



AZ OKTATÁS, A KUTATÁS ÉS A KÖZGYŰJTEMÉNYEK DIGITÁLIS TRANSZFORMÁCIÓJA FELSŐFOKON

NETWORKSHOP 2024
33. Országos Informatikai Konferencia

2024. április 3–5.
Eszterházy Károly Katolikus Egyetem, Eger

AZ OKTATÁS, A KUTATÁS ÉS A KÖZGYŰJTEMÉNYEK DIGITÁLIS TRANSZFORMÁCIÓJA FELSŐFOKON

NETWORKSHOP 2024
33. Országos Informatikai Konferencia

2024. április 3–5.
Eszterházy Károly Katolikus Egyetem, Eger

Szerkesztette: Tick József, Kokas Károly, Holl András

HUNGARNET Egyesület
Budapest, 2024



HUN-REN
Magyar Kutatási Hálózat

NETWORKSHOP

Szerkesztette: Tick József, Kokas Károly, Holl András

Tipográfia és tördelés: Vas Viktória

Korrektúra: Danyi Melinda

Angol nyelvi lektor: Cseresnyés Dorottya

Networkshop 2024 konferencia előadásainak közleményei

Eszterházy Károly Egyetem, Eger

2024. április 3–5.

ISBN 978-615-82243-2-1

DOI: <https://doi.org/10.31915/NWS.2024>

Kiadja a HUNGARNET Egyesület
az MTA Könyvtár és Információs Központ közreműködésével

Budapest

2024

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

TARTALOMJEGYZÉK

Előszó	5
Ungváry Rudolf A MARC21 formátum kettős szerkezete és a formátum felhasználói szintjének fordításai	7
Holl András, Andódy Katalin Adatbányászati gyakorlatok repozitóriumra és MTMT-re.....	16
Simon András Mesterséges intelligenciával támogatott adatgazdagítás a Nemzeti Levéltárban.....	22
Soós Gábor, Rövid András, Ormos Pál V2X – A járművek közötti kommunikáció kihívásai	29
Csernai Zoltán Egy online tanulást támogató portál kurzusának vizsgálata Big Data adatelemző módszerekkel.....	36
T. Nagy László, Németh Áron A mesterséges intelligencia (MI) teológiai kompetenciái	45
Mészáros Erika Kodaktól a jövőig – Egy könyvtári digitalizálás szinterei.....	55
Frankó Máté, Sándor Ákos Adatvizualizáció a könyvtári menedzsmentben: fejlesztések az SZTE Klebelsberg Könyvtár döntéstámogató rendszerében.....	63
Hernek István A felhasználóképzés szintjei az SZTE Klebelsberg Könyvtárban: az elsőévesektől a kutatókig	72
Némethi-Takács Margit, Borbély Mária Bibliográfiai kapcsolatok az általános megjegyzés adatmezőben.....	78
Dobás Kata, Tüskés Anna A magyar irodalomtörténet bibliográfiájának migrációja az ITidata szemantikus adatbázisba	87
Horváth Péter A kanonikus magyar költészet versformakeresője.....	96
Sebestyén Ádám, Sárközi-Lindner Zsófia Történeti források szemantikus feldolgozása – Az ELTEdata adatbázis új gyűjteményei	105

Bolya Mátyás

Lyukkártya és népdalrendezés – Egy mechanikus népzenei adatbázis digitális rekonstrukciójának lehetőségei.....112

Kovácsházy Tamás

Az idő, mint alapvető infrastruktúra, az idő szerepe az adatközpontban.....121

Albert Ágota Katalin

A mesterséges intelligencia használatának követelményei az oktatási szektorban, különös tekintettel a mesterséges intelligencia használatáról szóló rendelethez.....129

Varga Emese

Digitális szövegszerkesztés a dHUpa keretrendszerében135

Nemoda Zsuzsanna, Héjja Balázs, Nagy Andor, Tóth Máté

A Pest Megyei Digitális Könyvtár fejlesztése 141

Nagy Dóra, Sándor Ákos

Voice2text: a hanganyagátírás lehetőségei MI segítségével.....149

Kalcsó Gyula

Képek és metaadataik gyűjteményezése scrapingtechnológiával közösségi képmegosztó oldalról157

Péter Róbert, Szántó Zsolt, Biacsi Zoltán, Kocsis Zoltán, Berend Gábor, Bilicki Vilmos

Az AVOBMAT (Analysis and Visualization of Bibliographic Metadata and Texts) többnyelvű kutatási eszköz munkafolyamata és új funkciói163

Máray Tamás

Kvantum-számítástechnika: ez már a „jövő”?171

Fellegi Zsófia

Digitális kiadások migrációja: gépi és emberi intelligencia együttműködése.....177

Palkó Gábor

Posztmodern intertextualitás és digitális szövegkiadás184

Antal Dániel

A szlovák adatkicserélési tér magyarországi föderációjának lehetőségei.....192

Vass Johanna

Kutatási adatok megosztása a gyakorlatban – Adatrepozitóriumok használata az Ökológiai Kutatóközpont publikációiban199

Mihály Eszter, Micsik András, Nagy Kadosa

Irodalmi levélváltások nyomában TEI-vel és térképpel.....208

A kanonikus magyar költészet versformakeresője

A poem form searcher for Hungarian canonical poetry

Horváth Péter

ELTE Digitális Bölcsészeti Tanszék, Digitális Örökség Nemzeti Laboratóriumhorvath.peter@btk.elte.hu

Absztrakt

A tanulmány az 50 magyar költő összes versét tartalmazó ELTE Verskorpuszhoz fejlesztett Versformakeresőt mutatja be. A Versformakereső segítségével a korpusz verseit listázhatjuk rímképlet, szótagszám, időmértékes metrum és ütemhangsúlyos metrum alapján. A tanulmány az eszköz keresőfunkciói mellett röviden bemutatja a versek rímképletét és metrumát felismerő algoritmusok működésének a főbb elveit. Emellett egy rövid vizsgálat segítségével az eszköz kutatásban való hasznosíthatóságára is rámutat. A kvantitatív vizsgálat a Versformakeresővel kinyert adatok alapján igazolja azt a régóta létező irodalomtörténeti megfigyelést, miszerint az ütemhangsúlyos és a trochaikus metrum vonzza egymást.

Kulcsszavak: versformakereső, ELTE Verskorpusz, rímképlet, metrum, automatikus annotálás

Abstract

This paper presents a poem form searcher developed for the ELTE Poetry Corpus containing all the poems of 50 Hungarian authors. The poem form searcher allows the user to query poems on the basis of rhyme pattern, syllable number, quantitative meter, and qualitative meter. Besides the search functions of the tool, the paper presents the main principles of the algorithms recognizing the rhyme pattern and the meter of the poems. It also highlights the tool's usefulness in research through a short investigation. The quantitative analysis confirms the long-standing literary scholarly observation that qualitative and trochaic meters attract each other in Hungarian poetry.

Keywords: poem form searcher, ELTE Poetry Corpus, rhyme pattern, meter, automatic annotation

1. Bevezetés

A <https://versformakereso.elte-dh.hu/> oldalon elérhető versformakereső egy olyan eszköz, amelynek segítségével az ELTE Verskorpuszban szereplő 50 szerző több mint 13 ezer versét listázhatjuk a szótagszáma, rímelésre és metrumra vonatkozó formai jellemzők alapján.¹ Az eszköz a szerző reményei szerint nemcsak az irodalomtudományos kutatásokban, hanem akár az oktatásban is hasznosítható, hiszen a lekérdezőfelület segítségével az egyes rímképleteket és metrumtípusokat, illetve ezek különböző kombinációit megvalósító versekre egyszerűen hozhatunk szemléletes példákat. Mivel az eszköz az ELTE Verskorpusz

1 Horváth Péter, Kundráth Péter, Indig Balázs et al., „ELTE Verskorpusz – a magyar kanonikusköltészet gépileg annotált adatbázisa”, in Berend Gábor, Gosztolya Gábor és Vincze Veronika szerk., *XVIII. Magyar Számítógépes Nyelvészeti Konferencia*, 375–88 (Szeged: Szegedi Tudományegyetem TTIK, Informatikai Intézet, 2022). A korpusz innen letölthető: <https://github.com/ELTE-DH/poetry-corpus>

vershangzásra vonatkozó annotációiban keres, elsőként az ezeket létrehozó algoritmusokat mutatom be röviden. Ezt követően ismertetem a lekérdezőeszköz informatikai hátterét és keresőfunkcióit. Végezetül a trochaikus és ütemhangsúlyos metrum együttállását vizsgálom a Versformakereső használatával kinyert adatok segítségével, bemutattva hogy az eszköz irodalomtudományos kérdések kvantitatív vizsgálatára is használható.

2. A rímképletelemző algoritmus

A rímképletelemző algoritmus működését egy korábbi cikkben részletesen ismertettem,² így itt csak az algoritmus működésének főbb elveit foglalom össze. Az algoritmus nyolc szabálykészlet alapján elemzi a verset, amelyek az elsőtől a nyolcadik szabálykészletig haladva egyre kevesebb kritériumot állítanak fel arra vonatkozólag, hogy milyen esetben rímel két sor. Például az első, legszigorúbb szabálykészlet szerint csak akkor rímel két sor, ha azokban az utolsó és az utolsó előtti magánhangzók a hosszúságot nem számítva azonosak, az utolsó előtti szótagok hosszúsága azonos (rövid vagy hosszú), illetve a két sor megegyezik abban is, hogy a sorok utolsó magánhangzója után szerepel-e mássalhangzó vagy nem (nyílt vagy zárt szótag). Ezzel szemben a nyolcadik, leglazább szabálykészlet csupán azt a kritériumot tartalmazza, hogy a rímelő sorok utolsó magánhangzóinak a hosszúságot nem számítva azonosnak kell lennie. Az algoritmus célja olyan kimenet produkálása, amelyben a vers összes versszaka ugyanazt a rímképletet kapja. A program ennek érdekében egymás után futtatja a szabálykészleteket a legszigorúbbtól a leglazább felé haladva egészen addig, amíg el nem éri ezt a kimenetet. Amennyiben nem tud egységes kimenetet adni egyik szabálykészlet alapján sem, vagy a versszakok eltérő sorszáma miatt már eleve lehetetlen a versszakok egységes elemzése, akkor az alapbeállításként megadott szabálykészlettel elemzi le a verset.

3. Az időmértékes metrumot elemző algoritmus

Az időmértékes metrumot felismerő algoritmus működését úgyszintén részletesen bemutattam egy régebbi cikkben³, így itt csak az algoritmus működésének a főbb elveit és legfontosabb lépéseit foglalom össze. A program a jambikus, trochaikus, anapesztikus és daktilikus metrumokba sorolja be a verseket. A metrum típusa mellett megad egy 0 és 1 közötti szabályossági pontszámot is, amely azt jelzi, hogy a vers ritmusa milyen mértékben realizálja a felismert metrumot. A szabályossági pontszám alkalmazása mögött az a megfontolás áll, hogy a versek időmértékes metruma nem bináris (van vagy nincs), hanem fokozati jellemző, azaz egy vers különböző mértékben realizálhat egy adott metrumot. A program mind a négy metrum alapján elemzi a verset, és a vers azt a metrumcímét kapja meg, amelyre a legnagyobb szabályossági pontszám jött ki.

A szabályossági pontszám kiszámítása során a program egy egyszerű pontozásos módszert használ. Ennek során elsőként kiszámítja a verssorok szabályossági pontszámát a verssor-

2 Horváth Péter, „Egy automatikusan generált rímshótár fejlesztése és a magyar kanonikus költészet rímshavainak néhány jellemzője”, in Tick József, Kokas Károly és Holl András szerk., *Új technológiákkal, új tartalmakkal a jövő digitális transzformációja felé. 32 Networkshop: országos konferencia*, 77–84 (Budapest: HUNGARNET Egyesület, 2023).

3 Horváth Péter, „Két eljárás magyar nyelvű versek metrumának gépi felismertetéséhez”, *Digitális Bölcsészett 4* (2021): 79–103.

ban szereplő verslábak pontozásával.⁴ A sorok pontozása az alábbi képlet alapján történik, amelyben 'a' az alaplábak száma, 'b' az elsődleges helyettesítő lábak száma, 'c' a másodlagos helyettesítő lábak száma, 'd' pedig a teljes lábak száma:

$$\text{versorszabályosságipontszáma} = \frac{4a + 2b + c}{4d}$$

Ez tehát azt jelenti, hogy a sorban szereplő alapláb (pl. trochaikus metrum elemzése esetén a trocheus) négy pontot kap, az elsődleges helyettesítő láb (a spondeusz mind a négy metrumtípus elemzésekor) két pontot kap, a másodlagos helyettesítő láb (pl. trochaikus vagy jambikus metrum elemzése esetén a pirrichius) egy pontot kap, az alaplábbal ellentétes láb (pl. trochaikus metrum elemzése esetén a jambus) pedig nulla pontot kap. A sor szabályossági pontszáma a verslábak pontszámainak az átlaga. Ezt követően a program kiszámítja a vizsgált metrumra vonatkozóan az egész vers szabályossági pontszámát a sorok szabályossági pontszámainak az átlagolásával. Végezetül pedig összeveti a négy metrumra kapott szabályossági pontszámokat, és azzal a metrummal címkézi fel a verset, amelyre a legnagyobb szabályossági pontszámot kapta.

4. Az ütemhangsúlyos metrumot elemző algoritmus

Az ütemhangsúlyos metrumot elemző algoritmusnak itt megint csak a főbb működési elveit foglalom össze.⁵ A program azoknak a verseknek az ütemhangsúlyosságát ismeri fel, amelyekben minden sor ugyanazt az ütemosztást követi, illetve amelyekben a versszakok páratlan és páros sorára eltérő ütemosztás jellemző (abab... típusú versek). E két típus lefedi az ütemhangsúlyos versek túlnyomó részét. A program elemzésében az ütemek maximum hat szótagból állhatnak. Az ütemhangsúlyos metrum automatikus felismerése a szóeleji és nem szóeleji szótagok megkülönböztetésére épül, mivel a magyar nyelvben a szavak első szótagjai a hangsúlyos szótagok, vagyis ezek állhatnak az ütemek elején. Az elemzés során a program a szóeleji szótagokat 1 számjegyekkel, a nem szóeleji szótagokat pedig 0 számjegyekkel jelölő mátrixokat hoz létre. Például Arany János Szülőhelyem című négysoros versének a szóeleji és nem szóeleji szótagokat megkülönböztető mátrixát az 1. táblázat mutatja be.

Szülőhelyem, Szalonta,	1000100
Nem szült engem szalonba;	1110100
Azért vágyom naponta	1010100
Kunyhóba és vadonba.	1001100

1. táblázat: Szóeleji szótagokat jelölő mátrix

A program ezekből a mátrixokból egyszerű átlagolással kiszámítja, hogy a vers egyes szótaghelyeire milyen arányban esik szóeleji szótag. Amennyiben ez az arány eléri a 0,75-öt,

4 A verslábakra osztás módszerét lásd Horváth, „Két eljárás...”, 84–87.

5 A részletes leírást lásd Horváth, „Két eljárás...”, 93–98.

a program az adott szótaghelyen ütemkezdetet állapít meg. Például az Arany-vers esetében a szóeleji szótagok aránya az első szótaghely mellett az ötödik szótaghelyen éri el a 0,75-öt, vagyis a vers metruma négy és három szótagos ütemekből álló kétütemű hetes. Mivel a versek első szava sok esetben egyszótagú hangsúlytalan funkciószó (pl. és, de, a, az), amelyeket nem lenne észszerű külön ütemekként elemezni, a programba be van építve egy megszorítás, miszerint a második szótaghelyen nem lehet ütemkezdet. Az abab... szerkezetű versek elemzése során a program a versszakok páratlan és páros sorait külön mátrixokként elemzi, ugyanúgy 0,75-ös küszöbértékkel.

5. A Versformakereső informatikai háttere és funkciói

A Versformakereső a Python-nyelv FastApi keretrendszerében fejlesztett eszköz.⁶ Az eszköz egy SQLite-formátumú, a Verskorpusz XML-fájljaiból generált relációs adatbázisban keres. A relációs adatbázis egy darab adattáblából áll, amely az ELTE Verskorpusz összes versére vonatkozóan tartalmazza a versek metaadatait és formai jellemzőit, azaz a rímképletre, a metrumra, a sorok szótagszámára és a vers szószámára vonatkozó információkat.

A Versformakeresővel a relációs adatbázisban szereplő különböző formai jellemzők alapján szűrhetjük a verseket (1. ábra). Egyrészt kereshetünk rímképlet és a sorok szótagszámai alapján. Emellett azt is meghatározhatjuk, hogy a megadott rímképlet és/vagy szótagszámok milyen módon vonatkozzanak a versre. Három lehetőségből választhatunk: (1) a kilistázott verseknek van legalább egy olyan versszakuk, amely megfelel a megadott rímképletnek és/vagy szótagszámoknak; (2) a kilistázott verseknek az összes versszaka megismétli a megadott rímképletet és/vagy szótagszámokat; (3) a kilistázott versek egésze megfelel a megadott rímképletnek és/vagy szótagszámoknak.

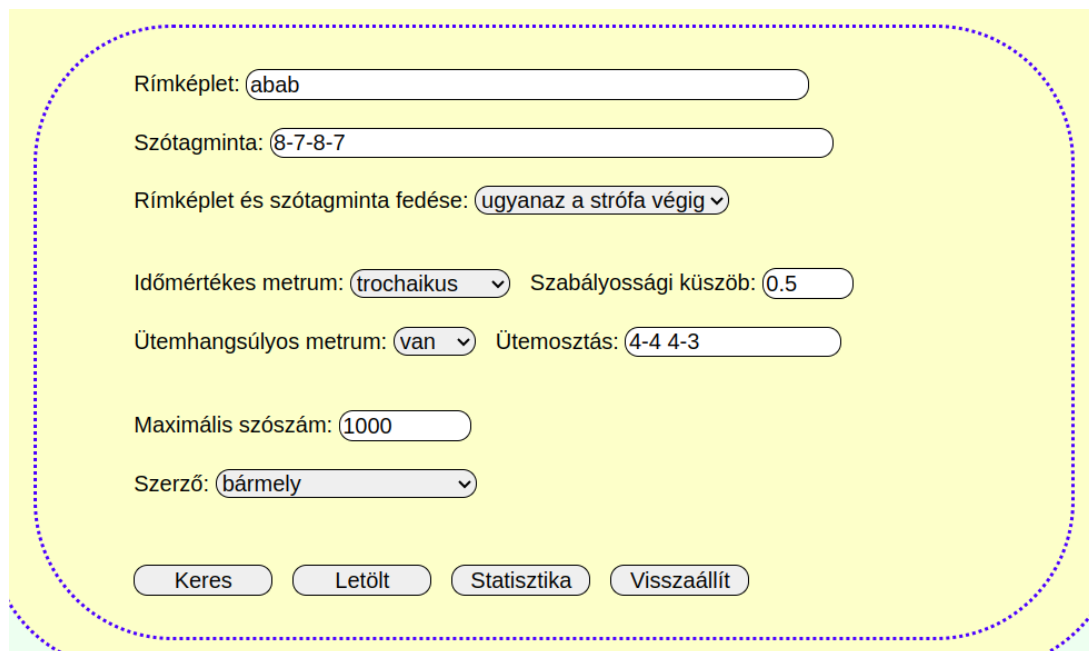
Az eszközzel specifikálhatjuk a versek metrumát is. Egyrészt megadhatjuk, hogy a kilistázandó verseknek legyen-e időmértékes metruma vagy ne, illetve az időmértékes metrumot le is szűkíthetjük a jambikus, trochaikus, daktilikus és anapesztikus metrumokra. Emellett az eszközzel állíthatjuk az időmértékes metrumra vonatkozó szabályossági pontszám küszöbértékét is, amely egy 0 és 1 közötti tetszőleges érték. Minél magasabbra állítjuk ezt az értéket, az időmértékes metrumot annál szabályosabban megvalósító versekre szűkítjük a keresést. Az alapbeállításként szereplő 0,5-ös érték már elfogadható találatokat ad.

A verseket szűrhetjük ütemhangsúlyos metrum alapján is. Egyrészt megadhatjuk, hogy legyen vagy ne legyen ütemhangsúlyos metruma a verseknek, másrészt specifikálhatjuk a verssorok ütemosztását is. Például kilistázhatjuk azokat a verseket, amelyekben a sorok két ütemű hetesek, négy szótagos első és három szótagos második ütemmel. Rákereshetünk olyan versekre is, amelyekben a versszakok páratlan és páros sorai eltérő ütemosztást követnek, például a páratlan sorok felező nyolcasok, míg a páros sorok két ütemű hetesek. Az időmértékes és az ütemhangsúlyos metrumra vonatkozó keresési funkciók együttes használatával megkaphatjuk a szimultán, azaz egyszerre időmértékes és ütemhangsúlyos verseket is.

6 Köszönöm Indig Balázsnak, hogy az elkészült programot a szükséges módosításokkal felrakta az ELTE Digitális Bölcsészeti Tanszékének szerverére, valamint Nagy Mihálynak a lekérdező fejlesztése kapcsán adott javaslatokat.

A verseket szűrhetjük szerző és szószám alapján is. Az eszköz alapbeállításban az 1000 szónál hosszabb verseket nem listázza. Ez a beállítás alkalmas a lírai műnemtől az epika irányába elmozduló hosszabb versek kiszűrésére.

A keresés eredményeként megkapjuk a megadott formai tulajdonságoknak megfelelő versek szerzőjét, címét, ELTE Verskorpuszbeli azonosítóját, valamint a formai jellemzőit (2. ábra). A vers azonosítójára rákattintva megnyithatjuk a verset a Verskorpusz felületén, ahol további automatikusan felismertetett jellemzők is szerepelnek. A találatokat TSV-formátumban lementhetjük a gépünkre is.



Rímképlet:

Szótagminta:

Rímképlet és szótagminta fedése:

Időmértékes metrum: Szabályossági küszöb:

Ütemhangsúlyos metrum: Ütemosztás:

Maximális szószám:

Szerző:

1. ábra: A Versformakereső lekérdezőfelülete

<<
>>

Találatok száma: 1-3 / 3

Gyulai Pál: ŐSZI DÉLUTÁN., [Gyulai_00665_0095](#). Szavak száma: 166, rímképlet: |abab|abab|abab|abab|abab|abab|abab|abab|abab|abab|, szótagminta: |8-7-8-7|8-7-8-7|8-7-8-7|8-7-8-7|8-7-8-7|8-7-8-7|8-7-8-7|8-7-8-7|, időmértékes metrum: trochaikus (0.69), ütemhangsúlyos metrum: 4-4_4-3

Gyulai Pál: II., [Gyulai_00665_0164](#). Szavak száma: 88, rímképlet: |abab|abab|abab|abab|abab|, szótagminta: |8-7-8-7|8-7-8-7|8-7-8-7|8-7-8-7|8-7-8-7|, időmértékes metrum: trochaikus (0.65), ütemhangsúlyos metrum: 4-4_4-3

Reviczky Gyula: XLIII. Nem voltál már ifju lányka, [Reviczky_01058_0052](#). Szavak száma: 65, rímképlet: |abab|abab|abab|abab|, szótagminta: |8-7-8-7|8-7-8-7|8-7-8-7|8-7-8-7|, időmértékes metrum: trochaikus (0.67), ütemhangsúlyos metrum: 4-4_4-3

Találatok száma: 1-3 / 3

<<
>>
↓

2. ábra: A keresési találatok megjelenítése

Az eszköznek van egy Statisztika funkciója is, amelynek használatával szerzőnként kilistázhatjuk a keresési feltételeknek megfelelő versek számát és a szerző összes verséhez képest megállapított arányát (3. ábra). A kapott adatokat többek között a szerzők születési évszáma, illetve a találatok abszolút és relatív gyakorisága alapján is rendezhetjük. A Statisztika funkció használatával megtudhatjuk, hogy mely szerzők írták a megadott formai jellemzőknek megfelelő legtöbb verset, illetve időbeli tendenciákat is kimutathatunk.

trochaikus, 0.5, ütemhangsúlyos, max. szószám: 100000				
Talált versek száma: 1037				
Szerző	Dátum	Összes vers száma	Talált versek száma	Talált versek aránya
Ady Endre	1877–1919	1116	70	0.063
Amade László	1704–1764	154	0	0.000
Ányos Pál	1756–1784	88	8	0.091
Arany János	1817–1882	417	113	0.271
Arany László	1844–1898	6	1	0.167
Babits Mihály	1883–1941	514	24	0.047
Bajza József	1804–1858	97	19	0.196
Balassi Bálint	1554–1594	107	4	0.037
Batsányi János	1763–1845	62	12	0.194
Berzsenyi Dániel	1776–1836	136	1	0.007
Bessenyei György	1747–1811	40	1	0.025
Bornemisza Péter	1535–1584	8	0	0.000
Csokonai Vitéz Mihály	1773–1805	394	11	0.028
Dayka Gábor	1769–1796	29	0	0.000
Dsida Jenő	1907–1938	1060	33	0.031

3. ábra: Az eszköz Statisztika funkciója

6. A trochaikus és időmértékes metrum együttállása

Az alábbiakban egy rövid esettanulmány segítségével mutatom be a Versformakereső kutatásban való hasznosíthatóságát. A magyar irodalomtörténetben régóta létezik az a kvalitatív alapú megfigyelés, miszerint az ütemhangsúlyos és trochaikus metrum vonzza egymást, azaz nagy eséllyel jelennek meg egyszerre.⁷ A vizsgálat kérdése az, hogy vajon ez az állítás igazolható-e a Verskorpuszban szereplő szerzők versei alapján, kvantitatív módon.

A kérdés megválaszolásához a Versformakereső Statisztika funkciójának a segítségével kinyertem, hogy az egyes szerzők hány darab jambikus, trochaikus, ütemhangsúlyos, valamint szimultán jambikus, azaz egyszerre jambikus és ütemhangsúlyos, illetve szimultán trochaikus, azaz egyszerre trochaikus és ütemhangsúlyos verset írtak. A jambikus és trochaikus metrumú versek számának a megállapításához 0,5-ös szabályossági küszöböt használtam. A nem szerzőkre lebontott, összesített adatokat a 2. táblázat mutatja be. A vizsgálat idején a korpuszban szereplő összes vers száma 13 362 volt.

Metrumtípus	Versek száma
jambikus	5792
trochaikus	1723
ütemhangsúlyos	3726
jambikus + ütemhangsúlyos	1327
trochaikus + ütemhangsúlyos	1037

2. táblázat. Különböző metrumokat megvalósító versek száma

Ezen adatok alapján kiszámítottam a jambikus és ütemhangsúlyos, valamint a trochaikus és ütemhangsúlyos metrumok együttállásának a kölcsönös információ értékét (MI-értékét).⁸ A 0 feletti MI-érték azt jelzi, hogy a két metrum többször jelenik meg egyszerre ugyanazon verseknél, mint ahogyan az a véletlenszerűség alapján várható lenne, azaz a két metrum vonzza egymást. A 0-nál alacsonyabb érték pedig azt jelzi, hogy a két metrum kevesebb-szer jelenik meg egyszerre ugyanazon verseknél, mint a véletlenszerűségen alapuló várható érték, azaz taszítják egymást. A mérés kiindulópontjaként használt várható érték tehát azt adja meg, hogy a vizsgált két metrum várhatóan hány-szor fordulna elő egy-

7 Vargyas Lajos, *Magyar vers – magyar nyelv. Verstani tanulmány* (Budapest: Szépirodalmi Könyvkiadó, 1966), 148–150; Horváth János, *Rendszeres magyar verstan* (Budapest: Akadémiai Kiadó, 1969), 128–133; Szepes Erika és Szerdahelyi István, *Verstan* (Budapest: Gondolat Kiadó, 1981), 267–269. Szuromi Lajos *A szimultán verselés* (Budapest: Akadémiai Kiadó, 1990) című könyvében rámutat arra, hogy számos szerző, különösen a 20. sz. elejétől, több szimultán jambikus verset írt, mint szimultán trochaikus (178–179, 190). Ez azonban abból az egyszerű tényből is következhet, hogy ezek a szerzők eleve jelentősen több jambikus verset írtak, mint trochaikus. A pusztán gyakorisági adatok ennek folytán félrevezetőek. Ugyanakkor Szuromi a korábbi megközelítésekhez hasonlóan ugyanúgy elfogadja, sőt magyarázataiban alapvető kiindulópontként használja a trochaikus és ütemhangsúlyos verselés hangzásbeli hasonlóságát (pl. 62–63, 68, 193).

8 Az MI-érték számításának nyelvészeti alkalmazásáról lásd Kenneth W. Church and Patrick Hanks, „Word association norms, mutual information, and lexicography”, *Computational Linguistics* 16, 1. sz. (1990): 22–29; Stefan Evert, „Corpora and collocations”, in Anke Lüdeling and Merja Kytö eds., *Corpus linguistics: An international handbook*. Vol. 2., 1212–1248 (Berlin, New York: Walter de Gruyter, 2009).

szere, ugyanazon verseknél, ha teljesen véletlenszerűen dőlné el, hogy az 5792 jambikus, az 1723 trochaikus és a 3726 ütemhangsúlyos metrumú vers melyik lenne a korpusz versei közül.⁹

Az eredményeket a 3. táblázat mutatja be. Az MI-értékeket kétféleképpen is kiszámítottam. Egyrészt megadtam a 2. táblázatban szereplő előfordulási adatok alapján, azaz anélkül, hogy figyelembe vettem volna a versek szerzőit. Másrészt megadtam a szerzőnként kiszámított MI-értékek átlagát is, illetve azt, hogy hány szerző esetében volt nagyobb, illetve kisebb ez az érték nullánál. A szerzőnként elvégzett számításba csak azokat a szerzőket vettem bele, akik legalább egy olyan verset írtak, amely egyszerre jambikus és ütemhangsúlyos, illetve trochaikus és ütemhangsúlyos.

	jambikus + ütemhangsúlyos	trochaikus + ütemhangsúlyos
Teljes korpusz	-0,283	1,11
A szerzőkre kapott MI-értékek átlaga	-0,105	0,987
Pozitív MI-értékű szerzők száma	14	39
Negatív MI-értékű szerzők száma	25	2

3. táblázat: Jambikus és ütemhangsúlyos, valamint trochaikus és ütemhangsúlyos metrumok MI-értékeik

A jambikus és ütemhangsúlyos metrum együttes előfordulásának negatív MI-értékeiből látható, hogy azok kevesebbszer jelennek meg együtt, mint ami a véletlenszerűség alapján várható lenne, azaz a két metrumtípus taszítja egymást. Az ELTE Verskorpusz 50 szerzőjéből 39 szerző írt legalább egy olyan verset, amely egyszerre jambikus és ütemhangsúlyos metrumú. Ebből a 39 szerzőből 25 esetében jelenik meg negatív MI-érték. A trochaikus és ütemhangsúlyos versek esetében mind a kétféleképpen elvégzett mérés pozitív MI-értéket eredményezett. Az ezt a verstípust legalább egyszer megvalósító 41 szerzőből 39 szerző esetében jelenik meg pozitív MI-érték, és csak kettő esetében negatív. A kapott MI-értékek alapján tehát megállapíthatjuk, hogy az ütemhangsúlyos és a trochaikus metrum közötti, a magyar irodalomtörténetben régóta megfigyelt együttállás jelensége a korpuszban szereplő kanonikus magyar költők versein végzett, kvantitatív méréssel is igazolható.

Ahogy a 2. táblázatból látható, trochaikus versből jelentősen kevesebb szerepel a korpuszban, mint ütemhangsúlyos versből (1723 vs. 3726). Ha véletlenszerűen kiválasztunk egy trochaikus verset a korpuszból, akkor 0,60 a valószínűsége annak, hogy az egyben ütemhangsúlyos is, ha viszont egy ütemhangsúlyos verset választunk ki véletlenszerűen, akkor jóval kevesebb, csak 0,28 a valószínűsége annak, hogy az egyben trochaikus is. Úgy tűnik tehát, hogy a trochaikus metrum kevésbé fordul elő önállóan, mint az ütemhangsúlyos metrum. Vagy máshogyan megfogalmazva: a trochaikus metrum jobban vonzza az ütemhangsúlyos metrumot, mint az ütemhangsúlyos a trochaikus metrumot.

9 Ha M azon versek száma, amelyben mind a két metrum előfordul, V a két metrum véletlenszerűségén alapuló várható együttes előfordulása, A azon versek száma, amelyben az egyik metrum előfordul, B azon versek száma, amelyben a másik metrum előfordul, C pedig az összes vers száma a korpuszban, akkor a kölcsönös információ érték kiszámításának a képletei a következők: $MI = \log_2 \frac{MV}{AB}$ $V = ABC$

Irodalomjegyzék

- Church, Kenneth W. and Hanks, Patrick. „Word association norms, mutual information, and lexicography”. *Computational Linguistics* 16, 1. sz. (1990): 22–29.
- Evert, Stefan. „Corpora and collocations”. In *Corpus linguistics: An international handbook*. Vol. 2., edited by Anke Lüdeling and Merja Kytö, 1212–1248. Berlin, New York: Walter de Gruyter, 2009.
- Horváth János. *Rendszeres magyar verstan*. Budapest: Akadémiai Kiadó, 1969.
- Horváth Péter. „Két eljárás magyar nyelvű versek metrumának gépi felismertetéséhez”. *Digitális Bölcsészet* 4 (2021): 79–103.
- Horváth Péter. „Egy automatikusan generált rímshótár fejlesztése és a magyar kanonikus költészet rímshavainak néhány jellemzője”. In *Új technológiákkal, új tartalmakkal a jövő digitális transzformációja felé: 32 Workshop: országos konferencia*, szerkesztette Tick József, Kokas Károly és Holl András, 77–84. Budapest: HUNGARNET Egyesület, 2023.
- Horváth Péter, Kundráth Péter, Indig Balázs, Fellegi Zsófia, Szlavich Eszter, Bajzát Tímea Borbála, Sárközi-Lindner Zsófia, Vida Bence, Karabulut Aslihan, Timári Mária, Palkó, Gábor. „ELTE Verskorpusz – a magyar kanonikus költészet gépileg annotált adatbázisa”. In *XVIII. Magyar Számítógépes Nyelvészeti Konferencia*, szerkesztette Berend Gábor, Gosztolya Gábor és Vincze Veronika, 375–88. Szeged: Szegedi Tudományegyetem TTIK, Informatikai Intézet, 2022.
- Szepes Erika és Szerdahelyi István. *Verstan*. Budapest: Gondolat Kiadó, 1981.
- Szuromi Lajos. *A szimultán verselés*. Budapest: Akadémiai Kiadó, 1990.
- Vargyas Lajos. *Magyar vers – magyar nyelv*. Verstani tanulmány. Budapest: Szépirodalmi Könyvkiadó, 1966.