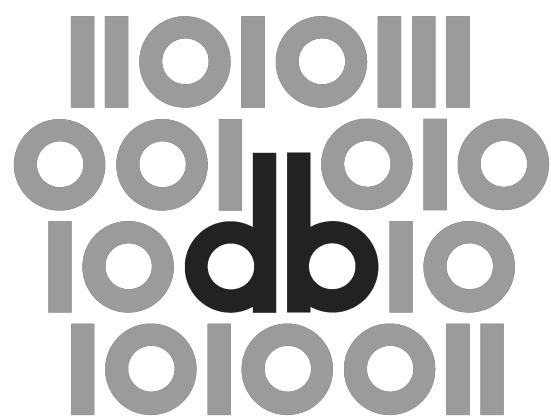


Digitális Bölcsészet
2021., negyedik szám

<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 Magyarország *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: Bucsecs Katalin

Tördelés: Hegedüs Béla

Grafika: Hegyi Gábor

<TANULMÁNYOK>

Horváth Péter  0000-0002-3517-5623

ELTE BTK TI Digitális Bölcsészet Tanszék

horvath.peeteer@gmail.com

Két eljárás magyar nyelvű versek metrumának gépi felismertetéséhez*

A tanulmány a magyar időmértékes és ütemhangsúlyos versmetrum gépi felismertetésének egy-egy eljárását mutatja be. A magyar időmértékes metrum elemzését végző algoritmus a verset besorolja a daktilikus, anapestikus, jambikus vagy trochaikus kategóriák valamelyikébe, és megadja a megvalósult versritmus szabályosságának mértékét. A magyar ütemhangsúlyos metrum elemzését végző algoritmus pedig megadja az ütemek szótagszámát. A tanulmány a két algoritmus alkalmazási lehetőségeit egy tesztkorpusz segítségével mutatja be, amely tizenegy, a 19. század közepén és második felében, valamint a 20. század első felében alkotó szerző verseit tartalmazza.

Kulcsszavak:

időmértékes metrum, ütemhangsúlyos metrum, gépi elemzés, korpuszvizsgálat, ELTE Verskorpusz



1. Bevezetés

A tanulmány a magyar időmértékes és ütemhangsúlyos metrum gépi felismertetésének egy-egy eljárását mutatja be. Az időmértékes metrum gépi elemzését elvégző algoritmus a négy nagy időmértékes metrumkategória, a daktilikus, az anapestikus, a jambikus és a trochaikus felismerésére, valamint a megvalósuló versritmus szabályosságának a mérésére képes. Természetesen a négy metrumkategórián belül további alkategóriák is vannak, mint amilyen például a hexameter, illetve olyan kötött képletű metrumok is léteznek, amelyek nem sorolhatók be ezekbe az általános kategóriákba. Ezen specifikusabb metrumok automatikus felismerésére az algoritmus nem képes. Az ütemhangsúlyos verselés felismerését végző algoritmus is az egyszerűbb, aaaa vagy abab szerkezetű versek felismerésére képes, azaz azon versek esetében állapít meg ütemhangsúlyos metrumot, ahol minden sorra ugyanaz az ütemosztás jellemző, vagy pedig kétfajta, a versszakok páratlan és páros soraira jellemző ütemosztás váltogatja egymást. Megjegyzendő ugyanakkor, hogy az ütemhangsúlyos verseknek a legnagyobb részét lefedi ez a két típus.

* A tanulmány az NKFIH K_21 137659 számú, „A személyjelölés konstrukcióinak korpuszalapú, kognitív poétikai vizsgálata” pályázat, valamint a Digitális Örökség Nemzeti Laboratórium keretében készült.

A 2. részben röviden bemutatok több hazai példát a vershangzás gépi elemzésére, valamint a metrum gépi felismertetésének néhány olyan nemzetközi példáját is ismertetem, amelyek bizonyos elgondolásait felhasználtam a két algoritmus megtervezése során. A 3. részben ismertetem az időmértékes metrum felismerésére, illetve a megvalósuló versritmus szabályosságának a mérésére létrehozott algoritmus főbb lépéseit. Az algoritmust Python nyelvben implementáltam, és az ELTE Verskorpusz részét képező tizenegy, a 19. század közepén és második felében, valamint a 20. század első felében alkotó költő versein futtattam le. Az így nyert, időmértékes metrumra vonatkozó kvantitatív adatokat a 4. rész mutatja be. Az 5. rész az ütemhangsúlyos verselés gépi felismerését elvégző algoritmus főbb lépéseit veszi sorra. Az úgyszintén Python nyelvben implementált algoritmusnak az említett tizenegy szerző versein való lefuttatásával kapott előfordulási adatokat a 6. részben ismertetem. Végezetül a 7. részben bemutatom azt is, hogy az időmértékes és az ütemhangsúlyos verselést elemző két algoritmus kimenetének együttes figyelembevételével hogyan vizsgálható a szimultán versmetrum.

2. A vershangzás és a metrum gépi elemzésének néhány példája

Magyar nyelvű versek hangzástulajdonságainak az automatizált, valamilyen számítógépes programmal végzett elemzésére viszonylag korai példákat is találhatunk. Voigt Vilmos 1972-es tanulmánya mutatja be az első kísérletet magyar nyelvű versek számítógépes ritmuselemzésére. A létrehozott programmal három Szabó Lőrinc-szonettnek ismertették fel a megvalósuló időmértékes ritmusát, azaz a program meghatározta az egyes szótagok hosszúságát egy négyfokozatú skálán, és összegezte, illetve átlagolta az egyes sorokra és szótaghelyekre kapott értékeket.¹ Saját korát megelőzte Jékel Pál és Papp Ferenc 1974-ben kiadott műve, amely Ady összes versének az algoritmikus úton előállított fonémastatisztikai adatait tartalmazza, valamint az adatokat elemző bevezető tanulmányt.² Ugyancsak a korai példák között tartható számon Jékel Pál és Szuromi Lajos 1980-ban megjelent műve, amely Petőfi 300 verse esetében tartalmazza a szótagok részben gépi úton előállított különböző típusú nyomatékértékeit, valamint a nyomatékértékek automatikusan előállított összegzését és különféle statisztikáit.³ Szorosan kapcsolódik a tanulmány témájához Lesi Zoltán kutatása is, aki tudomásom szerint elsőként hozott létre olyan többfunkciós programot, amely magyar nyelvű

¹ Voigt Vilmos, „Számítógépes ritmuselemzési kísérlet,” *Irodalomtörténeti Közlemények* 76, 2. sz. (1972): 203–211.

² Jékel Pál és Papp Ferenc, *Ady Endre összes költői műveinek fonémastatisztikája* (Budapest: Akadémiai Kiadó, 1974). Jékel és Papp kutatása nemcsak abban volt újszerű, hogy a korhoz képest viszonylag nagy mennyiségű szöveget dolgoztak fel automatikusan, hanem abban is, hogy a szöveghasonlóság mérésére egy olyan matematikai modellt alkalmaztak, amely csak harminc évvel később vált általánosan elterjedté a számítógépes nyelvészetben. Ez a szövegek hasonlóságának vektortérben történő, koszinusz távolságon alapuló mérése.

³ Jékel Pál és Szuromi Lajos, *Petőfi metrumai* (Debrecen: Kossuth Lajos Tudományegyetem, 1980). A kutatás módszertani problémáira hívja fel a figyelmet: Vadai István, „Számítógép a verstan szolgálatában: Módszertani megjegyzések egy számítógépes ritmuselemzési kísérlet kapcsán,” *Irodalomtörténeti Közlemények* 88, 1. sz. (1984): 74–86.

versek rímképletének, alliterációinak és metrumának a gépi elemzésére is alkalmas.⁴ Érdemes utalni Labádi Gergely Berzsenyi verseiről írott tanulmányára is, amelyben automatizált módszerekkel végzett, a versek szókészletére és fonémajellemzőire (szóhosszúság, magánhangzók és mássalhangzók eloszlása) vonatkozó vizsgálatok eredményei szerepelnek.⁵ Bár nem tartalmaz automatikusan előállított információkat, de megemlítendő még a Horváth Iván főszerkesztésével létrehozott, *A régi magyar vers leltára a kezdetektől 1600-ig (Répertoire de la poésie hongroise ancienne)* nevű adatbázis, amelyben többek között a versek rímképletére és metrumára vonatkozó adatok is szerepelnek kereshető formában.⁶

A versritmus és versmetrum gépi elemzésére számos nemzetközi példát is találhatunk. Például a 16. és 17. századi spanyol szonettekkel álló *Corpus de Sonetos del Siglo de Oro (Corpus of Spanish Golden-Age Sonnets)* minden verssor esetében tartalmazza a hangsúlyos és hangsúlytalan szótagok gépileg létrehozott – és a kétséges esetekben manuálisan ellenőrzött – annotációit.⁷ A *Cseh Versskorpusz (Korpus českého verse)* közel 80000 verset tartalmaz számos automatikusan létrehozott annotációs réteggel.⁸ Többek között a spanyol korpuszhoz hasonlóan szerepelnek benne a hangsúlyos és hangsúlytalan szótagok jelölései, de ezek mellett tartalmazza a hangsúlyos és hangsúlytalan szótagok annotációiból absztrahált metrumkategóriákat is. Az itt bemutatandó két algoritmusnak is az a célja, hogy ne csupán a versritmust ismerje fel, hanem a ritmusból, azaz a hosszú és rövid szótagok, illetve a hangsúlyos és hangsúlytalan szótagok sorából valamilyen metrumra is képes legyen absztrahálni.

A metrum gépi elemzésének tehát jellemzően két fő lépése van. Az első a ritmus elemzése, azaz az adott metrikai rendszertől függően a hangsúlyos és hangsúlytalan vagy a hosszú és rövid szótagok megállapítása a verssorokban. A második lépés pedig a megállapított ritmusból a metrumra történő absztrahálás. Az angol (és más indoeurópai nyelvek, például a már említett spanyol és cseh) esetében a versmetrum

⁴ Lesi Zoltán, „Automatikus verselemzés tanuló algoritmusok alkalmazásával,” in Alexin Zoltán és Csentes Dóra, szerk., *IV. Magyar Számítógépes Nyelvészeti Konferencia*, 402–407 (Szeged: Szegedi Tudományegyetem Informatikai Tanszékcsoport, 2006); Lesi Zoltán, „Automatikus formai verselemzés,” *Alkalmazott Nyelvtudomány* 8, 1–2. sz. (2008): 197–208. A nyilvánosan nem elérhető programot bemutató két tanulmányban sajnos csak néhány rövid megjegyzés és egy képernyőfotó utal a metrum felismertetésére. Uo., 205.

⁵ Labádi Gergely, „Az olvasó gép: Berzsenyi Dániel versei távolról,” *Digitális Bölcsészet* 1. sz. (2018): 17–34, <https://doi.org/10.31400/dh-hun.2018.1.126>.

⁶ *Répertoire de la poésie hongroise ancienne*, hozzáférés: 2021.04.08, <https://f-book.com/rpha>.

⁷ Borja Navarro-Colorado, „A Computational Linguistic Approach to Spanish Golden Age Sonnets: Metrical and Semantic Aspects,” in Anna Feldman, Anna Kazantseva, Stan Szpakowicz et al., eds., *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, 105–113 (Denver: Association for Computational Linguistics, 2015), <https://doi.org/10.3115/v1/W15-0712>; Borja Navarro-Colorado, Marí Ribes Lafoz and Noelia Sánchez, „Metrical Annotation of a Large Corpus of Spanish Sonnets: Representation, Scansion and Evaluation,” in Nicoletta Calzolari, Khalid Choukri, Thierry Declerck et al., eds. *Proceedings of the 10th Edition of the Language Resources and Evaluation Conference (LREC 2016)*, 4360–4364 (Portorož: ELRA, 2016).

⁸ Petr Plecháč and Robert Kolár, „The Corpus of Czech Verse,” *Studia Metrica et Poetica* 2, 1. sz. (2015): 107–118, <https://doi.org/10.12697/smp.2015.2.1.05>; Robert Ibrahim and Petr Plecháč, „Toward Automatic Analysis of Czech Verse,” in Barry P Scherr, James Baily and Evgeny V. Kazantsev, eds., *Formal Methods in Poetics: A Collection of Scholarly Works Dedicated to the Memory of Professor M.A. Krasnoperova*, 295–305 (Lüdenscheid: RAM, 2011).

nem a görög, a latin és a magyar időmértékes metrumra is jellemző hosszú és rövid szótagok, hanem a hangsúlyos és hangsúlytalan szótagok váltakozására épül. Például az angol nyelvű versekben a jambus egy hangsúlytalan és egy hangsúlyos szótagból áll. Mivel az angolban nincs általános szabály arra vonatkozólag, hogy egy szó melyik szótagja hangsúlyos, gyakori megoldás, hogy az angol nyelvű versek metrumát elemző algoritmusokba a szavak szótagjainak a hangsúlyértékét jelölő lexikális adatbázist építenek be. Ezzel a megoldással él például a *Scandroid* nevű program, amely a verseket a jambikus és anapestikus kategóriák valamelyikébe sorolja be.⁹ A *Scandroid*-nál több metrumtípus felismerésére is képes *AnalysePoem* és *Zeuscansion* eszközök is egy programba beépített adatbázist használnak a hangsúlyos és hangsúlytalan szótagok megállapításához.¹⁰ Az előbbi esetben ez egy felhasználó által bővíthető lexikai adatbázis, az utóbbiban pedig két kiejtési szótár, amelyek segítségével a program a szótárakban nem szereplő szavak hangsúlyviszonyaira is kínál elemzést, oly módon, hogy az elemzendő szó írásmódjára legjobban hasonlító szótári szó elemzését választja ki.

Szemben az angollal, magyar nyelvű versek esetében a rövid és hosszú szótagok váltakozására épülő időmértékes metrum elemzéséhez nem szükséges beépíteni az algoritmusba hangzásjellemzőkre vonatkozó lexikai adatbázist, hiszen néhány általános, könnyen algoritmizálható szabály alapján meghatározható, hogy mely szótagok tekintendők rövidnek, és melyek hosszúnak (ezeket a szabályokat lásd a következő részben). A magyar nyelvű versek ütemhangsúlyos metrumának a gépi felismertetésekor is szerencséje van a kutatónak, hiszen, szemben az angollal, a magyar kötött hangsúlyú nyelv, azaz a szavak hangsúlyos szótagjai mindig az első szótagok. Ez tehát azt jelenti, hogy az ütemhangsúlyos versek ütemeinek az első, hangsúlyos szótagjainak a megállapításához sem szükséges szóhangsúlyokat jelölő lexikai adatbázist beépíteni a programba.

A metrum szabályalapú elemzésének második fő lépése általában egy számos allépésből álló absztrakciós eljárás, amelynek során a program a megállapított versritmus alapján meghatározza a metrumot (pl. hexameter, jambikus, felező nyolcas). Ez az absztrakciós eljárás sok esetben valamilyen pontozásos módszert használ, amelynek során a legtöbb pontot kapó metrumalternatívát választja ki a program. Például az említett *AnalysePoem* nevű program hat metrumot képes megállapítani. A program a domináns metrum megállapításához számos tesztet végez el egymást követően, amelynek kimeneteként megbízhatósági értéket (*confidence number*) rendel a megállapított metrumhoz. Minél nagyobb a megbízhatósági érték, annál szabályosabban valószínűsíti meg a versritmus a metrumot, egy bizonyos megbízhatósági érték alatt pedig a vers nem tekinthető az adott metrumhoz tartozónak.¹¹ Ehhez hasonló pontozásos módszert alkalmaztam én is a magyar időmértékes metrum felismerésére létrehozott

⁹ Charles O. Hartman, *The Scandroid. Version 1.1. [User guide]* (2005), hozzáférés: 2020.01.19, <http://charlesohartman.com/verse/scandroid/ScandroidManual.pdf>.

¹⁰ Marc R. Plamondon, „Virtual Verse Analysis: Analysing Patterns in Poetry,” *Literary and Linguistic Computing* 21, 1. sz. (2006): 127–141, <https://doi.org/10.1093/llc/fq1011>; Manex Agirrezabal, Aitzol Astigarraga, Bertol Arrieta et al., „Zeuscansion: A Tool for Scansion of English Poetry,” *Journal of Language Modelling* 4, 1. sz. (2016): 3–28, <https://doi.org/10.15398/jlm.v4i1.102>.

¹¹ Plamondon, „Virtual Verse Analysis: Analysing Patterns in Poetry”.

algoritmusban. Az úgyszintén említett *ZeusScansion* nevű eszköz a versre általánosan jellemző metrumot oly módon állapítja meg, hogy kiszámítja az egyes szótaghelyekre eső szótagok átlagos hangsúlyértékét, majd az ebből absztrahált ritmusképletnek több alternatív, verslábakra tagolt elemzését is megadja, amelyekből egy pontozási rendszer alapján választja ki a legmegfelelőbbet (az algoritmus azonos szótagszámú sorokon működtethető).¹² A hangsúlyos szótagok egyes szótaghelyekre eső arányainak a számítását én is alkalmaztam a magyar ütemhangsúlyos metrum felismerésére létrehozott algoritmusban.

Bár a fenti, angol nyelvű versek metrumát elemző algoritmusokból bizonyos ötleteket, megközelítési módokat át tudtam venni, hangsúlyozandó, hogy a magyar időmértékes és ütemhangsúlyos versmetrum különbözik az angol és a többi ma is beszélt indoeurópai nyelv nagy részének a versmetrumától, vagyis ezen algoritmusokat nem lehet alkalmazni a magyar nyelvű versekre. A magyar versmetrum szabályalapú gépi felismertetéséhez a magyar nyelv, illetve a magyar időmértékes és ütemhangsúlyos metrum speciális jellemzőit figyelembe vevő algoritmusok kidolgozása szükséges. Az alábbiakban bemutatandó két algoritmus létrehozása során elsődleges verstani kézikönyvként a Szepes Erika és Szerdahelyi István által írt *Verstan* című összefoglaló monográfiát használtam.¹³

3. Az időmértékes metrum gépi felismertetésének lépései

A cél egy olyan algoritmus kialakítása volt, amely első lépésben a verssorokat, második lépésben pedig a verseket teszteli a négy időmértékes metrumra, a daktilikus, az anapesztikus, a jambikus és a trochaikus kategóriákra, mindegyik esetében egy 0 és 1 közötti szabályossági pontszámot rendelve először a verssorokhoz, majd pedig a versekhez. Minél nagyobb a tesztelt metrumra kapott szabályossági pontszám, annál inkább megvalósítja a vers az adott metrumot. A kapott szabályossági pontszámok alapján megtudhatjuk, hogy a négyből melyik metrumot valósítja meg leginkább az adott vers, illetve megadhatunk egy küszöbértéket is, amelyet ha nem halad meg a vers szabályossági pontszáma egyik tesztelt metrum esetében sem, akkor a vers nem sorolódik be egyik metrumkategóriába sem. Ez az eljárás tehát az adott metrumhoz tartozást nem igen/nem kérdésként, hanem fokozati jellemzőként ragadja meg, amely fokozati jellemző persze a küszöbértékek használatával átalakítható igen/nem bináris tulajdonsággá.¹⁴ Ennek az eljárásnak több előnye is van. Egyrészt az algoritmus futtatása során különböző küszöbértékekkel kísérletezhetünk, adott esetben a kutatási kérdés függvényében változtatva azt. Másrészt az időmértékes metrum vizsgálata nemcsak arra a kérdésre terjedhet ki, hogy hány darab jambikus, trochaikus stb. vers van egy adott korpuszban, hanem arra is, hogy egy adott metrumkategória megvalósulásaiaként felismert versek mennyire szabályosak, azaz milyen mértékben valósítják meg az adott metrum elvont ritmusképletét. Például a szabályossági pontszám alapján

¹² Agirrezabal et al., „*ZeusScansion: A Tool for Scansion of English Poetry*”.

¹³ Szepes Erika és Szerdahelyi István, *Verstan* (Budapest: Gondolat Kiadó, 1981).

¹⁴ Mindez persze nem jelenti azt, hogy ne lennének minimumfeltételei annak, hogy egy verssor egy adott metrumkategóriába besorolható-e.

vizsgálhatjuk azt, hogy a jambikus versek kötöttebb ritmusa hogyan lazult fel bizonyos szerzők vagy időszakok esetében.

A versek metrumának elemzése a verssorok ritmusának az elemzésére épül rá. A ritmus elemzésére a *hunpoem_analyzer-TEI* elnevezésű programot használtam, amelyet az ELTE Verskorpusz¹⁵ különböző, formailag egyszerűbben megragadható hangzsjellemzőinek az automatikus annotálásához írtam.¹⁶ Ennek a programnak az egyik funkciója a verssorok időmértékes ritmusának elemzése, azaz annak megállapítása, hogy a sorokban az egyes szótagok hosszúak vagy rövidek-e. A sorok hosszú és rövid szótagjainak az elemzése néhány egyszerű, a magyar verstanban jól ismert szabály alapján elvégezhető, így az előző részben bemutatott példáktól eltérően nem szükséges kiejtési szótárakat beépíteni az algoritmusba. Ezek a szabályok a következők:

- (1) a program rövid szótagként elemzi azokat a szótagokat, amelyekben rövid magánhangzó van, és közvetlenül a rövid magánhangzó után nem áll mássalhangzó, vagy csak egy rövid mássalhangzó áll;
- (2) a program hosszú szótagként elemzi azokat a szótagokat, amelyekben hosszú magánhangzó áll, valamint azokat a rövid magánhangzós szótagokat, amelyekben hosszú vagy egynél több mássalhangzó követi a magánhangzót;
- (3) a program figyelembe veszi azt a verstani szabályt is, miszerint a szó eleji mássalhangzó-torlódások (pl. *krákog*, *trottyos*, *strigula*) nem nyújtják meg az előtte lévő rövid magánhangzóra végződő szótagot, vagyis azok nem hosszúnak, hanem rövidnek számítanak.

Mivel a szótagok hosszúságának a megállapításában fontos információ az, hogy a magánhangzót egy vagy több mássalhangzó követi-e, szükséges volt a programba beépíteni annak vizsgálatát is, hogy egy két- vagy háromjegyű mássalhangzónak tűnő karaktersorozat valóban két- vagy háromjegyű mássalhangzónak, azaz egy fonémának tekintendő. Ehhez az e-magyar program morfológiai elemzőjét használtam.¹⁷ Amennyiben a két- vagy háromjegyű mássalhangzónak tűnő karaktersorozat közé morfémahatár esik, akkor az nem tekinthető egy fonémának. Az időmértékes metrum automatikus meghatározását elvégző algoritmus bemenetét a ritmuselemzés kimenete adja, amely minden verssor esetében egy 1 és 0 karakterekből álló karaktersor, amelyben a 0 karakter a rövid, azaz egymorás, az 1 pedig a hosszú, azaz kétmorás szótagokat reprezentálja (pl. „Nincsen apám, se anyám” – 1001001).

¹⁵ ELTE Verskorpusz, hozzáférés: 2021.09.01, <https://verskorpusz.elte-dh.hu/>.

¹⁶ Horváth Péter, „A verhangzás jellemzőinek automatikus feltárása József Attila verseiben,” *Digitális Bölcsészlet* 3. sz. (2020): 3–27, <https://doi.org/10.31400/dh-hun.2020.3.422>; Horváth Péter, „Az ELTE Verskorpusz automatikus annotációs eljárásai révén nyerhető kvantitatív adattípusok,” in Simon Gábor és Tolcsvai Nagy Gábor, szerk., *Nyelvtan, diskurzus, megismerés*, 313–332 (Budapest: Eötvös Kiadó, 2020).

¹⁷ Váradi Tamás, Simon Eszter, Sass Bálint et al., „Az e-magyar digitális nyelvfeldolgozó rendszer,” in Vincze Veronika, szerk., *XIII. Magyar Számítógépes Nyelvészeti Konferencia*, 49–60 (Szeged: Szegedi Tudományegyetem, Informatikai Intézet, 2017); Novák Attila, Rebrus Péter és Ludányi Zsófia, „Az emMorph morfológiai elemző annotációs formalizmusa,” in Vincze, *XIII. Magyar Számítógépes Nyelvészeti Konferencia*, 70–78; Indig Balázs, Sass Bálint, Simon Eszter et al., „emtsv – Egy formátum mind felett,” in Berend Gábor, Gosztolya Gábor és Vincze Veronika, szerk., *XV. Magyar Számítógépes Nyelvészeti Konferencia*, 235–247 (Szeged: Szegedi Tudományegyetem TTIK, Informatikai Intézet, 2019).

Mivel az időmértékes ritmus elemzését elvégző *hunpoem_analyzer-TEI* programnak az ELTE Verskorpusz annotációjában is szereplő kimenete 0 (rövid szótag) és 1 (hosszú szótag) karakterekből áll, az időmértékes ritmusból metrumra absztraháló algoritmus a metrumok tesztelése előtt átkonvertálja a beolvasott karaktorsorok 0 és 1 karaktereit a moraszámnak megfelelő karakterekké, azaz a 0 karaktereket 1 karakterekre, az 1 karaktereket pedig 2 karakterekre változtatja (pl. „Nincsen apám, se anyám” – 2112112). Ez az átalakítás megkönnyíti a moraszámokon alapuló verslábakra bontást.

A metrumok tesztelése során az algoritmus első lépésben verslábakra bontja a sorok ritmusát reprezentáló, 1 és 2 karakterekből álló karaktorsorokat. Ez az anapestikus és daktilikus, valamint a jambikus és trochaikus metrumok tesztelése esetében eltérő módon valósul meg. Az anapestikus és daktilikus metrumok esetében a teljes lábak mindig négymorásak, hiszen nemcsak az alaplábak, a daktilus (– uu), illetve az anapestus (uu –), hanem az ezeket gyakran helyettesítő spondeus (– –) és elvéve helyettesítő proceleuzmatikus (uuuu) is négymorás, azaz időértékük négy rövid szótag időértékét teszi ki. Az anapestikus és daktilikus metrumok tesztelése során tehát az algoritmus első lépésben megpróbálja a verssorok hosszú és rövid szótagjait reprezentáló, 1 és 2 karakterekből álló karaktorsorokat négy moránként verslábakra bontani, azzal az engedménnyel, hogy az utolsó láb lehet csonka láb vagy olyan teljes láb, ahol a négymorás érték úgy jön ki, hogy a sorban szereplő utolsó, rövid szótag konvencionálisan hosszú szótagot, vagy a sorban szereplő utolsó, hosszú szótag rövid szótagot helyettesít. Ez alapján tehát a négymorás metrumú verssorok utolsó lába lehet egy hosszú vagy rövid szótagból álló csonka láb, illetve egy négymorás teljes lábként funkcionáló, anapestusnak (uu –) számítható tribrachisz (uuu), egy úgyszintén teljes láb értékű, spondeusnak (– –) számítható trocheus (– u), valamint egy daktilusnak (– uu) számítható krétikus (– u –) és egy proceleuzmatikusnak (uuuu) számítható negyedik paión (uuu –). Ha a verssor négymorás lábakra való felbontása nem megvalósítható, akkor az azt jelenti, hogy a verssor nem lehet anapestikus vagy daktilikus, vagyis a sor szabályosságai pontszáma mind az anapestikus, mind a daktilikus metrum esetében 0 lesz.

Bár az itt bemutatandó algoritmusnak nem célja, hogy felismerje azokat a kötöttebb metrumképleteket, ahol nemcsak a sor végén, hanem a sor belsejében is szerepel csonka láb, a pentameter esetében mégis kivételt tettem, hiszen a pentameter a hexameter mellett a daktilikus metrumok legtipikusabb fajtája, a hexameter és a pentameter együttes előfordulása adja a magyar költészetben is gyakori disztichont. A hexametert az algoritmus a fent bemutatott módszer alapján be tudja sorolni a daktilikus kategóriába, hiszen a hexameterben hat darab négymorás teljes láb szerepel. A pentameterben azonban a harmadik versláb egy fél láb, egy darab hosszú szótag, amely után ráadásul cezúra, azaz szóhatár van.¹⁸ A daktilikus metrum tesztelésébe így külön beépítettem egy funkciót, amely a verssort megpróbálja pentameterként elemezni, és csak ezt követően, ennek sikertelensége esetén kezdi meg az algoritmus a sor ritmusát jelölő karaktorsor négymorás verslábakra szabdalását. A pentameter felismeréséhez az algoritmus egyrészt egy egyszerű illesztéses módszerrel megvizsgálja,

¹⁸ Szepes és Szerdahelyi két hémiepesz kólon kapcsolataként írja le a pentametert. Szepes és Szerdahelyi, *Verstan*, 219–220. Mivel az algoritmus elemzésének alapegysége a versláb, megmaradtam a pentameter hagyományos megközelítésénél.

hogyan az előzetesen listázott, pentameternek tekinthető ritmusképletek valamelyikének megfelel-e az adott sor ritmusa, majd pedig, amennyiben az illesztés sikeres volt, megvizsgálja, hogy a sorban a harmadik, csonka lábat követően új szó kezdődik-e. Ha mind a két feltételnek megfelel a sor, akkor az algoritmus a pentameternek megfelelő tagolással, azaz egy soron belüli fél lábbal osztja verslábakra a sort.

A jambikus és trochaikus metrumú versek esetében a teljes lábak általában két szótagosak. A lábak moraszámára nincsen megkötés: a két metrum alaplábát adó jambus (u –), illetve trocheus (– u) hárommorás, az ezeket helyettesítő spondeus (– –) négymorás, de úgyszintén megjelenhet helyettesítő lábként a kétmorás pirrichius (uu) is. A jambikus és trochaikus metrumok tesztelésénél az algoritmus első lépésben két szótagonként felosztja a verssorok hosszú és rövid szótagjait reprezentáló, 1 és 2 karakterekből álló karaktorsorokat. Az utolsó láb ebben az esetben is lehet egy szótagú csonka láb. Speciális esetekben megjelenhetnek a jambikus és trochaikus sorokban három szótagú verslábak is: jambust helyettesítő anapesztus (ciklikus anapesztus) és trocheust helyettesítő daktilus (ciklikus daktilus). Az ilyen sorokat az algoritmus nem tudja helyesen elemezni.

Az algoritmus második lépése annak ellenőrzése, hogy a verssor megfelel-e bizonyos, a tesztelt metrumhoz kapcsolódó minimumfeltételeknek. Az anapesztikus és daktilikus metrumok tesztelése esetében, amennyiben a sort fel lehetett osztani négymorás verslábakra, az algoritmus megvizsgálja, hogy a verslábak között van-e amphibrachisz (u – u). Az amphibrachisz helyettesítő lábként való megjelenése az anapesztikus vagy daktilikus metrumoknál nem jellemző, az amphibrachisz jelenléte így általában arra utal, hogy annak ellenére, hogy a verssort négymoránként verslábakra lehet tagolni, a sor anapesztikus vagy daktilikus metrumúként való elemzése rossz irány, és az sokkal inkább lesz jambikus vagy trochaikus (vagy pedig nem jellemző rá semmilyen időmértékes metrum). Amennyiben tehát szerepel amphibrachisz a vizsgált verssorban, a verssor szabályossági pontszáma 0 lesz mind az anapesztikus, mind a daktilikus metrum tesztelése esetén.

A jambikus és trochaikus metrumok tesztelése esetében a verssor két szótagonként történő verslábakra osztását követően az algoritmus megvizsgálja a sor utolsó teljes lábát. A jambikus metrum minimumfeltétele ugyanis az, hogy amennyiben a sor teljes lábra végződik, az utolsó láb jambus (u –) vagy pirrichius (uu), amennyiben csonka lábra végződik, akkor pedig az utolsó előtti lábnak, azaz az utolsó teljes lábnak jambusnak kell lennie. A trochaikus metrum minimumfeltétele a sor teljes lábra végződése esetén az utolsó láb trocheus (– u) vagy spondeus (– –) volta, csonka lábra végződés esetén pedig az utolsó előtti láb, azaz az utolsó teljes láb trocheus volta. A jambikus vagy trochaikus kategóriába való besorolásnak e minimumfeltételei követik a verstanoknak azt az elképzelését, miszerint a jambikusság, illetve trochaikusság elsődleges kritériuma az utolsó teljes láb jambus vagy trocheus volta, ahol a jambust konvencionálisan pirrichius, a trocheust pedig konvencionálisan spondeus helyettesítheti, amennyiben az utolsó teljes lábat nem követi egy további csonka láb. Ha a verssor a tesztelt jambikus vagy trochaikus metrum e minimumfeltételeit nem teljesíti, akkor a szabályossági pontszáma a sornak az adott metrum esetében 0 lesz.

Az alábbiakban felsorolásszerűen összegzem, hogy az algoritmus a négy metrum esetében milyen minimumfeltételeket érvényesít. Amennyiben a verssor nem felel

meg a tesztelt metrum alább megadott minimumfeltételeinek, akkor a verssornak az adott metrumra vonatkozó szabályossági pontszáma 0.

Anapesztikus és daktilikus metrum minimumfeltételei:

- A verssor teljes lábai négymorásak.
- A verslábak között nincs amphibrachisz.

Jambikus metrum minimumfeltételei:

- Ha a verssor teljes lábra végződik, akkor az utolsó láb jambus vagy pirrichius, ha csonka lábra végződik, akkor az utolsó előtti láb, azaz az utolsó teljes láb jambus.

Trochaikus metrum minimumfeltételei:

- Ha a verssor teljes lábra végződik, akkor az utolsó láb trocheus vagy spondeus, ha csonka lábra végződik, akkor az utolsó előtti láb, azaz az utolsó teljes láb trocheus.

Amennyiben az elemzett verssor megfelel a tesztelt metrum minimumfeltételeinek, akkor az algoritmus kiszámolja a sor szabályossági pontszámát, azaz megad egy 0 és 1 közötti értéket, amely azt jelzi, hogy az adott sor ritmusa milyen mértékben felel meg a tesztelt metrumnak. Minél közelebb van a szabályossági pontszám az 1-hez, a tesztelt metrum szempontjából annál szabályosabb a sor. A szabályossági érték számolásának alapja a verslábak pontozása. Minden teljes láb 0-tól 4-ig terjedő pontszámot kap attól függően, hogy melyik metrumra teszteli az algoritmus a verssort. Négy pontot azok a verslábak kapnak, amelyek a tesztelt metrumkategóriának az alaplábai. A daktilikus metrum esetében a daktilus, az anapesztikus metrum esetében az anapesztus, a jambikus metrum esetében a jambus, a trochaikus metrum esetében pedig a trocheus. Két pontot kap a spondeus, amely mind a négy metrum esetében konvencionálisan használható helyettesítő láb. Egy pontot kapnak a kevésbé konvencionális helyettesítő lábak: a daktilikus és anapesztikus metrum tesztelése esetén a proceleuzmatikus, a jambikus és trochaikus metrum tesztelése esetén pedig a pirrichius. Nulla pontot kapnak a tesztelt metrum alaplábalal ellentétes lábak, azaz daktilikus metrum tesztelése esetén az anapesztus, anapesztikus metrum tesztelése esetén a daktilus, jambikus metrum tesztelése esetén a trocheus, trochaikus metrum tesztelése esetén pedig a jambus. A sorvégi fél lábakat, illetve pentameter esetében a sor végi és soron belüli fél lábat az algoritmus nem veszi figyelembe a sorok szabályossági pontszámainak a számítása során. Az algoritmus minden sor esetében összeadja a verslábak pontszámait, majd az összeget elosztja a sorban szereplő teljes lábak számának a négyszeresével. Az így kapott pontszám a sornak egy adott metrumra vonatkozó szabályossági pontszáma. Az algoritmus tehát az alábbi képlet alapján számítja ki a sorok szabályossági pontszámát.

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

s = a verssor szabályossági pontszáma

a = alaplábak száma

b = elsődleges helyettesítő lábak száma

c = másodlagos helyettesítő lábak száma

d = teljes lábak száma (d = a + b + c)

Ahogy azt fentebb már írtam, a sorvégi rövid szótagok konvencionálisan hosszúnak, a sorvégi hosszú szótagok pedig rövidnek is számíthatnak. Így az algoritmus a daktilikus és anapestikus metrum tesztelésénél az utolsó lábánál szereplő tribrachiszt (UUU) anapestusnak (UU -), a trocheust (- U) spondeusnak (- -), a krétikust (- U -) daktilusnak (- UU), a negyedik paiónt (UUU -) pedig proceleuzmatikusnak (UUUU) veszi, és ennek megfelelően pontozza őket. A jambikus és trochaikus metrum tesztelésekor pedig az utolsó lábánál szereplő pirrichiust (UU) jambusnak (U -), a spondeust (- -) pedig trocheusnak (- U) veszi, és a pontozás is ennek megfelelően történik. A fenti általános képletet az algoritmusban természetesen az egyes metrumokra specifikálva alkalmaztam, ezek a képletek az alábbi módon adhatók meg.

daktilikus:

$$s = \frac{\text{daktilusszám} \times 4 + \text{spondeusszám} \times 2 + \text{proceleuzmatikusszám}}{\text{verslábszám} \times 4}$$

anapestikus:

$$s = \frac{\text{anapestusszám} \times 4 + \text{spondeusszám} \times 2 + \text{proceleuzmatikusszám}}{\text{verslábszám} \times 4}$$

jambikus:

$$s = \frac{\text{jambusszám} \times 4 + \text{spondeusszám} \times 2 + \text{pirrichiusszám}}{\text{verslábszám} \times 4}$$

trochaikus:

$$s = \frac{\text{trocheusszám} \times 4 + \text{spondeusszám} \times 2 + \text{pirrichiusszám}}{\text{verslábszám} \times 4}$$

A képletekből láthatjuk, hogy amennyiben egy sor összes teljes lába a tesztelt metrum alaplába, például jambikus metrum tesztelése esetén a sor összes teljes lába jambus, akkor a számlálóba és a nevezőbe ugyanaz a szám kerül, és így a sor jambikus metrumra kapott szabályossági pontszáma 1 lesz. Minél több a verssorban a helyettesítő versláb (például jambikus metrum tesztelése esetén spondeus vagy pirrichius), illetve a metrum alaplábával ellentétes láb (például jambikus metrum tesztelése esetén trocheus), annál alacsonyabb szám kerül a számlálóba, és így annál alacsonyabb lesz a sor szabályossági pontszáma. Az alábbi *1. táblázat* néhány négy lából álló sornak a daktilikus és jambikus metrum tesztelése esetén előálló szabályossági pontszámát mutatja be. A szabályossági pontszámokat két tizedesjegyre kerekítettem.

1. táblázat. Különböző négy lábú ritmusok szabályossági pontszáma daktilikus és jambikus metrumok tesztelése esetén

Daktilikus		Jambikus	
Ritmusképlet	Szabályossági pontszám	Ritmusképlet	Szabályossági pontszám
- uu - uu - uu - uu	1,00	u - u - u - u -	1,00
- uu - uu - uu - -	0,88	u - u - - - u -	0,88
- uu - uu - - - -	0,75	- - u - - - u -	0,75
- - - uu - uu - -	0,75	- - - - - - u -	0,63
- - - uu - - - -	0,63	u - u - uu u -	0,81
uuuu - uu - - - -	0,56	- - u - uu u -	0,69
uuuu - uu uuuu - -	0,50	uu u - uu u -	0,63
- uu - uu - uu uu -	0,75	uu - - uu u -	0,50
- uu - - - uu uu -	0,63	- u - - uu u -	0,44
- uu uuuu - uu uu -	0,56	- u - u - u u -	0,25

A táblázatból látható, hogy minél kevésbé követi egy verssor ritmusa a tesztelt metrum elvont ritmusképletét, annál kisebb a szabályossági pontszám, egy bizonyos pontszám alatt pedig az adott sor már nem igazán tekinthető a daktilikus vagy jambikus metrum megvalósulásának.

Miután az algoritmus kiszámolta egy vers minden sorának a tesztelt metrumhoz kapcsolódó szabályossági pontszámát, megadja az egész versnek a tesztelt metrumra vonatkozó szabályossági pontszámát is. Ez egy egyszerű számítás alapján történik: az algoritmus összeadja az egyes sorok szabályossági pontszámait, majd az összeget elosztja az összes sor számával. A képlet tehát a következő (v a vers szabályossági pontszáma, s a verssorok szabályossági pontszáma, n a sorok száma):

$$v = \frac{\sum_{i=1}^n s_i}{n}$$

A számítás eredménye a verssorok szabályossági pontszámához hasonlóan egy 0 és 1 közötti szám, ahol az 1 érték a versnek a tesztelt metrum szempontjából teljesen szabályos voltát jelzi. Az 1 érték akkor jöhet ki, ha minden sor szabályossági pontszáma 1. Minél alacsonyabb a sorok szabályossági pontszáma, annál alacsonyabb lesz a vers egészére kapott szabályossági pontszám is.

Az algoritmus utolsó lépése az elemzett vers négy metrumra kapott szabályossági pontszámainak az összevetése. A verset az algoritmus abba a metrumkategóriába sorolja be, amelyre a legnagyobb szabályossági pontszám jött ki. Az algoritmus futtatása előtt megadható egy küszöb is. Ez egy olyan 0 és 1 közötti érték, amelyet meg kell haladnia a legnagyobb szabályossági pontszámnak ahhoz, hogy a verset az algoritmus valóban besorolja az adott metrumkategóriába. Mindenképpen érdemes küszöbértéket megadni, mivel az időmértékes metrum fokozati megközelítése miatt, még ha alacsony szabályossági pontszámmal is, de gyakorlatilag minden vers besorolódik a négyből

valamelyik metrumkategóriába.¹⁹ A küszöbérték megadásával biztosíthatjuk, hogy az algoritmus kimeneteként csak a valóban időmértékes metrumú versek sorolódnak be valamelyik időmértékes metrum kategóriájába.

Az alábbi felsorolás pontokba szedve összegzi a fent bemutatott, az időmértékes metrum gépi felismertetésére kidolgozott algoritmus főbb lépéseit:

1. A szótagok hosszúságának (rövid vagy hosszú) megállapítása
2. A vers tesztelése a négy metrumra
 - a. A verssorok verslábakra osztása a tesztelt metrumnak megfelelően
 - b. A tesztelt metrum minimumfeltételeinek ellenőrzése
 - c. A verssorok szabályossági pontszámának kiszámítása a tesztelt metrumnak megfelelően
 - d. A vers szabályossági pontszámának kiszámítása a sorok szabályossági pontszáma alapján
3. A vers négy metrumra kapott szabályossági pontszámának és a megadott küszöbértéknek az összevetése

A felsorolásban az (a)–(d) lépéseket beljebb szedtem. Ezzel azt próbáltam jelezni, hogy az (a)–(d) olyan, a (2) nagyobb lépés részét képező ismétlődő lépéssorozat, amelyet az algoritmus egy vers elemzésekor mind a négy metrum tesztelésénél elvégez. Ezen ismétlődő lépéssorozat eredményeképpen áll elő a verset jellemző, négy metrumra kapott négy szabályossági pontszám, amelyeket az algoritmus a (3) lépésben összevet egymással, illetve az előzetesen megadott küszöbértékkel.

4. Az időmértékes metrumot elemző algoritmus alkalmazása tizenegy költő versein

Az előző részben bemutatott algoritmust Python nyelvben implementáltam²⁰ oly módon, hogy annak bemenetét az ELTE Verskorpusz automatikus annotálására írt, úgyszintén Pythonban fejlesztett *hunpoem_analyzer-TEI* program TEI XML kimenete adja, amely többek között tartalmazza a verssoroknak a hosszú és rövid szótagokat megkülönböztető ritmuselemzését. Az időmértékes metrum felismerésére létrehozott programot tizenegy, a 19. század közepén, illetve második felében, valamint a 20. század első felében alkotó költőnek az ELTE Verskorpuszból²¹ vett versein futtattam le.

¹⁹ Megjegyzendő, hogy van elvi lehetősége annak, hogy egy vers mind a négy metrum esetében 0 szabályossági pontszámot kapjon. Ez abban az esetben következhetne be, ha a versnek egy olyan sora sincsen, amely a négy metrumból legalább az egyik minimumfeltételeit teljesítené.

²⁰ A kód a tanulmány mellékleteként letölthető: <https://doi.org/10.31400/dh-hun.2021.4.2361>.

²¹ Horváth Péter, Kundráth Péter, Indig Balázs et al., „ELTE Verskorpusz – a magyar kanonikus költészet 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); *ELTE Poetry Corpus*, hozzáférés: 2021.09.01, <https://github.com/ELTE-DH/poetry-corpus>. A level3 mappában szerepel a vershangzásra vonatkozó annotációkat is tartalmazó TEI XML kimenet.

A tizenegy költő születésük időrendjében a következő: Tompa Mihály, Arany János, Petőfi Sándor, Gyulai Pál, Vajda János, Reviczky Gyula, Komjáthy Jenő, Ady Endre, Babits Mihály, Kosztolányi Dezső, József Attila.

A 2. táblázat a vizsgált tesztkorpusz időmértékes metrumú verseinek az előfordulási adatait mutatja be. A metrumok szabályossági pontszámához kapcsolódó küszöböt 0,5 értékben adtam meg.

2. táblázat. Időmértékes metrumú versek gyakorisága (küszöbérték: 0,5)

Szerző	Összes	Időm.	Arány	Jambikus		Trochaikus	
				Előford.	Arány	Előford.	Arány
Tompa Mihály	491	429	0,87	307	0,72	116	0,27
Arany János	417	285	0,68	115	0,40	156	0,55
Petőfi Sándor	839	600	0,72	352	0,59	240	0,40
Gyulai Pál	156	112	0,72	43	0,38	69	0,62
Vajda János	199	156	0,78	85	0,54	69	0,44
Reviczky Gyula	335	331	0,99	271	0,82	57	0,17
Komjáthy Jenő	246	213	0,87	188	0,88	21	0,10
Ady Endre	1116	369	0,33	260	0,70	109	0,30
Babits Mihály	514	314	0,61	247	0,79	59	0,19
Kosztolányi Dezső	630	534	0,85	503	0,94	26	0,05
József Attila	599	309	0,52	249	0,81	50	0,16

A táblázat második oszlopában szerepel a szerzőkhöz tartozó versek száma, a harmadik oszlopban pedig azt tüntettem fel, hogy ebből a program mennyit elemzett időmértékes metrumúnak, azaz hány vers szabályossági pontszáma haladta meg a 0,5 értéket a négyből legalább az egyik metrum esetében.²² A negyedik oszlop az időmértékesként elemzett versek arányát mutatja be az összes vershez képest. Az utolsó négy oszlop a jambikusként és trochaikusként felismert versek számát és az időmértékesként felismert versekhez viszonyított arányát mutatja be. A táblázatból látható, hogy az elemzőprogram jól meg tudta ragadni azt az ismert tendenciát, hogy míg a 19. század első harmadában született szerzők (Tompa, Arany, Petőfi, Gyulai, Vajda) nagy arányban írtak trochaikus metrumú verseket – Aranytól és Gyulainál az így elemzett versek meghaladják az időmértékesként elemzett versek 50%-át –, addig a 19. század utolsó harmadától a trochaikus metrum visszaszorul, a jambikus metrumú versek aránya pedig még nagyobb lesz. A jambikus versek Kosztolányinál fordulnak elő a legnagyobb arányban: az időmértékesként azonosított versek 94%-át ismerte fel jambikusként a program. A táblázat harmadik oszlopából azt is láthatjuk, hogy a program Reviczkynek szinte az összes versét időmértékesként elemezte, ezzel szemben Ady esetében mindössze a versek 33%-át ismerte fel időmértékesként.

²² Ha egy vers több metrumra is pontosan ugyanazt a szabályossági pontszámot kapja, a program nem sorolja be a verset egyik időmértékes metrumkategóriába sem, és azt az általános „időmértékes” kategóriába sem sorolja be. Ez az eljárás a kapott előfordulási adatokat csak minimálisan módosítja. 0 küszöbérték esetén 10 vers kapott egynél több metrumra azonos szabályossági pontszámot, 0,3 küszöbérték esetén 5 vers, 0,5 küszöbérték esetén pedig nem volt ilyen vers.

A programot lefuttattam 0,3 küszöbértékkel is. Az így nyert előfordulási adatokat a 3. táblázat mutatja be. Látható, hogy ezzel a küszöbértékkel már Ady verseinek a jelentős részét is időmértékesként ismeri fel a program. Mindez azt mutatja, hogy Ady versritmusa a tesztkorpuszban szereplő többi költő versritmusához képest kevésbé szabálykövető, vagy legalábbis nem azokat a szabályokat követi, mint a többi szerző versritmusa, és amelyek a programban az időmértékes metrum tipikus jellemzőiként implementálva lettek.²³

3. táblázat. Időmértékes metrumú versek gyakorisága (küszöbérték: 0,3)

Szerző	Összes	Időm.	Arány	Jambikus		Trochaikus	
				Előford.	Arány	Előford.	Arány
Tompa Mihály	491	483	0,98	326	0,67	150	0,31
Arany János	417	396	0,95	136	0,34	226	0,57
Petőfi Sándor	839	791	0,94	416	0,53	351	0,44
Gyulai Pál	156	148	0,95	49	0,33	96	0,65
Vajda János	199	196	0,98	99	0,51	95	0,48
Reviczky Gyula	335	334	1,00	271	0,81	60	0,18
Komjáthy Jenő	246	243	0,99	198	0,81	36	0,15
Ady Endre	1116	910	0,82	694	0,76	209	0,23
Babits Mihály	514	429	0,83	295	0,69	117	0,27
Kosztolányi Dezső	630	606	0,96	535	0,88	61	0,10
József Attila	599	504	0,84	366	0,73	126	0,25

A 4. táblázat az anapestikusként, illetve daktilikusként elemzett versek számát mutatja be 0,5 és 0,3 küszöbértékkel. Babitsnál 0,5 küszöbértékkel a következő verseket ismerte fel a program daktilikusként: *Extasis*, *A sziget nem elég magas*, *Ekloga*, *Klasszikus álmok*, *Új Leoninusok*, *Vágy ered*, *Május huszonhárom Rákospalotán*, *Márciusi reggelen*. Ezek között a versek között disztichonok is vannak (*Új Leoninusok*, *Május huszonhárom Rákospalotán*). A disztichon egy hexameter és egy pentameter váltakozására épül. Mivel a programba külön beépítettem a pentameter felismerésének funkcióját, a disztichonban írt verseket a program fel tudja ismerni daktilikusként.

4. táblázat. Anapestikus és daktilikus versek előfordulása (küszöbérték: 0,5 és 0,3)

Szerző	Anapestikus		Daktilikus	
	> 0,5	> 0,3	> 0,5	> 0,3
Tompa Mihály	3	3	3	4
Arany János	1	8	13	26
Petőfi Sándor	5	12	3	12
Gyulai Pál	0	3	0	0

²³ Ady speciális versritmusára több szerző is felhívta már a figyelmet, például: Horváth János, *Rendszeres magyar verstan* (Budapest: Akadémiai Kiadó, 1969), 164–165; Gáldi László, *Ismerjük meg a versformákat!* (Budapest: Móra Ferenc Ifjúsági Könyvkiadó, 1961), 103–106; Vargyas Lajos, *Magyar vers – magyar nyelv: Verstani tanulmány* (Budapest: Szépirodalmi Könyvkiadó, 1966), 158–165.

Vajda János	1	1	1	1
Reviczky Gyula	2	2	1	1
Komjáthy Jenő	0	1	4	8
Ady Endre	0	3	0	4
Babits Mihály	0	7	8	10
Kosztolányi Dezső	3	6	2	4
József Attila	1	2	9	10

Mivel a program nem csupán besorolja egy metrumkategóriába az időmértékesként felismert verset, hanem minden vers esetében egy szabályossági pontszámot is megad, vizsgálhatjuk azt is, hogy az egyes szerzők esetében milyen eltéréseket mutatnak a szabályossági pontszámok középértékei. Az 5. táblázat a 0,5 küszöbértékkel jambikusként elemzett versek szabályossági pontszámainak az átlagát és mediánját mutatja be. Minél nagyobb egy szerző esetében az átlag, illetve a medián, annál szabálykövetőbben alkalmazta az adott szerző a jambikus metrumot. Látható, hogy a tesztkorpusz szerzői közül Reviczky és Komjáthy esetében a legmagasabbak ezek a középértékek, vagyis az ő jambikus versritmusuk követte leginkább a jambikus metrum elvont ritmusképletét. A legalacsonyabb középértékeket Adynál találjuk, ami megint csak megerősíti azt, hogy Ady versritmusa elüt a tesztkorpuszban szereplő többi szerző versritmusától, kevésbé követte azokat a szabályokat, amelyeket a többiek.

5. táblázat. Jambikus versek szabályossági pontszámainak középértékei (küszöbérték: 0,5)

Szerző	Átlag	Medián
Tompa Mihály	0,72	0,73
Arany János	0,71	0,72
Petőfi Sándor	0,68	0,69
Gyulai Pál	0,67	0,66
Vajda János	0,65	0,65
Reviczky Gyula	0,77	0,78
Komjáthy Jenő	0,77	0,77
Ady Endre	0,58	0,57
Babits Mihály	0,67	0,68
Kosztolányi Dezső	0,74	0,75
József Attila	0,63	0,61

5. Az ütemhangsúlyos metrum gépi felismertetésének lépései

Az alábbiakban bemutatandó második eljárás célja az egyszerűbb, aaaa, illetve abab szerkezetű ütemhangsúlyos metrumú versek felismerése, tehát azon ütemhangsúlyos versek felismerése, amelyeknek minden sorára ugyanaz az ütemosztás jellemző, vagy pedig két típusú, a versszakok páratlan és páros soraira jellemző ütemosztás váltogatja

egymást. Az ütemhangsúlyos verselés alapja az ütemeleji hangsúlyos szótagok szabályszerű visszatérése. A magyar kötött hangsúlyú nyelv, azaz a szavaknak – speciális helyzetektől eltekintve – csupán az első szótagja lehet hangsúlyos, a többi szótag pedig hangsúlytalan. Bár a különféle egy szótagos funkciószavak (névelők, kötőszavak stb.) általában hangsúlytalanok, ezek a magyar költészeti hagyományban adhatják az ütem elején álló hangsúlyos szótagokat, így az algoritmusnak nem szükséges megkülönböztetnie a funkciószavakat a többi szótól. Az ütemhangsúlyos verselés automatikus elemzését elvégző algoritmus első lépése a verssorok szó eleji hangsúlyos szótagjainak megállapítása. Ennek a lépésnek a kimenete az időmértékes ritmus elemzéséhez hasonlóan minden sor esetében egy 0 és 1 számokból álló karaktorsor. Ezúttal azonban az 1 a hangsúlyos, azaz a szó eleji szótagokat, a 0 pedig a hangsúlytalan, azaz a nem szó eleji szótagokat reprezentálja.

Az algoritmus a kapott hangsúlyértékek összesítése alapján állapítja meg, hogy az adott versre jellemző-e valamilyen ütemhangsúlyos metrum. Ennek alapja az, hogy egy vers sorainak a hangsúlyértékeit reprezentáló, 0 és 1 számokból álló karaktorsorok olyan mátrixot adnak ki, amelyben minden oszlop egy adott sorszámú szótaghelyre vonatkozik. Az oszlopokban szereplő, szó eleji szótagokat jelölő egyesek összeadásával megállapítható, hogy a vers adott sorszámú szótaghelyére esik-e ütemkezdet vagy nem.²⁴ Természetesen egy ütemhangsúlyos vers esetében is vannak olyan verssorok, amelyben valamelyik ütem elejére nem esik szó eleji szótag. Emiatt az algoritmust úgy alakítottam ki, hogy a program futtatásának kezdetén a felhasználó megadhatja, hogy a verssorok minimum hány százalékában legyen egy adott szótaghelyen szó eleji szótag (azaz a mátrixban 1-es karakter) ahhoz, hogy a program ütemkezdetet állapítson meg. Én a programot 0,75 értékkel futtattam, vagyis a program egy adott szótaghelyen akkor állapított meg ütemkezdetet, ha legalább a sorok 75%-ában szó eleji szótag esett az adott szótaghelyre.

Az algoritmusba nincsen beépítve a hosszú szótagokra vonatkozó ütemalkotási licencia, miszerint az ütem eleji hangsúlyos szótagokat helyenként helyettesíthetik hosszú szótagok.²⁵ Ugyanígy nem építettem be az összetett szavakra vonatkozó ütemalkotási licenciát sem, amely szerint az ütem eleji hangsúlyos szótag az összetett szavak második vagy további összetételi tagjainak az első szótagja, illetve nem elváló igekötős igék esetében az igekötő utáni első szótag is lehet.²⁶ Az ütemalkotási licenciák

²⁴ Megjegyzendő, hogy már a Voigt-féle 1972-es első magyar gépi ritmuselemzés is ilyen mátrixok alapján számolta az egyes szótaghelyekre eső átlagos értékeket, azzal a különbséggel, hogy a szótagok hosszúságát jelölték a mátrixban egy négyfokozatú skálán. Voigt, „Számítógépes ritmuselemzési kísérlet”. Jékel és Szuromi ugyancsak hasonló módszert használt az egyes szótaghelyekre eső szótagnyomatékok átlagolására. Jékel és Szuromi, *Petőfi metrumai*.

²⁵ Szepes és Szerdahelyi, *Verstan*, 375–376.

²⁶ Uo., 373–375. Az algoritmusnak létrehoztam egy olyan változatát, amely az *e-magyar* morfológiai elemzése alapján figyelembe veszi az összetételi tagok első szótagjait is, de ezt a verziót végül elvettem, mivel az *e-magyar* (és minden más szabadon felhasználható, magyar nyelvre alkalmazható elemző) a szavak egyszerűsített morfológiai címkéi alapján végzi el az egyértelműsítést, azaz a több lehetséges morfoszintaktikai elemzésből az adott mondatkontextusban a legvalószínűbb kiválasztását. Ezeknek az egyszerű címkéknek azonban abban az esetben több összetett, a morfológiai felbontást is tartalmazó elemzés is megfeleltethető, ha a szótó felbontható összetételi tagokra. Ökölszabálynak jónak tűnt, hogy ha az egyértelműsített egyszerű címkének több morfológiai felbontás is megfeleltethető, akkor azt a morfológiai felbontást hasznosítsa az algoritmus, amelyben

figyelman kívül hagyását kompenzáltam azzal, hogy viszonylag alacsonyan, 75%-ban határoztam meg az ütem elejére eső szó eleji szótagok minimális arányát.

Az ütemek megállapításánál két megszorítást is érvényesítettem: (1) a második szótaghely nem lehet ütemkezdet; (2) egy ütem nem lehet hosszabb egy előre megadott szótagszámnál. Az (1) megszorítást azért volt szükséges bevezetni, mert a verssorok első szava sok esetben egy szótagú szó, ugyanakkor verstani szempontból nem lenne észszerű egy szótagos ütemeket feltételezni a sorok elején, főleg, hogy az itt szereplő egy szótagú szavak sok esetben hangsúlytalan kötőszavak vagy egyéb típusú funkciószavak.²⁷ A (2) megszorítás összhangban áll a magyar verstani kutatások azon gyakori megállapításával, miszerint egy bizonyos szótagszám felett nem beszélhetünk ütemről. Ugyanakkor ezt a maximális szótagszámot a felhasználó maga adhatja meg a program futtatásának kezdetén. Én a programot hat szótagos maximális ütemhosszúsággal futtattam. Az ütemek hosszúságára vonatkozó megszorításból az is következik, hogy az ütemekre megadott maximális szótagszámmal rendelkező vagy annál rövidebb verssorok a program elemzésében lehetnek ütemosztást nem tartalmazó önálló ütemek.²⁸

a legtöbb morfémára lett felbontva a szó, hiszen az összetételi tagok ebben az esetben állapíthatók meg. Ez a megoldás azonban jelentősen túlgenerálta az összetett szavakat, és például olyan szavak is összetett szavakként értelmeződtek, mint a *szerelem*, a *tanács* vagy a *babér* (*szerelem*, *tanács*, *babér*). Emellett ha az algoritmus minden igekötős ige esetében hangsúlyosnak venné az igekötő utáni első szótagot, akkor valószínűleg jelentősebben megnőne annak az esélye, hogy a program olyan helyen is ütemhatárt állapítson meg, ahol valójában nem kezdődik új ütem. Például, ha az említett Weöres-vers esetében az igekötős igék igekötő utáni szótagját is automatikusan hangsúlyosnak venné az algoritmus – azaz 1-es karakterrel reprezentálná –, akkor a vers harmadik szótagja esetében is ütemkezdetet állapítana meg, ami félrevinné a vers metrumának elemzését.

²⁷ Felvethető, hogy a sorok elején álló egy szótagos szavak felütéseknek tekintendők. Ugyanakkor az általam áttekintett verstani szakirodalomban csak megjegyzésszerű, meglehetősen bizonytalan, az algoritmizálhatóság szempontjából nem eléggé egyértelmű megállapításokat találtam az ütemhangsúlyos metrumú versekben megjelenő esetleges felütésekről, ráadásul a szerzők többnyire a jelenség kifejezetten ritka voltát hangsúlyozzák. Gáldi László az ütemelőzőnek a magyar verstani elemzésekben „a magyar versrendszertől *teljesen idegen* és zeneileg is aránylag modern fogalmához csak kivételesen” folyamodik (Gáldi, *Ismerjük meg a versformákat!*, 19). Szepes és Szerdahelyi úgyszintén kuriózumként tartja számon a jelenséget, amely csak „bizonyos – kivételes – esetekben” jelenik meg ütemhangsúlyos verselésünkben (Szepes és Szerdahelyi, *Verstan*, 371). Megjegyzendő az is, hogy felütések leginkább csak hangsúlytalan funkciószavak lehetnének, és szinte kizárólag olyan pozícióban, ahol közvetlenül utánuk nem egy hozzájuk hasonló hangsúlytalan funkciószó, hanem egy hangsúlyos első szótaggal rendelkező szó áll. Az itt bemutatott algoritmus ugyanakkor nem vizsgálja, hogy a versben szereplő szavak funkciószavak vagy nem. A Weöres-vers is jól példázza, hogy hiába van a sor elején sok egy szótagú szó, ellenkező a ritmusérzékünkkel, ha azokat megpróbálnánk felütésekként ritmizálni.

²⁸ A verstani szakirodalom alapján az korántsem egyértelmű, hogy mi lenne az ütemek maximális szótagszáma. Vargyas Lajos megközelítésében például az ütem maximálisan négy szótagos lehet, bár elemzéseiben kivételes esetekben szerepelnek négynél több szótagból álló ütemek is (Vargyas, *Magyar vers – magyar nyelv*, 9–14). Gáldi László úgyszintén alapvetően négy szótagban állapítja meg a maximális szótagszámot, de megjegyzi, hogy az kivételesen lehet öt is (Gáldi, *Ismerjük meg a versformákat!*, 27). Horváth János a szokványos és a gyors ütemtípust különíti el, az előbbi maximális szótagszáma négy, az utóbbié hat (Horváth, *Rendszeres magyar verstan*, 18–51). Arany János leírásában is az ütemek maximum négy szótagosak, ugyanakkor az ütemek nagyobb, sormetszetekkel elválasztott tagokat alkothatnak, amelyek maximum hét szótagosak lehetnek (Arany János, „A magyar nemzeti vers-idomról,” in Arany János, *Tanulmányok és kritikák I.* [második,

Az alábbi 6. táblázat Weöres Sándor *A medve töprengése* című, tizenkét hét szótagos sorból álló versének a szó eleji és nem szó eleji szótagjait mutatja be. Az 1 jelöli a szó eleji, a 0 pedig a nem szó eleji szótagokat.

6. táblázat. Weöres Sándor *A medve töprengése* című versének szó eleji szótagjai

	1. szótag	2. szótag	3. szótag	4. szótag	5. szótag	6. szótag	7. szótag
1. sor	1	1	1	0	1	1	1
2. sor	1	0	1	0	1	0	0
3. sor	1	0	0	1	1	0	0
4. sor	1	1	1	0	1	0	0
5. sor	1	1	0	0	1	0	0
6. sor	1	1	0	0	1	0	0
7. sor	1	1	0	0	1	1	0
8. sor	1	1	0	0	1	0	0
9. sor	1	0	0	1	0	0	1
10. sor	1	1	1	0	0	0	1
11. sor	1	1	1	0	1	0	0
12. sor	1	1	1	0	1	1	0
Összesen	12	9	6	2	10	3	3
Arány	100%	75%	50%	17%	83%	25%	25%

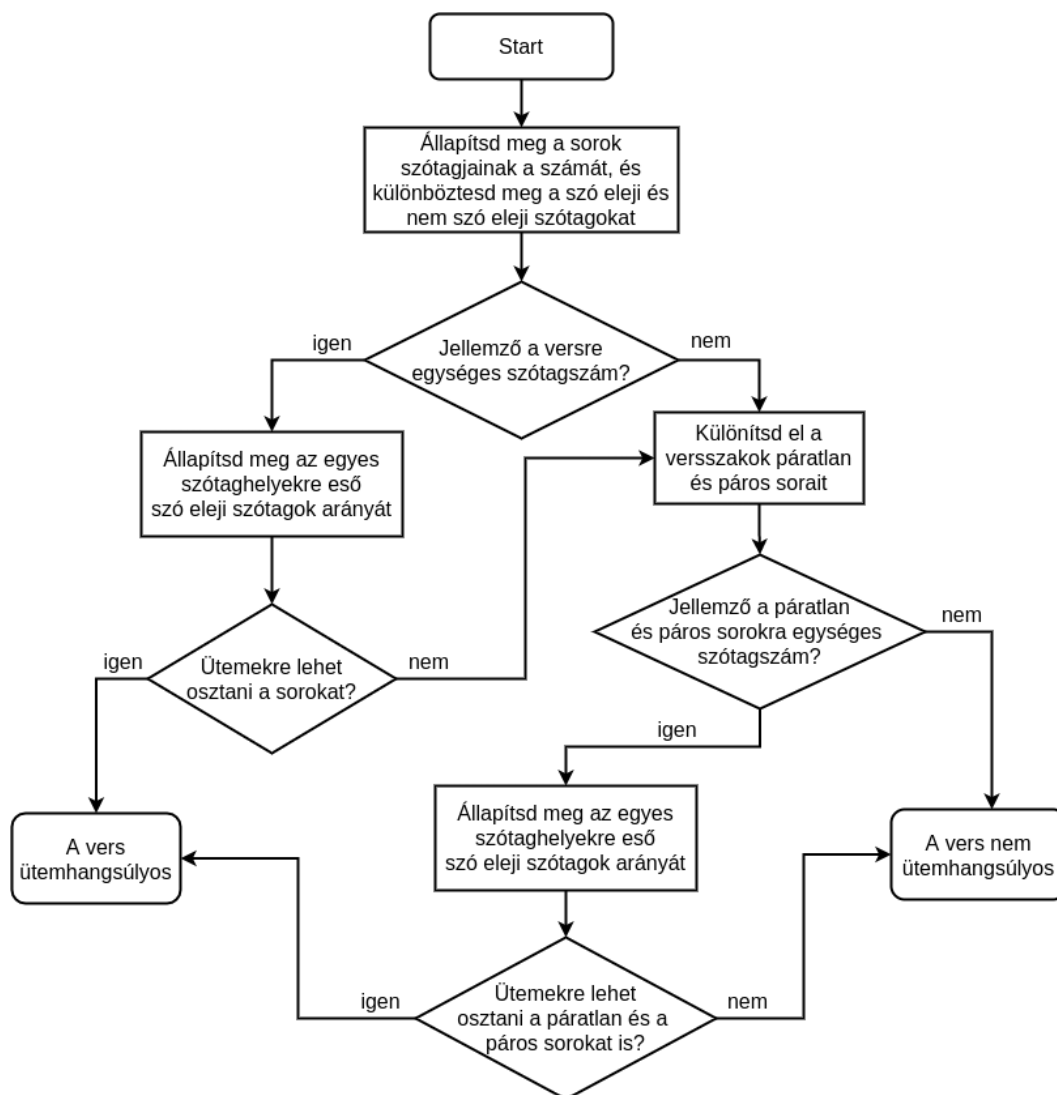
Ahogy a 6. táblázattal igyekeztem szemléltetni, az algoritmus által előállított, a vers egyes soraira vonatkozó, 1 és 0 karaktereket tartalmazó karaktorsorok olyan mátrixként értelmezhetők, amelyben a karaktorsorok adott sorszámú karakterei a verssorok adott sorszámú szótaghelyére vonatkoznak. Egy meghatározott szótaghelyen, azaz a mátrix egy adott oszlopában szereplő számok összeadásával, majd az így kapott összegnek a vers sorainak a számával való elosztásával megkaphatjuk, hogy az adott szótaghelyre milyen arányban esik szó eleji szótag. Amennyiben ez az arány eléri a megadott küszöbértéket – az általam használt beállítással 75%-ot –, akkor a szótaghelyen az algoritmus ütemkezdetet állapít meg. A 2. táblázatból például látható, hogy ezen eljárás alapján a program az első szótaghely mellett az ötödik szótaghelynél állapítana meg ütemkezdetet, hiszen ezen a szótaghelyen a sorok 83%-ában szó eleji szótag áll, vagyis a vers metruma kétütemű hetes, ötödik szótagra eső ütemkezdettel. A második szótaghelyen is a szótagok 75%-a szó eleji, az algoritmusba beépített megszorítás miatt azonban itt nem lehet ütemkezdet.

javított kiadás], 288–320 [Debrecen: Debreceni Egyetemi Kiadó, 2012]). Szepes és Szerdahelyi a négy szótagos ütemek mellett az öt és hat szótagos ütemeket a magyar ütemhangsúlyos metrumú versek leggyakoribb ütemtípusainak tartja, de hét, nyolc, kilenc, sőt tíz szótagon felüli ütemek létezése mellett is érvelnek, és nem állapítanak meg felső határt az ütemek maximális szótagszámára vonatkozóan (Szepes és Szerdahelyi, *Verstan*, 365–371). A program futtatásánál alkalmazott hat szótagos ütemhatár mögött elsősorban praktikus szempontok állnak. Bár én sem tartom kizártnak a hat szótagnál hosszabb ütemek létezését, az ütemhangsúlyos metrumú versek azonosítását minden bizonnyal félrevinné, ha például a program hét vagy nyolc szótagos ütemekből álló ütemhangsúlyos versekként elemezné a hét vagy nyolc szótagú sorokból álló, sormetszetet nem tartalmazó összes verset.

Ütemhangsúlyos verselést csak azokban a versekben lehet megállapítani, amelyekben a sorok szótagszáma valamilyen szabályosságot mutat, hiszen azonos ütembeosztású sorokról csak azonos szótagszámú sorok esetében beszélhetünk. Ennek folytán az ütemhangsúlyos metrumot elemző algoritmus a szó eleji és nem szó eleji szótagok meghatározása mellett azt is megvizsgálja, hogy a vers esetében megadható-e olyan szótagszám, amely általánosan jellemzi a vers sorait. Az algoritmust úgy alakítottam ki, hogy a program futtatásának kezdetén a felhasználó megadhatja, hogy a sorok minimum hány százalékának kell azonos szótagszámúnak lennie ahhoz, hogy ezt a szótagszámot a program a vers általános szótagszámának tekintse. Az algoritmus jelenlegi implementációjában ez mindig ugyanaz a 0 és 1 közötti érték, mint az ütemkezdetek megállapításához megadandó, szó eleji szótagokra vonatkozó minimum arány. Ahogyan arra már többször is utaltam, 0,75 értékkel futtattam a programot, vagyis a verssorok minimum 75%-ának kell azonos szótagszámúnak lennie ahhoz, hogy a program az ezen sorokra jellemző szótagszámot a vers alapszótagszámának tekintse, és e szótagszám alapján elvégezze az ütemhangsúlyos metrum elemzését a fent jelzett módon. Az algoritmus a megállapított alapszótagszámnál hosszabb vagy rövidebb sorok szó eleji szótagjait nem számolja bele az egyes szótaghelyekre eső szó eleji szótagok összegébe, ugyanakkor ezek a sorok is beleszámítanak a vers összes sorába az egyes szótaghelyekre eső szó eleji szótagok arányainak a számítása során. Ez tehát azt jelenti, hogy az alapszótagszámnál hosszabb vagy rövidebb sorokat az algoritmus úgy tekinti, mintha a mátrixban lenne egy csupa nulla karakterből álló sor.

Amennyiben nem jellemző a versre valamilyen általános szótagszám, vagy bár megállapítható általános szótagszám, de ütemhangsúlyosságot nem sikerült megállapítania az algoritmusnak, akkor az algoritmus különválasztja a versszakok páratlan és páros sorait, és megvizsgálja, hogy a szétválasztott sorokra külön-külön jellemző-e valamilyen általános, a sorok előzetesen megadott arányára – az általam használt beállításban a sorok minimum 75%-ára – jellemző szótagszám, illetve valamilyen ütemosztás. Ezzel az eljárással az abab szerkezetű ütemhangsúlyos metrumokat lehet felismertetni, azaz azokat az eseteket, amelyekben a versszakok páratlan és páros sorainak ütemosztása eltérő (pl. abab–abab–abab, aba–aba–aba, aba–abab–ab²⁹). Ha a program a szétválasztott sorok esetében sem tudott általános szótagszámot, majd ütemhangsúlyosságot megállapítani, akkor a verset nem ütemhangsúlyosként kategorizálja. Az ütemhangsúlyos verselést elemző algoritmus főbb lépéseit az 1. ábra mutatja be.

²⁹ A versszakoknak nem muszáj ugyanannyi sorból állniuk ahhoz, hogy az algoritmus soronként váltakozó ütemosztást állapítson meg.



1. ábra. Az ütemhangsúlyos verselést felismerő algoritmus főbb lépései

Az, hogy a program egy vers esetében nem tud megállapítani ütemhangsúlyos metrumot, nem feltétlenül jelenti azt, hogy a versre valóban nem jellemző valamilyen ütemhangsúlyos metrum, hiszen léteznek a fentieknél bonyolultabb szerkezetű ütemhangsúlyos versek is, amelyek felismerésére a bemutatott algoritmus nem képes.

6. Az ütemhangsúlyos metrumot elemző algoritmus alkalmazása tizenegy költő versein

A fent bemutatott, az ütemhangsúlyos metrum felismertetésére kitalált algoritmust úgyszintén implementáltam Python programnyelvben,³⁰ és lefuttattam ugyanazon a tesztkorpuszon, mint amelyen az időmértékes metrum felismerését elvégző programot. A 7. táblázat mutatja be a program futtatásával kapott előfordulási adatokat.

³⁰ A kód a tanulmány mellékleteként letölthető: <https://doi.org/10.31400/dh-hun.2021.4.2361>.

A második oszlopban szerepel a tesztkorpusz egyes szerzőihez tartozó összes vers száma, a harmadik oszlopban pedig a program által ütemhangsúlyosként felismert versek száma. Az utolsó oszlop mutatja be az ütemhangsúlyos versek arányát az adott szerző összes verséhez viszonyítva.

7. táblázat. Ütemhangsúlyos versek gyakorisága (megegyező szótagszámú sorok és ütemkezdetre eső szó eleji szótagok minimális aránya: 0,75; ütemek maximális szótagszáma: 6)

Szerző	Összes vers	Ütemhangsúlyos	Arány
Tompa Mihály	491	203	0,41
Arany János	417	217	0,52
Petőfi Sándor	839	405	0,48
Gyulai Pál	156	73	0,47
Vajda János	199	58	0,29
Reviczky Gyula	335	89	0,27
Komjáthy Jenő	246	45	0,18
Ady Endre	1116	238	0,21
Babits Mihály	514	114	0,22
Kosztolányi Dezső	630	82	0,13
József Attila	599	148	0,25

A program futtatásával kapott előfordulási adatok megfelelnek az irodalomtörténeti tudásunknak, miszerint a 19. század közepének népies, „magyaros” megszólalásmódjaival is kísérletező költői (Tompa, Arany, Petőfi, Gyulai) jóval nagyobb mértékben használtak ütemhangsúlyos metrumokat, mint az őket követő költők.

Az ütemhangsúlyosságot elemző program nem csupán azt mondja meg egy versről, hogy az ütemhangsúlyos-e vagy sem, hanem azt is megállapítja, hogy pontosan milyen ütemhangsúlyos metrumot követ az adott vers (azzal a már említett megszorítással, hogy a program csak aaaa, illetve abab típusú ütemhangsúlyos metrumokat képes felismerni). Ennek köszönhetően megvizsgálható az is, hogy egy adott szerző milyen típusú ütemhangsúlyos metrumokat használt a leggyakrabban. A 8. táblázatban Petőfi leggyakoribb tíz ütemhangsúlyos metrumát tüntettem fel a program elemzése alapján. A táblázatban látható két leggyakoribb metrum a kétütemű tízes és a felező nyolcas. Az előbbiben az első ütem négy, a második pedig hat szótagból áll. A táblázat negyedik helyén szereplő metrumnál csupán egy szám szerepel: 6. Ezek olyan versek, amelyekben a vers sorai hat szótagosak, a sorokon belül pedig nincsen ütemosztás. Mivel a programot hat szótagos maximális ütemhosszúsággal futtattam, a program a megegyező szótagszámú, hat szótagnál nem hosszabb verssorokból álló verseket akkor is ütemhangsúlyosként elemzi, ha a sorokban nincsen ütemosztás. Megemlítendő még az ötödik helyen álló 4 | 6 és 4 | 5 típusú metrum, ahol a kétféle kimenet arra utal, hogy a versszakok páratlan és páros sorai eltérő ütemosztást követnek. A páratlan sorok kétütemű tízesek, négy és hat szótagból álló ütemekkel, a páros sorok pedig kétütemű kilencesek, négy és öt szótagos ütemekkel.

8. táblázat. Petőfi leggyakoribb ütemhangsúlyos metrumai (megegyező szótagszámú sorok és ütemkezdetre eső szó eleji szótagok minimális aránya: 0,75; ütemek maximális szótagszáma: 6)

	Metrum	Előfordulás
1.	4 6	39
2.	4 4	38
3.	6 6	36
4.	6	35
5.	4 6 és 4 5	24
6.	5 5	20
7.	4 4 3	18
8.	4 4 és 4 3	13
9.	4 4 és 3	8
10.	4 4 és 6	6

7. A szimultán metrumú versek gépi felismertetése

A két algoritmus együttes alkalmazásával lehetőségünk van a szimultán metrumú versek gépi felismertetésére is. A szimultán versmetrum megközelítésében Szepes és Szerdahelyi értelmezését követem, amely szerint azokban az esetekben beszélhetünk szimultaneitásról, „amikor ugyanazok a költemények egyszerre felelnek meg valamilyen időmértékes és ütemhangsúlyos képletnek, tehát kétféle ritmusúak”.³¹ A szerzők hangsúlyozzák, hogy „a szimultán versek szövegeiben nem egy – a hangsúlyos és időmértékes metrumok összegzéséből adódó – sajátos metrum érvényesül, hanem e szövegekben a hangsúlyos és időmértékes nyomatekrend egymás mellett áll, s így nem összegezhető.”³² Az időmértékes és az ütemhangsúlyos metrumot elemző két algoritmus együttes futtatásával lehetőség van azon versek gépi felismertetésére, amelyekben egyszerre érvényesül az időmértékes metrum és az ütemhangsúlyos metrum, vagyis amelyek Szepes és Szerdahelyi definíciója alapján szimultán metrumú verseknek tekinthetők.

A 9. táblázat a tesztkorpusz szerzői esetében bemutatja az ütemhangsúlyosként felismert versek számát, a szimultánként (ütemhangsúlyosként és időmértékesként) felismert versek számát, valamint ez utóbbiak arányát az adott szerző összes verséhez képest. A táblázat utolsó két oszlopa azt mutatja be, hogy a szimultán versek közül a program elemzése alapján mennyi a jambikus és a trochaikus időmértékes metrumot érvényesítő vers. A program futtatása során az időmértékes metrumokra vonatkozó küszöböt 0,5 értékben adtam meg.

³¹ Szepes és Szerdahelyi, *Verstan*, 510.

³² Uo., 513. Szepes és Szerdahelyi ugyanakkor a szimultán ritmusú versek mellett megkülönbözteti a kevert ritmusú és az ötvözött ritmusú verseket is. Az előbbi esetében a versben ütemhangsúlyos és időmértékes részek váltogatják egymást soronként vagy néhány soros szakaszonként, az utóbbi esetben pedig a versnek csak egy metruma van, ez azonban az időmértékes és ütemhangsúlyos rendszer elveinek az összegyúrásából jön létre. Uo., 513–517. Ezek gépi azonosítása nem volt célom.

9. táblázat. Szimultán metrumú versek gyakorisága (megegyező szótagszámú sorok és ütemkezdetre eső szó eleji szótagok minimális aránya: 0,75; ütemek maximális szótagszáma: 6; időmértékes metrumok küszöbértéke: 0,5)

Szerző	Összes vers	Ütemhangsúlyos	Szimultán	Arány	Szimultán jambikus	Szimultán trochaikus
Tompa Mihály	491	203	165	0,34	80	84
Arany János	417	217	148	0,35	31	116
Petőfi Sándor	839	405	261	0,31	102	158
Gyulai Pál	156	73	51	0,33	10	41
Vajda János	199	58	42	0,21	14	28
Reviczky Gyula	335	89	86	0,26	48	38
Komjáthy Jenő	246	45	32	0,13	19	12
Ady Endre	1116	238	122	0,11	49	73
Babits Mihály	514	114	73	0,14	47	25
Kosztolányi Dezső	630	82	69	0,11	62	7
József Attila	599	148	90	0,15	61	28

A táblázatból látható, hogy a szimultán metrumú versek a 19. század első harmadában született, az ütemhangsúlyos metrumot amúgy is nagyobb mértékben használó költők esetében jelennek meg a legnagyobb arányban (Tompa, Arany, Petőfi, Gyulai). Ugyancsak feltűnő, hogy annak ellenére, hogy Aranyon és Gyulain kívül az összes többi vizsgált szerző nagyobb arányban használta a jambikus metrumot, mint a trochaikus metrumot (lásd a 2. táblázatban), a szimultán versek esetében az említett két szerző mellett Tompánál, Petőfinél, Vajdánál és Adynál is nagyobb számban jelennek meg a trochaikus metrumú versek, mint a jambikusak. Ez összefügghet azzal az elképzeléssel, miszerint az ütemhangsúlyos metrum természeténél fogva vonzza a trochaikus lüktetést.³³

8. Összegzés

A tanulmány magyar nyelvű versek metrumának gépi felismertetésére mutatott be két szabályalapú algoritmust. Az első algoritmus célja a versek besorolása a négy nagy időmértékes metrumkategória, az anapestikus, a daktilikus, a jambikus vagy a trochaikus metrumok valamelyikébe. Az algoritmus először a verssorok, majd a versek szabályossági pontszámát számítja ki mind a négy metrumra vonatkozóan. A szabályossági pontszám mind a sorok, mind a versek esetében egy 0 és 1 közötti szám, amely azt jelzi, hogy a verssor, illetve a vers milyen mértékben valósítja meg az adott metrum elvont ritmusképletét. A versekhez rendelt szabályossági pontszám lehetővé teszi, hogy a kutatói elemzésből kizárjuk azokat verseket, amelyek ritmusa a felismert metrum tekintetében nem ér el egy meghatározott szabályossági szintet. A bemutatott másik algoritmus célja az aaaa, illetve abab szerkezetű ütemhangsúlyos

³³ Szepes és Szerdahelyi, *Verstan*, 267–269; Horváth, *Rendszeres magyar verstan*, 128–133; Vargyas, *Magyar vers – magyar nyelv*, 148–150.

metrumú versek felismerése, azaz az algoritmus azon verseknél állapít meg ütemhangsúlyos metrumot, amelyekben minden sorra ugyanaz az ütemosztás jellemző, vagy pedig két típusú, a versszakok páratlan és páros soraira jellemző ütemosztás váltogatja egymást. Ennek az algoritmusnak az alapja az egyes szótaghelyekre eső szó eleji szótagok arányának a kiszámítása.

A két algoritmust Python nyelvben implementáltam, és az ELTE Verskorpuszból kiszedett tesztkorpuszon futtattam le, amely tizenegy, a 19. század közepén és második felében, valamint a 20. század első felében alkotó költő verseit tartalmazta. A korpuszvizsgálatokban többek között bemutattam, hogy az időmértékes metrum gépi elemzése során a küszöbérték módosítása miképpen befolyásolja a kapott előfordulási adatokat, hogy a kimenetként kapott szabályossági pontszámok középértékei hogyan használhatóak fel a versritmusra irányuló kutatásokban, illetve hogy az ütemhangsúlyos metrumot elemző algoritmus hogyan használható az ütemhangsúlyos metrumtípusok gyakorisági listájának az előállítására. Azt is bemutattam, hogy a két algoritmus együttes használata miképpen járulhat hozzá a szimultán metrumú versek vizsgálatához.

Bár a két algoritmus csupán az egyszerűbb metrumok felismerésére képes, és az elemzés nem közelíti meg az emberi elemzés pontosságát és árnyaltságát, arra alkalmasnak tűnnek, hogy kvantitatív vizsgálatokhoz használható adatokat állítsanak elő. A metrumra vonatkozó, automatizáltan előállított kvantitatív adatok persze nem feltétlenül az olyan nagymértékben kanonikus szerzők esetében hozhatnak új eredményeket, mint akiket a tanulmány tesztkorpuszában is szerepeltettem, hiszen e szerzők nagy részének a versritmusával sokat foglalkozott az irodalomtörténet. A kanonikus szerzők szerepeltetése a tanulmányban elsősorban azt a célt szolgálta, hogy teszteljem, illetve bemutassam a kialakított két algoritmus használhatóságát: mivel a 19. és 20. század költészetét reprezentáló, a versmetrum kézi annotációit tartalmazó magyar verskorpusz nem létezik, az algoritmusok használhatóságát úgy tudtam valamelyest feltérképezni, hogy megnéztem, a kanonikus szerzőkön lefuttatott program kimenete összhangban van-e azzal, amit ezen költők versritmusa kapcsán lehetett tudni. Ugyanakkor a nem vagy kevésbé kanonikus költészet vizsgálatához a metrum gépi elemzése már jóval produktívabb módon járulhat hozzá, hiszen ez olyan, a költészeti kánonnál sokszorosán nagyobb szövegkorpusz, amelynek az elolvasására, illetve a szoros olvasás technikáin alapuló elemzésére nem igazán van lehetőség. A bemutatott két algoritmus természetesen több ponton továbbfejleszthető vagy akár teljesen újragondolható. A magyar időmértékes és ütemhangsúlyos metrum gépi felismertetésére minden bizonnyal számos további, más elveket és számításokat alkalmazó eljárás is kitalálható.

Two Computational Methods for Detecting Meter in Hungarian Poetry

This paper presents two methods for the machine recognition of meter in Hungarian poetry, one for quantitative and another for qualitative meter. The algorithm analysing the former classifies the poems as dactylic, anapestic, iambic or trochaic, and defines the degree of regularity of the rhythm in the poems. The algorithm detecting the Hungarian qualitative meter returns the number of syllables of each beat. The paper also demonstrates the applicability of the two algorithms by using a test corpus containing the poems of eleven authors written between 1838 and 1941.

Keywords:

quantitative meter, qualitative meter, machine analysis, corpus research, ELTE Poetry Corpus