

**ZÁRT FOGALMI TARTOMÁNYOK A MAGYAR KANONIKUS
KÖLTÉSZETBEN – EGY KORPUSZNYELVÉSZETI MEGKÖZELÍTÉS¹**

**CLOSED CONCEPTUAL DOMAINS IN HUNGARIAN
CANONICAL POETRY – A CORPUS LINGUISTIC APPROACH**

HORVÁTH PÉTER²

Absztrakt: A tanulmány azt mutatja be, hogy zárt fogalmi tartományokhoz tartozó fogalmak gyakorisági megoszlásának a mérése hogyan járulhat hozzá költészeti korpuszok kvantitatív elemzéséhez. A módszer ismertetése 21 kanonikus magyar költő verseinek a vizsgálatával, három esettanulmány segítségével történik. Az első esettanulmány a NAPSZAKOK fogalmi tartományához tartozó ÉJSZAKA fogalmának a gyakorisági elemzését mutatja be. A második esettanulmány a négy évszakefogalom gyakoriságát, a harmadik esettanulmány pedig a színfogalmak gyakoriságát vizsgálja. Az elemzett fogalmak gyakoriságában történő időbeli változás tükrözi a magyar költészetnek a 20. század elején bekövetkező közismert poétikai változását. A tanulmány azt is bemutatja, hogy a vizsgált kanonikus magyar költők versei esetében az ÉJSZAKA, az ÉVSZAKOK és a SZÍNEK előfordulásának gyakorisága között erős pozitív korreláció van, amely azt jelezheti, hogy az ezekre történő utalás ugyanannak a poétikai eljárásnak más-más aspektusa, amelynek célja a versekben megképződő fizikai világ jellemzőiről való érzékszervi benyomások előtérbe helyezése. Végezetül a tanulmány bemutatja, hogy a három fogalmi tartomány fogalmai közül mely fogalompárok fordulnak elő együtt, ugyanazon versekben magasabb kölcsönös információértékkel.

Kulcsszavak: zárt fogalmi tartomány, költészet, korpusznyelvészet, korreláció, MI-érték, napszak, évszak, szín

Abstract: The paper shows how the distribution of different concepts within a closed conceptual domain can be used for quantitative analysis of poetic corpora. The method is presented in three case studies based on the poems of 21 canonical Hungarian poets. The first case study analyzes the frequency of the concept NIGHT, which is part of the conceptual domain TIMES OF DAY. The second case study analyzes the frequencies of the four season concepts, and the third case study assesses the frequencies of color concepts. The change in the frequencies of the concepts analyzed seems to reflect the well-known poetic change in Hungarian poetry at the beginning of the 20th century. The paper also demonstrates that in canonical Hungarian authors' poetry, there is a strong positive correlation between the frequencies of NIGHT, SEASONS and COLORS, which may mean that referring to these is part of the same

¹ A tanulmány a Digitális Örökség Nemzeti Laboratórium és az NKFIH K_21 137659 számú, *A személyjelölés konstrukcióinak korpuszalapú, kognitív poétikai vizsgálata* pályázat keretében készült.

² DR. HORVÁTH PÉTER
Eötvös Loránd Tudományegyetem BTK
horvath.peeteer@gmail.com

poetic toolkit aiming to highlight sensory impressions of the physical setting of poems. Finally, the paper shows which concepts from the three different conceptual domains co-occur in the same poems with a higher mutual information score.

Keywords: *closed conceptual domain, poetry, corpus linguistics, correlation, MI-score, time of day, season, color*

1. Bevezetés

A tanulmány versek vizsgálatának egy kvantitatív, korpusznyelvészeti módszerét mutatja be, amelynek az alapja zárt fogalmi tartományokhoz tartozó fogalmak gyakoriságának a mérése. A bemutatandó módszer hozzájárulhat a költészet távoli olvasását megcélzó különböző eljárások szélesítéséhez (a távoli olvasásról lásd Moretti 2000, 2005). Eddig viszonylag kevés magyar költészetre vonatkozó korpusznyelvészeti vizsgálatot végeztek. Ezek vagy egy magyar költő összes versének (lásd Jékel–Papp 1974, Jékel–Szuromi 1980, Lesi 2008, Labádi 2018) vagy pedig több költő néhány versének a vizsgálatára irányultak (lásd Fónagy 1959, Zsilka 1974, Simon 2020, Maróthy–Seláf–Plecháč 2021). Az utóbbi évek technológiai fejlődése és a digitalizáció előrehaladása lehetővé teszi, hogy számos magyar költő összes versét vizsgáljuk korpusznyelvészeti eszközökkel. Az alább bemutatandó vizsgálatnak is az az egyik célja, hogy az eddigieknél jóval nagyobb korpusz használatával mutasson rá a magyar költészet kvantitatív vizsgálatában rejlő új lehetőségekre.

A zárt fogalmi tartomány terminust olyan jól körülhatárolható fogalmi mezőkre használom, amelyek feloszthatók véges számú fogalomra. Például a színek zárt fogalmi tartományát a színnevekkel kifejezhető véges számú fogalom alkotja. A zárt fogalmi tartományokhoz tartozó fogalmak előfordulásának a mérése alapján gyakorisági adatokhoz juthatunk a zárt fogalmi tartománynak egy adott költészeti korpuszban való megjelenési módjáról, és különböző alkorpuszokat hasonlíthatunk össze ezen gyakorisági adatok alapján. A korpusznyelvészeti vizsgálatokban fogalmak előfordulásának kvantitatív vizsgálata nem újdonság. Például Rayson (2008) a kulcsszavak vizsgálatára kidolgozott módszert használta fogalmi tartományok előfordulásának mérésére (lásd még McIntyre 2013). Zárt fogalmi tartományok fogalmainak a megoszlása versek egy adott csoportjában olyan jellemzőkre irányíthatja a figyelmünket, amelyek a közeli olvasás technikáival nem tárhatók fel. Ezek a jellemzők többnyire a versekben megképződő fiktív világ fizikai, társas és mentális szférájának az általános tulajdonságaival függnek össze (vö. Tátrai 2011: 171–189, 2015).³

A tanulmány a módszer bemutatása során három zárt fogalmi tartomány fogalmainak a gyakoriságát vizsgálja 21 magyar költő 300 szónál rövidebb verseiben. A három zárt fogalmi tartomány a következő: NAPSZAKOK, ÉVSZAKOK és SZÍNEK. Ezek a tartományok a versek fizikai világához kapcsolódnak, az elbeszélt jelenetek, illetve a fiktív megnyilatkozó és az általa megszólított fiktív befogadó kommunikációjának

³ Tátrai (2011, 2015) ezt a hármasfelosztású modelljét narratív szövegekre alkalmazza, de véleményem szerint a modell lírai szövegek esetében is hasznosítható.

a keretét adó fiktív lírai beszédhelyzet fizikai jellemzőinek a megalkotásához járulnak hozzá (a lírai beszédhelyzet aposztrofikus jellegéről lásd Culler 1981: 135–154, Tátrai 2015). A három fogalmi tartomány vizsgálatával ennél fogva a vizsgált versekben megképződő fizikai világról juthatunk ismeretekhez, különböző verscsoportoknak a három tartomány gyakorisági adatai alapján való összevetése révén pedig a verscsoportok fizikai világának tipikus és atipikus mintázatait tárhatjuk fel.

A 2. rész bemutatja a vizsgálati korpusz legfontosabb jellemzőit, valamint a gyakorisági vizsgálatához használt eszközöket. A 3. rész a NAPSZAKOK zárt fogalmi tartományhoz tartozó ÉJSZAKA fogalmának a gyakorisági adatait mutatja be. A 4. rész az ÉVSZAKOK, az 5. rész pedig a SZÍNEK zárt fogalmi tartományhoz kapcsolódó gyakorisági adatokat ismerteti. A 6. rész bemutatja, hogy az ÉJSZAKA, az ÉVSZAKOK és a SZÍNEK előfordulásának gyakorisága között erős pozitív korreláció van. A 7. rész a három fogalmi tartomány fogalmain végzett együttes előfordulási elemzés eredményeit ismerteti. Végezetül a 8. rész röviden összefoglalja a korpuszvizsgálat eredményeit.

2. A korpusz és a vizsgálat eszközei

A gyakorisági elemzésekhez használt vizsgálati korpusz 21 magyar költő verseiből áll össze. Ezt a vizsgálati korpuszt az ELTE Verskorpuszból⁴ szedtem ki, amely jelenleg 49 kanonikus magyar költő összes versét tartalmazza (Horváth et al. 2022). Az ELTE Verskorpuszban szereplő azon költők verseit használtam, akiknek legalább 100 darab 300 szónál rövidebb versük van a korpuszban, illetve akik Csokonainál nem születtek régebben.⁵ A 300 szavas határral az volt a célom, hogy kiszűrjem az elbeszélő költeményeket a vizsgálati korpuszból, mivel a vizsgálat a magyar kanonikus líra jellemzőinek a feltárására irányult. A korpusz formátuma TEI XML (TEI Consortium 2019). A TEI XML fájlok nem csupán a versek szövegeit tartalmazzák, hanem egyéb típusú annotációk mellett a szavak automatikusan felismertett lemmáját, szófaját és morfoszintaktikai jellemzőit is. Ezek a grammatikai jellemzők az e-magyar programmal lettek felismertetve, amely egy NLP-eszköz magyar nyelvű szövegek grammatikai jellemzőinek az automatikus elemzésére (Váradai et al. 2018, Indig et al. 2019). A 21 szerző verseit tartalmazó vizsgálati korpusz 10 015 versből és 1 055 726 szóból áll. Az 1. táblázat mutatja be a korpuszban szereplő 21 költőt a születési és halálozási évükkel, valamint a szerzőkhöz tartozó versek és szavak számát. A szerzők születési sorrendben szerepelnek, a későbbi táblázatokban is ezt a sorrendet követi a szerzők feltüntetése.

A zárt fogalmi tartományok fogalmainak a gyakorisági adatait egy Python programozási nyelvben írt szkript segítségével nyertem ki a korpuszból. A szkripttel ki-nyert gyakorisági adatok további elemzését táblázatkezelő programban végeztem el. A vizsgált fogalmakra vonatkozó keresési kifejezéseket manuálisan gyűjtöttem össze. Ehhez a saját kutatói intuícióm mellett különböző szinonimaszótárakat használtam.

⁴ <https://github.com/ELTE-DH/poetry-corpus>.

⁵ Az ELTE Verskorpuszban Csokonai az első szerző, aki a nyelvtörténetben szimbolikus korszakhatárnak számító 1772-es év után született, és 100-nál több 300 szónál rövidebb verse szerepel a korpuszban.

1. táblázat. A vizsgálati korpusz tartalma

Szerző	Születési és halálozási év	Versek száma	Szavak száma
Csokonai Vitéz Mihály	1773–1805	292	33 174
Berzsenyi Dániel	1776–1836	125	15 258
Kölcsey Ferenc	1790–1838	140	13 878
Vörösmarty Mihály	1800–1855	596	47 490
Arany János	1817–1882	316	30 441
Tompa Mihály	1817–1868	273	43 419
Petőfi Sándor	1823–1849	787	87 572
Madách Imre	1823–1864	250	27 049
Gyulai Pál	1826–1909	144	17 549
Vajda János	1827–1897	149	21 236
Reviczky Gyula	1855–1889	305	34 919
Komjáthy Jenő	1858–1895	206	31 593
Ady Endre	1877–1919	1088	107 830
Somlyó Zoltán	1882–1937	360	38 268
Juhász Gyula	1883–1937	1267	108 152
Babits Mihály	1883–1941	458	62 450
Kosztolányi Dezső	1885–1936	591	62 680
Tóth Árpád	1886–1928	425	50 321
Reményik Sándor	1890–1941	644	73 168
József Attila	1905–1937	578	54 933
Dsida Jenő	1907–1938	1021	94 346
Összes	1773–1941	10 015	1 055 726

3. A NAPSZAKOK fogalmi tartománya

A napokat különböző időintervallumokra osztjuk fel. A napot két részre osztó két legnagyobb időintervallum a nappal és az éjszaka, amelyekhez a Nap pozíciójából adódó eltérő fizikai jellemzők tartoznak. Fogalmi rendszerünkben alapbeállításnak feltételezhetően a NAPPAL tekinthető, hiszen életünk aktívabb része jellemzően erre a napszakra esik. A NAPPAL alapbeállítás voltát az is mutatja, hogy amikor valaki egy vele megesett történetet, eseményt beszél el, és nem adja meg a történet idejét, hajlamosak vagyunk azt gondolni, hogy az eseményekre nappal került sor, nem pedig éjszaka. Ez alól persze kivételt jelentenek az olyan speciális események, amelyek tipikusan este vagy éjszaka zajlanak. A NAPPAL alapbeállítás jellege feltehetőleg versek olvasása során is érvényesül. Azaz ha a fiktív lírai megnyilatkozó nem specifi-

kálja a napszakot, általában olvasóként nem feltételezzük, hogy a fiktív lírai beszédhelyzet vagy a lírai megnyilatkozó által elbeszélte események ideje éjszakára esik. Ezt többnyire csak abban az esetben gondoljuk, ha a vers fiktív megnyilatkozója explicitte teszi valamilyen nyelvi kifejezés segítségével az ÉJSZAKA napszakját. Az ÉJSZAKA napszakjának a versbeli megjelenítése így tartalmilag jelöltte teszi a verseket, hiszen mind a fiktív megnyilatkozás, mind az elbeszélte események idejének éjszaka volta eltérés az idő alapbeállítás szerinti értelmezésétől. Érdekes kérdés, hogy a magyar kanonikus költészet szerzői az ÉJSZAKA napszakjának nyelvi jelölésével milyen mértékben tértek el költészetükben az időre vonatkozó alapbeállítástól.

Az első esettanulmány azt vizsgálja, hogy a vizsgálati korpuszban az egyes szerzők esetében milyen gyakorisággal jelennek meg azok a versek, amelyekben szerepel az ÉJSZAKA fogalma, azaz amelyekben eltérés jelentkezik az idő alapbeállítás szerinti értelmezéséhez képest. A vizsgálatban keresési kifejezéseként az ÉJSZAKA fogalmára vonatkozó szinonim szavakat használtam. Mivel a korpuszban a szavak lemmája is meg van adva, keresési kifejezéseként elég volt a szavak lemmáját, azaz a szótári alakját listázni. A listázott lemmákat az (1)-ben adtam meg. A sztenderd alakok mellett listáztam a régies alakokat, illetve a bevett írásmódtól eltérő alakokat is. Úgyszintén szerepeltettem a keresési kifejezések között az igék, melléknévi ige-nevek és az igékből képzett *-ás/-és* képzős főnevek *be-* és *rá-* igekötős alakjait (pl.: *beesteledik, ráesteledik*). Ezeket a lista áttekinthetősége érdekében nem tüntettem fel az (1)-ben.

(1) *este, estve, est, éjszaka, éjjszaka, éjjel, éjel, éj, éjj, éjjél, éjjfél, alkony, alkon-
nyat, alkonyulat, szürkület, estefelé, estefele, esti, estvei, éjszakai, éjjszakai, éjjeli,
éjeli, éji, éjji, éjféli, éjjféli, alkonyi, alkonyati, alkonyulati, szürkületi, esteledik,
estveledik, alkonyodik, alkonyul, sötétedik, esteledő, estveledő, alkonyodó, alko-
nyuló, esteledett, estveledett, alkonyodott, alkonyult, esteledés, estveledés, alko-
nyodás, alkonyulás, sötétedés*

Ahogy az (1) lista is mutatja, a módszer grammatikai kategóriákat nem vesz figyelembe. A listában főnevek, melléknévek, melléknévi ige-nevek és igék egyaránt szerepelnek.⁶ A cél az volt, hogy lehetőség szerint (közel) az összes ÉJSZAKÁRA vonatkozó lemmát listázzam, függetlenül a lemma szófaji kategóriájától. A gyakorisági adatok kinyeréséhez írt szkript a futtatás során végigmegegy a korpusz verseinek az XML-fájljain, és megnézi, hogy a vers tartalmazza-e bármelyik lemmát az (1)-ben felsoroltak közül. A szkript kimenetként minden szerző esetében megadja azoknak a verseknek a számát és arányát, amelyek legalább egyet tartalmaznak az (1) listában szereplő lemmák közül. Ezek a versekre vonatkozó gyakorisági adatok jelzik azt, hogy az adott szerző milyen mértékben tért el az idő alapbeállítás szerinti értelmezésétől. A kapott gyakorisági adatokat a 2. táblázat mutatja be.

⁶ A határozói ige-nevek lemmája az e-magyar elemzésében az igei alak, úgyhogy ezeket nem kellett külön listázni a keresőkifejezések között.

2. táblázat. ÉJSZAKÁRA utaló versek gyakorisága

Szerző	Versek száma	Versek száma, amelyekben szerepel ÉJSZAKA	ÉJSZAKA %	Rangsor
Csokonai	292	41	14,0%	20
Berzsenyi	125	24	19,2%	15
Kölcsey	140	54	38,6%	4
Vörösmarty	596	83	13,9%	21
Arany	316	52	16,5%	17
Tompa	273	114	41,8%	2
Petőfi	787	180	22,9%	12
Madách	250	40	16,0%	18
Gyulai	144	43	29,9%	8
Vajda	149	36	24,2%	10
Reviczky	305	54	17,7%	16
Komjáthy	206	32	15,5%	19
Ady	1088	220	20,2%	14
Somlyó	360	145	40,3%	3
Juhász	1267	475	37,5%	6
Babits	458	131	28,6%	9
Kosztolányi	591	270	45,7%	1
Tóth	425	163	38,4%	5
Reményik	644	150	23,3%	11
József	578	118	20,4%	13
Dsida	1021	378	37,0%	7
Átlag			26,7%	
Medián			23,3%	

A táblázat 3. oszlopában látható azoknak a verseknek a száma, amelyekben megjelenik az ÉJSZAKA fogalma. Ezeknek a verseknek az arányát az adott szerző összes verséhez képest a 4. oszlop mutatja be. Az utolsó oszlopban azt tüntettem fel, hogy az egyes szerzők az ÉJSZAKÁRA utaló versek arányai alapján létrehozott rangsorban hányadik helyen vannak. A táblázatból látható, hogy arányaiban Kosztolányinak van a legtöbb ÉJSZAKÁRA utaló verse, a versek 45,7%-a tartalmaz legalább egyet az (1)-ben listázott lemmák közül. A második helyen Tompát találjuk, akinél a versek 41,8%-a tartalmaz ÉJSZAKÁRA vonatkozó utalást. Arányaiban a legkevesebb ÉJSZAKÁRA utaló verset Vörösmarty írta. Az ő verseinek csupán a 13,9%-a sorolódott ebbe a kategóriába. Az átlagot és a mediánt úgyszintén feltüntettem a táblázatban. Az átlag 26,7%, a medián

pedig 23,3%. A táblázatból látható, hogy az ÉJSZAKÁRA utaló versek aránya a 20. század eleji költőknél jellemzően magasabb, mint a korábbi költőknél. Somlyó, Juhász, Babits, Tóth és Dsida esetében is meghaladják az átlagot és a mediánt ezek az arányszámok. Meglepő módon Ady viszont arányaiban jóval kevesebb ÉJSZAKÁRA utaló verset írt, mint az említett szerzők. Ady költészete ebből a szempontból sokkal inkább hasonlít Vajda, Reviczky és Komjáthy 19. század végi költészetére.

4. Az ÉVSZAKOK fogalmi tartománya

A második rövid esettanulmány az ÉVSZAKOK fogalmi tartományának a megjelenését vizsgálja. Ebben az esetben nem beszélhetünk valós tapasztalatokban gyökerező alapbeállításról. Nem igazán mondhatjuk, hogy a négyből az egyik évszak jelentősebb szerepet játszik az emberi életben, mint a többi. Ugyanakkor az viszont lehetséges, hogy egy adott költészeti hagyományban vagy egy költészeti hagyomány bizonyos időszakaiban valamelyik évszak kitüntetett szerepet kap, azaz gyakrabban történik rá utalás, mint a többi évszakra. Az is vizsgálható, hogy egy adott költészeti hagyományban létezik-e valamilyen tipikus gyakorisági sorrendje az évszakok versekben való használatának, és persze azt a kérdést is felvethetjük, hogy mely szerzők térnek el az évszakok használatának tipikus mintázataitól. Az ilyen eltérések a szerzők poétikájának egyedi, a versek szoros olvasáson alapuló elemzésével nem feltétlenül feltárható jellemzőire irányítják a figyelmet. Az évszakokra vonatkozó gyakorisági vizsgálat során a négy évszakra utaló, (2)-ben feltüntetett lemmákat használtam keresési kifejezéseként. A lekérdezésnél nemcsak a lemmákra, hanem a zárójelben megadott szófaji kategóriákra is szűrtem. Ennek köszönhetően az *ősz* lemmának a melléknévi, a haj színére vonatkozó használatait ki tudtam szűrni.

(2)

TAVASZ: *tavas*z (főnév), *tavaszi* (melléknév), *tavas*zodik (ige, határozószó⁷), (*határozószó*), *tavas*zodó (melléknév⁸), *tavas*zodott (melléknév), *tavas*zodás (főnév), *kitavas*zodik (ige, határozószó), *kitavas*zodó (melléknév), *kitavas*zodott (melléknév), *kitavas*zodás (főnév), *kikelet* (főnév), *kikeleti* (melléknév)

NYÁR: *nyár* (főnév), *nyári* (melléknév)

ŐSZ: *ősz* (főnév), *őszi* (melléknév)

TÉL: *tél* (főnév), *téli* (melléknév)

A szkript végigment minden egyes szerző versein, és mindegyik évszak esetében megadta, hogy az adott szerzőnek hány olyan verse van, amely legalább egy, az adott évszakra vonatkozó lemmát tartalmaz a (2)-ben feltüntetett lemmák közül. Természetesen egy vers több évszakkategóriába is besorolódhatott, amennyiben a versben szereplő lemmák több évszakra is utalnak. A kapott gyakorisági adatokat a 3. táblázat mutatja be.

⁷ Az e-magyar a határozói igeneveket határozószóként elemzi, de lemmaként az igei alakot adja meg.

⁸ Az e-magyar a melléknévi igeneveket melléknevekként címkézi fel.

3. táblázat. ÉVSZAKOKRA utaló versek gyakorisága

Szerző	Versek száma	Évszakok %	1.		2.		3.		4.	
			évszak	%	évszak	%	évszak	%	évszak	%
Csokonai	292	13,7%	tavas	8,9%	nyár	4,5%	tél	4,1%	ősz	1,7%
Berzsenyi	125	12,0%	tavas	9,6%	ősz	2,4%	nyár tél	0,8%		
Kölcsy	140	15,7%	nyár	7,1%	tavas	6,4%	tél	4,3%	ősz	1,4%
Vörösmarty	596	9,7%	tavas	6,0%	tél	4,4%	ősz	2,0%	nyár	1,3%
Arany	316	13,6%	tél	7,3%	nyár	6,3%	tavas	5,1%	ősz	3,8%
Tompai	273	34,4%	tavas	19,8%	tél	12,5%	ősz	12,1%	nyár	8,4%
Petőfi	787	16,0%	tavas	9,5%	ősz	5,8%	tél	5,6%	nyár	2,4%
Madách	250	19,6%	tavas	13,2%	ősz	5,2%	nyár	4,4%	tél	4,0%
Gyulai	144	24,3%	tavas	17,4%	ősz	9,7%	nyár	5,6%	tél	3,5%
Vajda	149	18,1%	nyár	9,4%	tavas tél	6,0%	ősz	2,0%		
Revczky	305	25,2%	tavas	14,1%	nyár	9,5%	ősz	7,2%	tél	3,6%
Komjáthy	206	13,1%	nyár	7,3%	tavas	4,9%	tél	1,9%	ősz	1,5%
Ady	1088	16,3%	tavas	5,9%	nyár	5,1%	ősz	5,0%	tél	4,7%
Somlyó	360	28,3%	ősz tél	10,8%	nyár	8,6%	tavas	6,1%		
Juhász	1267	30,9%	tavas	14,7%	ősz	11,5%	nyár	8,3%	tél	5,8%
Babits	458	22,5%	tavas	9,4%	tél	9,0%	nyár	7,0%	ősz	5,9%
Kosztolányi	591	22,5%	nyár	8,8%	ősz	8,6%	tél	8,0%	tavas	3,7%
Tóth	425	21,2%	tavas	8,2%	ősz	7,8%	nyár	5,6%	tél	4,0%
Reményik	644	19,6%	ősz	10,6%	tavas	6,2%	tél	5,0%	nyár	4,7%
József	578	13,8%	nyár	5,2%	tél	4,7%	ősz	4,0%	tavas	3,3%
Dsida	1021	26,7%	ősz	12,4%	tavas	9,7%	nyár	7,2%	tél	4,9%
Átlag		19,9%								
Medián		19,6%								
I.			tavas (12)		ősz (7)		tél (7), nyár (7)		tél (7)	
II.			nyár (5)		nyár (5), tél (5), tavasz (5)		ősz (6)		ősz (5)	
III.			ősz (3)				tavasz (2)		nyár (4)	
IV.			tél (2)						tavasz (2)	

A 3. táblázat második oszlopában tüntettem fel az adott szerzőnek a korpuszban szereplő összes versét. A harmadik oszlopban látható azoknak a verseknek az aránya, amelyekben legalább egy ÉVSZAKRA vonatkozó szó szerepel. A középértékeket úgy-szintén feltüntettem. Az átlag 19,9%, a medián pedig 19,6%. A gyakorisági adatok azt mutatják, hogy a 20. század előtti költők többnyire az átlag és a medián alatt vannak. Ezeknek a szerzőknek a nagy része arányaiban kevesebb ÉVSZAKRA utaló

verset írt, mint az őket követő 20. századi szerzők többsége. A legalacsonyabb gyakorisági értéket Vörösmarty és Berzsenyi esetében találjuk. Arányaiban a legtöbb évszakra utaló verset Tompa írta, a második legtöbbet pedig Juhász.

A 3. táblázat azt is bemutatja, minden szerző esetében csökkenő sorrendben, hogy a szerzők versei milyen arányban utalnak az egyes évszakokra. Például Csokonai esetében a versek 8,9%-a utal TAVASZRA, 4,5%-a utal NYÁRRÁ, 4,1%-a utal TÉLRE és 1,7%-a utal ŐSZRE. Az utolsó négy sorban az egyes évszakok rangsorát tüntettem fel a négy gyakorisági pozícióban. Ezekből az adatokból látható, hogy 12 szerző esetében a TAVASZ, 5 szerző esetében pedig a NYÁR jelenik meg a legtöbb versben.⁹ Ezekből az adatokból az is látszik, hogy a meleg évszakok népszerűbbek, mint a hideg évszakok. Négy szerző különbözik el ettől a mintázattól: Arany, Somlyó, Reményik és Dsida. Az ő esetükben a két hideg évszak valamelyike, a TÉL vagy az ŐSZ jelenik meg a legtöbb versben, illetve Somlyónál a két hideg évszak azonos gyakorisággal áll az élen. Feltehető, hogy a négy szerzőből három 20. századi szerző. Úgyszintén érdemes megjegyezni, hogy a vizsgált szerzők esetében a 19. század közepéig többnyire a TAVASZ volt a legnépszerűbb évszak. Ezt követően a leggyakoribb évszakok változatosabb képet mutatnak. A második gyakorisági pozícióban az ŐSZ áll az élen, amely hét szerző esetében is a második leggyakrabban használt évszak.

Végezetül megemlítendő, hogy az évszakoknak az egyes szerzőknél megfigyelhető gyakorisági sorrendjei tipikus mintázatot mutatnak: amennyiben egy szerzőnél a leggyakrabban használt évszak meleg évszak (TAVASZ vagy NYÁR), akkor a második leggyakrabban használt évszak jellemzően hideg évszak (TÉL vagy ŐSZ), illetve amennyiben egy szerzőnél a leggyakrabban használt évszak hideg évszak, akkor a második leggyakrabban használt évszak meleg évszak. Ez a mintázat a vizsgált 21 szerzőből 15-re érvényes. Az, hogy az első két leggyakrabban használt évszak a legtöbb szerzőnél ellentétpárt alkot, minden bizonnyal azzal magyarázható, hogy az évszakok sok esetben a versekben egymással szembeállított, egymás ellentétéként értelmezett időszakok, élethelyzetek, érzelmek metaforikus, illetve allegorikus megjelenítésére szolgálnak. Megjegyzendő, hogy a mintázatot nem követő hat szerzőből három, Vajda, Reviczky és Komjáthy azonos időszakban alkotott, illetve a mintázattól úgyszintén eltérő Ady időben követi ezeket a szerzőket, ami felveti a lehetőségét annak, hogy e négy szerző esetében a mintázattól való elkülönülődést ne a szerzők egyéni poétikájával, hanem az adott irodalomtörténeti korszak poétikájával magyarázzuk. Az ő esetükben egyébként a két leggyakrabban használt évszak egyaránt meleg évszak, illetve Vajda esetében második helyen azonos előfordulással szerepel a TAVASZ és a TÉL.

5. A SZÍNEK fogalmi tartománya

A harmadik esettanulmány a SZÍNEK fogalmi tartományhoz tartozó fogalmak gyakoriságát vizsgálja. A megválaszolandó kérdések hasonlóak, mint az előző két esetben:

⁹ Amennyiben egy szerzőnél két évszak is ugyanazzal a gyakorisággal jelenik meg, akkor mind a két évszakot hozzászámoltam az adott gyakorisági pozícióban való évszak-előfordulásokhoz.

Mely szerzők használtak nagyobb mértékben színfogalmakat a költészetükben? Vannak-e a magyar kanonikus költészetben tipikusabban használt színek? Mi az időbeli megoszlása a leggyakrabban használt színeknek? A vizsgált, színekre utaló lemmákat a (3)-ban tüntettem fel. A sztenderd alakok mellett listáztam az archaikus alakokat és az írásváltozatokat is. A keresőkifejezéseként listázott lemmák között szerepeltettem a (3)-ban felsorolt igéknek az *el-*, *meg-*, *be-*, *ki-*, *át-* és *össze-* igekötőkkel való előfordulásait (pl. *megfeketedik*, *átfeketül*), az igekötő nélküli és igekötős igékből képzett folyamatos és befejezett melléknévi igeneveket (pl. *feketedő*, *megfeketedő*, *feketedett*, *megfeketedett*), valamint az igekötő nélküli és igekötős igékből képzett *-ás/-és* képzős főneveket is (pl. *feketülés*, *megfeketülés*, *elfeketedés*). Ezeket a lemmákat a (3) listában a jobb áttekinthetőség érdekében nem tüntettem fel.

(3)

FEKETE: *fekete*, *feketeség*, *feketés*, *feketesség*, *feketül*, *feketedik*, *feketél*, *feketél*, *feketellik*, *feketít*, *feketit*

FEHÉR: *fehér*, *fehérség*, *fehéres*, *fehéresség*, *fehérül*, *fehéredik*, *fehérel*, *fehérell*, *fehérellik*, *fehérlík*, *fehérít*, *fehérít*

SZÜRKE: *szürke*, *szürkesség*, *szürkés*, *szürkesség*, *szürkül*, *szürkülődik*, *szürkél*, *szürkél*, *szürkellik*, *szürkít*, *szürkít*

PIROS: *piros*, *pirosság*, *pirosas*, *pirosasság*, *pirosul*, *pirosodik*, *pirosol*, *pirosoll*, *pirosollik*, *piroslik*, *pirosít*, *pirosít*, *vörös*, *vörösség*, *vöröses*, *vörösség*, *vörösül*, *vörösödik*, *vörösöl*, *vörösöll*, *vörösölik*, *vörösít*, *vörösít*, *vörösedik*, *vörösel*, *vörösell*, *vörösellik*, *vöröslík*, *vöres*, *vöresség*, *vöreses*, *vöresség*, *vöresül*, *vöresedik*, *vöresel*, *vöresell*, *vöresellik*, *vöreslik*, *vöresít*, *vöresít*, *veres*, *veresség*, *vereses*, *veresség*, *veresül*, *veresedik*, *veresel*, *veresell*, *veresellik*, *vereslik*, *veresít*, *veresít*

KÉK: *kék*, *kékség*, *kékes*, *kékesség*, *kékül*, *kékel*, *kékell*, *kékellik*, *kéklík*, *kékít*, *kékít*

ZÖLD: *zöld*,¹⁰ *zöldes*, *zöldesség*, *zöldül*, *zöldel*, *zöldell*, *zöldellik*, *zöldlik*, *zöldít*, *zöldít*

SÁRGA: *sárga*, *sárgaság*, *sárgás*, *sárgasság*, *sárgul*, *sárgál*, *sárgáll*, *sárgállik*, *sárgít*, *sárgít*

BARNA: *barna*, *barnaság*, *barnás*, *barnásság*, *barnul*, *barnál*, *barnáll*, *barnállik*, *barnít*, *barnít*

Az alkalmazott módszer ugyanaz volt, mint a két korábbi fogalmi tartomány vizsgálata esetében. A szkript egy verset akkor számolt hozzá egy adott szín előfordulásához, ha a vers legalább egy, az adott színre vonatkozó szót tartalmaz. Egy vers több szín előfordulásába is beleszámolódhatott, ha az több színfogalomnak a szavát is tartalmazza. A 4. táblázat mutatja be a kapott gyakorisági adatokat, hasonló módon, mint az ÉVSZAKOK esetében.

¹⁰ A *zöldség* (zöld[ADJ] + *ség*) lemmát az elsődleges jelentése miatt nem vettem bele a keresési kifejezések közé.

4. táblázat. SZÍNEKRE utaló versek gyakorisága

Szerző	Versek száma	Színek %	1.		2.		3.		4.	
			szín	%	szín	%	szín	%	szín	%
Csokonai	292	14,7%	kék	4,5%	piros	4,1%	fekete	3,4%	barna	2,7%
Berzsenyi	125	25,6%	zöld	12,0%	barna	7,2%	kék	4,8%	fekete piros sárga	2,4%
Kölcsy	140	32,9%	zöld	20,0%	kék	9,3%	barna	7,9%	piros	2,9%
Vörösmarty	596	13,6%	zöld	3,9%	barna	3,2%	piros	2,5%	fekete kék	2,3%
Arany	316	18,0%	zöld	8,9%	piros	3,5%	kék	2,8%	fehér sárga	1,9%
Tompai	273	45,4%	zöld	21,6%	kék	9,5%	fehér sárga	8,8%	piros	8,4%
Petőfi	787	20,6%	piros	6,4%	kék	5,0%	zöld	4,8%	fekete	3,4%
Madách	250	15,2%	kék	3,6%	sárga barna	3,2%	zöld	2,8%	fehér piros	2,4%
Gyulai	144	26,4%	zöld	11,1%	kék	9,7%	piros	6,2%	fehér	4,9%
Vajda	149	17,4%	zöld	6,0%	piros	4,7%	fekete kék	3,4%	fehér	2,7%
Revczky	305	18,0%	kék	7,9%	zöld	3,9%	piros	3,3%	fehér	2,6%
Komjáthy	206	10,2%	kék	3,9%	fekete	3,4%	zöld	1,9%	fehér szürke barna	0,5%
Ady	1088	23,3%	piros	8,5%	fehér	6,8%	fekete	4,1%	kék	3,2%
Somlyó	360	46,1%	fehér	13,6%	fekete	12,2%	piros	8,1%	kék	6,1%
Juhász	1267	33,5%	szürke	9,2%	kék	7,3%	piros	6,9%	fekete	6,4%
Babits	458	38,9%	kék	10,3%	fehér	9,6%	zöld	9,0%	piros	8,3%
Kosztolányi	591	45,7%	fehér	15,1%	fekete	10,0%	piros	9,8%	kék	9,6%
Tóth	425	38,8%	sárga	11,3%	kék	8,9%	fekete	8,7%	zöld	7,5%
Reményik	644	28,0%	fekete	7,1%	fehér	7,0%	szürke	6,1%	kék	5,3%
József	578	27,5%	piros	7,8%	kék	6,2%	fehér	5,5%	fekete zöld barna	4,5%
Dsida	1021	46,6%	fehér	15,8%	fekete	12,1%	piros	10,4%	kék	8,8%
Átlag		27,9%								
Medián		26,4%								
I.			zöld (7), kék (5)							
II.			kék (12), zöld (8), fehér (6), piros (6)							
III.			kék (15), piros (13), zöld (12), fekete (9), fehér (8)							

A táblázat harmadik oszlopában szerepel azoknak a verseknek az aránya, amelyek a nyolc színfogalomból legalább egyre utalnak. Az arányszámok szóródása meglehetősen nagy. A legnagyobb arányt, 46,6%-ot, Dsidánál találhatjuk. Tompa, Somlyó és Kosztolányi úgyszintén a gyakorisági lista élén áll, több mint 45%-os arányszámmal. Az adatok azt mutatják, hogy a 20. század elejének a szerzői szívesebben használtak színfogalmakat, mint a korábbi szerzők (Juhász, Babits, Tóth és Reményik úgyszintén az átlag és a medián felett van, és József Attila is meghaladja a mediánt). A 20. századi költők esetében a SZÍNEKRE, az ÉVSZAKOKRA és az ÉJSZAKÁRA utaló versek magasabb aránya a magyar irodalomban a 20. század elején bekövetkező, a klasszikus modernizmus megjelenéseként számontartott poétikai változást tükrözheti. Megjegyzendő ugyanakkor, hogy József Attila esetében a színekre utaló versek gyakorisága alacsonyabb, mint a többi 20. századi szerző esetében (Adyt leszámítva). József Attilánál hasonló mintázatot láthattunk az ÉJSZAKA és az ÉVSZAKOK fogalmaira vonatkozó gyakorisági adatok terén is. A vizsgálati korpuszban olyan szerzők is vannak, akik a verseikben jóval ritkábban utalnak SZÍNEKRE. Komjáthy verseinek például csak a 10,2%-a tartalmaz SZÍNEKRE vonatkozó kifejezést, de Csokonai és Vörösmarty esetében is 15% alatt vannak ezek a versek.

Egy további érdekes eredmény, hogy az Ady esetében talált gyakorisága a valamilyen színfogalomra utaló verseknek sokkal inkább hasonlít a 19. századi költők esetében talált gyakorisági adatokhoz, mint a 20. századi költőknél megfigyelt gyakorisági adatokhoz. Hasonló gyakorisági adatokat láthattunk Adynál az ÉJSZAKA fogalma esetében is. Ha a költészeti korszakokat kizárólag a SZÍNEK fogalmi tartományának és az ÉJSZAKA fogalmának a gyakorisága alapján kéne meghatározni, akkor Ady nem a többi 20. század eleji költővel kerülne egy csoportba, hanem a 20. század előtti szerzőkkel.

A táblázat alsó három sora azt mutatja be, hogy mely színek jelennek meg leggyakrabban az első, az első és második, valamint az első, második és harmadik leggyakoribb színek között.¹¹ Látható, hogy a legtöbb szerző a ZÖLD SZÍNT használta leggyakoribb színeként, a második legtöbb szerző pedig a KÉK SZÍNT használta leggyakoribb színeként (I. sor). Ha az első és második leggyakrabban használt színeket egy csoportként vizsgáljuk (II. sor), akkor a KÉK SZÍN használata kerül az élre.

A leginkább használt színek időbeli megoszlásában egy markáns váltást figyelhetünk meg a 19. és a 20. század fordulóján. A ZÖLD SZÍN Madách kivételével az összes 19. századi költő esetében megjelenik a három leginkább használt szín között, a tíz szerzőből hét esetében ráadásul ez a szín szerepel az első helyen. Ezzel szemben a kilenc 20. századi költőből nyolc esetében a ZÖLD SZÍN nem jelenik meg a három leggyakoribb szín között, és a fennmaradó egy szerzőnél, Babitsnál is csak harmadik

¹¹ Ebben az esetben tehát a táblázat alsó három sora különbözik az ÉVSZAKOK gyakorisági adatait bemutató táblázat alsó négy sorától, hiszen míg az ÉVSZAKOKNÁL mind a négy gyakorisági pozíció esetében külön megadtam, hogy az egyes évszakok az adott gyakorisági pozícióban hány szerzőnél fordulnak elő, addig itt a gyakorisági pozíciók összevonása révén előálló kategóriák alapján adtam meg, hogy mely színfogalmak fordulnak elő a legtöbb szerzőnél.

helyen szerepel. A FEHÉR SZÍN esetében fordított tendencia figyelhető meg. A 20. század előtt alkotó 12 szerző közül csak Tompánál jelenik meg a FEHÉR a három leggyakrabban használt szín között, és ott is csak harmadik helyen. Ezzel szemben a kilencből hét 20. századi szerzőnél is megjelenik a FEHÉR a három leggyakrabban használt szín között, hat költő esetében pedig az első két leginkább használt szín között szerepel. Úgy tűnik, hogy a ZÖLD SZÍN háttérbe szorulása és a FEHÉR SZÍN előtérbe kerülése úgyszintén tükrözi a 20. század elején a magyar irodalomban bekövetkező poétikai korszakváltást.

További kérdésként vehető fel, hogy kik azok a költők, akik a verseikben sok színt használtak nagyobb mértékben. A kérdésre úgy válaszolhatunk a legegyszerűbben, ha minden költő esetében kiszámoljuk a vizsgált nyolc szín gyakoriságának az átlagát és szórását. Ezeket az értékeket az 5. táblázat mutatja be.

5. táblázat
*A különböző színekre utaló versek
gyakoriságának átlaga és szórása*

Szerző	Átlag	Szórás
Csokonai	2,27%	1,69%
Berzsenyi	4,0%	3,94%
Kölcsey	5,01%	7,13%
Vörösmarty	1,96%	1,33%
Arany	2,85%	2,59%
Tompa	8,38%	6,16%
Petőfi	3,56%	1,86%
Madách	2,3%	1,26%
Gyulai	4,34%	4,35%
Vajda	2,7%	2,11%
Reviczky	2,7%	2,44%
Komjáthy	1,34%	1,55%
Ady	3,65%	2,68%
Somlyó	7,64%	3,42%
Juhász	5,89%	2,19%
Babits	7,54%	2,12%
Kosztolányi	8,32%	3,81%
Tóth	7,36%	2,3%
Reményik	4,44%	2,38%
József	4,88%	1,64%
Dsida	8,24%	4,57%

A magas átlaghoz tartozó alacsony szórás azt jelzi, hogy az adott szerző sok szintet használt hasonlóan magas gyakorisággal. Például az 5. táblázatból látható, hogy Babits esetében az átlag magas, 7,54%, a szórás azonban csak 2,12%, ami a többi magas átlaghoz tartozó szóráshoz képest meglehetősen alacsony érték. Mindebből levonható a következtetés, hogy Babits költészetében nem csupán egy vagy két szín kerül nagyobb gyakorisággal előtérbe, hanem a szerző számos szintet használt a verseiben hasonlóan magas gyakorisággal, ami egy impresszionisztikusabb, az érzékszervi benyomásokra fogékonyabb poétikai attitűdöt jelez. A színek gyakoriságának legnagyobb átlaga Tompánál jelenik meg (8,38%), ugyanakkor az ő esetében találjuk a legnagyobb szórást is (6,16%), ami arra utal, hogy Tompa egyenetlen gyakorisággal használta a színeket, és az átlagot bizonyos színek, mindenekelőtt a ZÖLD szín kiugró gyakorisága húzza fel.

6. Az ÉJSZAKA, az ÉVSZAKOK és a SZÍNEK gyakoriságának korrelációja

Az ÉJSZAKA fogalmának és az ÉVSZAKOK és SZÍNEK fogalmi tartományának a gyakorisági adatai hasonló tendenciákat mutatnak. Láthattuk például, hogy mind a három fogalom, illetve fogalmi tartomány előfordulásának a gyakorisága nagyobb a 20. századi költők jelentős részénél, mint számos korábbi költőnél. Ezen eredmények alapján megfogalmazható az a feltételezés, hogy a vizsgált kanonikus magyar költők esetében az ÉJSZAKÁRA, az ÉVSZAKOKRA és a SZÍNEKRE utaló versek arányai között erős pozitív korreláció van. A hipotézis igazolásához az ÉJSZAKA, az ÉVSZAKOK és a SZÍNEK gyakoriságának mind a három lehetséges párosítása esetében kiszámítottam a Pearson-féle korrelációs együtthatót. A korrelációs együttható lehetséges szélső értékei -1 és 1 . A 0 érték arra utal, hogy a két gyakorisági adatsor között nincsen korreláció, az 1 érték tökéletes pozitív, a -1 érték pedig tökéletes negatív korrelációra utal. A kiszámított korrelációs együtthatókat a 6. táblázat mutatja be.

6. táblázat
ÉJSZAKÁRA, ÉVSZAKOKRA és SZÍNEKRE utaló
versek gyakoriságának korrelációja

	Pearson-féle r
ÉJSZAKA – ÉVSZAKOK	0,72
ÉJSZAKA – SZÍNEK	0,91
ÉVSZAKOK – SZÍNEK	0,69

Mind a három korrelációs együttható $0,5$ felett van, ami arra utal, hogy az ÉJSZAKA, az ÉVSZAKOK és a SZÍNEK gyakorisága között erős pozitív korreláció van. A $0,91$ -es érték különösen erős korrelációra utal. A gyakoriságok közötti erős pozitív korreláció azt jelenti, hogy ha egy adott szerző esetében az egyik fogalmi tartomány/fogalom előfordulásának a száma magas, akkor jellemzően a másik két fogalmi tartomány/fogalom előfordulásának a száma is magas, illetve ha egy adott szerző esetében az egyik fogalmi tartomány/fogalom előfordulásának a száma alacsony, akkor jellemzően a másik két fogalmi tartomány/fogalom előfordulási száma is alacsonyabb.

Vagyis azoknak a szerzőknek a nagy része, akik a verseikben nagyobb mértékben utalnak éjszakára, úgyszintén nagyobb mértékben utalnak évszakokra és színekre, illetve azoknak a szerzőknek a többsége, akik a verseikben nagyobb mértékben utalnak évszakokra, úgyszintén nagyobb mértékben utalnak színekre. Az éjszaka, az évszakok és a színek gyakorisága közötti erős pozitív korreláció azt jelezheti, hogy az ezekre való utalás ugyanannak a poétikai eljárásnak más-más aspektusa, amelynek célja a lírai beszédhelyzet, illetve a fiktív megnyilatkozó által elbeszélte jelenetek fizikai jellemzőiről való érzékszervi benyomások előtérbe helyezése.

7. A három fogalmi tartomány fogalmainak együttes előfordulásai

A három zárt fogalmi tartomány fogalmai egy adott versben különböző kombinációkban fordulhatnak elő együtt. A magyar kanonikus költészet verseiben megképződő fizikai világ tipikus és atipikus mintázatainak a vizsgálatához a három fogalmi tartomány fogalmainak tipikus és atipikus együttes előfordulásai is szempontokkal szolgálhatnak. A tipikus és atipikus kombinációkat a fogalompárok együtt-előfordulásának MI-értéke, azaz kölcsönös információ (mutual information) értékük kiszámolása alapján határoztam meg. Bár a nyelvészetben a kölcsönös információt eredetileg a felszíni közelség és a szintaktikai együtt-előfordulás mérésére szolgáló aszszociációs metrikaként vezették be (Church–Hanks 1990), az használható fogalmak egyazon versen belül való együttes előfordulásának a mérésére is (ez a textuális együttelőfordulás esete, lásd Evert 2009: 1220–1224). A 0-nál nagyobb MI-érték azt jelzi, hogy a két fogalom egyazon versen belül többször fordul elő, mint az pusztán a véletlen alapján várható lenne. A 0-nál kisebb MI-érték pedig azt jelzi, hogy a két fogalom kevesebbszer fordul elő ugyanabban a versben, mint az a véletlen alapján várható lenne. A vizsgálat során a 21 szerző verseit tartalmazó alkorpuszokat összevonva, egy közös korpuszként kezeltem, és ezen egybevont korpusz alapján végeztem el a fogalompárok MI-értékének a kiszámítását. Az MI-érték számításának a képletét a (4) mutatja be. A képletben A azon versek száma, amelyekben a vizsgált fogalompár első fogalma előfordul, B azon versek száma, amelyekben a fogalompár második fogalma előfordul, C a korpuszban szereplő összes vers száma, M pedig azon versek száma, amelyekben a vizsgált fogalompár mindkét fogalma megjelenik.

$$(4) \quad MI = \log_2 \frac{M}{\left(\frac{A \times B}{C}\right)}$$

Az MI-értékeket csak különböző fogalmi tartományokból származó fogalmak együttes előfordulásaira számítottam ki. Ez tehát azt jelenti, hogy például a PIROS és a ZÖLD SZÍN vagy a TAVASZ és a TÉL együttelőfordulását nem vizsgáltam, mivel ezek ugyanahhoz a fogalmi tartományhoz tartoznak. A nyolc színfogalom, a négy évszadfogalom és az ÉJSZAKA fogalma összesen 44 fogalmi párt alkothat abban az esetben, ha azokat a kombinációkat kizárjuk, amelyek esetében a két fogalom ugyanahhoz a fogalmi tartományhoz tartozik. Minél magasabb az MI-érték, annál erősebb az aszszociáció a két fogalom között, azaz annál valószínűbben fordulnak elő ugyanabban

a versben. A negatív MI-érték azt jelzi, hogy a két fogalom tendenciaszerűen nem fordul elő együtt egyazon versben. Az MI-értékek kiszámításához egy szkriptet használtam, amely az összes eltérő fogalmi tartományból származó fogalompár esetében megszámolta, hogy a fogalompár fogalmai hány versben fordulnak elő, illetve hogy a két fogalom hányszor fordul elő együtt, ugyanabban a versben, majd pedig kiszámolta a fogalompárra vonatkozó MI-értéket a (4)-ben megadott képlet alapján.

A 7. táblázat mutatja be a 20 legmagasabb MI-értékkel rendelkező fogalompárt. Ezek a fogalompárok tehát nagyobb valószínűséggel fordulnak elő ugyanabban a versben, mint a többi lehetséges kombináció. Úgy is fogalmazhatunk, hogy a vizsgált három fogalmi tartománynak ezek azok a fogalmai, amelyek a magyar kanonikus költészetben nagyobb mértékben vonzzák egymást. A táblázat harmadik oszlopa mutatja be az MI-értékeket, a negyedik oszlop pedig azoknak a verseknek a számát, amelyekben a két fogalom egyaránt előfordul.

7. táblázat

Ugyanazon versekben előforduló fogalompárok MI-értéke

	Fogalom 1	Fogalom 2	MI-érték	Előfordulás
1	ŐSZ	SÁRGA	1,84	86
2	TAVASZ	ZÖLD	1,27	112
3	NYÁR	SÁRGA	1,02	40
4	TÉL	ZÖLD	0,98	59
5	ŐSZ	SZÜRKE	0,97	53
6	NYÁR	ZÖLD	0,83	57
7	ŐSZ	ZÖLD	0,8	68
8	TAVASZ	KÉK	0,78	95
9	TÉL	FEHÉR	0,77	66
10	TAVASZ	SÁRGA	0,75	48
11	ŐSZ	PIROS	0,73	81
12	ÉJSZAKA	FEHÉR	0,68	309
13	ÉJSZAKA	SZÜRKE	0,67	164
14	ŐSZ	KÉK	0,67	74
15	NYÁR	PIROS	0,63	62
16	NYÁR	KÉK	0,63	59
17	ÉJSZAKA	FEKETE	0,61	247
18	ÉJSZAKA	BARNA	0,57	137
19	NYÁR	FEHÉR	0,56	61
20	TÉL	SÁRGA	0,56	27

Érdemes megjegyezni, hogy az ÉJSZAKA fogalma egyik évszakegységgel sem áll együtt 0,5-nél nagyobb MI-értékkel. Úgy tűnik, hogy a magyar kanonikus költészetben az ÉJSZAKA és az ÉVSZAKOK nem vonzzák egymást olyan mértékben, mint az ÉJSZAKA és bizonyos színek, illetve mint az ÉVSZAKOK és bizonyos színek. Úgyis szintén feltűnő, hogy az ÉJSZAKA fogalma nem jelenik meg a 11 legmagasabb MI-értékkel rendelkező fogalompár között. Ez feltételezhetően azzal magyarázható, hogy az érzékszervi információk, különösen a színek éjszaka háttérbe szorulnak. Ezt a magyarázatot az is támogatja, hogy a 20 fogalompár között szereplő négy, ÉJSZAKÁT tartalmazó fogalompárban „színtelen” szín szerepel: FEHÉR, SZÜRKE, FEKETE és BARNA. A 44 fogalompárból egy pár, a TAVASZ–BARNA kapott negatív MI-értéket (MI: -0,21, előfordulás: 26). A negatív MI-érték azt jelzi, hogy a két fogalom taszítja egymást, azaz tendenciaszerűen nem fordulnak elő ugyanabban a versben.

8. Összegzés és néhány zárómegjegyzés

A tanulmány a költészet vizsgálatának egy kvantitatív, korpusznyelvészeti módszerét mutatta be, amelynek alapja zárt fogalmi tartományokhoz tartozó fogalmak előfordulásának mérése. A zárt fogalmi tartományok olyan jól körülhatárolható fogalmi mezők, amelyek feloszthatók véges számú fogalomra. A módszer bemutatása során három zárt fogalmi tartományt, a NAPSZAKOK, az ÉVSZAKOK és a SZÍNEK tartományának a gyakorisági jellemzőit vizsgáltam 21 magyar kanonikus költő 300 szónál rövidebb versében. A fogalmi tartományokhoz tartozó fogalmak gyakorisági vizsgálatával a szerzők verseiben megképződő fizikai világról juthatunk ismeretekhez. A vizsgálat rámutatott arra az általános tendenciára, hogy a NAPSZAKOK fogalmi tartományhoz tartozó ÉJSZAKA fogalmának, valamint az ÉVSZAKOK és a SZÍNEK fogalmi tartományának a gyakorisága is többnyire magasabb a 20. századi szerzők esetében, mint a korábbi szerzőknél. Azt is láthattuk, hogy míg a 19. század végéig a ZÖLD SZÍN volt a legnépszerűbb, addig a 20. századi költőknél ennek a színnek a használata nagymértékben háttérbe szorult. Ezzel szemben a FEHÉR SZÍN használata fordított tendenciát mutat: a 20. század előtt nem alkotta részét a leggyakrabban használt színeknek, a századfordulótól azonban előtérbe került a használata. Úgy tűnik, a gyakoriságoknak ez a változása tükrözi a magyar irodalomban a 20. század elején bekövetkező, a klasszikus modernség megjelenéseként számon tartott poétikai változást. Egy másik, némileg meglepő eredmény, hogy Ady esetében az ÉJSZAKA és a SZÍNEK használatának a gyakorisága jobban hasonlít a 19. századi szerzőknél talált gyakorisági adatokhoz, mint a 20. századi szerzőknél talált gyakorisági adatokhoz.

A tanulmány azt is bemutatta, hogy a vizsgált magyar kanonikus szerzők esetében erős pozitív korreláció van az ÉJSZAKÁRA, az ÉVSZAKOKRA és a SZÍNEKRE utaló versek arányai között. Ez azt jelenti, hogy azon szerzők esetében, ahol az egyik fogalmi tartományra/fogalomra utaló versek száma magasabb, jellemzően a másik két fogalmi tartományra/fogalomra utaló versek száma is magasabb. Az ÉJSZAKA, az ÉVSZAKOK és a SZÍNEK gyakoriságai közötti erős pozitív korreláció azt jelezheti, hogy az ezekre történő utalás ugyanannak a poétikai eljárásnak más-más aspektusa,

amelynek célja a lírai beszédhelyzet, illetve a fiktív megnyilatkozó által elbeszélte jelenetek fizikai jellemzőiről való érzékszervi benyomások előtérbe helyezése.

Végezetül a kölcsönös információérték kiszámítása alapján vizsgáltam a különböző fogalmi tartományokhoz tartozó fogalmak együttlőfordulását. E módszerrel azt vizsgálhattuk, hogy melyek azok a fogalompárok, amelyek gyakrabban fordulnak elő ugyanabban a versben, mint az a puszta véletlen alapján várható lenne. Az eredményekből láthattuk, hogy a magyar kanonikus költészetben az ÉJSZAKA fogalma és az évszakefogalmak kevésbé vonzzák egymást, mint az ÉJSZAKA és bizonyos színek, illetve mint az ÉVSZAKOK és bizonyos színek, illetve azt is láthattuk, hogy az ÉJSZAKA fogalma a színek közül színtelen színekkel áll együtt nagyobb kölcsönös információértékkel. A tanulmányban vizsgált fogalmak mellett más típusú fogalmak is bevonhatók egy együttlőfordulást vizsgáló kutatásba. Például tanulságos lenne megnézni, hogy az itt vizsgált, elsősorban a versek fizikai világának a megképzésében részt vevő fogalmak milyen, többnyire igékkel kifejezett tipikus eseményekkel állnak jellemzően együtt.

Megjegyzendő, hogy a három fogalmi tartománynak erős metaforikus potenciálja van. Például az ÉJSZAKA fogalma jellemzően a SZOMORÚSÁG, MAGÁNY, HALÁL, INAKTIVITÁS, NEM LÉTEZÉS stb. metaforikus forrástartománya. A TÉL úgyszintén hasonló fogalmak forrástartományaként szokott megjeleníteni. A TAVASZ és a NYÁR ezzel szemben a SZERELEM, BOLDOGSÁG, ÉLET, AKTIVITÁS stb. céltartományoknak a tipikus forrástartománya (a fogalmimetafora-elméletről lásd Lakoff–Johnson 1980, Lakoff 1992). A bemutatott korpusznyelvészeti módszer a vizsgált fogalmaknak a metaforákban és allegóriákban játszott forrástartomány-jellegét nem tudta figyelembe venni. A későbbi kutatások feladata olyan módszerek kidolgozása, amelyek a fogalmakra utaló nyelvi kifejezések gyakoriságának kvantitatív mérését összekapcsolják a nyelvi kifejezések versekben betöltött szemantikai funkciójának vizsgálatával.

Irodalom

- Church, Kenneth W. – Hanks, Patrick 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics* 16/1, 22–29.
- Culler, Jonathan 1981. *The pursuit of signs: Semiotics, literature, deconstruction*. London, New York: Routledge.
- Evert, Stefan 2009. Corpora and collocations. In: Lüdeling, Anke – Kytö, Merja (eds.): *Corpus linguistics: An international handbook*. Vol. 2. Berlin, New York: Walter de Gruyter. 1212–1248.
- Fónagy Iván 1959. *A költői nyelv hangtanából*. Budapest: Akadémiai Kiadó.
- Horváth Péter – Kundráth Péter – Indig Balázs – Fellegi Zsófia – Szilávi Eszter – Bajzát Tímea Borbála – Sárközi-Lindner Zsófia – Vida Bence – Karabulut Aslihan – Timári Mária – Palkó Gábor 2022. ELTE Verskorpusz – a magyar kanonikus költészet gépileg annotált adatbázisa. In: Berend Gábor – Gosztolya Gábor –

- Vincze Veronika (szerk.): *XVIII. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi Tudományegyetem TTIK, Informatikai Intézet. 375–388.
- Indig Balázs – Sass Bálint – Simon Eszter – Mittelholcz Iván – Vadász Noémi – Makrai Márton 2019. One format to rule them all – The emtsv pipeline for Hungarian. In: Friedrich, Annemarie – Zeyrek, Deniz – Hoek, Jet (eds.): *Proceedings of the 13th Linguistic Annotation Workshop*. Stroudsburg: Association for Computational Linguistics. 155–165.
- Jékel Pál – Papp Ferenc 1974. *Ady Endre összes költői műveinek fonémastatisztikája*. Budapest: Akadémiai Kiadó.
- Jékel Pál – Szuromi Lajos 1980. *Petőfi metrumai*. Debrecen: Kossuth Lajos Tudományegyetem.
- Labádi Gergely 2018. Az olvasó gép: Berzsenyi Dániel versei távolról. *Digitális Bölcsészet* 1, 17–34.
- Lakoff, George 1992. The contemporary theory of metaphor. In: Ortony, Andrew (ed.): *Metaphor and thought*. Cambridge: Cambridge University Press. 202–251.
- Lakoff, George – Johnson, Mark 1980. *Metaphors we live by*. Chicago: The University of Chicago Press.
- Lesi Zoltán 2008. Automatikus formai verselemzés. *Alkalmazott Nyelvtudomány* 8/1–2, 197–208.
- Maróthy Szilvia – Seláf Levente – Plecháč, Petr 2021. Rhyme in 16th-Century Hungarian Historical Songs: A Pilot Study. In: Plecháč, Petr – Kolár, Robert – Bories, Anne-Sophie – Říha, Jakub (eds.): *Tackling the Toolkit: Plotting Poetry through Computational Literary Studies*. Prague: Institute of Czech Literature of the Czech Academy of Sciences. 43–58.
- McIntyre, Dan 2013. Language and style in David Peace’s 1974: a corpus informed analysis. *Études de stylistique anglaise* 4, 133–146.
- Moretti, Franco 2000. Conjectures on world literature. *New Left Review* 1, 54–68.
- Moretti, Franco 2005. *Graphs, maps, trees: Abstract models for literary history*. London, New York: Verso.
- Rayson, Paul 2008. From key words to key semantic domains. *International Journal of Corpus Linguistics* 3/4, 519–49.
- Simon Gábor 2020. Géppel mért műfajiság: Esettanulmány a modern magyar elégia korpuszalapú vizsgálatához. In: Kulcsár-Szabó Zoltán (szerk.): *Hagyomány és innováció a magyar és világirodalomban*. Budapest: Eötvös Kiadó. 45–64.
- Tátrai Szilárd 2011. *Bevezetés a pragmatikába*. Budapest: Tinta Könyvkiadó.

- Tátrai, Szilárd 2015. Context-dependent vantage points in literary narratives: A functional cognitive approach. *Semiotica: Revue publiée par l'association internationale des semiotique = Journal of the international association for semiotic studies*. 9–37.
- Tátrai Szilárd 2015. Apostrophic fiction and joint attention in lyrics: A social cognitive approach. *Studia Linguistica Hungarica* 30, 105–117.
- TEI Consortium 2019. *TEI P5: Guidelines for electronic text encoding and interchange: Version 3.5.0*. <https://tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf>.
- Váradai Tamás – Simon Eszter – Sass Bálint – Mittelholcz Iván – Novák Attila – Indig Balázs – Farkas Richárd – Vincze Veronika 2018. e-magyar – A digital language processing system. In: Calzolari, Nicoletta – Choukri, Khalid – Cieri, Christopher – Declerck, Thierry – Hasida, Koiti – Isahara, Hitoshi – Maegaard, Bente – Mariani, Joseph – Moreno, Asuncion – Odijk, Jan – Piperidis, Stelios – Tokunaga, Takenobu (eds.): *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Paris: European Language Resources Association. 1307–1312.
- Zsilka Tibor 1974. *Stilisztika és statisztika*. Budapest: Akadémiai Kiadó.