eredményeként a családtagok száma a Móricz ágon 35-ről 62 főre nőtt, az anyai, Pallagi-ágon 25-ről 90 főre gyarapodott. Holics Janka oldalán a Holics ág 45-ről 142-re, a Szklenár család pedig 55 főről 101-re növekedett. Így a családfa összesen 409 nevet tartalmaz. Ennek megjelenítésére a [Gephi](#) nyílt forráskódú, hálózatelemző és -ábrázoló szoftvert választottuk, amely gráf alapú adatvizualizációk készítésére alkalmas. Ez az adatvizualizációs eszköz egy kapcsolatrendszer olyan egyszerűsített ábrázolását teszi lehetővé, mely az aktuális szempontoknak megfelelően kiemeli az összefüggéseket, és ez által az eredmények könnyebben értelmezhetővé válnak⁵. Előnye, hogy interaktívá, dinamikussá tehető, kutatható, nagy adatmennyiség esetén is jól átlátható (lásd: 1. ábra), a kapcsolódási pontok mentén többféle irányból elindulva lehet végigjárni a hálózatot és összerakni a puzzle darabkait.

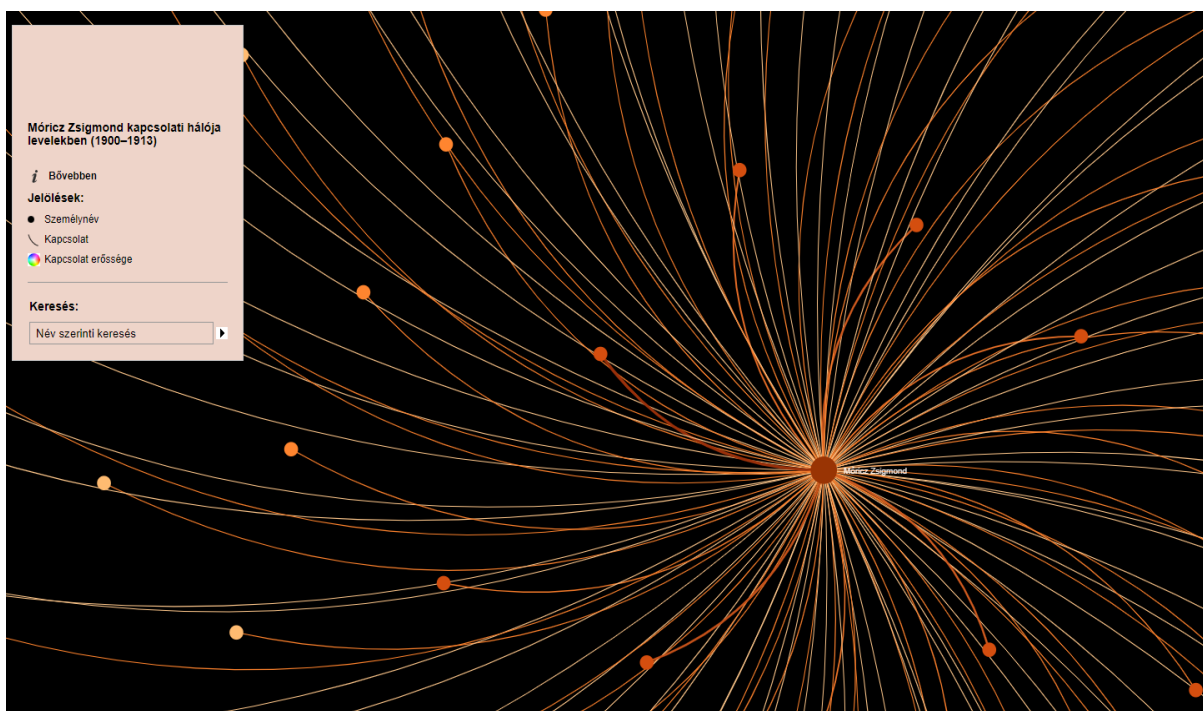


1. ábra: Móricz családfa vizualizáció
<https://dhupla.hu/s/dataviz/moricz-csalad/>

A családtagok között kétfajta kapcsolatot jelenítettünk meg: a gyerek-szülő és házastársi viszonyt. Számokkal súlyoztuk a kapcsolat erősségét, melyet a vizualizáción az élek vastagsága mutat - a gyerek-szülő kapcsolatot a vékonyabb, a házastársi kapcsolatot a vastagabb él jelöli. A bal oldali panelben külön rászűrhetünk a családi ágakra - a Móricz, a Pallagi, a Szklenár és a Holics család mellett Móricz szűk családi ágát is külön klaszterbe rendeztük. Emellett konkrét

5 KRÉSZ Miklós – TÓTH Attila, *Gráf alapú adatbányászat és vizualizáció: egy esettanulmány = Tudományos és művészeti műhelymunkák: új utakon a Pedagógusképző Kar című konferencia előadásai*, szerk. Tóth Szergej, Szegedi Tudományegyetem, Szeged, 2009, 358. Lásd: <http://acta.bibl.u-szeged.hu/id/eprint/64233>.

személynevekre is rákereshetünk. Minden családtag rendelkezik egy személyi adatlappal, ami a jobb oldali panelben ugrik elő, ahol az illető személyi adatai, a születési hely- és ideje, a halálozási hely- és ideje, valamint a közvetlen hozzátartozóinak a nevei láthatóak.



2. ábra: Mórícz kapcsolati háló adatvizualizáció
<https://dhupla.hu/page/kreativ/moricz-kapcsolatihalo>

Hasonló séma alapján, ugyancsak a Gephi szoftverben készült a kapcsolati háló is. (Lásd 2. ábra) 520 kézirat alapján 136 levelezőpartner állítja fókuszba az író, amely a pályakezdés tizenhárom évét sűríti egybe. A középpontban Mórícz Zsigmond látható, körülötte a levelező pályatársak állnak.

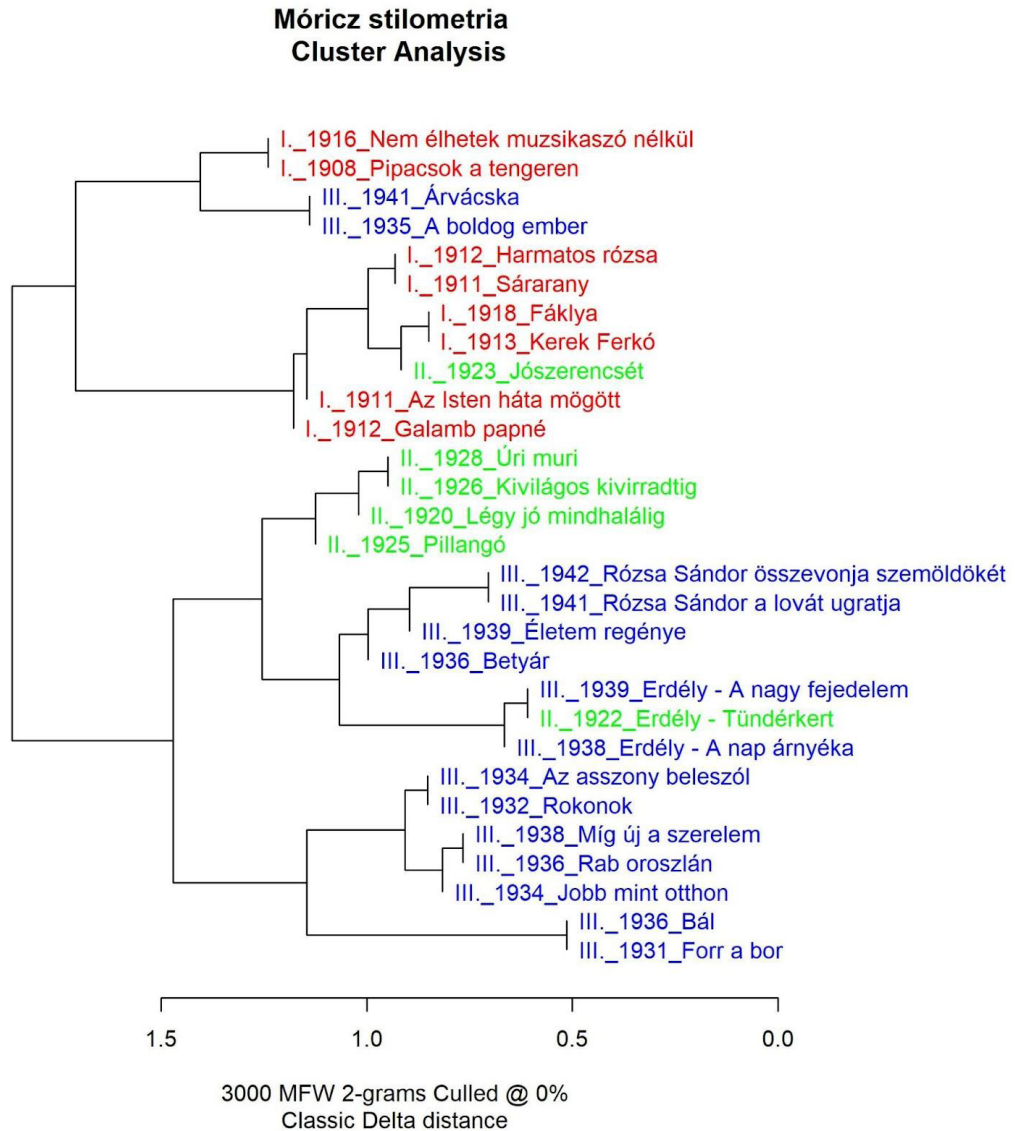
Stilometriai elemzés

A stilometria nagyméretű szövegtörzsek stílusának statisztikai alapú elemzésére alkalmas, melyben a szövegek összehasonlítása számszerűsített nyelvi jellemzők és különféle távolságmértékek alapján történik. Ez által tetten érhető az ún. „szerzői ujjlenyomat”, mely a szövegek tematikus szintje alatt létezik.⁶

Az elemzés szorosan kapcsolódik a Digitális Bölcsészeti Központban korábban készült, Mórícz-levelezésen alapuló [stilometriai vizsgálat](#)hoz, melyet továbbgondolva készítettük el a regénykorpuszon alapuló elemzést. Magyarországon találtunk példát korpuszok gépi elemzésére, stilometriai vizsgálatokra is, viszont olyan vizsgálat, amely egyetlen szerző életművének feldolgozásával foglalkozik, csak elvétve olvasható. Ebből kiindulva, arra tettünk kísérletet, hogy Mórícz Zsigmond regényeit statisztikai alapon elemezzük, és megnézzük, hogy felállítható-e egy hozzávetőleges időrend, láthatóvá válnak-e eltérések az életműben, megcélozható-e valamiféle korszakolás? Emellett, méréseink eredményét összevetettük a Mórícz-szakirodalom vonatkozó részeivel, abból a tapasztalatból kiindulva, hogy a legtöbb stilometriai elemzés megáll a vizualizációnál és az irodalomtörténeti következtetések csak ritkán kerülnek előtérbe.

6 SZEMES Botond, *l. m.*

Az R programozási nyelv erre alkalmas *stylo* csomagjával végeztük az elemzést, mely távolságmétrikák alapján állít fel szógyakorisági listákat, majd ezeket egymáshoz viszonyítva hoz létre egy dendrogramot. Ahhoz, hogy a vizsgálatunk minél megbízhatóbb eredményeket adjon, többféle beállítási paraméterrel kísérleteztünk. A legjobb eredményeket végül a klasszikus delta, az Eder delta és a Canberra távolságmértékek adták.

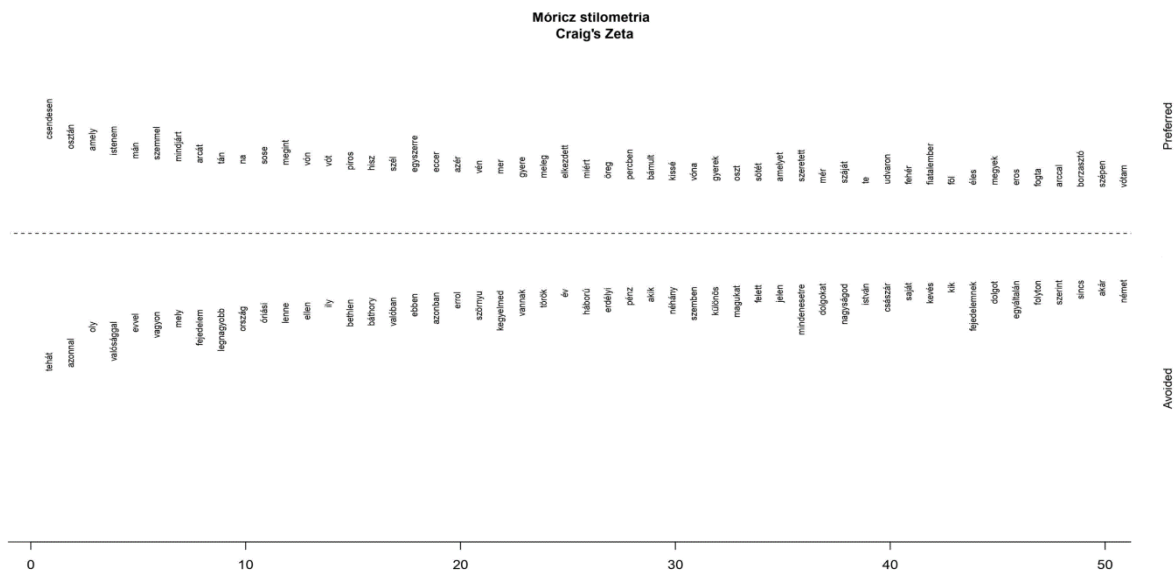


3. ábra: Klaszteranalízis klasszikus delta távolsággal

Az itt látható, 3-as ábrán a szerzőazonosítási célra létrehozott klasszikus delta távolsággal mértünk. A beállított paraméterek szerint a 3000 leggyakoribb szó-bigramot (tehát egymást követő lexémákat) vette figyelembe a szoftver, illetve nem használtuk a cullingot, ami azt jelenti, hogy nem szűrtünk ki egyetlen szó-bigramot sem a korpuszból.

Előzetesen a korpuszt három alkorpuszra bontottuk, ez tükröződik a színek szerinti besorolásban. Az 1908 és 1918 között írt szövegek pirossal, az 1920-as években írt szövegek zölddel, az 1930 után írt regények késsel jelöltek. A mérés után nagyjából kialakul egy időrend, a 10-es, 20-as, 30-as évek szerint. Az 1908 és 1923 között kiadott regények szorosan egymás mellett helyezkednek el, ugyanígy az 1930-as évek szövegei is egy nagyobb tömbbe rendeződnek. Érdekes viszont, hogy nem mindegyik szöveg került közel a "kortársához", illetve egészen eltérő tematikájú szövegek is egymás mellé rendeződtek.

A 3-as számú ábra alapján az látszik, hogy *A boldog ember* (1935) és az *Árvácska* (1941) című szövegek több rokonságot mutatnak a korai művekkel, mint a későbbiekkel. A móríci írásmódra mindvégig jellemző az elhangzott és lejegyzett dialógusokból való építkezés, mely módszer végletekig vitt megtestesítője lett végül *A boldog ember* és az *Árvácska*.⁷ Az a tény, hogy a tárgyalt szövegek közelebb kerültek az életmű első szakaszában írt regényekhez, arra enged rákérdezni, hogy a korai művek erőteljesebben közvetítik-e a hangzó nyelvet, az élőbeszédet. Emellett megfigyelhető még, hogy a történelmi regények (Erdély-trilógia és Rózsa Sándor regények), szorosan illeszkedtek egymáshoz, melynek oka a létrehozott archaizáló nyelvben keresendő.



4. ábra: Stilometriai elemzés Oppose függvényvel

Ehhez a kísérlethez a korpuszunkat az eddigi eredményeink alapján osztottuk két részre, vagyis a korai és a kései művekre, ahol a korszakhatár az 1920-as évek közepén húzódik meg. Az R-ben az *oppose* függvényt használtuk, amely két szövegtörzset összehasonlítására alkalmas. A függvény két szólistát állít elő egy diagramon, az egyik az első korpusz kifejezetten preferált szavait tartalmazza a másik korpuszhoz képest. A másik szólista pedig az első korpusz kifejezetten került szavait tartalmazza a második szövegtörzshöz képest.

A mérésből láthatóvá válik, hogy az első csoportba tartozó művek preferált lexémái között gyakoriak a tájnyelvi elemek: 'mán', 'vót', 'eccer', 'azér', 'oszt', 'má' stb. Ezek nem kizárólag tájnyelvi jellegük miatt érdekesek, hanem a hangzó nyelv móríci alkalmazása szempontjából is, mely szövegépítésének egyik jellemzője.

A második csoportban túlsúlyba kerülnek az anyagi világhoz, történelemhez, társadalomhoz kapcsolható kifejezések. Előfordulnak a 'pénz', 'fejedelem', 'ország', 'kegyelmed', 'háború', 'erdélyi', 'nagyságod', 'császár', 'dolog' kifejezések, melyek különböző szemantikai mezőkön helyezhetők el, mégis összekapcsolhatóak. Mintha az írói tekintet a premier plámban készült portré műfajától eltávolodva, átkapcsolt volna a nagytörténelmi történő ábrázolásmódra.

⁷ Cséve Anna, *A tükör mint hallucinációs fal (Móríci Zsigmond kéziratos könyveiről)*, *Alföld* 2005/9., 74–82., 79.; Szilágyi Zsófia, *Móríci Zsigmond*, Kalligram, Pozsony, 2013, 221.



Kitekintés

A digitális szövegkiadások metodikája és a szövegelemző szoftverek párhuzamosan fejlődnek, kölcsönhatásban állnak egymással, éppen ezért a kutatás során felmerült kérdésekhez újabb megoldási lehetőségek is társultak. Tervezzük a munka folytatását, pontosabban azt, hogy a digitális kiadások készítésével párhuzamosan a stilometriai vizsgálatokat is folytassuk a Móricz szövegeken. Célunk egy konszenzusháló létrehozása a Móricz-korpusz elemeiből, melyhez új eszközöket, illetve a *stylo* újabb funkcióit fogjuk tesztelni. Ugyanígy a levelezéskiadás alapján készült kapcsolati háló bővítése és az entitástár gazdagítása is a további munkafolyamatok részét képezi.

Bibliográfia

- CSÉVE Anna, *A tükör mint hallucinációs fal (Móricz Zsigmond kéziratok könyveiről)*, *Alföld* 2005/9., 74–82., 79.
- CSÉVE Anna, MAKKAI T. Csilla, TÖRÖK Sándor Mátyás, VARGA Emese, VÉTEK Bence: *Rokonok a gráfon*. <https://pim.hu/hu/digitalis-bolcseszeti-kozpont/moricz/rokonok-a-grafon>
- David HOLMES – Judit KARDOS, *Who Was the Author? An Introduction to Stylometry*, *Chance* 16. (2003/2.), 6. lásd: <https://id.art1lib.org/ireader/42465305>.
- KRÉSZ Miklós – TÓTH Attila, *Gráf alapú adatbányászat és vizualizáció: egy esettanulmány = Tudományos és művészeti műhelymunkák: új utakon a Pedagógusképző Kar című konferencia előadásai*, szerk. Tóth Szergej, Szegedi Tudományegyetem, Szeged, 2009, 358. Lásd: <http://acta.bibl.u-szeged.hu/id/eprint/64233>.
- MAKKAI T. Csilla, VARGA Emese, Vétek Bence: *Egy életmű metszete. Stilometriai vizsgálat Móricz Zsigmond regényein* Szépirodalmi Figyelő, 2022/1.
- MARÓTHY Szilvia. 2020. *Digitális bölcsészet. A szövegtől az adatig*, Akadémiai, Budapest, 2020. Lásd: <https://doi.org/10.1556/9789634545781>.
- SZEMES Botond, *A digitális bölcsészet mint kódfejtés. A nyelv statisztikai vizsgálatának hagyományáról*, *Műút* 2021/079. Lásd: <https://www.muut.hu/archivum/37522>.
- SZILÁGYI Zsófia, *Móricz Zsigmond*, Kalligram, Pozsony, 2013, 221.