

HORVÁTH PÉTER

Szintaktikai ismétlődés a magyar kanonikus költészetben¹

horvath.peter@btk.elte.hu
ORCID: 0000-0002-3517-5623

HELIKON

Syntactic Repetition in Canonical Hungarian Poetry

Abstract

This paper uses corpus linguistic methods to investigate syntactic repetition in canonical Hungarian poetry. The research corpus consists of grammatically and metrically annotated poems of 24 canonical Hungarian authors from the 19th and the first half of the 20th century. To measure the degree of syntactic repetition, the number of different part-of-speech 2-grams, 3-grams, and 4-grams is counted in different window sizes. By comparing the syntactic repetition scores of poems with the scores obtained for a reference corpus of Hungarian novels, the paper demonstrates that syntactic repetition is an important characteristic of canonical Hungarian poetry. The paper also investigates whether there is a correlation between syntactic repetition and phonological repetition in the poems. On the one hand, it argues that the increase in syntactic repetition in the 20th century compensates for the decrease in phonological repetition. On the other hand, it presents that qualitative meter poems based on bars have higher syntactic repetition scores than iambic poems based on feet. In addition, by calculating the overlap coefficient between the part of speech n-grams of lines, the paper shows that consecutive rhyming lines are more similar syntactically than consecutive non-rhyming lines.

Keywords: canonical Hungarian poetry, ELTE Poetry Corpus, ELTE Novel Corpus, syntactic repetition, poetic function, moving-average type-token ratio, part of speech n-gram, meter, rhyme

¹ A tanulmány az NKFIH K 137659 sz. pályázat (*A személyjelölés konstrukcióinak korpusz-alapú, kognitív poétikai vizsgálata*) keretében, illetve a Digitális Örökség Nemzeti Laboratórium támogatásával készült.

1. BEVEZETÉS

A ritmussal, a rímeléssel, az alliterációval és akár egyéb eszközökkel megvalósított fonológiai ismétlődés a költészet egyik jellegadó, bár nem feltétlenül szükséges tulajdonsága. A fonológiai ismétlődés mellett ugyanakkor az ismétlődés más típusai is fontos szerepet játszhatnak. A 18. század közepén, az óhéber költészetet vizsgálva már Robert Lowth rámutatott a párhuzamosságok egyéb típusaira.² Megfigyelte például, hogy az óhéber költészetben a párhuzamosság egyik gyakori megvalósítása egy korábbi, rendszerint az előző sor szintaktikai szerkezetének a megismétlése. Ahogyan arra Horváth Iván felhívta a figyelmet, a fonológiai ismétlődés mellett az ismétlődés egyéb típusait meghatározónak valló elképzelések a magyar verselméletben is már igen korán megjelentek.³ Ennek első példájaként Földi János 1792-es írására utal, amely Lowth elgondolását mutatja be.⁴ Ebben Földi amellett érvel, hogy az óhéber költészet fogalmi és szintaktikai párhuzamosságokra épülő verselése az időmértékes és a rímes verselésnél alkalmasabb technika, illetve hogy a Lowth által megállapított párhuzamosságok egyik típusa a latin nyelvű költészetben is fellelhető.⁵ Horváth Arany Jánosnak *A magyar nemzeti vers-idomról* című verstani munkáját is a Lowth-féle elgondolás folytatásának tekinti. Az ismétlődés szisztematikusabb vizsgálata azonban csak a 20. századi strukturalista irodalomtudományban vált hangsúlyosabbá. Jakobson például több írásában is a grammatikai ismétlődést és variációt bizonyos versek, illetve költészeti hagyományok egyik legfontosabb jellemzőjének tekinti.⁶

A tanulmány egyrészt arra keresi a választ, hogy kvantitatív, korpusznyelvészeti eszközökkel igazolható-e a grammatikai ismétlődés egy fajtájának, a Lowth által is vizsgált szintaktikai ismétlődésnek a kitüntetett szerepe a magyar kanonikus költészetben. A tanulmány emellett azt is vizsgálja, hogy a versekben megjelenő szintaktikai ismétlődés mértéke összefügg-e a versek metrumával és a rímeléssel. Ez utóbbi vizsgálatok tehát arra az általános kér-

² Robert LOWTH, *Lectures on the Sacred Poetry of the Hebrews*, ford. G. Gregory (Boston–New York: Codman Press, 1829), 154–166.

³ HORVÁTH Iván, „A grammatikai szemlélet kezdetei a magyar verselméletben (Földitől Aranyig)”, *Irodalomtörténeti Közlemények* 76, 3. sz. (1972): 290–306; lásd még Horváth Iván, *A vers* (Budapest: Gondolat Kiadó, 1991), 50–88.

⁴ FÖLDI János, „Elmélkedés a ’Sidó Vers-írásról”, *Magyar Hírmondó* 1, 2. köt. (1792): 905–914.

⁵ Uo., 910–911.

⁶ Roman JAKOBSON, „A grammatika poétikája és a poétika grammatikája”, in Roman JAKOBSON, *Hang – Jel – Vers*, 258–277 (Budapest: Gondolat Kiadó, 1969); Roman JAKOBSON, „Grammatikai párhuzamosság a népköltészetben”, in JAKOBSON, *Hang – Jel – Vers*, 347–371.

désre keresik a választ, hogy a szintaktikai ismétlődés és a fonológiai ismétlődés megvalósulása egymástól függetlenül vagy pedig egymással összefüggésben jön-e létre. Vagy még általánosabban megfogalmazva, arra próbálnak választ adni, hogy van-e valamilyen motivációs viszony a hangzás és a grammatika szintjén megvalósuló ismétlődés között. Mindezek mellett a tanulmány a szintaktikai ismétlődés időbeli alakulását is vizsgálja, amit összefüggésbe hoz a metrumnak, azaz a versekben megjelenő fonológiai ismétlődés egyik fő típusának az időbeli alakulásával.

Nem fonológiai jellemzők magyar nyelvű versekben való ismétlődését már vizsgálták automatizáltan, kvantitatív módszertannal: Seláf Levente és szerzőtársai 16. századi históriás énekekben karakter 2-gramok alapján vizsgálták a versekben megjelenő epikus formulákat, amelyek az értelmezésükben egy sornál nem hosszabb, szintaktikailag és lexikailag azonos vagy hasonló, sok esetben szó szerinti vagy közel szó szerinti ismétlődő szövegrészek.⁷ Az alább bemutatandó vizsgálatok az ismétlődés jelenségét absztraktabb szinten, a szintaxis szintjén próbálják megragadni. A vizsgálatok így eltérő módszertant kívánnak meg, hiszen az a cél, hogy a szintaktikai ismétlődésnek azokat az eseteit tegyük mérhetővé, amelyek nem a szó szerinti vagy közel szó szerinti ismétlődésből következnek. Ez utóbbiak itt inkább a mérést torzító, és ezért valamiképpen kiszűrendő tényezőkként jelentkeznek.

Az irodalomnak és más művészeti ágaknak is egyik meghatározó jellemzője az egyes alkotások között, illetve az egyes alkotásokon belül megvalósuló ismétlődés (az előbbire példa a műfaji konvenciók ismétlése).⁸ Ugyanakkor az ismétlődés nem csupán az irodalomnak, hanem magának a nyelvnek is az egyik alapjellemezője. Ugyanazok a szavak, szókapcsolatok akár egy rövidebb szövegen belül is gyakran többször megismétlődnek. Ha a szavaknál absztraktabb nyelvi kategóriák, például szófaji kategóriák szövegbeli előfordulását nézzük, akkor az ismétlődés értelemszerűen még szembetűnőbbé válik. Az ismétlődés nem csupán a nyelv lexikai és grammatikai szintjén, hanem a nyelv hangzásszintjén is megjelenik. A fonológiai ismétlődés olyan eljárásai, mint

⁷ SELÁF Levente, Petr PLECHÁČ, VIGYIKÁN Villő és KISS Margit, „A belső formulák a 16. századi históriás énekekben”, in *A históriás ének: Poétikai és filológiai kérdések*, szerk. SELÁF Levente (Budapest: Gépeskönyv, 2023), hozzáférés: 2025.10.14, <https://f-book.com/book/2023/A-historias-enek-kerdesek/index.php?chapter=8>. A cikkben szerepel egy részletes definíció arra vonatkozóan, hogy mit tekintenek a szerzők formulának.

⁸ Az ismétlődés fajtáiról és jelentőségéről például lásd SZEGEDY-MASZÁK Mihály, „A művészi ismétlődés néhány változata az irodalomban és a zenében”, in *Ismétlődés a művészetben: Tanulmányok*, szerk. HORVÁTH Iván és VERES András, 77–159 (Budapest: Akadémiai Kiadó, 1980).

a rím vagy az alliteráció, nemcsak irodalmi szövegekben, hanem a hétköznapi nyelvhasználatban is tetten érhetők.⁹ Mivel a nyelvre az ismétlődés, illetve annak itt vizsgált típusa, a szintaktikai ismétlődés általánosan jellemző, a versekben megjelenő szintaktikai ismétlődést önmagában nem vizsgálhatjuk, hiszen más típusú szövegekkel való összevetés nélkül nem tudjuk megmondani, hogy a versek esetében megállapított mértéke az ismétlődésnek a vers szöveg-típusához kapcsolható specifikusabb jellemző, vagy csupán a nyelv általános tulajdonságát tükrözi. Ezt a problémát egy regényeket tartalmazó referenciakorpusz használatával kezelem. A referenciakorpusz alapján már elhelyezhetjük a versek szintaktikai ismétlődésének a mértékét jelző értékeket.

A 2. rész röviden bemutatja a vizsgálatokhoz használt korpuszokat és a korpuszok felhasznált annotációs rétegeit. A 3. rész ismerteti a vizsgálatokban használt módszereket. A 4. rész mutatja be, hogy a vizsgálati korpuszban szereplő 24 kanonikus költőre milyen mértékben jellemző a szintaktikai ismétlődés jelensége. Az eredményekből kiderül, hogy a versekben általában nagyobb mértékben jelenik meg a szintaktikai ismétlődés, mint a regényekben, illetve az is kirajzolódik, hogy a 20. század első felének a szerzőinél többnyire nagyobb mértékben jelenik meg a szintaktikai ismétlődés, mint a 19. században alkotó szerzőknél. Ez utóbbi összefüggésbe hozható a versmetrum 20. századi fellazulásával: amellet érvelek, hogy a metrumban megnyilvánuló fonológiai ismétlődés mértékének a csökkenését a szintaktikai ismétlődés mértékének a növekedése kompenzálja. Az 5. és a 6. rész a fonológiai ismétlődés és a szintaktikai ismétlődés közötti további összefüggéseket vizsgálja. Az 5. rész bemutatja, hogy a szerzők többsége esetében az ütemhangsúlyos metrumú versekben a szintaktikai ismétlődés erőteljesebben jelenik meg, mint a jambikus metrumú versek esetében, ami az ütemek és a verslábak eltérő jellemzőivel magyarázható. A 6. rész pedig rámutat arra, hogy a közvetlenül egymást követő rímelő sorok szintaktikai szempontból jobban hasonlítanak egymásra, mint a közvetlenül egymás után álló nem rímelő sorok, ami a szintaxis szintjén érvényesülő absztrakt tartalmi jellemzők és a rímeléssel megvalósuló fonológiai jellemzők közötti ikonikus viszonyal magyarázható. A 7. rész röviden összegzi a vizsgálat eredményeit.

⁹ Például lásd Réka BENCZES, *Rhyme over Reason: Phonological Motivation in English* (Cambridge: Cambridge University Press, 2019), <https://doi.org/10.1017/9781108649131>.

2. A VIZSGÁLAT KORPUSZAI

A vizsgálat 24 magyar kanonikus költő összes versén alapul. A szerzők a 19. században és a 20. század első felében alkottak. A verseket a hozzájuk tartozó annotációkkal együtt az ELTE Verskorpusból szedtem ki.¹⁰ Az ELTE Verskorpuszban szereplő azon 19. és 20. századi szerzőknek a verseit vettem bele a vizsgálatba, akik legalább 100 verset írtak. A vizsgálati korpuszban szereplő költőket és a verseik számát az 1. táblázat mutatja be. A táblázatban külön oszlopokban feltüntettem a legalább 100, legalább 50 és legalább 25 tokenből álló versek számát is. A tokenszám a versekben található szavak és központosítások együttes számára utal. Azért tüntettem fel külön e három terjedelmi kategóriába tartozó versek számát, mert a szintaktikai ismétlődés mérését minimum 100, 50, illetve 25 tokenes ablakméretekkel végeztem el. (Ennek módszertanát a következő rész mutatja be.)

1. táblázat. Vizsgálati korpusz

Költő	Versek száma				Költő	Versek száma			
	Min. 100	Min. 50	Min. 25	Össz.		Min. 100	Min. 50	Min. 25	Össz.
Csokonai Vitéz Mihály	277	338	384	394	Komjáthy Jenő	205	220	236	246
Berzsenyi Dániel	104	122	130	136	Ady Endre	700	1075	1116	1116
Kisfaludy Károly	78	109	120	123	Kaffka Margit	94	101	102	102
Kölcsy Ferenc	79	129	145	150	Somlyó Zoltán	260	377	379	379
Vörösmarty Mihály	292	424	542	662	Juhász Gyula	571	1190	1251	1278
Arany János	252	315	375	417	Babits Mihály	431	502	512	514
Tómpa Mihály	436	474	490	491	Kosztolányi Dezső	405	582	615	630
Petőfi Sándor	529	738	823	839	Tóth Árpád	310	407	435	451
Madách Imre	208	245	283	319	Reményik Sándor	475	639	665	667
Gyulai Pál	114	152	156	156	József Attila	345	552	581	599
Vajda János	168	190	196	199	Dsida Jenő	583	987	1049	1060
Revczky Gyula	238	317	334	335	Radnóti Miklós	175	272	296	298

Az ELTE Verskorpusz tartalmazza a versek szövegeit és a szövegek automatikusan létrehozott annotációit. A korpuszban annotálva vannak a versek szerkezeti egységei, azaz a versszakok és a verssorok, a szavak grammatikai

¹⁰ HORVÁTH Péter, KUNDRÁTH Péter és INDIG Balázs, „ELTE Verskorpusz – a magyar kanonikus költészet gépileg annotált adatbázisa”, in *XVIII. Magyar Számítógépes Nyelvészeti Konferencia*, szerk. BEREND Gábor, GOSZTOLYA Gábor és VINCZE Veronika, 375–388 (Szeged: Szegedi Tudományegyetem TTIK, Informatikai Intézet, 2022). A korpusz innen letölthető: <https://github.com/ELTE-DH/poetry-corpus>.

jellemzői, valamint vershangzásra vonatkozó jellemzők. A szavak grammatikai annotációi az e-magyar programmal lettek létrehozva.¹¹ A korpusz többek között minden szó esetében szerepelteti a szavak lemmáját, azaz a szótári alakját, a szófaji címkéjét és a morfoszintaktikai jellemzőit a Universal Dependenciesnek a magyar nyelvre alkalmazott taxonómiája alapján.¹² A szintaktikai ismétlődés mérésének alapját ezek a szófaji és morfoszintaktikai jellemzők adják. A vershangzásra vonatkozó, úgyszintén automatizáltan létrehozott annotációkból a rímképletre¹³ és a metrumra¹⁴ vonatkozó annotációkat használja a vizsgálat. A versek négy időmértékes metrumtípus valamelyikébe, a daktilikus, az anapesztikus, a jambikus vagy a trochaikus metrumokba sorolódnak, és a metrumcímke mellett rendelkeznek egy 0 és 1 közötti szabályossági pontszámmal is, amely azt fejezi ki, hogy a vers ritmusa milyen mértékben realizálja a vers elvont metrumképletét. A vizsgálati kérdéstől függően 0,3 vagy 0,5 küszöbértéket használtam az időmértékes versek csoportjának kijelöléséhez. A korpuszban az időmértékes metrum mellett a versek esetleges ütemhangsúlyos metruma is fel van tüntetve.¹⁵

A vizsgálat referenciakorpuszát az ELTE Regénykorpusz adta, amely a vizsgálat idején 399 regényt tartalmazott a 19. századból és a 20. század első felé-

¹¹ VÁRADI Tamás, SIMON Eszter és SASS Bálint, „Az e-magyar digitális nyelvfeldolgozó rendszer,” in *XIII. Magyar Számítógépes Nyelvészeti Konferencia*, szerk. VINCZE Veronika, 49–60 (Szeged: Szegedi Tudományegyetem, Informatikai Intézet, 2017); NOVÁK Attila, REBRUS Péter és LUDÁNYI Zsófia, „Az emMorph morfológiai elemző annotációs formalizmusa,” in *XIII. Magyar Számítógépes Nyelvészeti Konferencia*, szerk. VINCZE Veronika, 70–78; INDIG Balázs, SASS Bálint és SIMON Eszter, „emtsv – Egy formátum mind felett,” in *XV. Magyar Számítógépes Nyelvészeti Konferencia*, szerk. BEREND Gábor, GOSZTOLYA Gábor és VINCZE Veronika, 235–247 (Szeged: Szegedi Tudományegyetem TTIK, Informatikai Intézet, 2019).

¹² VINCZE Veronika, SIMKÓ Katalin, SZÁNTÓ Zsolt és FARKAS Richárd, „Universal Dependencies and Morphology for Hungarian – and on the Price of Universality,” in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, 356–365 (Valencia, Spain: Association for Computational Linguistics, 2017), <https://doi.org/10.18653/v1/E17-1034>.

¹³ HORVÁTH Péter, „Egy automatikusan generált rímszótár fejlesztése és a magyar kanonikus költészet rímszavainak néhány jellemzője,” in *Új technológiákkal, új tartalmakkal a jövő digitális transzformációja felé. 32 Workshop: országos konferencia*, szerk. TICK József, KOKAS Károly és HOLL András, 77–84 (Budapest: HUNGARNET Egyesület, 2023), <https://doi.org/10.31915/NWS.2023.12>.

¹⁴ HORVÁTH Péter, „Két eljárás magyar nyelvű versek metrumának gépi felismertetéséhez”, *Digitális Bölcsészet* 4 (2021): 79–103, <https://doi.org/10.31400/dh-hun.2021.4.2361>.

¹⁵ Az időmértékes és ütemhangsúlyos metrum felismertetésének, valamint az időmértékes metrumhoz kapcsolódó szabályossági pontszám kiszámításának részleteit lásd HORVÁTH Péter, „Két eljárás...”

ből.¹⁶ A korpuszban a regények szövegei és szerkezeti annotációi mellett szerepelnek a szavak lemmájára, szófajára és morfoszintaktikai jellemzőire vonatkozó annotációk is. Ezek az ELTE Verskorpuszhoz hasonlóan az e-magyar programmal lettek létrehozva, ugyanúgy a Universal Dependencies címkekészlete alapján. A két korpusz azonos programmal, illetve azonos címkekészlettel megvalósított annotálása lehetővé teszi ezek kvantitatív alapú összevetését.

3. A VIZSGÁLAT MÓDSZEREI

A szintaktikai ismétlődés mérésének alapja a szövegben egymást követő szavak szófajai által alkotott szófajcsoportok ismétlődésének a mérése. Például, ha a névelő-melléknév-főnév szófajcsoport vagy az ige-névelő-főnév szófajcsoport egy szövegrészen belül többször előfordul, akkor az növeli a szintaktikai ismétlődésnek a mért értékét. Ezt a megközelítést tehát valójában pontosabb lenne úgy definiálni, mint amely a szöveg szintagmatikus tengelyén méri szófajok szomszédsági viszonyainak az ismétlődését. Ugyanakkor a kutatás abból indul ki, hogy szófajok közötti szomszédsági viszonyok ismétlődése sok esetben mondattani értelemben vett szintaktikai viszonyok, azaz függőségi viszonyok ismétlődésére utal. Ha pedig bizonyos morfoszintaktikai jellemzőkkel, például az esetalakkal kiegészítjük a szófaji címkéket, akkor a szófajok közötti szomszédsági viszonyok és a függőségi viszonyok ismétlődése közötti kapcsolat még közvetlenebbé tehető. Mindazonáltal hangsúlyozandó, hogy a szófajok szomszédsági viszonyain alapuló eljárás azokban az esetekben, ahol a függőségi viszonyok különböző szórendben is megjelenhetnek (például az ige és különböző funkciójú főnévi bővítményei között), a szintaktikai ismétlődést akkor képes csak hatékonyan megragadni, ha nem csupán a függőségi viszonyok, hanem (valamilyen szinten) a szórend terén is megvalósul az ismétlődés. Ez ugyanakkor ebben a vizsgálatban nem feltétlenül hátrány, hiszen versek esetében a szintaktikai szerkezet ismétlésének poétikai eszköze általában nem csupán a függőségi viszonyok, azaz az absztrakt szintaktikai szerkezet megismétlését, hanem a szórend (bizonyos szintű) megismétlését is megkívánja, mivel ez utóbbi nélkül az olvasó jó eséllyel nem érzékelné az ismétlődést.

Az alábbiakban részletesen bemutatom a kutatásban használt, a szintaktikai ismétlődés mérésére kidolgozott módszereket. A kialakított módszerek alapján megvalósított méréseket automatizáltan, programkód segítségével végeztem el.

¹⁶ A korpusz innen letölthető: <https://github.com/ELTE-DH/regenykorpusz>.

3.1. A szintaktikai ismétlődés mértékének mérése

A vizsgálat fő módszerét a típus-token arány mozgó átlagát mérő eljárás módosított, szófaji címkékre és n-gramokra alkalmazott verziója adja. A típus-token arány mozgó átlagának a mérését a szókincs változatosságának a mérésére dolgozták ki. Egy szöveg szókinszének a változatosságát sokféle módszerrel próbálták már mérni. Ezek közül a legegyszerűbb a típus-token arány megadása,¹⁷ azaz a szövegben szereplő különböző szavak számának az elosztása a szövegben szereplő összes szó számával. Nem nehéz ugyanakkor belátni, hogy ezzel a módszerrel eltérő hosszúságú szövegeket nem lehet összevetni, hiszen minél hosszabb egy szöveg, annál kisebb eséllyel fognak benne új, a szöveg korábbi pontján még nem szereplő szavak szerepelni.

Ennek a problémának az egyik megoldása, hogy felosztják a szöveget egymást követő, egyenlő szószámú szegmensekre, a szegmenseknek külön kiszámolják a típus-token arányait, majd az így kapott értékeket átlagolják.¹⁸ Ennek az eljárásnak két problémája van. Egyrészt mivel a szövegek szószáma többnyire nem egész számú szorzata a szegmensek előre definiált szószámának, a szöveg vége kimarad a vizsgálatból. Másrészt a szöveg lexikailag eltérő szerkezeti részei és a szegmensek közötti véletlen megfelelés jelentősen torzíthatja az eredményt. E problémákat oldja meg a típus-token arány mozgó átlagának (*moving-average type-token ratio*) a kiszámítása.¹⁹ Ebben az esetben is ugyanolyan hosszú, azonban egymást átfedő szegmenseknek számolják ki a típus-token arányait, majd az így kapott értékeket átlagolják. Például 500 szavas szegmensek használata esetén kiszámolják a szöveg 1-től az 500. szóig, majd a 2-től az 501. szóig, majd a 3-tól az 502. szóig stb. terjedő szegmenseinek a típus-token arányait, és a kapott értékeket átlagolják. Mivel a szegmensek mindig csak egy szóval tolódnak el, a számításból nem marad ki a szöveg vége sem, illetve mivel az összes lehetséges szegmensen számításba veszi az eljárás, a szöveg szerkezeti egységei és a szegmensek esetleges egybeesésének a torzító hatása is kiküszöbölhető.

¹⁷ A magyar és más erőteljesebben ragozó nyelvek esetében a típus-token arány mérése helyett jellemzően a lemma-token arány mérését alkalmazzák.

¹⁸ Wendell JOHNSON, „Studies in Language Behavior: I. A Program of Research”, *Psychological Monographs* 56, 2 sz. (1944): 1–15, <https://doi.org/10.1037/h0093508>; Mike SCOTT, *Oxford WordSmith Tools: Version 4.0*. (Oxford: Oxford University Press, 2007).

¹⁹ Michael A. COVINGTON és Joe D. McFALL, „Cutting the Gordian Knot: The Moving-Average Type-Token Ratio (MATTR)”, *Journal of Quantitative Linguistics* 17, 2. sz. (2010): 94–100, <https://doi.org/10.1080/09296171003643098>.

A módszer azonban nem csupán a szókincs változatosságának, hanem – ahogyan arra a szerzők is utalnak²⁰ – kis ablakmérettel szavak egymáshoz közeli ismétlődésének a mérésére is alkalmas. Emellett nem csupán szavak ismétlődésének a mérését teszi lehetővé, hanem n-gramoknak, azaz szósoroknak vagy egyéb szavakhoz rendelt címkéknek, például szófajszorozatoknak az ismétlődését is képes mérni. A mérés elvégzéséhez két változót kell előre definiálnunk: az n-gramok méretét és az ablak méretét. Másként megfogalmazva: azt kell megadnunk, hogy a vizsgált n-gramoknak mekkora legyen az elemszáma, illetve, hogy az egyes szegmenseknek, amelyben az egymástól különböző n-gramok száma meghatározásra kerül, mekkora legyen az elemszáma. A szintaktikai ismétlődés mérésére kialakított eljárást az (1)-ben szereplő szófaji címkék alapján mutatom be. Tételezzük fel, hogy az (1) egy rövid, 14 szóból álló vers szavainak a szófajait és központozásait (két vessző és a végén egy pont) tünteti fel.

(1) ige névelő melléknév főnév , ige névelő melléknév főnév , határozószó ige névelő főnév .

Amennyiben az ismétlődést 12 elemszámú ablakokban vizsgáljuk, akkor az első ablakba a (2), a második ablakba a (3), a harmadik ablakba a (4), a negyedik ablakba pedig az (5) szófaji címkék, illetve írásjelek kerülnek.

(2) ige névelő melléknév főnév , ige névelő melléknév főnév , határozószó ige

(3) névelő melléknév főnév , ige névelő melléknév főnév , határozószó ige névelő

(4) melléknév főnév , ige névelő melléknév főnév , határozószó ige névelő főnév

(5) főnév , ige névelő melléknév főnév , határozószó ige névelő főnév .

Ha 3-gramok, azaz három elemből álló n-gramok alapján vizsgáljuk az ismétlődést, akkor a (2) 12-elemű ablak elemeit a (6)-ban felsorolt 3-gramokra bonthatjuk.

²⁰ Uo., 97–98.

(6) ige névelő melléknév
 névelő melléknév főnév
 melléknév főnév ,
 főnév , ige
 , ige névelő
 ige névelő melléknév
 névelő melléknév főnév
 melléknév főnév ,
 főnév , határozószó
 , határozószó ige

A (2) ablakban tehát összesen tíz darab 3-gram szerepel, amelyekből az ige-névelő-melléknév, a névelő-melléknév-főnév és a melléknév-főnév-vessző 3-gramok kétszer szerepelnek, vagyis a különböző 3-gramok száma hét. Ha az ablakban szereplő különböző 3-gramok számát elosztjuk az ablakban szereplő összes 3-grammal, akkor a típus-token arány mintájára kapunk egy 0 és 1 közötti értéket, amely kifejezi, hogy az ablakban szereplő n-gramok mennyire változatosak, illetve mennyire ismétlődnek. Minél közelebb van ez a szám a 0-hoz, annál kisebb az n-gramok változatossága, azaz annál nagyobb az n-gramok ismétlődésének a mértéke az ablakon belül. A (2) esetében ez az érték $7/10$, azaz $0,70$. Ha ezt az értéket az összes, azaz a (3), a (4) és az (5) ablakra is kiszámoljuk, akkor a kapott értékek átlagolásával adhatjuk meg a szöveg ismétlődési pontszámát, amely $0,85$. Mivel a típus-token arány mérésének az általános használatával szemben ebben az esetben nem a változatosság nagyságának, hanem az ismétlődés nagyságának mérése a cél, érdemes a kapott átlagot 1-ből kivonni annak érdekében, hogy a nagyobb érték ne az ismétlődés kisebb, hanem nagyobb mértékére utaljon, azaz az érték és az ismétlődés nagysága között ne fordított viszony legyen. A példa esetében tehát a szintaktikai ismétlődés mértéke $1 - 0,85$, azaz $0,15$.

A kapott pontszámokat befolyásolja az eljárás során használt n-gramok mérete. Ha 2-gramok alapján vizsgáljuk az ismétlődést, akkor a (2) ablak szó-faji címkéi, illetve írásjelei esetében a (7)-ben szereplő 2-gramokból álló listát kapjuk.

(7) ige névelő
 névelő melléknév
 melléknév főnév
 főnév ,

, ige
 ige névelő
 névelő melléknév
 melléknév főnév
 főnév ,
 , határozószó
 határozószó ige

A (7)-ben 11 2-gram szerepel, amelyekből négy kétszer jelenik meg, a különböző n-gramok száma tehát ebben az esetben is hét. A 7/11 osztást elvégezve viszont már nem 0,70-et, hanem két tizedesjegyre kerekítve 0,64-et kapunk, az egész szövegre pedig két tizedesjegyre kerekítve 0,70-et. A szintaktikai ismétlődés mértékének a nagysága tehát ebben az esetben $1 - 0,70$, azaz $0,30$. Vagyis, ha 2-gramok alapján végezzük el a számítást, akkor erőteljesebb ismétlődésre utaló értéket kapunk, mint ha 3-gramokkal végeznénk el azt.

A kapott értékeket ugyancsak befolyásolja az ablak mérete. Például, ha 3-gramok alapján végezzük el a vizsgálatot nyolc elemszámú ablakokkal, akkor az egész szövegre kapott ismétlődési érték kerekítve $1 - 0,94$, azaz $0,06$ lenne. De ha például hat elemszámú ablakokkal vizsgálnánk 3-gramokat, akkor mindegyik ablakra 1 értéket kapnánk, és így az egész vers átlagos értéke is 1 lenne, az ismétlődés mértéke pedig $1 - 1$, azaz 0 . Vagyis a túl kicsi ablak miatt az ismétlődő 3-gramok egyszer se kerülnének egy ablakon belülre, és ezért nem lehetne szintaktikai ismétlődést kimutatni.

A szintaktikai ismétlődés mérésekor problémaként jelentkezhet az, hogy szófaj n-gramok ismétlődésének oka lehet konkrét szövegrészek, például szókapcsolatok vagy akár teljes sorok megismétlése a versben (például refrének esetében). Mivel ez a kutatás kifejezetten csak a szövegek absztraktabb, szintaktikai szintjén megvalósuló ismétlések vizsgálatára irányul, a szövegrészek megismétlésével előálló szintaktikai ismétlődést valamiképpen figyelmen kívül kell hagyni. Ennek érdekében azt az eljárást követtem, hogy a fent bemutatott módszerrel kiszámítottam lemma n-gramok alapján is az ismétlődési értéket, majd ezt kivontam a szófaj n-gramokra kapott ismétlődési értékből. A lemma a szavak szótári alakja. Például az *asztal*, *asztalon*, *asztalokkal* szóalakok lemmája az *asztal* szóalak. Ezzel a módszerrel mind a szó szerinti, mind a lemma szerinti, azaz a közel szó szerinti ismétlések egyik típusa révén előálló szintaktikai ismétlődést figyelmen kívül lehetett hagyni a szintaktikai ismétlődés értékének megállapítása során. A módszer ugyanakkor azokat az ismét-

léseket nem hagyja figyelmen kívül, ahol az n -gramok szavai között legalább egy lemma szerinti eltérés van (például: *Kell a zsoltár*; *kell a fobász*; *Kell a bit*).²¹

3.2. Egymást követő sorok szintaktikai hasonlóságának mérése

A versekben megjelenő szintaktikai ismétlődés mértéke mellett a kutatás azt is vizsgálja, hogy a közvetlenül egymást követő rímelő sorok, illetve a közvetlenül egymást követő nem rímelő sorok között milyen mértékben valósul meg a szintaktikai ismétlődés jelensége. Magyar nyelvű históriás énekek esetében Seláf és szerzőtársai már tettek kísérletet a rímelő sorok grammatikai párhuzamosságának vizsgálatára az egymással rímelő sorokban szereplő utolsó négy szó grammatikai azonosságának automatizált meghatározásával.²² A szerzők beszámolója szerint azonban a morfológiai elemző a 16. századi szövegeken túl nagy hibaarányal működött, ezért végül elvetették a módszer alkalmazását. Az általam követett módszer lényege, hogy a sorpárokat alkotó két sor n -gramjai közötti átfedési együtthatót (*overlap coefficient*) számítottam ki, majd az így kapott értékeket átlagoltam az egyes szerzők esetében. Az átfedési együtthatót két halmaz közötti hasonlóság számszerűsítésére használják. Értékét úgy kapjuk meg, hogy elosztjuk a két halmaz metszésének elemszámát a kisebb halmaz elemszámával. A kapott eredmény ebben az esetben is egy 0 és 1 közötti szám, ahol a 0 érték azt jelzi, hogy a két halmaznak nincs közös eleme, az 1 érték pedig arra utal, hogy a kisebb halmaz összes eleme megtalálható a nagyobb halmazban is. Például ha van két halmazunk, amelyből az egyik A, B, C és D elemeket tartalmazza, a másik pedig A, C, D, E és F elemeket, akkor a két halmaz metszése A, C és D elemeket tartalmazza. Az átfedési együttható értéke pedig $3/4$, azaz $0,75$.

Az itt bemutatandó vizsgálatban a két halmazt a két összehasonlítandó, egymást követő sor szófaj n -gramjai alkotják. Ugyanakkor mivel egy sorban ugyanaz a szófaj n -gram többször is előfordulhat, az átfedési együttható számításának azt a változatát használtam, amely multihalmazok hasonlóságát adja meg, azaz olyan halmazok hasonlóságát, amelyben ugyanaz az elem többször is megjelenhet. Ennek számítása során a két halmaz metszésének az elemszámát úgy kell megadni, hogy az elemeknek mindig abban a halmazban való előfordulási számát vesszük figyelembe, ahol azok kevesebbszer fordulnak elő. Például tételezzük fel, hogy szófaji címkék 3-gramjai alapján vizsgáljuk a (8)-ban és a (9)-ben szereplő sorok hasonlóságát multihalmazok átfedési együtthatója alapján.

²¹ Ady Endre: *A papok istene*.

²² SELÁF, „A belső formulák a 16. századi históriás énekekben”.

(8) melléknév főnév , melléknév főnév , melléknév főnév ,

(9) ige melléknév főnév , melléknév főnév ,

A (8)-ban és a (9)-ben szereplő szófaji címkék, illetve írásjelek a (10)-ben és a (11)-ben szereplő 3-gramokra oszthatók.

(10) melléknév főnév ,

főnév , melléknév

, melléknév főnév

melléknév főnév ,

főnév , melléknév

, melléknév főnév

melléknév főnév ,

(11) ige melléknév főnév

melléknév főnév ,

főnév , melléknév

, melléknév főnév

melléknév főnév ,

A (10) és (11)-ben szereplő 3-gramok közül a melléknév-főnév-vessző, a főnév-vessző-melléknév és a vessző-melléknév-főnév 3-gram szerepel mindkét sorban. A két multihalmaz metszésében lévő elemszám megadásához azt kell megnéznünk, hogy ez a három n-gram hányszor szerepel az egyes sorokban (multihalmazokban), és az alacsonyabb előfordulási számokat kell összeadnunk. A melléknév-főnév-vessző 3-gram háromszor szerepel az első sorban és kétszer a másodikban, így itt a 2 értéket tartjuk meg. A főnév-vessző-melléknév és a vessző-melléknév-főnév 3-gram egyaránt kétszer szerepel az első sorban és egyszer a második sorban, így itt mind a két n-gram esetében 1 értékkel számolunk. A két multihalmaz metszésében lévő 3-gramok száma tehát $2+1+1$, azaz 4. Az átfedési együttható kiszámításához ezt az összeget kell elosztanunk a kisebb multihalmaz elemszámával. A példa esetében a kisebb multihalmaz a második sor, amelyben a 3-gramok száma 5. A két sor átfedési együtthatója így $4/5$, azaz 0,8, amely a két sor 3-gramjainak kifejezetten nagymértékű egyezését jelzi.²³

²³ A leírt eljárásnak a képlete a következő (A és B a két multihalmaz): átfedési együttható:

$$= \frac{|A \cap B|}{\min(|A|, |B|)}$$

Itt is jelentkezik az a probléma, hogy ha a második sor szó szerint vagy közel szó szerint megismétli az első sort, akkor az a szófaj n-gramok esetében is magas átfedési együtthatót eredményez, miközben a vizsgálat a sorpárok-nak a szintaxis, azaz elsősorban a szófaj n-gramok szintjén megvalósuló hasonlóságát kívánja mérni. A problémát ebben az esetben is úgy oldottam meg, hogy a szófaj n-gramok alapján számolt átfedési együttható mellett kiszámítottam lemma n-gramok alapján is a sorpárok átfedési együtthatóját, majd ez utóbbi értéket kivontam a szófaj n-gramok alapján kapott értékből.

3.3. Kiterjesztett szófaji címkék használata

A vizsgálatokban a szófaj n-gramoknak egy olyan verzióját is használok, ahol a főnevek, a névmások és a melléknevek esetében nemcsak a szófaji címke, hanem az esetalak és az egyes szám harmadik személyű birtokos személyjel esetleges megléte is fel van tüntetve, az igék esetében pedig jelezve van, ha főnévi igenévről van szó.²⁴ Ezeket a továbbiakban kiterjesztett szófaji címkéknek nevezem. A (13)-ban szereplő 3-gram például egy nominatívusz esetalakú melléknévből, egy nominatívusz esetalakú főnévből és egy szublatívusz esetalakú, egyes szám harmadik személyű birtokos személyjellel ellátott főnévből áll (például *sötét gályánk szőnyegére*).²⁵ A (14) 3-gram pedig egy igét, egy akkuzatívusz esetalakú főnevet, valamint egy főnévi igenetet tartalmaz (például *segítettem koronát leütni*).²⁶

²⁴ Ez a megközelítés a mozaik-ngram módszerhez hasonlóan tehát abból indul ki, hogy a szintaktikai szerkezetek pontosabb megragadásához szükséges lehet az n-gramok egyes elemeinek eltérő specifikációs szinten történő definiálása. A mozaik n-gram módszerről lásd: Balázs INDIG, „Mosaic N-grams: Avoiding Combinatorial Explosion in Corpus Pattern Mining for Agglutinative Languages”, in *Human Language Technologies as a Challenge for Computer Science and Linguistics*, szerk. Zygmunt VETULANI, Patrick PAROUBEK és Marek KUBIS, 147–151 (Poznań: Adam Mickiewicz University, 2017); INDIG Balázs, LAKI László és PRÓSZÉKY Gábor, „Mozaik nyelvmodell az AnaGramma elemzőhöz”, in *XII. Magyar Számítógépes Nyelvészeti Konferencia*, szerk. TANÁCS Attila, VARGA Viktor és VINCZE Veronika, 260–270 (Szeged: Szegedi Tudományegyetem, TTIK, Informatikai Intézet, 2016); BAJZÁT Tímea és INDIG Balázs, „Személyjelölési mintázatok feltárása mozaik n-gram alapon három segédige + főnévi igenévi típus konstrukcióiban”, in *Személyjelölés és grammatika*, szerk. LACZKÓ Krisztina, TÁTRAI Szi-lárd és TOLCSVAI NAGY Gábor, 85–120 (Budapest: ELTE Eötvös Kiadó, 2024).

²⁵ Ady Endre, *Ima Baál istenhez*.

²⁶ Petőfi Sándor, *A Honvéd*.

(13) melléknév|nominatívusz főnév|nominatívusz főnév|szublatívusz|-
birtokviszony E/3

(14) ige főnév|akkuzatívusz igelinfinitívusz

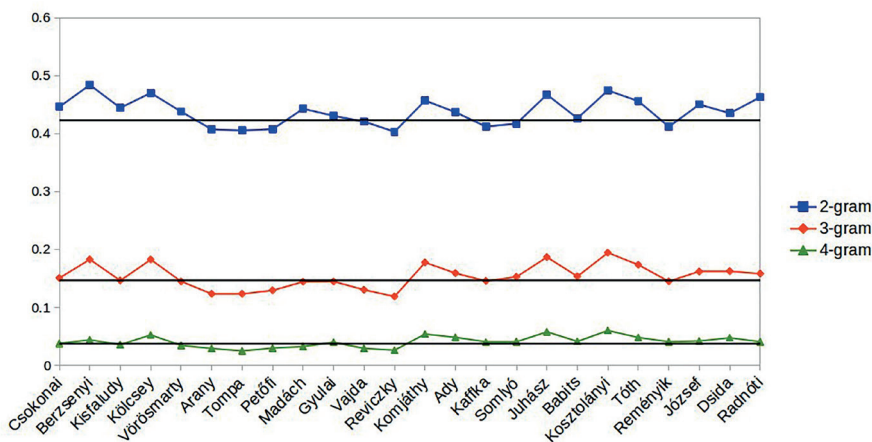
A főnevek, névmások és melléknevek esetalakjának a szerepeltetése lehetővé teszi, hogy a vizsgálat differenciáltabban vegye figyelembe a különböző szintaktikai szerkezeteket, hiszen az igei konstrukciók, illetve mondatkonstrukciók egyik lényegi tulajdonsága, hogy milyen esetalakú bővítményekkel jelennek meg. Ahogyan a (14) is példázza, a különböző számban és személyben megjelenő birtokos személyjelek közül csak az egyes szám harmadik személyű birtokos személyjelet vettem figyelembe. Ennek oka az, hogy a szintaktikai szempontból nagyobb relevanciával bíró birtokos jelzős szerkezeteknek ez a jelölője. A birtokszón első vagy második személyű birtokos személyjelek csak akkor fordulnak elő, ha a birtokos jelző első vagy második személyű személyes névmás (például az én cipőm, a te lakásod). Az ilyen névmási birtokos jelzők azonban jóval ritkábbak, mint a főnévi birtokos jelzők. Ugyanakkor versekben első és második személyű birtokos személyjelek számos esetben előfordulnak egyéb, a szintaktikai szerkezet szempontjából nem igazán releváns funkcióban, ami miatt az első és második személyű birtokos személyjelek figyelembevétele félrevezethetné a szintaktikai ismétlődésre fókuszáló vizsgálatokat. A főnévi igenév megkülönböztetése az igétől pedig azért indokolt, mert a főnévi igenév használatához több, kifejezetten gyakori szintaktikai konstrukció is kapcsolódik (például segédigés szerkezetek). Az esetalakhoz, birtokossághoz és főnévi igenévhez kapcsolódó morfoszintaktikai információk figyelembevétele tehát összességében érzékenyebbé teszi a kidolgozott n-gram alapú módszert a specifikusabb szintaktikai szerkezetekre.

4. SZINTAKTIKAI ISMÉTLŐDÉS A MAGYAR KANONIKUS KÖLTÉSZETBEN

Elsőként azt az általánosabb kérdést vizsgálom, hogy a magyar kanonikus költészetben nagyobb mértékben jelenik-e meg a szintaktikai ismétlődés jelensége, mint a magyar regényirodalomban. Amennyiben a kérdésre pozitív válasz adható, az arra utal, hogy a magyar kanonikus költészetben a fonológiai ismétlődés mellett a szintaktikai ismétlődésnek is kitüntetett szerepe van. Emellett azt is vizsgálom, hogy a magyar kanonikus költészetben milyen időbeli tendenciát mutat a szintaktikai ismétlődés mértékének az alakulása.

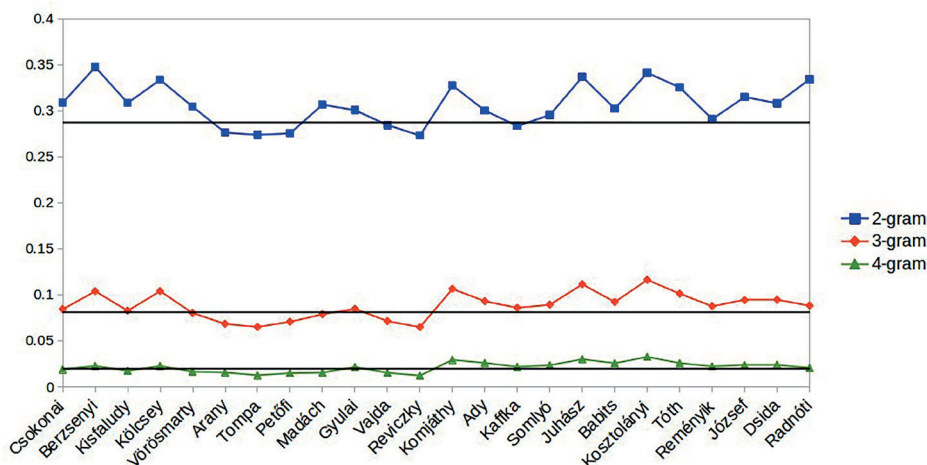
A két kérdés megválaszolásához első lépésben a kutatási korpusz 24 költőjének versei és a regénykorpusz esetében számítottam ki 100, 50 és 25 tokenes ablakbeállítással, kettő, három és négy elemszámú szófaj n-gramokra a szintaktikai ismétlődés mértékét. A tokenek száma az ablak elemszámára, azaz az ablakban található szavak és központozások együttes előfordulási számára utal. Az egyes költők esetében a számítást minden versre külön elvégeztem, majd a kapott értékeknek számítottam ki a mediánját.²⁷ A regénykorpusznál első lépésben a korpuszban szereplő 120 szerző esetében külön-külön kiszámítottam az egyes szerzők műveinél megállapított ismétlődési értékek mediánját, majd második lépésben az így kapott 120 mediánértéknek számítottam ki a mediánját. Erre az eljárásra azért volt szükség, hogy az egyes szerzők ugyanolyan súllyal vegyenek részt az egész regénykorpuszt reprezentáló mediánérték kiszámításában, függetlenül attól, hogy hány regénnyel szerepelnek a korpuszban. (Az egyes költőkre és a regénykorpuszra kapott mediánértékeket lásd a Függelék 7. táblázatában.)

A 100, 50 és 25 tokenes ablakméretekre kapott mediánértékeket az 1., 2. és 3. ábra mutatja be. Az egyes ablakméretekhez tartozó ábrák egyszerre jelenítik meg a 2-, 3- és 4-gramokra kapott értékeket a szerzők születési sorrendjében. A regényekre kapott mediánértékeket a vízszintes vonalak jelölik.

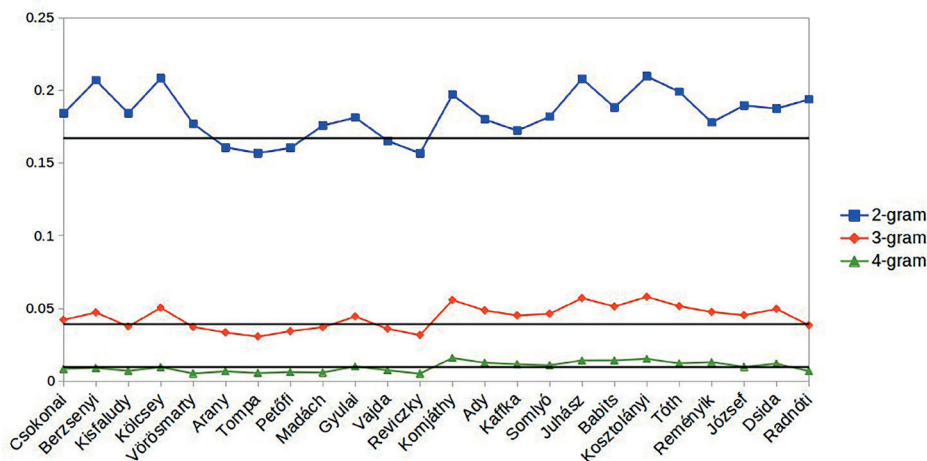


1. ábra. A szintaktikai ismétlődés mértékének mediánjai szófaji címkék 2-gramjai, 3-gramjai és 4-gramjai alapján, 100 tokenes ablakbeállítással

²⁷ A medián az átlaghoz hasonlóan egy számokból álló adatsor középértékét adja meg. Egy számokból álló adatsor mediánja a növekvő sorrendbe rendezett számokból a középső, páros számú adatsor esetén pedig a középső kettőnek az átlaga. A medián az átlagnál általában megbízhatóbb középérték, mivel nem érzékeny a kiugró adatokra.



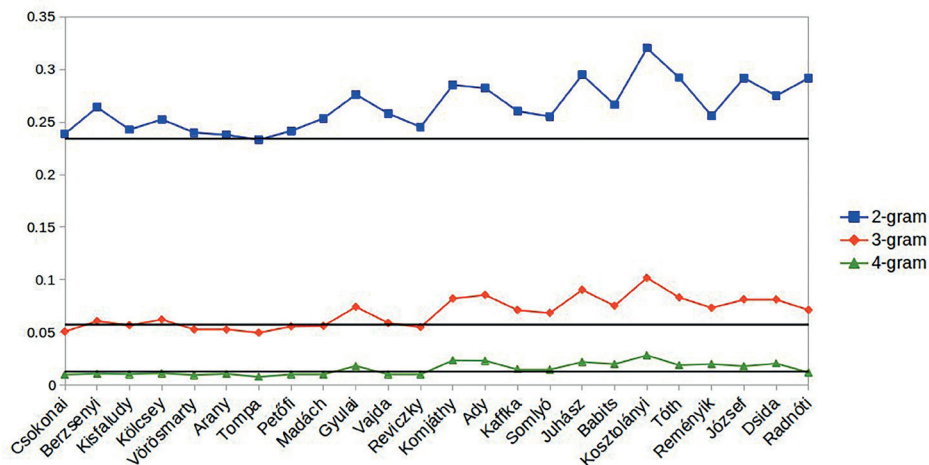
2. ábra. A szintaktikai ismétlődés mértékének mediánjai szófaji címkék 2-gramjai, 3-gramjai és 4-gramjai alapján, 50 tokenes ablakbeállítással



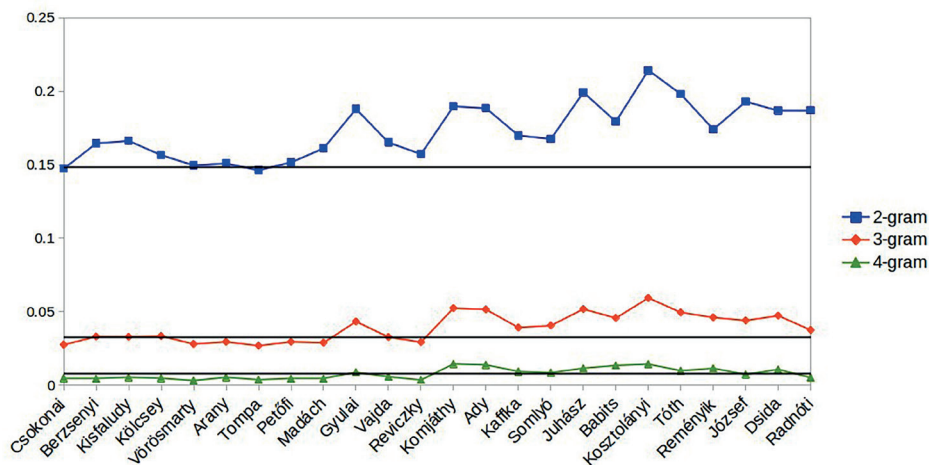
3. ábra. A szintaktikai ismétlődés mértékének mediánjai szófaji címkék 2-gramjai, 3-gramjai és 4-gramjai alapján, 25 tokenes ablakbeállítással

Ugyanezeket a méréseket elvégeztem kiterjesztett szófaji címkék alapján is. (Az egyes költőkre és a regénykorpuszra kapott mediánértékeket lásd a *Függelék* 8. táblázatában.) Ezek a mérések tehát a szófaji címke mellett az esetalakot, az egyes szám harmadik személyű birtokos személyjelet, illetve igéknél

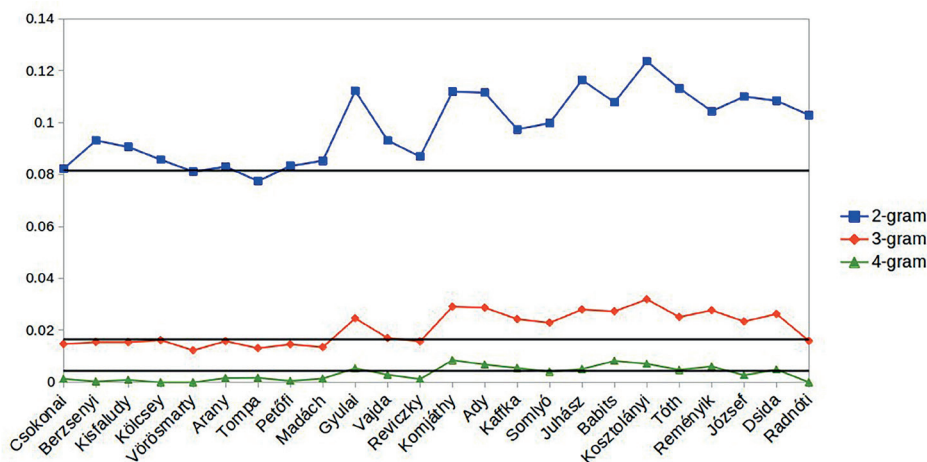
a főnévi igenévi jelet is figyelembe veszik, ami a szintaktikai viszonyok differenciáltabb megragadását teszi lehetővé. A mérések eredményeit a 4., 5. és 6. ábra mutatja be.



4. ábra. A szintaktikai ismétlődés mértékének mediánjai kiterjesztett szófaji címkék 2-gramjai, 3-gramjai és 4-gramjai alapján, 100 tokenes ablakbeállítással



5. ábra. A szintaktikai ismétlődés mértékének mediánjai kiterjesztett szófaji címkék 2-gramjai, 3-gramjai és 4-gramjai alapján, 50 tokenes ablakbeállítással



6. ábra. A szintaktikai ismétlődés mértékének mediánjai kiterjesztett szófaji címkék 2-gramjai, 3-gramjai és 4-gramjai alapján, 25 tokenes ablakbeállítással

A hat ábrából látható, hogy a legtöbb mérési beállítás esetén a költők többsége a regényekre kapott ismétlődési értéket meghaladja. A 2. táblázat mutatja be, hogy az egyes mérési beállításoknál a 24-ből hány szerző haladja meg a regényekre kapott értéket. A regényekre kapott értéket meghaladó legtöbb költőt kiterjesztett szófaji címkék 2-gramjainál találjuk. 50 és 25 tokenes ablakméretnél a 24-ből 22 szerző, 100 tokenes ablakméretnél pedig 23 szerző haladja meg a regényeknél mért értéket. Szürke cellaszínezéssel emeltem ki azokat a mérési beállításokat, ahol a regényekre kapott ismétlődési értéket csupán a vizsgált költők fele vagy kevesebb mint a fele haladja meg. Látható, hogy mind a négy ilyen eset a 4-gramos beállításhoz kapcsolódik. Ezekből három a kiterjesztett szófaji címkék használatánál jön elő.

2. táblázat. A regényekre kapott szintaktikai ismétlődési értékek mediánját meghaladó költők száma az egyes mérési beállításoknál (összes költő = 24)

	Szófaji címkék			Kiterjesztett szófaji címkék		
	2-gram	3-gram	4-gram	2-gram	3-gram	4-gram
100 token	16	13	15	23	16	12
50 token	18	17	15	22	17	11
25 token	19	15	12	22	13	10

A 3. táblázat a költőkre kapott mediánértékek regények mediánjától való eltéréseinek (pozitív vagy negatív szám) összegét mutatja be. Egy darab, szűr-kével jelölt mérési beállítást leszámítva mindenhol pozitív érték szerepel, ami úgyszintén azt jelzi, hogy a vizsgált költők verseiben a szintaktikai ismétlődés mértéke jellemzően nagyobb, mint a magyar regényirodalomban.

3. táblázat. A költőkre kapott mediánok regények mediánjától való eltéréseinek összege az egyes mérési beállításoknál

	Szófaji címkék			Kiterjesztett szófaji címkék		
	2-gram	3-gram	4-gram	2-gram	3-gram	4-gram
100 token	0.3461	0.1681	0.0691	0.7393	0.2670	0.0557
50 token	0.4632	0.1727	0.0296	0.5967	0.1647	0.0047
25 token	0.3927	0.1175	0.0065	0.4053	0.1006	-0.0246

A szófaji, illetve kiterjesztett szófaji címkék alapján végzett mérésekből az derül ki, hogy a magyar kanonikus költészetben általánosságban nagyobb mértékben jelenik meg a szintaktikai ismétlődés jelensége, mint a magyar regényirodalomban. Úgy tűnik tehát, hogy a fonológiai ismétlődés mellett a verseket valóban jellemzi egy másik típusú, a szintaktikai szerkezetek használatában testet öltő ismétlődés is. Ahogyan arra korábban utaltam, az ismétlődés, beleértve a szintaktikai ismétlődést is, a nyelv alapjellemezője. Az ismétlődés nyelvbeli fontosságát jól mutatja, hogy egymással szemben álló nyelvészeti irányzatokban is elsődleges, a nyelv működésében nélkülözhetetlen mechanizmusként tűnik fel. A konstrukciós nyelvtani megközelítések például a grammatikát olyan konstrukciók összességeként ragadják meg, amelyek az ismétlődő, egymáshoz fonológiaiilag és szemantikailag hasonló nyelvi megnyilatkozásokból absztrahálódnak, illetve amelyek egyúttal ezeknek az egymáshoz hasonló megnyilatkozásoknak az ismételt létrehozását motiválják.²⁸ De a konstrukciós megközelítésekkel szemben álló generatív nyelvészet is a nyelvi képesség alapjának tekinti az ismétlés egy típusát, a rekurziót, amely lehetővé teszi, hogy ugyanazt a műveletet újra és újra elvégezve egy nyelvi szerkezetbe ismételten beágyazzuk ugyanazt

²⁸ Lásd például Adele E. GOLDBERG, *Constructions: A Construction Grammar Approach to Argument Structure* (Chicago, London: The University of Chicago Press, 1995); Adele E. GOLDBERG, *Constructions at Work: The Nature of Generalization in Language* (Oxford: Oxford University Press, 2005), <https://doi.org/10.1093/acprof:oso/9780199268511.001.0001>; William CROFT és Alan D. CRUSE, *Cognitive Linguistics* (Cambridge: Cambridge University Press, 2004), 257–290, <https://doi.org/10.1017/CBO9780511803864>.

a szerkezetet.²⁹ A versekben megjelenő fonológiai és szintaktikai ismétlődés tehát nem egy új, a hétköznapi beszédben nem létező minőséget ad hozzá a nyelvhez, hanem a nyelv működésének, lényegének egyik alapjellemezőjét teszi láthatóbbá. A szintaktikai ismétlődés erőteljesebb megjelenését így a fonológiai ismétlődés erőteljesebb megjelenéséhez hasonlóan magyarázhatjuk a Jakobson által leírt poétikai funkcióval.³⁰ Mindkettő a versszöveg nyelviségére, a szöveg referenciája helyett magára a nyelvi mechanizmusra hívja fel a figyelmet azáltal, hogy a nyelvet lehetővé tevő, ugyanakkor a hétköznapi megnyilatkozásainkban kevésbé szembeűnő eljárást, az ismétlést állítja a figyelem középpontjába. A poétikai funkció szempontjából a felfokozott szintaktikai ismétlődés csupán annyiban különbözik a rímtől, a ritmustól vagy éppen az alliterációtól, hogy nem a hangzás, hanem a szerkezet szintjén mutatja fel az ismétlődést, mint a nyelv egyik jellegadó tulajdonságát.

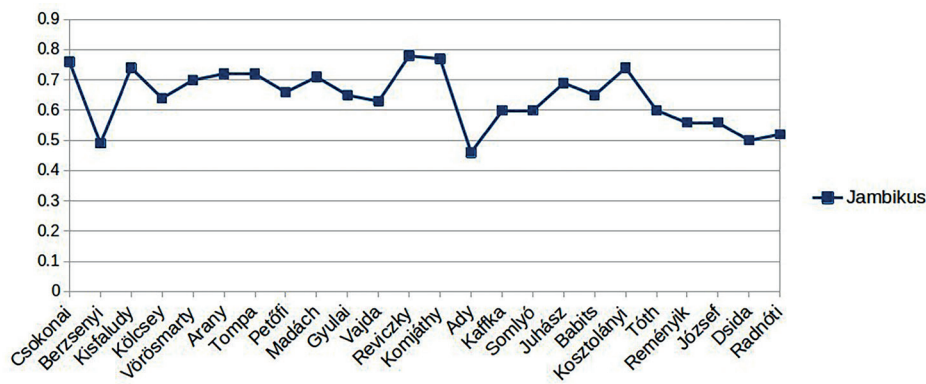
Az ábrákból az is jól látszik, hogy a szintaktikai ismétlődés mértéke időben változik. A 19. század utolsó harmadában alkotó Komjáthynál, illetve a 20. században alkotó szerzők esetében általánosan megnő a szintaktikai ismétlődés mértéke a korábbi időszakhoz képest. Egy kevésbé erőteljes mintázat is megfigyelhető: a 19. század elején, Berzsenyi, Kisfaludy, Kölcsey, illetve a sima szófaji címkéken alapuló beállításoknál Csokonai esetében a szintaktikai ismétlődés erőteljesebben jelenik meg, mint az őket követő, a 19. század közepén alkotó költőknél. Sima szófaji címkék esetén a 19. század második felében alkotó költők nagy része is alatta marad az említett 19. század eleji költőknek. A kiterjesztett szófaji címkék használatánál érdekes módon Gyulai esetében a 19. századi értékekhez képest kifejezetten magas az ismétlődés mértéke. Ennek pontos okait egy a szövegek szoros olvasását elvégző kvalitatív vizsgálat tárhatná fel.

Felvethető a kérdés, hogy a 19. század végén, illetve a 20. század elején a szintaktikai ismétlődés mértékének a megnövekedése a magyar költészet egyéb poétikai dimenzióitól függetlenül megy-e végbe, vagy pedig összefüggésbe hozható valamilyen más poétikai változással. Máshogyan megfogalmazva a kérdés az, hogy a szintaktikai ismétlődés erőteljesebbé válása pusztán többletminőségként jelenik meg a magyar költészetben, ami érintetlenül hagyva a már meglévő poétikai eszköztárat mintegy hozzáadódik ahhoz, vagy pedig a vers egyéb poétikai dimenzióira is kiható rendszerváltozás eredményeként áll

²⁹ Lásd például Marc D. HAUSER, Noam CHOMSKY és W. Tecumseh FITCH, „The Faculty of Language: What is It, Who Has It, and How Did It Evolve?”, *Science* 298, 5598 sz. (2002): 1569–1579, <https://doi.org/10.1126/science.298.5598.1569>.

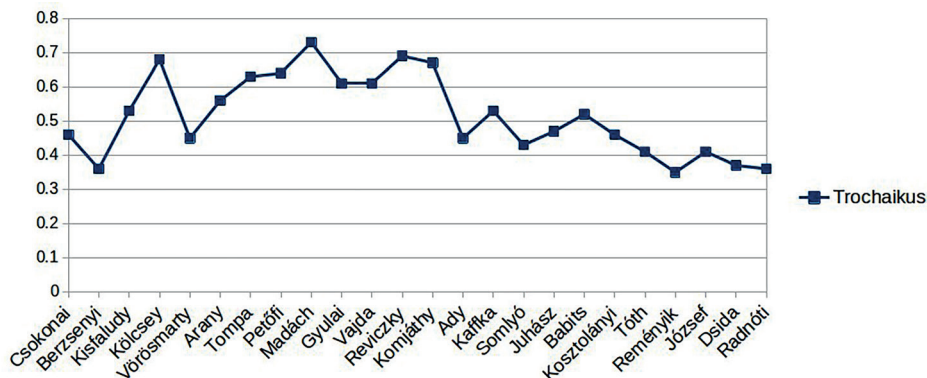
³⁰ Roman JAKOBSON, „Nyelvészet és poétika”, in JAKOBSON, *Hang – Jel – Vers*, 211–257.

elő. Amennyiben Lowth-nak, Földinek és Jakobsonnak a tanulmány elején hivatkozott elgondolásai alapján abból indulunk ki, hogy a versszerűség érzetét nemcsak a fonológiai ismétlődés, hanem egyéb típusú ismétlődések is megerősíthetik, akkor megfogalmazható az a hipotézis, hogy a szintaktikai ismétlődés erőteljesebbé válása a fonológiai ismétlődés visszaesését kompenzálja.³¹ Azaz a metrumnak a 20. századra bekövetkező fellazulása eredményezte egy másik típusú ismétlődés, a szintaktikai ismétlődés erőteljesebbé válását. A poétikai rendszerben tehát az egyik típusú ismétlődés gyengülését egy másik típusú ismétlődés megerősödése egyenlíti ki. A metrumnak a 20. század elejére bekövetkező fellazulását mind a jambikus, mind a trochaikus metrumot megvalósító versek esetében megfigyelhetjük. A 7. ábra a jambikus, a 8. ábra a trochaikus metrumú versek szabályossági pontszámainak a mediánját mutatja be az egyes szerzők esetében (a pontos mediánértékeket lásd a *Függelék* 9. táblázatában). A következő részben használt 0,5 helyett ezúttal 0,3 szabályossági pontszám felett vettem bele az egyes verseket a vizsgálatba annak érdekében, hogy a mediánértékek ne csak a jambikus és trochaikus metrumot szigorúbban megvalósító verseken alapuljanak. A jambikus és trochaikus versekbe nem vettem bele a szimultán metrumú verseket, azaz azokat, amelyekben ütemhangsúlyos metrum is megjelenik az időmérték mellett.



7. ábra. A jambikus metrumú versek szabályossági pontszámainak mediánjai szerzőnként (szabályossági küszöbérték: 0,3)

³¹ Köszönettel tartozom Szemes Botondnak, aki a tanulmány első változatát olvasva felvetette az ötletet, hogy a szintaktikai ismétlődés növekedése összefügghet a fonológiai ismétlődés fellazulásával.



8. ábra. A trochaikus metrumú versek szabályossági pontszámainak mediánjai szerzőnként (szabályossági küszöbérték: 0,3)

A 7. és 8. ábrából láthatjuk, hogy míg a 19. század költőinél többnyire szabályosabban valósul meg a jambikus és a trochaikus metrum, addig a 20. század első felében alkotó szerzőknél a szabályosság mértéke visszaesik. Ez pont fordítottja a szintaktikai ismétlődés esetében megfigyelt trendnek, ami megerősíti a hipotézist, hogy olyan rendszerszintű változásról van szó, ahol a metrumhoz kapcsolódó fonológiai ismétlődés mértékének a visszaesését kompenzálja a szintaktikai ismétlődés mértékének a növekedése. Megjegyzendő persze, hogy nem feltétlenül érdemes a fonológiai ismétlődés mértékét független változónak, a szintaktikai ismétlődést pedig ettől függő változónak tekinteni, azaz okra és okozatra felosztani a két jelenséget. Az ismétlődés e két típusának a változásai minden bizonnyal kölcsönösen hatottak egymásra. Természetesen egyéni kivételek vannak. Például Komjáthy és Kosztolányi esetében a metrum megvalósításának a szabályossága és a szintaktikai ismétlődés mértéke is nagy fokú.

Érdemes kiemelni, hogy a szófaj n-gramokkal megragadott szintaktikai jellemzők valójában szemantikai jellemzőknek az absztrakt kifejeződései. A szófajok megfeleltethetők sematikus szemantikai kategóriáknak,³² a szófajok közötti viszonyok pedig sok esetben sematikus szemantikai viszonyoknak.

³² Lásd: Ronald W. LANGACKER, *Cognitive Grammar: A Basic Introduction* (Oxford: Oxford University Press, 2008), 93–160, <https://doi.org/10.1093/acprof:oso/9780195331967.001.0001>; TOLCSVAI Nagy Gábor, „Jelentés”, in TOLCSVAI Nagy Gábor, szerk., *Nyelvtan*, 310–390 (Budapest: Osiris Kiadó, 2017).

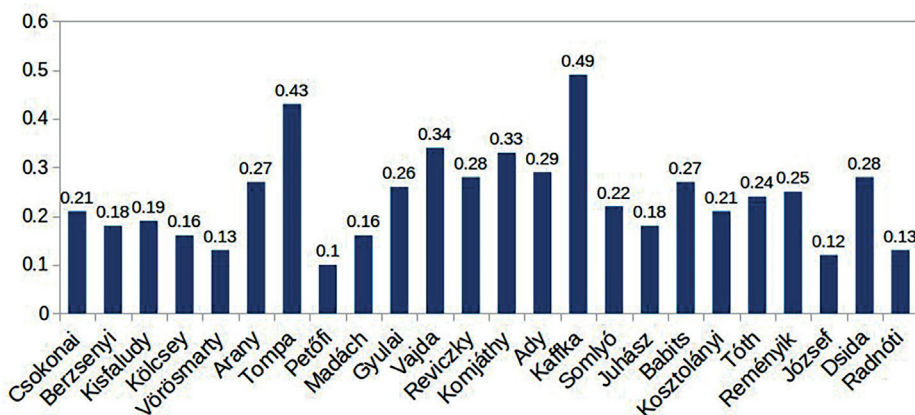
Például a melléknév-főnév kapcsolat általában a következő sematikus jelentéssel írható le: entitások egy adott tulajdonsággal rendelkező típusa. A szintaxis szintjén megragadott ismétlődés jelensége tehát nem más, mint absztraktabb szinten megvalósuló szemantikai ismétlődés. Amennyiben a szintaktikai ismétlődést absztrakt szinten érvényesülő tartalmi ismétlődéssel azonosítjuk, akkor levonhatjuk a következtetést, hogy a 20. század elejének a kanonikus szerzői nagyobb mértékben hoztak létre elvont, azaz a szintaxis szintjén tetten érhető tartalmi párhuzamokat, mint a 19. század szerzőinek a nagy része.

4.1. Szintaktikai ismétlődés és halmozás

A vizsgálatokhoz használt, átfedő n-gramokon alapuló módszer akkor is szintaktikai ismétlődést állapít meg, hogyha legalább háromszor szerepel egymás után ugyanaz a szófaji címke, központosítással elválasztva vagy központosítások nélkül. Ezek az esetek általában halmozásnál jelennek meg, vagyis amikor ugyanolyan szintaktikai funkcióban szereplő szavak fordulnak elő egymást követően. Ennek legtipikusabb esete a jelzőhalmozás (például *A ködből hulltak a szívek, / Csúnya, kicsiny, kemény szívek*).³³ A melléknév-vessző-melléknév-vessző-melléknév címkesor 2-gramjai a következők: melléknév-vessző, vessző-melléknév, melléknév-vessző, vessző-melléknév. Látható tehát, hogy a melléknév-vessző és a vessző-melléknév esetén is ismétlődő n-gramokat kapunk. Ha 3-gramokra bontjuk a címkesort, akkor csak a melléknév-vessző-melléknév n-gram fordul elő kétszer. Természetesen az azonos szintaktikai szerepű szavak ismétlése úgyszintén tekinthető a szintaktikai ismétlődés egy fajtájának. Ugyanakkor felmerülhet a kérdés, hogy a verseknek a regényekhez képest mért nagyobb mértékű szintaktikai ismétlődését, illetve a szintaktikai ismétlődés mértékének a növekedését a 20. század elején nem pusztán csak a halmozás gyakoribb használata okozza. A 9. ábra a vizsgált költők esetében mutatja be azoknak a verseknek az arányát, amelyekben legalább egy darab, minimum három elemből álló melléknév-, főnév- vagy ige-halmozás szerepel egy verssoron belül. Főneveknél és melléknéveknél azokat az eseteket tekintetem halmozásnak, ahol egy verssoron belül egymást követően vesszővel elválasztva vagy vessző nélkül minimum három főnév vagy melléknév szerepel ugyanolyan esetben úgy, hogy vagy egyik szónál sincs birtokos jelölve, vagy pedig mindegyik szó esetében ugyanolyan számú és személyű birtokos van jelölve. Igéknél csak azokat az

³³ Ady Endre, *A Léda szíve*.

eseteket tekintettem halmozásnak, ahol egy verssoron belül legalább három ige követi egymást teljesen megegyező morfoszintaktikai jellemzőkkel.

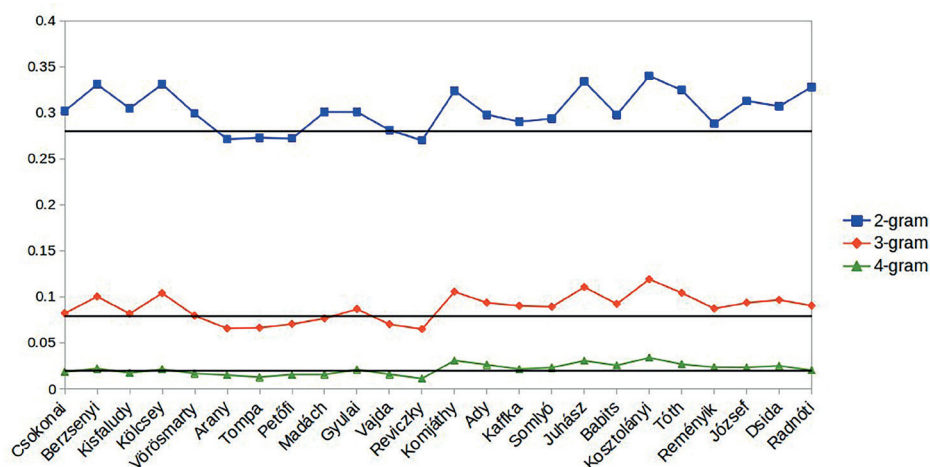


9. ábra. Minimum három elemből álló főnév-, melléknév- vagy igehalmozást tartalmazó versek aránya

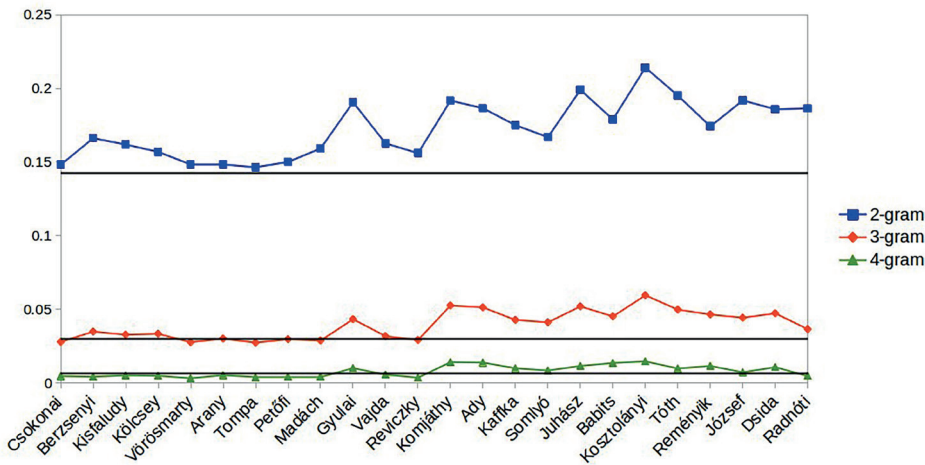
A 9. ábra oszlopdiagramjából láthatjuk, hogy viszonylag nagy arányban fordulnak elő legalább három elemből álló főnév-, melléknév- vagy igehalmozást tartalmazó versek az egyes életművekben. Tompa és Kaffka esetében ezeknek a verseknek az aránya megközelíti a verseik felét, de Arany, Gyulai, Vajda, Reviczky, Komjáthy, Ady, Babits és Dsida esetében is 25% felett van ezeknek a verseknek az előfordulása. Ebben a tanulmányban nincs mód a jelenség részletesebb vizsgálatára, a jövőben azonban mindenképpen érdemes lenne a halmozás különböző típusaival és ezek poétikai funkciójával behatóbban is foglalkozni.

Annak érdekében, hogy megbizonyosodhassunk arról, hogy a szintaktikai ismétlődésnek az előző részben megállapított, regényeknél magasabb értékeit, illetve 20. századi megemelkedését nem csupán a halmozás jelensége okozza, a méréseket elvégeztem úgy is, hogy a szintaktikai ismétlődés számításakor csak azokat az ablakokat vettem figyelembe, amelyekben nem szerepel egymást követően központozással elválasztva vagy központozás nélkül három vagy több ugyanolyan szófaji címke. Ez tehát azt jelenti, hogy szemben a 9. ábra előállításához használt módszerrel, csak a szófaji címkéket vettem figyelembe, illetve nemcsak a melléknév-, főnév- és igehalmozásokat tartalmazó

ablakokat zártam ki a vizsgálatból, hanem azokat az ablakokat is, amelyekben bármilyen más szófajú szóból szerepel egymást követően legalább három. Emellett az azonos szófajú szavakat elválasztó írásjelet nem korlátoztam a vesszőre, illetve nem vettem figyelembe a sorhatárokat sem (a regények esetében erre lehetőség sem lenne). Az így kizárt ablakokban szereplő halmozások jelentős része természetesen grammatikai értelemben nem tekinthető halmozásnak, a vizsgálatok eredményeit ugyanakkor a tényleges halmozásokhoz hasonlóan torzíthatják ezek az esetek is. A 10. ábra az 50 tokenes ablakmérettel futtatott mérések eredményeit mutatja be szófaji címkék 2-gramjai, 3-gramjai és 4-gramjai alapján. A 11. ábra ugyanezt mutatja be kiterjesztett szófaji címkék alapján. Ebben az esetben a halmozást tartalmazó ablakok kizárását is a kiterjesztett szófaji címkék alapján végeztem el. A 4. táblázat azt mutatja be, hogy az egyes mérési beállításoknál a 24-ből hány szerző haladja meg a regényekre kapott értéket. Az 5. táblázat pedig a költőkre kapott mediánértékek regények mediánjától való eltéréseinek összegét mutatja be a különböző mérési beállításoknál. Az egyes mérési beállításoknál nem kerültek bele a vizsgálatba azok a versek, amelyek esetében egy olyan ablak sem volt, amely nem tartalmazott halmozást.



10. ábra. A szintaktikai ismétlődés mértékének mediánjai szófaji címkék 2-gramjai, 3-gramjai és 4-gramjai alapján, 50 tokenes ablakbeállítással, a halmozást tartalmazó ablakok kizárásával



11. ábra. A szintaktikai ismétlődés mértékének mediánjai kiterjesztett szófaji címkék 2-gramjai, 3-gramjai és 4-gramjai alapján, 50 tokenes ablakbeállítással, a halmozást tartalmazó ablakok kizárásával

4. táblázat. A regényekre kapott szintaktikai ismétlődési értékek mediánját meghaladó költők száma az egyes mérési beállításoknál, a halmozást tartalmazó ablakok kizárásával (összes költő = 24)

	Szófaji címkék			Kiterjesztett szófaji címkék		
	2-gram	3-gram	4-gram	2-gram	3-gram	4-gram
100 token	17	16	15	24	20	14
50 token	20	18	15	24	18	12
25 token	20	16	13	24	19	11

5. táblázat. A költőkre kapott mediánok regények mediánjától való eltéréseinek összege az egyes mérési beállításoknál, a halmozást tartalmazó ablakok kizárásával

	Szófaji címkék			Kiterjesztett szófaji címkék		
	2-gram	3-gram	4-gram	2-gram	3-gram	4-gram
100 token	0.4595	0.2165	0.0560	0.9003	0.3397	0.0815
50 token	0.5607	0.2287	0.0441	0.7133	0.2284	0.0270
25 token	0.4611	0.1570	0.0125	0.5103	0.1425	-0.0109

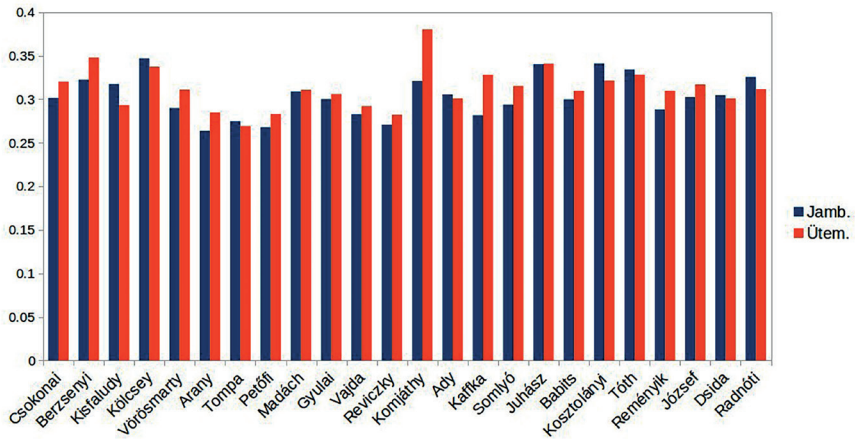
A kapott eredmények érdekes módon azt mutatják, hogy a halmazásokat tartalmazó ablakok kizárásával egyáltalán nem csökkent a regények és a versek között talált különbség a szintaktikai ismétlődés mértékében, sőt az még markánsabbá vált. Kiterjesztett szófaji címkék 2-gramjai esetén például az összes költő meghaladja a regényeknél mért értéket. A 10. és a 11. ábra alapján láthatjuk azt is, hogy a halmazást tartalmazó ablakok kizárásával kapott értékek alapján is megállapítható, hogy a szintaktikai ismétlődés mértéke megemelkedik a 20. század eleji költészetben a korábbi időszakhoz képest.

5. A METRUM TÍPUSA ÉS A SZINTAKTIKAI ISMÉTLŐDÉS ÖSSZEFÜGGÉSEI A MAGYAR KANONIKUS KÖLTÉSZETBEN

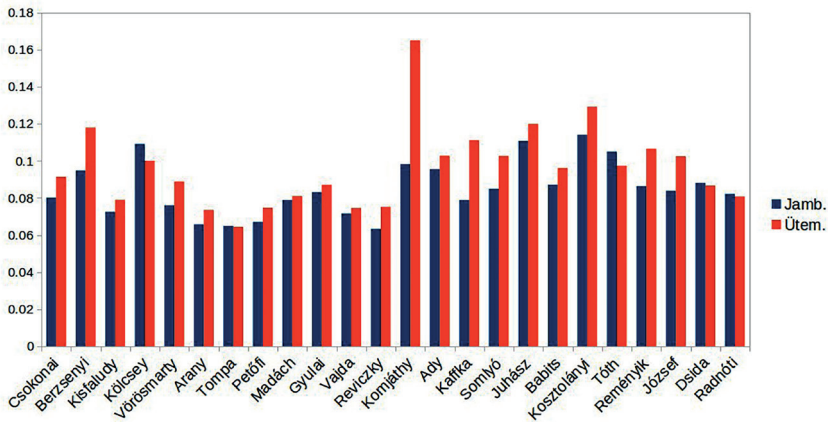
Az előző részben láthattuk, hogy a szintaktikai ismétlődés és a fonológiai ismétlődés időbeli alakulása feltehetően nem független egymástól. Ugyanakkor felvethető a kérdés, hogy a fonológiai ismétlődés különböző típusai, például a versben realizálódó metrumtípusok nem befolyásolják-e eltérő módon a szintaktikai ismétlődés megvalósulását. Az alábbiakban azt vizsgálom, hogy a metrum típusa és a szintaktikai ismétlődés mértéke között felfedezhető-e valamilyen összefüggés. A vizsgálat során a magyar költészet két leggyakoribb metrumtípusába, a jambikus és az ütemhangsúlyos metrumtípusba sorolható versek szintaktikai ismétlődését vetettem össze. A vizsgálatban azokat a verseket tekintettem jambikusnak, amelyeknek a szabályossági pontszáma meghaladja a 0,5 értéket. Ezekben a versekben a vers ritmusa jól érzékelhetően valósítja meg a jambikus metrumot, ami nagy valószínűséggel a szerző tudatos törekvésének az eredménye. A *Függelék* 10. táblázata mutatja be szerzőnként, hogy a metrumelemző program hány verset sorolt 0,5 küszöbértékkel a jambikus és a trochaikus, valamint az ütemhangsúlyos, a szimultán jambikus és a szimultán trochaikus versek közé. Ebben a vizsgálatban az ütemhangsúlyos kategória részének tekintettem a szimultán trochaikus verseket is, azaz azokat a verseket, amelyek egyszerre trochaikusak és ütemhangsúlyosak. Ugyanakkor a szimultán jambikus, azaz egyszerre jambikus és ütemhangsúlyos verseket kiszűrtem a vizsgálatból, vagyis ezeket se a jambikus, se az ütemhangsúlyos csoport részeként nem vettem számításba.³⁴ A *Függelék* 11. táblázata mutatja be a vizsgált szerzők esetében a jambikus és az ütemhangsúlyos metrumú versekre kapott szintaktikai ismétlődés értékei-

³⁴ A metrumelemző program működését részletesen bemutatja HORVÁTH Péter, „Két eljárás magyar nyelvű versek metrumának gépi felismertetéséhez”. A szimultán metrum Szepes és Szerdahelyi értelmezését követem, amely szerint azokban az esetekben beszélhetünk

nek a mediánját szófaji címkék 2-gramjai, 3-gramjai és 4-gramjai alapján, 50 tokenes ablakmérettel. A 12., 13. és 14. ábra oszlopdiagramjai a három n-gram-méretre kapott mediánértékeket tüntetik fel.

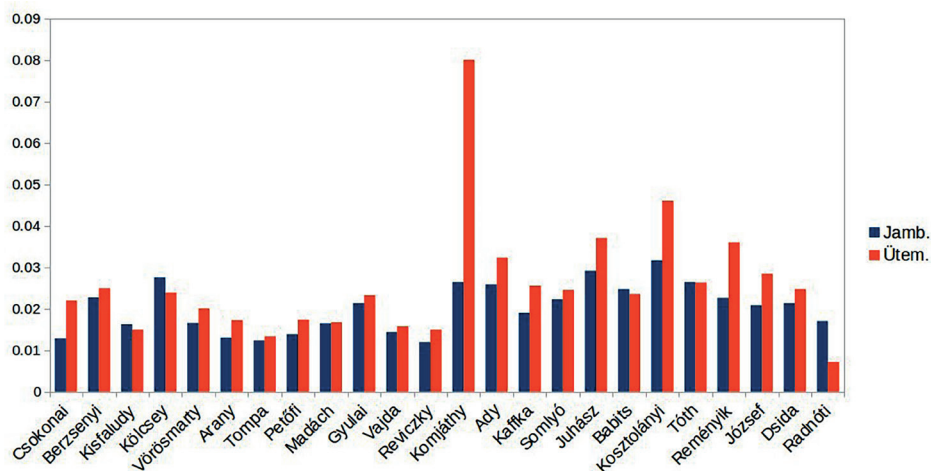


12. ábra. A szintaktikai ismétlődés mértékének mediánjai jambikus és ütemhangsúlyos verseknél, szófaji címkék 2-gramjai alapján, 50 tokenes ablakbeállítás



13. ábra. A szintaktikai ismétlődés mértékének mediánjai jambikus és ütemhangsúlyos verseknél, szófaji címkék 3-gramjai alapján, 50 tokenes ablakbeállítás

szimultaneitásról, „amikor ugyanazok a költemények egyszerre felelnek meg valamilyen idő-mértékes és ütemhangsúlyos képletnek, tehát kétféle ritmusúak”. SZEPES Erika és SZERDAHELYI István, *Verstan* (Budapest: Gondolat Kiadó, 1981), 510.



14. ábra. A szintaktikai ismétlődés mértékének mediánjai jambikus és ütemhangsúlyos verseknél, szófaji címkék 4-gramjai alapján, 50 tokenes ablakbeállítással

Ugyanezeket a méréseket lefuttattam kiterjesztett szófaji címkék alapján is. A kapott értékek úgyszintén a *Függelék* 11. táblázatában szerepelnek. A 6. táblázat mutatja be, hogy az egyes mérési beállításoknál a 24 szerzőből hány esetben nagyobb az ütemhangsúlyos versek szintaktikai ismétlődésének a mediánja a jambikus versek szintaktikai ismétlődésének a mediánjánál. A táblázatból láthatjuk, hogy minden mérési beállítás esetén a költők többségénél ez a mintázat jelenik meg.

6. táblázat. Az egyes mérési beállításoknál azon költők száma, akiknél az ütemhangsúlyos versekre kapott szintaktikai ismétlődés mediánja meghaladja a jambikus versekre kapott szintaktikai ismétlődés mediánját (összes költő = 24)

	Szófaji címkék			Kiterjesztett szófaji címkék		
	2-gram	3-gram	4-gram	2-gram	3-gram	4-gram
50 token	16	19	19	13	15	18

Az ütemhangsúlyos versek nagyobb mértékű szintaktikai ismétlődéséből arra következtethetünk, hogy az ütemhangsúlyos verselés nagyobb mértékben motiválja a szintaktikailag megegyező, illetve hasonló szerkezetek alkalmazását, mint a jambikus verselés. Ennek az lehet az oka, hogy míg a jambi-

kus verselés verslábai általában két szótagosak és nem feltétlenül szó elején kezdődnek, addig az ütemhangsúlyos versek ütemei jellemzően hosszabbak (például felező nyolcas esetében négy szótagosak, felező tizenkettes esetében pedig hat szótagosak), illetve az ütemek első szótagja általában szóeleji szótagra esik. Az ütemek nagyobb méretüknél fogva, illetve szavakat nem kérésztbe metsző jellegüknél fogva egy teljes szintaktikai frázist is magukban foglalhatnak. A sok esetben szintaktikai frázisoknak megfeleltethető ütemeknek egy meghatározott mintázat szerint való ismétlődése így motiválhatja a szintaktikai frázisok ismétlődését. Mindez a jambikus metrumról sokkal kevésbé mondható el, hiszen a jellemzően két szótagos és sok esetben nem szó eleji szótagnál kezdődő verslábak általában nem feleltethetők meg szintaktikai frázisoknak. Vagyis a verslábak valamilyen rendszer szerint megvalósuló ismétlődéséből kevésbé fog következni a szintaktikai frázisok valamilyen rendszer szerint megvalósuló ismétlődése.

Megjegyzendő, hogy az ütemek és a szintaktikai frázisok közötti kapcsolatra más terminológiát használva vagy éppen implicit módon régóta utalnak a magyar verstani művek. Már Arany János megfogalmazza *A magyar nemzeti vers-idomról* című művében, „hogy nem elég a verssort, úgy nevezett lábak és metszetek szerint, külsőleg feltagolni, azaz nem elég, hogy bizonyos helyeken végződjék a szó: hanem szükséges, hogy e külső felosztásnak megfelelően a benső: hogy a mondatnak (s így a gondolatnak is) azon részei csoportuljanak egy ütembe, egy, a sormetszet által elkülönzött ízbe, melyek egymással legszorosab kapcsolatban vannak, mert csak így lesz tökélyes a ritmusus”.³⁵ Horváth János úgyszintén felhívja a figyelmet arra, hogy „ha félsorból félsorba, vagy éppen sorból sorba törik át egy gondolati egység, kivált, ha az egyszersmind hangtani és mondattani egység (szólam, szószerkezet): a verstani határnak ezt az áthágását (az ú. n. *enjambement*-t) szabálysértésnek érezzük”.³⁶ Az ütemhangsúlyos metrum leírása során Vargyas Lajos is a szólamtagolás elvéből indul ki.³⁷ Megközelítésében a szólamok „a mondatnak, illetve a beszédnek szorosan összetartozó részei”,³⁸ amelyek lényegében szintaktikai frázisoknak feleltethetők meg. Preskriptívnek ható leírásában úgy fogalmaz, hogy „[ö]sszetartozó szólamokat széttagolhat [...] a vers, mégis jó lehet a ritmus,

³⁵ ARANY János, „A magyar nemzeti vers-idomról,” in ARANY János, *Tanulmányok és kritikák I.*, 288–320 (Debrecen: Debreceni Egyetemi Kiadó, 2012²), 296.

³⁶ HORVÁTH János, *Rendszeres magyar verstan* (Budapest: Akadémiai Kiadó, 1969 [1951]), 25–26.

³⁷ VARGYAS Lajos, *Magyar vers – magyar nyelv: Verstani tanulmány* (Budapest: Szépirodalmi Könyvkiadó, 1966).

³⁸ Uo., 11.

csak össze nem tartozókat ne kényszerítsen egy ütembe; vagyis két különböző szólamba tartozó szó találkozása egy ütemen belül hibás.”³⁹ Szuromi Lajos úgyszintén hangsúlyozza, hogy a több szóból álló ütemek általában szólam-egészt alkotnak és „fokozott értelmi tagoló” funkcióval rendelkeznek.⁴⁰ A szintaktikai megközelítés egyik legkövetkezetesebb alkalmazója Lotz János volt, aki a hangsúly szerepét másodlagosnak tekinti, és az ütemek határait mondattani határokkal hozza összefüggésbe.⁴¹ Hangsúlyozandó, hogy mindez nem jelenti azt, hogy az ütemeknek és a szintaktikai frázisoknak feltétlenül egybe kellene esniük. Az ütemhatárok keresztbe metszhetnek szintaktikai frázisokat anélkül, hogy azt valamiféle metrikai hibának kéne felfogni. Simon Gábor kognitív nyelvészeti kiindulópontú elemzésében részletesen bemutatja, hogy a szintaktikai frázisok és az ütemek közötti eltérés hogyan járulhat hozzá a vers komplex jelentésének a megképződéséhez.⁴²

6. A RÍMELES ÉS A SZINTAKTIKAI ISMÉTLŐDÉS ÖSSZEFÜGGÉSEI A MAGYAR KANONIKUS KÖLTÉSZETBEN

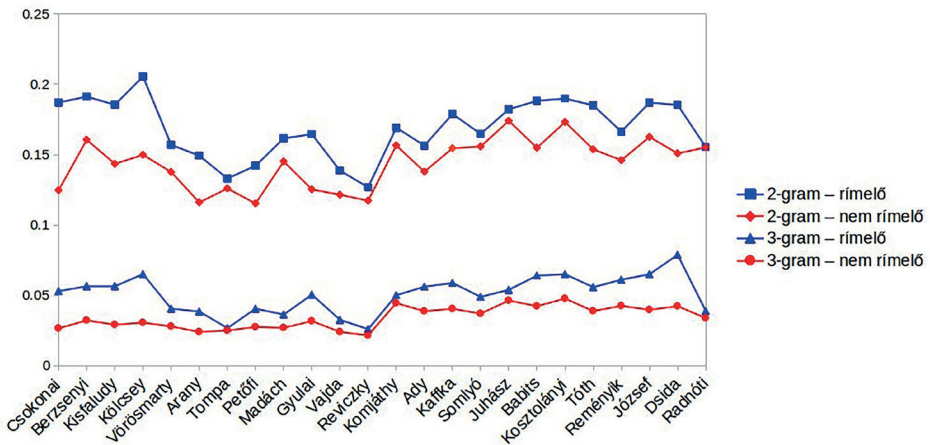
A vizsgálat utolsó kérdése az, hogy vajon a metrum mellett a fonológiai ismétlődésnek egy másik tényezője, a rímelés mutat-e összefüggést a szintaktikai ismétlődés mértékével. A kérdés megválaszolásához a közvetlenül egymást követő egymással rímelő sorok és a közvetlenül egymást követő egymással nem rímelő sorok átfedési együtthatóját számítottam ki sima szófaji címkék és kiterjesztett szófaji címkék 2- és 3-gramjai alapján. A rímelő és nem rímelő sorpárokra kapott értékeket minden szerző esetében átlagoltam. Az így kapott értékeket a *Függelék* 12. táblázata, illetve a 15. és a 16. ábra mutatja be. A vizsgálatba csak az 1000 szónál rövidebb verseket vettem bele annak érdekében, hogy az átlagok számítását ne torzítsák a hosszú elbeszélő költemények.

³⁹ Uo., 18.

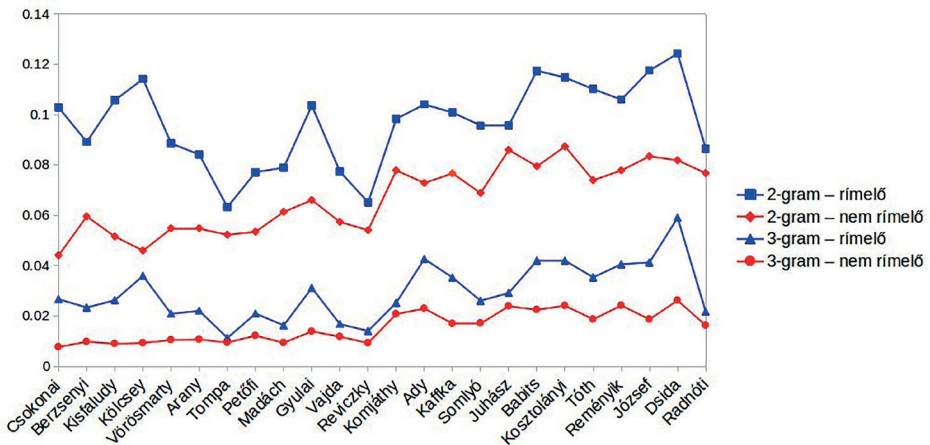
⁴⁰ SZUROMI Lajos, *A szimultán verselés* (Budapest: Akadémiai Kiadó, 1990), 58.

⁴¹ LOTZ János, „Az Uráli verselésről”, in LOTZ János, *Szonettkoszorú a nyelvről*, 237–257 (Budapest: Gondolat Kiadó, 1976), 238.

⁴² SIMON Gábor, „Az ütemhangsúlyos verselés kognitív poétikai magyarázata”, *Magyar Nyelvőr* 138, 4. sz. (2014): 460–478.



15. ábra. Egymást követő rímelő és nem rímelő sorok átfedési együtthatójának átlaga szófaji címkék 2-gramjai és 3-gramjai alapján



16. ábra. Egymást követő rímelő és nem rímelő sorok átfedési együtthatójának átlaga kiterjesztett szófaji címkék 2-gramjai és 3-gramjai alapján

Ahogy az a 12. táblázatból és az ábrákból is kiderül, a közvetlenül egymást követő rímelő sorok átfedési együtthatóinak az átlaga mindig meghaladja a közvetlenül egymást követő nem rímelő sorok átfedési együtthatóinak az átlagát. Mindez azt mutatja, hogy a rímeléssel megvalósuló fonológiai ismétlődés motiválja a szintaktikai ismétlődés megvalósulását. A vizsgálat eredménye tehát egy újabb bizonyítékát szolgáltatja annak, hogy a versekben megjelenő fonológiai és szintaktikai ismétlődés nem független egymástól.

Amennyiben a vizsgált szintaktikai ismétlődést sematikus szinten megvalósuló szemantikai ismétlődéssel azonosítjuk, akkor a rímelő sorpárok esetében a rímeléssel megjelenő fonológiai hasonlóság a sorok szintaktikai hasonlóságában megmutatkozó sematikus szemantikai hasonlóságnak az ikonikus leképeződése. Az utóbbi évtizedekben számos vizsgálat rámutatott a nyelv ikonikus természetére, azaz arra, hogy a nyelvi forma és a jelentés közötti szimbolikus kapcsolat nem teljesen önkényes, a forma, vagyis a nyelv fonológiai oldala valamilyen módon hasonlít az általa kifejezett tartalomra, azaz a nyelv szemantikai oldalára. A nyelv ikonikus természete nyilvánul meg például abban, hogy a szemantikailag közelebb álló elemek alakilag is közelebb helyezkednek el egymáshoz, hogy az információközlés szempontjából fontosabb elemek a mondatban, illetve a szövegben kitüntetett pozícióban állnak, vagy hogy a szemantikailag kevésbé komplex fogalmakat alakilag is kevésbé komplex módon fejezzük ki.⁴³ A rímelő sorpárok esetében a rímelés révén megjelenő fonológiai párhuzamosság és a szintaxison keresztül megmutatkozó szemantikai párhuzamosság tehát forma és tartalom egymást tükröző, ikonikus megfelelésre vezethető vissza. Ebben a viszonyban a rímelés egyszerre motiválja és egyszerre le is képezi a két sor szintaktikai hasonlóságában tetten érhető sematikus szemantikai hasonlóságot. Az eredmények tehát megerősítik azokat az elképzeléseket, miszerint a rím nem csupán fonológiai, hanem fonológiai és szemantikai jelenség.⁴⁴

⁴³ A nyelv ikonikus természetéről például lásd: John HAIMAN, „The Iconicity of Grammar: Isomorphism and Motivation”, *Language* 56, 3. sz. (1980): 515–540, <https://doi.org/10.2307/414448>; Claus D. PUSCH, „Ikonizität”, in *Language Typology and Language Universals: Handbücher zur Sprach- und Kommunikationswissenschaft*, Band 20.1., szerk. Martin HASPELMATH, Ekkehard KÖNIG, Wulf OESTERREICHER és Wolfgang RAIBLE, 369–384 (Berlin, New York: Walter de Gruyter, 2001); Willy van LANGENDONCK, „Iconicity”, in *The Oxford Handbook of Cognitive Linguistics*, szerk. Dirk GEERAERTS és Hubert CUYCKENS, 394–418 (Oxford: Oxford University Press, 2007); BENCZES, Rhyme over Reason...

⁴⁴ A rím jelentéskezdeményező szerepéről lásd SIMON Gábor, *Egy kognitív poétikai rímelmélet megalapozása* (Budapest: Tinta Könyvkiadó, 2014).

7. ÖSSZEGZÉS

A szófaji címkék n-gramjaira épülő vizsgálatokból kiderült, hogy a magyar kanonikus költészetnek a felfokozott fonológiai ismétlődés mellett a felfokozott szintaktikai ismétlődés is egy fontos sajátossága, hiszen a vizsgált szerzők többsége esetében a szintaktikai ismétlődés mértéke meghaladja a regények esetében kapott értéket. A vizsgálatokból azt is láthattuk, hogy a 20. század elején alkotó szerzők esetében megnő a szintaktikai ismétlődés mértéke a 19. században alkotó szerzők nagy részéhez képest. Összevetve a szintaktikai ismétlődés mértékének időbeli alakulását a jambikus és trochaikus metrumú versek metrikai szabályosságának időbeli alakulásával, amellelt érveltem, hogy a szintaktikai ismétlődés mértékének a növekedése a metrikai fellazulást, azaz a fonológiai ismétlődés mértékének a csökkenését kompenzálja. Más szavakkal a poétikai rendszer a versszerűséget azzal őrzi meg, hogy az ismétlődésnek egy adott dimenzióban történő visszaesését egy másik dimenzióban megvalósuló felerősítésével egyenlíti ki.

Emellett azt is sikerült igazolni, hogy a szintaktikai ismétlődés mértéke összefügg a versekben megjelenő fonológiai ismétlődés típusával. Láthattuk, hogy az ütemhangsúlyos metrumú versekben nagyobb mértékben jelenik meg a szintaktikai ismétlődés, mint a jambikus versekben. Ennek magyarázata az lehet, hogy szemben az időmértékes metrumok verslábaival, az ütemek sok esetben teljes szintaktikai frázisoknak feleltethetők meg, az ütemek ismétlődése így motiválhatja a szintaktikai frázisok ismétlődését.

Ugyancsak sikerült igazolni, hogy a fonológiai ismétlődés egy másik megvalósulása, a rímelés is összefüggésbe hozható a szintaktikai ismétlődéssel, hiszen a közvetlenül egymást követő rímelő sorok szintaktikailag jobban hasonlítanak egymásra, mint az egymást követő nem rímelő sorok. Ez a jelenség a rímelés és a szintaxis irányából is magyarázható. Egyrészt a rímeléssel megvalósuló fonológiai párhuzamosság feltételezhetően motiválja a szintaktikailag párhuzamos szerkesztésmódot. Másrészt, ha a szintaxisra sematikus szemantikaként tekintünk, akkor a fonológiai párhuzamosság a szintaktikai hasonlóságban testet öltő szemantikai párhuzamosságnak az ikonikus leképeződése.

A szintaktikai ismétlődés, csakúgy, mint a fonológiai ismétlődés, nem a költészet sajátossága, hanem magának a nyelvnek a sajátossága. Amellett érveltem, hogy a szintaktikai ismétlődés fokozott jelenléte a költészetben – hasonlóan a rímeléssel, ritmussal és egyéb eszközökkel megvalósított fonológiai ismétlődés felerősítéséhez – a nyelv poétikai funkcióját valósítja meg. A poétikus szöveg

esetében a fonológia, a szintaxis és minden bizonnyal a nyelv egyéb, itt nem vizsgált dimenzióiban is megvalósuló felfokozott ismétlés a nyelv egyik leg-alapvetőbb, ugyanakkor a hétköznapi nyelvhasználatban kevésbé szembetűnő mechanizmusát helyezi a figyelem előterébe.

FÜGGELÉK

7. táblázat. A regényekre és versekre kapott szintaktikai ismétlődés értékeinek mediánja szófaji címkék n-gramjai alapján

Szerző/ alkorpusz	100 token			50 token			25 token		
	2-gram	3-gram	4-gram	2-gram	3-gram	4-gram	2-gram	3-gram	4-gram
Regények	0.4237	0.1472	0.0383	0.2871	0.0814	0.0205	0.1671	0.0392	0.0096
Csokonai	0.4465	0.1511	0.0383	0.3087	0.0846	0.0194	0.1841	0.0422	0.0085
Berzsenyi	0.4843	0.1831	0.0444	0.3474	0.1041	0.0232	0.2069	0.0473	0.0091
Kisfaludy	0.4447	0.1468	0.0362	0.3085	0.083	0.0178	0.1843	0.0377	0.0071
Kölcsey	0.4704	0.1829	0.0529	0.3336	0.1042	0.0232	0.2083	0.0505	0.0096
Vörösmarty	0.4381	0.1451	0.0348	0.3044	0.0805	0.0169	0.1771	0.0373	0.0054
Arany	0.4076	0.1238	0.0295	0.2763	0.0687	0.0161	0.1608	0.0335	0.0068
Tómpa	0.4057	0.1238	0.0254	0.2738	0.0654	0.013	0.1569	0.0307	0.0057
Petőfi	0.408	0.1298	0.0302	0.2755	0.071	0.0153	0.1604	0.0344	0.0063
Madách	0.4432	0.1447	0.0329	0.3067	0.0792	0.0159	0.1758	0.0371	0.006
Gyulai	0.4309	0.145	0.0403	0.3006	0.085	0.022	0.1813	0.0445	0.0101
Vajda	0.421	0.1307	0.0297	0.2845	0.0717	0.0157	0.1653	0.036	0.0075
Reviczky	0.4035	0.1192	0.0266	0.2731	0.0652	0.0126	0.1568	0.0317	0.0053
Komjáthy	0.4573	0.1779	0.0545	0.3275	0.1067	0.0298	0.197	0.0558	0.016
Ady	0.4372	0.1594	0.0488	0.3002	0.0933	0.0264	0.18	0.0487	0.0127
Kaffka	0.4123	0.146	0.0403	0.2835	0.0862	0.0222	0.1723	0.0452	0.0117
Somlyó	0.4175	0.1531	0.0416	0.2954	0.0895	0.0239	0.1819	0.0464	0.011
Juhász	0.4676	0.1871	0.0583	0.3368	0.1117	0.0305	0.208	0.0571	0.0141
Babits	0.4267	0.1539	0.0417	0.3023	0.0925	0.0261	0.1883	0.0514	0.0142
Kosztolányi	0.4748	0.1948	0.0606	0.3412	0.1165	0.033	0.2096	0.058	0.0154
Tóth	0.4563	0.1738	0.0485	0.3253	0.1015	0.0262	0.199	0.0516	0.0123
Reményik	0.4122	0.1453	0.0412	0.291	0.0879	0.0227	0.178	0.0476	0.0131
József	0.4505	0.1622	0.0425	0.315	0.0945	0.0241	0.1897	0.0454	0.01
Dsida	0.4354	0.1628	0.0479	0.3081	0.0949	0.0246	0.1875	0.0497	0.0121
Radnóti	0.4632	0.1586	0.0412	0.3342	0.0885	0.021	0.1938	0.0385	0.0069

8. táblázat. A regényekre és versekre kapott szintaktikai ismétlődés értékeinek mediánja kiterjesztett szófaji címkék n-gramjai alapján

Szerző/ alkorpusz	100 token			50 token			25 token		
	2-gram	3-gram	4-gram	2-gram	3-gram	4-gram	2-gram	3-gram	4-gram
Regények	0.2340	0.0578	0.0130	0.1481	0.0324	0.0078	0.0815	0.0166	0.0044
Csokonai	0.2388	0.0509	0.01	0.1473	0.0276	0.0049	0.0822	0.0147	0.0013
Berzsenyi	0.264	0.0609	0.0109	0.1645	0.0332	0.0045	0.0931	0.0155	0.0003
Kisfaludy	0.2428	0.0569	0.0105	0.1662	0.033	0.0055	0.0906	0.0154	0.0009
Kölcsey	0.2525	0.0623	0.0113	0.1565	0.0335	0.0049	0.0857	0.0162	0

Szerző/ alkorpusz	100 token			50 token			25 token		
	2-gram	3-gram	4-gram	2-gram	3-gram	4-gram	2-gram	3-gram	4-gram
Vörösmarty	0.24	0.0529	0.0096	0.1496	0.0281	0.0033	0.0811	0.0123	0
Arany	0.2378	0.0528	0.0107	0.1508	0.0296	0.0055	0.083	0.0158	0.0016
Tómpa	0.2333	0.0498	0.0079	0.1463	0.027	0.0039	0.0775	0.0131	0.0017
Petőfi	0.2416	0.0558	0.0103	0.1516	0.0296	0.0045	0.0833	0.0146	0.0005
Madách	0.2536	0.0563	0.0101	0.1611	0.029	0.0045	0.0853	0.0135	0.0014
Gyulai	0.2762	0.0744	0.0183	0.1879	0.0434	0.0089	0.1122	0.0246	0.0054
Vajda	0.2581	0.0591	0.0104	0.1651	0.0328	0.0059	0.0931	0.017	0.0029
Reviczky	0.2453	0.0551	0.0103	0.1572	0.0294	0.0037	0.0869	0.0157	0.0012
Komjáthy	0.2853	0.0823	0.0236	0.1897	0.0525	0.0146	0.1119	0.0291	0.0085
Ady	0.2821	0.0857	0.0233	0.1884	0.0516	0.014	0.1115	0.0287	0.0068
Kaffka	0.2604	0.0713	0.0152	0.1699	0.0393	0.0095	0.0974	0.0243	0.0055
Somlyó	0.2553	0.0686	0.015	0.1675	0.0407	0.0087	0.0998	0.0229	0.0041
Juhász	0.2949	0.0905	0.0221	0.1992	0.0519	0.0117	0.1164	0.028	0.0051
Babits	0.2665	0.0754	0.02	0.1793	0.0458	0.0136	0.1078	0.0273	0.0082
Kosztolányi	0.3202	0.102	0.0284	0.2142	0.0594	0.0145	0.1236	0.0319	0.0071
Tóth	0.292	0.0833	0.019	0.198	0.0496	0.01	0.1132	0.0251	0.0048
Reményik	0.2559	0.0735	0.0201	0.174	0.0462	0.0116	0.1044	0.0277	0.0061
József	0.2918	0.0816	0.018	0.1929	0.0441	0.0075	0.11	0.0234	0.0027
Dsida	0.2751	0.0814	0.0206	0.1869	0.0474	0.0109	0.1084	0.0263	0.0049
Radnóti	0.2918	0.0714	0.0121	0.187	0.0376	0.0053	0.1029	0.0159	0

9. táblázat. A nem szimultán jambikus és trochaikus versek száma és szabályossági pontszámainak mediánja 0,5 és 0,3 szabályossági küszöbértékkel

Szerző	Összes	Jambikus				Trochaikus			
		0,5 <		0,3 <		0,5 <		0,3 <	
		db	Medián	db	Medián	db	Medián	db	Medián
Csokonai	394	46	0.83	64	0.76	50	0.72	118	0.46
Berzsenyi	136	20	0.7	45	0.49	1	0.56	19	0.36
Kisfaludy	123	19	0.74	21	0.74	11	0.69	22	0.53
Kölcsy	150	44	0.66	58	0.64	21	0.69	23	0.68
Vörösmarty	662	64	0.76	94	0.7	38	0.78	82	0.45
Arany	417	84	0.72	96	0.72	40	0.65	63	0.56
Tómpa	491	226	0.72	239	0.72	32	0.65	38	0.63
Petőfi	839	249	0.68	292	0.66	81	0.65	102	0.64
Madách	319	97	0.72	107	0.71	45	0.74	49	0.73
Gyulai	156	33	0.66	37	0.65	27	0.64	41	0.61
Vajda	199	71	0.64	82	0.63	41	0.63	54	0.61
Reviczky	335	223	0.78	223	0.78	19	0.7	20	0.69
Komjáthy	246	169	0.77	178	0.77	9	0.7	14	0.67
Ady	1116	203	0.57	585	0.46	35	0.6	107	0.45
Kaffka	102	38	0.66	52	0.6	4	0.56	6	0.53
Somlyó	379	155	0.63	202	0.6	10	0.58	40	0.43
Juhász	1278	666	0.69	720	0.69	14	0.6	37	0.47
Babits	514	200	0.67	246	0.65	33	0.61	62	0.52
Kosztolányi	630	440	0.74	471	0.74	19	0.76	44	0.46
Tóth	451	185	0.65	274	0.6	10	0.55	44	0.41
Reményik	667	458	0.59	605	0.56	1	0.61	4	0.35
József	599	184	0.61	287	0.56	21	0.58	69	0.41
Dsida	1060	295	0.58	634	0.5	9	0.66	67	0.37
Radnóti	298	60	0.75	112	0.52	4	0.54	32	0.36

10. táblázat. A minimum 50 tokenből álló versek száma az egyes metrumkategóriákban szerzőnként, 0,5 küszöbértékkel

	Összes	Jambikus	Trochaikus	Ütemh.	TrochSzim.	JambSzim.
Csokonai	338	37	45	123	8	6
Berzsenyi	122	20	0	14	0	14
Kisfaludy	109	16	11	28	15	4
Kölcsey	129	42	19	5	13	8
Vörösmarty	424	56	33	25	55	49
Arany	315	77	35	54	84	14
Tómpa	474	222	31	38	83	73
Petőfi	738	224	79	130	143	85
Madách	245	93	44	3	54	37
Gyulai	152	33	26	22	38	10
Vajda	190	69	39	14	27	14
Reviczky	317	216	18	3	35	41
Komjáthy	220	164	7	8	9	11
Ady	1075	197	32	111	66	48
Kaffka	101	37	4	6	8	3
Somlyó	377	155	10	28	10	62
Juhász	1190	649	12	39	84	315
Babits	502	196	32	40	24	44
Kosztolányi	582	417	17	10	6	54
Tóth	407	169	9	20	22	29
Reményik	639	440	1	14	6	28
József	552	178	19	53	24	51
Dsida	987	286	8	70	30	67
Radnóti	272	59	3	9	2	12

11. táblázat. 50 szavas ablakkal szófaji címkék és kiterjesztett szófaji címkék 2-, 3-, és 4-gramjaira kapott ismétlődési értékek mediánja metrumkategóriánként

Szerző	Szófaji címkék						Kiterjesztett szófaji címkék					
	2-gram		3-gram		4-gram		2-gram		3-gram		4-gram	
	Jamb.	Ütem.	Jamb.	Ütem.	Jamb.	Ütemh.	Jamb.	Ütem.	Jamb.	Ütem.	Jamb.	Ütemh.
Csokonai	0.3014	0.3203	0.0802	0.0914	0.0128	0.0219	0.1536	0.1489	0.0351	0.0276	0.0013	0.0055
Berzsenyi	0.3224	0.3478	0.0948	0.118	0.0227	0.0249	0.1578	0.1557	0.029	0.0378	0.005	0.0052
Kisfaludy	0.3175	0.293	0.0725	0.079	0.0162	0.0149	0.1436	0.1735	0.0273	0.0293	0.0052	0.0024
Kölcsey	0.3469	0.3373	0.1091	0.1001	0.0275	0.0238	0.1618	0.156	0.0405	0.0322	0.007	0.0064
Vörösmarty	0.2898	0.3109	0.0761	0.0888	0.0165	0.02	0.1575	0.1612	0.0336	0.0317	0.0056	0.007
Arany	0.2637	0.2848	0.0657	0.0735	0.013	0.0172	0.148	0.1525	0.028	0.0315	0.0049	0.006
Tómpa	0.2747	0.2691	0.0649	0.0644	0.0123	0.0133	0.144	0.143	0.0269	0.0249	0.0038	0.0039
Petőfi	0.2679	0.2831	0.067	0.0747	0.0138	0.0173	0.1493	0.1559	0.0274	0.0342	0.0042	0.0053
Madách	0.3089	0.3107	0.0789	0.081	0.0164	0.0167	0.1623	0.1579	0.0306	0.027	0.004	0.0046
Gyulai	0.3002	0.306	0.0831	0.087	0.0213	0.0232	0.185	0.1933	0.0417	0.0459	0.0079	0.0088
Vajda	0.2828	0.292	0.0716	0.0745	0.0143	0.0157	0.1752	0.1668	0.0341	0.0307	0.0058	0.0042
Reviczky	0.2709	0.2822	0.0633	0.0752	0.0119	0.0149	0.1552	0.1627	0.0295	0.0399	0.0039	0.0065
Komjáthy	0.321	0.3802	0.0983	0.1649	0.0264	0.08	0.1835	0.2195	0.0471	0.0629	0.0122	0.0279
Ady	0.3056	0.3008	0.0955	0.1027	0.0258	0.0323	0.1949	0.1802	0.0565	0.0554	0.0148	0.017
Kaffka	0.2816	0.328	0.0789	0.1111	0.019	0.0255	0.1617	0.1725	0.0372	0.0399	0.0091	0.0074
Somlyó	0.2938	0.3152	0.0849	0.1026	0.0222	0.0245	0.162	0.1785	0.0367	0.0452	0.0076	0.0096
Juhász	0.3402	0.3408	0.1108	0.1199	0.0291	0.037	0.1981	0.2041	0.0495	0.0546	0.0107	0.0109
Babits	0.2999	0.3096	0.0871	0.0961	0.0247	0.0235	0.1789	0.1807	0.0421	0.0514	0.011	0.015
Kosztolányi	0.341	0.3213	0.1141	0.1292	0.0316	0.046	0.2143	0.209	0.0579	0.0589	0.0141	0.0135

	Szófaji címkék						Kiterjesztett szófaji címkék					
	2-gram		3-gram		4-gram		2-gram		3-gram		4-gram	
Szerző	Jamb.	Ütem.	Jamb.	Ütem.	Jamb.	Ütemh.	Jamb.	Ütem.	Jamb.	Ütem.	Jamb.	Ütemh.
Tóth	0.334	0.3282	0.105	0.0974	0.0264	0.0263	0.2024	0.1922	0.0496	0.0527	0.009	0.0106
Reményik	0.2883	0.3097	0.0864	0.1065	0.0226	0.036	0.1737	0.1902	0.0452	0.0697	0.0113	0.0213
József	0.3025	0.3169	0.0838	0.1025	0.0208	0.0284	0.1828	0.1981	0.0407	0.0476	0.0073	0.0087
Dsida	0.3046	0.3008	0.0881	0.0867	0.0213	0.0247	0.1833	0.1711	0.0448	0.0419	0.0091	0.0111
Radnóti	0.3255	0.3117	0.0822	0.0808	0.017	0.0071	0.195	0.1874	0.0383	0.0329	0.0057	0.004

12. táblázat. A rímelő és nem rímelő sorpárok átfedési együtthatóinak átlaga az 1000 szónál rövidebb versekben

Szerző	Szófaji címkék				Kiterjesztett szófaji címkék			
	2-gram		3-gram		2-gram		3-gram	
	Rímelő	Nem rímelő	Rímelő	Nem rímelő	Rímelő	Nem rímelő	Rímelő	Nem rímelő
Csokonai	0.1871	0.1247	0.053	0.0265	0.1029	0.0441	0.0267	0.0077
Berzsenyi	0.1913	0.1606	0.0566	0.0322	0.0892	0.0596	0.0234	0.0098
Kisfaludy	0.1855	0.1435	0.0566	0.0291	0.1058	0.0516	0.0263	0.009
Kölcsey	0.2056	0.1499	0.0651	0.0306	0.1142	0.046	0.036	0.0093
Vörösmarty	0.157	0.1377	0.0405	0.0279	0.0886	0.0549	0.0209	0.0105
Arany	0.1493	0.1161	0.0385	0.024	0.0842	0.0549	0.0221	0.0107
Tómpa	0.1331	0.126	0.0267	0.025	0.0633	0.0523	0.0113	0.0095
Petőfi	0.1422	0.1153	0.0405	0.0275	0.0772	0.0535	0.021	0.0122
Madách	0.1616	0.1452	0.0363	0.0269	0.079	0.0614	0.0163	0.0094
Gyulai	0.1646	0.1253	0.0505	0.0317	0.1037	0.0661	0.0312	0.0139
Vajda	0.1387	0.1215	0.0326	0.024	0.0774	0.0574	0.0168	0.0118
Revczky	0.1268	0.1173	0.0259	0.0215	0.0652	0.0541	0.0141	0.0093
Komjáthy	0.1692	0.1566	0.0501	0.0445	0.0984	0.0779	0.0253	0.0208
Ady	0.1565	0.1379	0.0561	0.0387	0.1041	0.0729	0.0427	0.023
Kaffka	0.1789	0.1546	0.0588	0.0404	0.1009	0.0767	0.0353	0.017
Somlyó	0.1648	0.1558	0.0489	0.037	0.0956	0.0689	0.0261	0.0172
Juhász	0.1823	0.1741	0.0539	0.0463	0.0957	0.086	0.0292	0.0239
Babits	0.1883	0.1549	0.064	0.0423	0.1175	0.0795	0.0421	0.0225
Kosztolányi	0.1899	0.1732	0.065	0.0477	0.1148	0.0874	0.042	0.0241
Tóth	0.1851	0.1538	0.0558	0.0388	0.1102	0.074	0.0354	0.0187
Reményik	0.1664	0.1461	0.0611	0.0425	0.106	0.0779	0.0406	0.0242
József	0.187	0.1626	0.065	0.0397	0.1176	0.0835	0.0413	0.0187
Dsida	0.1854	0.1509	0.0788	0.0423	0.1243	0.0819	0.0591	0.0262
Radnóti	0.1554	0.1552	0.0388	0.0338	0.0865	0.0768	0.0217	0.0163