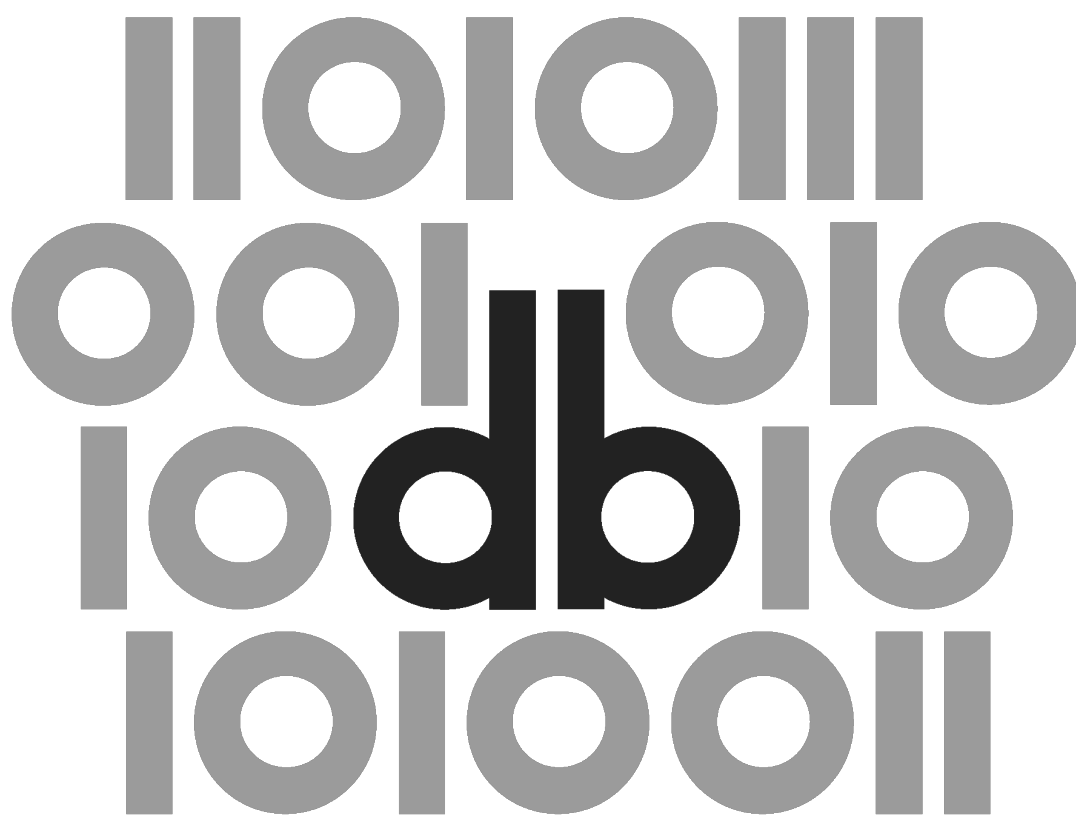
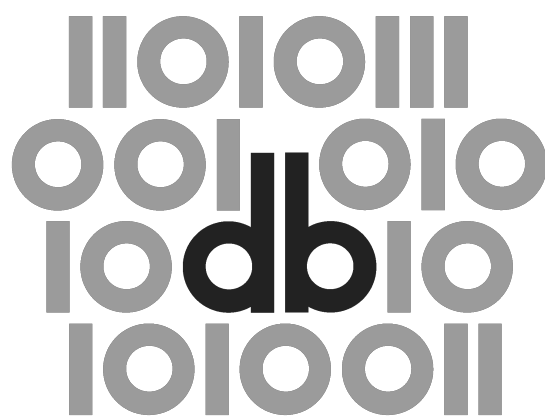




Digitális Bölcsészet
2021., negyedik szám

<DIGITÁLIS BÖLCSÉSZET>



4 (2021)

Felelős szerkesztő:

Maróthy Szilvia

Szerkesztőség:

Kokas Károly, Parádi Andrea

Rovatvezetők:

Tanulmányok: Kiss Margit

Műhely: Péter Róbert

Kritika: Almási Zsolt

Labor: Mártonfi Attila

Tanácsadó testület:

Bartók István, Fazekas István, Golden Dániel, Horváth Iván, Palkó Gábor, Pap Balázs,
Sass Bálint, Seláf Levente

Korábbi munkatársaink:

Bartók Zsófia Ágnes (szerkesztő, rovatvezető), Fodor János (szerkesztő),

†Labádi Gergely (szerkesztő, rovatvezető), †Orlovsky Géza (tanácsadó testület)

ISSN 2630-9696

DOI: 10.31400/dh-hun.2021.4

Kiadja a Bakonyi Géza Alapítvány és az ELTE BTK Régi Magyar
Irodalom Tanszéke (1088 Budapest, Múzeum krt. 4/A).

Felelős kiadó az ELTE BTK Régi Magyar Irodalom Tanszék vezetője.

Megjelenik az Open Journal Systems (OJS) v. 3. platformon, melynek
működtetését az ELTE Egyetemi Könyvtár- és Levéltár biztosítja.



Ez a mű a Creative Commons *Nevezd meg! – Ne add el! – Így add tovább! 2.5 Magyarországi Licenc* (<http://creativecommons.org/licenses/by-nc-sa/2.5/hu/>) feltételeinek megfelelően felhasználható.

Honlap: <http://ojs.elte.hu/digitalisbolcseszett>

Email cím: dbfolyoirat@gmail.com

Olvasószerkesztő: Bucsics Katalin

Tördelés: Hegedüs Béla

Grafika: Hegyi Gábor

<KRITIKA>

Seláf Levente  0000-0002-6052-9841

ELTE BTK Régi Magyar Irodalom Tanszék

selaf.levente@btk.elte.hu

Petr Plecháč. *Versification and Authorship Attribution*. Prague: Institute of Czech Literature–Karolinum Press, 2021. ISBN 9788024648712. 98 oldal. <https://doi.org/10.14712/9788024648903>*

E rövidke, vékony, kevesebb mint 100 oldalas kötet témája egy új kutatási módszer, a verselési jellemzők felhasználása versben írt szövegek szerzői attribúciós problémáinak megoldásához.

Az elmúlt bő száz évben a szerzőség megállapításának, ellenőrzésének módszere nagyon sokat változott. Alapvetően stilometriai, a szerző szókincsét, nyelvhasználatát vizsgáló eljárásokat használnak a kétes hitelű, kérdéses szerzőségű, esetleg hamisított szövegek kiszűrésére és pontos szerzőjük azonosítására. Plecháč könyve azt mutatja be, hogy a költői szövegek ritmikai és verselési jellemzői ugyanolyan jól használhatók az ilyen jellegű feladatoknál, mint a szókincs elemei.

A négy fejezet közül az első a kvantitatív alapú szerzőségi vizsgálatok történetét vázolja röviden. Általában a stilometriai elemzések módszertanának változásait tekinti át, a matematikai és statisztikai alapú elemzések első alkalmazásaitól egészen a 21. századi módszerekig: bemutatja, hogyan finomodott a vizsgálat a szókincs általános elemzésétől a funkciószavak variációinak összevetéséig, majd azt, hogy a különféle statisztikai módszerek hogyan tudták egyre pontosabban meghatározni egy mű szerzőségét. Egészen a legfrissebb, jelenkori módszerekig jut el, Burrows deltáján (2002, 2003) és annak módosított változatain és a különféle vektorgeometriai technikákon keresztül egészen a szupport vektor gépekig (SVM), melyek két szöveg hasonlóságát (vagy több szöveg hasonlóságán keresztül egy adott szerző stílusát) ábrázolni tudják. A fejezet utolsó része a verselési tulajdonságokon alapuló attribúciós kísérletek történetét vázolja röviden, (melyek egyébként, a Shakespeare-kutatást leszámítva, egészen későn, csak az elmúlt évtizedekben születtek meg) bemutatva, hogy míg a szógyakoriság statisztikai vizsgálata rengeteget fejlődött, finomodott az elmúlt évtizedekben, ez a módszertani áttörés a verselési tulajdonságok vizsgálatában nem következett be.

A második, rövid fejezet egy verstani áttekintés, amely rögzíti azokat az alapfogalmakat, paramétereket, melyeknek segítségével a verselés kvantitatív jellemzői vizsgálhatóvá válnak. Három kategóriát vesz fel, ezek a ritmus, a rím és az eufónia. A ritmus vizsgálatának egyik módja a ritmikai profil létrehozása: ez a hangsúlyos verselésű szövegekben az egyes metrikai pozíciókra eső hangsúlyok gyakoriságát vizsgálja. Hátránya, hogy nem tud mit kezdeni azzal, ha egy pozícióba két szótag is jut, ami nem ritka például az angolban, vagy a ritmusképleten felüli, esetleg a hiányzó szótagokkal.

* A publikáció az Innovációs és Technológiai Minisztérium Nemzeti Kutatási Fejlesztési és Innovációs Alapból nyújtott támogatásával, a K 135631 számú, *A régi magyar költészet számítógépes metrikai és stilometriai vizsgálata* című pályázati program keretében készült.

A ritmustípus (*rhythmic type*) ezzel ellentétben lekottázza az adott sor ritmusát, nem az absztrakt képlettel veti össze, ennek pedig az a hátránya, hogy az adatok nehezen lesznek mintába rendezhetők. Ezekhez képest Plecháč a ritmikus n-gramok vizsgálatát javasolja: ez a sort kisebb egységekre osztja, és ezek gyakoriságát vizsgálja, a két és három hangsúlytalan szótagból álló elemtől a csak egy hangsúlyosat tartalmazóig. A cseh és a spanyol költészetben ez tűnik a legjobbnak, míg a német, hangsúlyos verselésben a második módszert tartja célravezetőnek. A rímek vizsgálatában hét különböző jegyet vesz figyelembe két egymásra rímelő szó viszonyának meghatározásához: a morfológiai jegyeket, a rím��avak szótagszámát, az utolsó hangsúlyos szótag utáni szótagok számát, az utolsó szótag mássalhangzózárlatát, az utolsó magánhangzót, és gyenge rímeknél az utolsó magánhangzó előtti mássalhangzót és az utolsó előtti (hangsúlyos) magánhangzót. Az eufónia pedig bizonyos hangok gyakoriságát jelenti – ez gyakorlatilag az alliteráció vizsgálatának kiterjesztése teljes művekre.

A következő fejezet a sztichometriai módszerek kísérleti vizsgálatát tartalmazza három nagy korpuszon, a cseh, német és spanyol költészetből, ezek a következők: *The Corpus of Czech Verse, Metricalizer – the Corpus of German Verse, Corpus de sonetos del Siglo de Oro*. A kísérlet célja annak megállapítása, hogy e verseknél a gépi tanulással mennyire felismertethetők egyes szerzők szövegei stilometriai, illetve sztichometriai adatok alapján. Ez a vizsgálat sok érdekes, bár megjósolható részeredménnyel is járt, így például a nagyon termékeny szerzők esetében, akik változatos stílusban írtak, a szövegeik felismerése kevésbé hatékonyan sikerült az algoritmusnak. Izgalmas, hogy melyik tulajdonságok voltak a legjellemzőbbek egy-egy adott alkorpuszban. A cseh verseknél például a rímek fonetikai összetétele, a német versekben a morfológiai jegyek, míg a spanyol szövegek valahol a kettő között foglaltak helyet. A ritmikus n-gramok közül a négyeleműek jellemezték a legjobban a versek szerzőit az ebből a szempontból releváns cseh és spanyol anyagban. A fejezet konklúziója megállapítja, hogy a verselési jegyek alapján nagyon jó arányban felismerhetők a szerzők, sőt, néha ezek felülmúlják a lexikai elemekből kinyerhető jellemzők pontosságát. A két szempont kombinációja pedig nagyon biztos eredményt ad az szerzőiség szempontjából.

A kötet utolsó része két esettanulmányt tartalmaz. Az egyik egy klasszikus attribúciós probléma tisztázása, William Shakespeare és John Fletcher *Két nemes rokon* című drámájában állapítja meg a szerzőpáros tagjainak hozzájárulását a műhöz. Már 1812-ben született elemzés, amely megpróbálta tisztázni, pontosan melyikük melyik jelenetet, felvonást írta, és azóta is sok, eltérő metodikájú és szempontú elemzés született. Plecháč tanulmánya kombinálja a szókincs és a metrikai jellemzők gyakoriságának vizsgálatát, és így nagyon pontos eredményt kap, főleg a csúszó vagy gördülő attribúció (*rolling attribution*) módszerével (itt a szöveget százsoros részekben vizsgálja, de ötsoronként előrehaladva (1–100, 5–105, 10–110 stb.), ami nagyon jó százalékban kiszűri a hibákat. Végeredményben a tanulmány megerősíti az elmúlt évtizedek filológiai álláspontját a két szerző által írt részekről. Nagyon meggyőző bizonyítása ez annak, hogy a verselés ugyanolyan egyedi, szerzőspecifikus jellemző, mint a szóhasználat. A verselés szempontjából egyébként az adott esetben a legnagyobb különbség az volt, hogy Shakespeare sokkal több hangsúlyos végződésű sort írt, míg Fletcher-nél többségben voltak a hangsúlytalan sorvégek.

A kötet utolsó tanulmányának társszerzője Artjoms Šeļa, a tartui egyetem kutatója. Ebben a vizsgálatban egy hamisított szövegkorpusz beazonosítása a cél. 1978-ban A. A. Iljusin publikálta G. S. Batyenko költő, katonatiszt és dekabrista lázadó összes költeményét, köztük számos olyan verssel, amelyeket állítása szerint Batyenko hosszú fogsága és szibériai száműzetése során írt. E versek hitelességét sokan megkérdőjelezték, a legalaposabb vizsgálatnak M. I. Sapir vetette őket alá, de számítógép használata nélkül. Sapir konklúziója a vizsgálatból az volt, hogy lehetetlen stilisztikai és verselési tulajdonságok alapján meghatározni egy szerzőt. Ezzel a túlzó általánosítással száll szembe a tanulmány, mely meggyőzően mutatja ki Batyenko és kortársai verseinek elemzésével, és ezeket összevetve a Pszeudo-Batyenko-féle korpuszsal, sőt, Iljusin saját verseivel, hogy igenis, nagy pontossággal tisztázni lehet, hogy melyek az Iljusin által írt, hamisított versek.

A rövid, de bravúros könyv tehát több szempontból is bemutatja vizsgálati módszere érvényességét. A magyar filológiában nincs sok olyan kérdéses szerzőségű verses szöveg, amelyek ilyen komplex, stilo- és sztichometriai elemzése komoly meglepetéseket tartogathat. De azért vannak esetek, amikor érdemes ezzel megpróbálkozni, például az *Eurialus és Lucretia* című széphistória esetében. Emellett a módszer pontosságát magyar versanyagon olyan, már bizonyított hamisítások korpuszának elemzésével lehetne ellenőrizni, mint Thaly Kálmán kuruc versei. Izgalmas lehet a többféle költői személyiséget és stílust kidolgozó Kovács András Ferenc verseinek összevetése is, hogy azonosítsuk, mely jegyek jellemzik az alteregóit, s melyek, megcáfolhatatlanul, a közös szerzőt.