

## Adalékok a birtokos szerkezetekbe sűrített hasonlatok számítógépes azonosításához

Bajzát Tímea Borbála<sup>1,2</sup>, Szemes Botond<sup>1</sup>, Szlávich Eszter<sup>1</sup>

<sup>1</sup>Eötvös Loránd Tudományegyetem, Digitális Örökség Nemzeti Laboratórium  
1088 Budapest, Múzeum krt. 4., 4–6.

boboszemes@gmail.com  
szlavich.eszter@btk.elte.hu

<sup>2</sup>ELKH Irodalomtudományi Intézet, Digitális Örökség Nemzeti Laboratórium  
1118 Budapest, Ménesi út 11–13.  
bajzat.timea@abtk.hu

**Kivonat:** A dolgozat a birtokos szerkezetekbe sűrített hasonlatok automatikus azonosítására kifejlesztett módszerünk bemutatását, valamint az azonosított esetek alapján a konstrukció mint poliszém hálózat elemzését tartalmazza. Ez a különböző típusú sűrített hasonlatok kvantitatív és kvalitatív vizsgálatára irányuló kutatásunk megalapozó lépésének tekinthető. Egy ilyen kutatást a szerkezetek sajátos stílusértéke motivál, hiszen azokra egyszerre jellemző a díszítettség és a tömörség a több tagmondat által kidolgozott hasonlatokhoz képest. A dolgozatban bemutatott konstrukció formai elkülönítése komoly nehézséget jelent, amelyet egy többlépcsős, reguláris kifejezéseken alapuló, valamint a szavak grammatikai tulajdonságait is figyelembe vevő módszer kifejlesztésével kívánunk megoldani.

### 1. Bevezetés

Az alábbi dolgozat egy tágabb, kvantitatív kutatás megelőző lépéseiről számol be. Ez a kutatás a névszavakból álló szókapcsolatokba sűrített hasonlatok azonosítását tűzi ki célul, hogy összevesse ezek gyakoriságát a több tagmondat által kidolgozott, formailag konvencionálisabb, „hagyományos” hasonlatok gyakoriságával különböző korok irodalmi és hétköznapi szövegeiben. A kutatásunk nyelvelméleti kiindulópontja a funkcionális kognitív elméleti keret (Langacker 1987, 2008, Tolcsvai Nagy szerk. 2017). A több tagmondatból álló, konvencionálisabbnak nevezett hasonlatok alatt azokat az eseteket értjük, amikor a hasonlítást kidolgozó mellékmondat a főmondat által nyelvileg színre vitt elemi jelenettel (elemi jelenet vö. Imrényi 2017: 666) „párhuzamba állítható” (Kugler 2017: 850).

A jelen tanulmány a birtokos szerkezetbe sűrített hasonlatok számítógépes azonosítására kifejlesztett módszerünk bemutatását, valamint a konkrét nyelvi példányok vizsgálatával a konstrukció jelentéshálózatának leírását kezdeményezi. Ez a különböző típusú sűrített hasonlatok kvantitatív és kvalitatív vizsgálatára irányuló kutatásunk megalapozó lépésének tekinthető. A kutatást a szerkezetek sajátos stílusértéke motiválja, hiszen azokra egyszerre jellemző a figurák részletezése és a tömörség a több tagmondat által kidolgozott hasonlatokhoz képest. A dolgozatban

bemutatott konstrukció formai elkülönítése összetett kihívást jelent, amelyet egy többlépcsős, reguláris kifejezéseken alapuló, valamint a szavak grammatikai tulajdonságait is figyelembe vevő módszer kifejlesztésével kívánunk megoldani. A hasonlítás tömörsége önmagában is fontos, stilisztikai potencialitással bíró elem a vizsgált szövegekben, amit tovább erősítenek a szerkezetekben egymás mellé kerülő jelzős főnevek nagy száma. Példaként említhetők a következő mondatpárok:

- (1a) A puha takaró egy meghitt, adventi ölelésként vár a nehéz nap után.  
(1b) Úgy vár a puha takaró a nehéz nap után, mint egy meghitt, adventi ölelés.

- (2a) A színpadra belibbenő balett-táncosok üde méltóságával kínálta helytel az idős hölgyet.  
(2b) Olyan üde méltósággal kínálta helytel az idős hölgyet, mint a színpadra belibbenő balett-táncosok.

A vizsgálat egy szöveg fontos stílusértékének tekinti, ha ezek (lásd (1a) és (2a)) a szerkezeti mintázatok nagy számban fordulnak elő benne, vagy ha gyakorinak tekinthetők a tagmondatkapcsolatokkal kidolgozott hasonlatokhoz viszonyítva – továbbá rákérdez az egyes korszakokban való eloszlásukra is. Ezen eloszlás kvantitatív, számítógépes vizsgálata lehetőséget nyújt számunkra ahhoz, hogy a feltárt nyelvi mintázatok alapján hozzájáruljunk a magyar prózairodalom stílustörténeti vizsgálatához.

A fent idézett második mondatpár ((2a) és (2b)) rávilágíthat az egyes konstrukciók automatikus azonosításának nehézségeire is – ami pedig elengedhetetlen a későbbi kvantitatív vizsgálatok számára. A birtokos szerkezetekbe sűrített hasonlatok ugyanis – amelyek sajátos stílusértékére (a konstrukció tömörségére és a jelzőhalmozás lehetőségére) már Herczeg Gyula is felfigyel (Herczeg 1958, 1981) – nehezen különíthetők el más birtokos szerkezetektől, amelyek nem dolgoznak ki hasonlítást: „*A színpadra belibbenő balett-táncosok öltözőjében kínálta helytel az idős hölgyet.*” Az eddigi munkánk során egyértelműen ez jelentette a legnagyobb akadályt a többi típushoz (a *-ként* essivus rag és *-szerű* produktív képzőszerű utótag, valamint a *gyanánt* névutóval alkotott névszói szerkezetekhez) képest. Az alábbiakban ezért elsősorban a hasonlítást kifejező birtokosok azonosítására kifejlesztett módszerünket ismertetjük, illetve – a kereshetőségnek köszönhetően – a konstrukció poliszémiájának szemléltetésére is hozunk a korpuszunkból példányokat, hogy felvázolhassuk a konstrukció jelentéshálózatát.

A kutatásunk célja tehát egy olyan funkcionális kognitív nyelvészeti kiindulópontot érvényesítő módszer kidolgozása, amellyel formai és szemantikai kritériumok alapján nagymennyiségű szövegből bizonyos megszorításokkal konstrukciók nyerhetők ki reguláris kifejezések alkalmazásával. Az általunk kidolgozott metódus és a meghatározott megszorítások korpuszadatok elemzésével, s Herczeg Gyula megfigyelései alapján, dinamikusan, több fázisban szerveződtek. A referenciakorpuszunk egy olyan 100 szépirodalmi regényből álló szövegtömeg volt, amely összeállításánál során törekedtünk arra, hogy a magyar nyelvű regényirodalom történetiségére vonatkozóan reprezentatívnak bizonyuljon (a korpuszépítés elvét az 3. fejezet mutatja be részletesebben). De a későbbiekben az általunk kialakított vizsgálati metódust alkalmazni kívánjuk az ELTE regénykorpusz adatbázisának szövegeire

(<https://regenykorpusz.elte-dh.hu/>) is, mivel ezen korpusz csaknem 400 regény nyelvileg annotált, kereshető elérését biztosítja. A kutatáshoz azért használtunk szépirodalmi szövegeket, mert ez a vizsgált konstrukció – a fent is megfogalmazott feltevéseink alapján – stílusspecifikus, az irodalmi nyelv eszközeként funkcionálhat. A vizsgálatot az R programozási nyelven és szoftverkörnyezetben végeztük.

A célzott konstrukciót sűrített hasonlatnak neveztük el (Herczeg 1958: 313 alapján), ezzel a terminussal fogunk hivatkozni rá; a konstrukció felépítése pedig: DET – N<sub>GEN</sub> – N<sub>INS</sub> – V, azaz egy névelő; egy birtokos esetű főnév; egy birtokos személyjeles, instrumentalis esetű főnév és egy ige alkotja: *az előkelők magabiztosságával megállt*. A konstrukció tetszőlegesen bővíthető a főnevek jelzőivel: *az orosz táncosnők finom eleganciájával lépett be*. A szerkezetet azért tartjuk sűrített hasonlatnak, mert a legtöbb esetben parafrázálhatóak hasonlító összetett mondatokkal: *olyan magabiztosan állt meg, mintha előkelő lenne / olyan magabiztosan állt meg, mint egy előkelő; olyan finom eleganciával lépett be, mintha orosz táncosnő lenne / olyan finom eleganciával lépett be, mint egy orosz táncosnő*.

Ha a sűrített hasonlatokat a hasonlító összetett mondat parafrázisokkal rokonítjuk, akkor a hasonlító összetett mondat semlegesebb stílusértékű. Továbbá a sűrített hasonlatok esetében kiemelt jelentőségű az, hogy a tömörítés eszközeként funkcionálnak, hiszen egy tagmondatban teszik elérhetővé a hasonlító és a hasonlított közötti viszonyt integráltabb kidolgozásban. A hasonlítás műveletét végrehajtó nyelvi szerkezetek forma–jelentés pár szimbolikus kapcsolata alapján történő elemzésének kezdeményezését támogatja a funkcionális kiindulópont is, hiszen a nyelvi konstruálás során az eltérő kidolgozási formák a használat és a jelentés felől motiváltak (Langacker 1999, Tolcsvai Nagy 2017b: 224 Langacker 2008: 55–91 alapján).

A tanulmányunk 5. fejezetében az általunk sűrített hasonlatnak nevezett séma konstrukciójának poliszém hálózatával foglalkozunk, ugyanis amellet, hogy ezen szerkezet képes egy integráltabb jelentésviszonyban végrehajtani a hasonlítás műveletét, további jelentéssűrítő potenciállal is bír, amely stilisztikailag szintén jelöltebb (azaz nemsemleges) szerkezeteket hoz létre. Ezen funkciója korántsem periferikus az irodalmi korpuszból hozott eredmények alapján, továbbá nehezen határolható el a hasonlítás végrehajtó példányoktól, azonban kevésbé parafrázálhatóak a hasonlító összetett mondatokkal, valamint a kompozitum birtokos esetű főneve rendszerint valamilyen absztrakt entitás (lásd (3)).

(3) [...], *akkor sokkal kiméletesebben és főleg a megfontoltság ravaszságával kell eljárnia.*<sup>1</sup>

A forma–jelentés pár szimbolikus viszonyából kiindulva motivált az, hogy az előzőekben tárgyalt két jelentést/jelentésárnyalatot nem élesen elválasztva egymástól tárgyaljuk. Az efféle integrált, jelentéssűrítő funkcióit aktiváló szerkezetekre Herczeg Gyula (1958: 313) is felfigyelt, elkülönítve kezelte őket a sűrített hasonlatoktól. Az előbbi parafrázálhatóságát abban ragadja meg, hogy két összekapcsolt módhatározóvá alakítja a komponensszerkezeteket (Herczeg 1958: 313). (Az előző példát átalakítva: “...*megfontoltan és ravaszsan kell eljárnia.*” A két típus nem szigorú elhatárolását mutatja azonban, hogy – bár kissé erőltetetten hat –, ám nem tűnik teljesen

<sup>1</sup> Nadas Péter – Emlékiratok könyve (1986)

hibásnak e példamondat hasonlattá alakítása sem: *“olyan ravaszan, mint a megfontoltság.”*) Azonban fontos megjegyeznünk azt, hogy a parafrázálhatóság mind ebben az esetben, mind pedig a sűrített hasonlat jelentésben pusztán csak részleges, hiszen az integráció miatt sajátos jelentésárnyalattal (és stíluspotenciállal) bírnak ezen példányok.

## 2. A vizsgálat módszere

A kutatás ezen fázisa tehát egy olyan keresési eljárást kíván létrehozni, amely lehetővé teszi azoknak a birtokos szerkezeteknek az automatikus azonosítását, amelyek a fent vázolt módon sűrített jelentést dolgoznak ki. Mint említettük, ezek elkülönítése más birtokos szerkezetektől komoly kihívást jelent, amennyiben nem határozható meg egy egyértelmű formai különbség közöttük. A különböző eljárások kézi kiértékelése után, egy többlépcsős, reguláris kifejezéseken alapuló módszert hoztunk létre, amely hatékonyan azonosította a kívánt szerkezetet a keresések során. Fontos, hogy a hatékonyság alatt azt értjük, hogy az általunk felépített reguláris kifejezés nagy arányban teszi szűrhetővé és ezáltal korpuszalapúan vizsgálhatóvá a birtokos szerkezettel kidolgozott jelentéssűrítő kifejezéseket, viszont számolnunk kellett azzal is, hogy az így lekért mintákban olyan példányok is előfordultak, amelyek ugyan formailag illeszkedtek a keresett mintázatra, de nem jelentéssűrítő funkcióban fordultak elő. A módszer segítségével nagyszámú példát tudtunk hozni a vizsgált konstrukcióra a magyar regényirodalom több, mint 170 évéből, ami a szerkezet jobb megértését és leírását biztosítja. Az így nyert belátások pedig a jövőben integrálhatók lehetnek a keresést meghatározó szempontrendszerbe, amitől a módszer hatékonyságának növelését reméljük.

A keresések az e-magyar (Indig és mtsai 2019, Simon és mtsai 2020) nyelvi elemzővel feldolgozott szövegeken futnak le. Az, hogy szófaji és morfológiai elemzésnek vetjük alá a korpuszunkat alkotó szövegtömeget, azért indokolt, mert később, a reguláris kifejezés futtatásakor tudjuk csökkenteni ezzel a számunkra nem kedvező példányok előfordulási számát, mivel figyelembe tudjuk venni a tokenek és azok lemmáinak grammatikai tulajdonságait. Az emtsv-vel való feldolgozás kimenete alapján a szövegeket táblázatos formában kezeljük, amely táblázatokban az oszlopokat 1. a szöveg szavai, 2. a lemmák, 3. a szófajcímkék és 4. a morfológiai elemzés alkotják. A keresések egyes fázisai az adott szempontok szerint más-más oszlopra vonatkoznak, ám a találatok pozíciója (azaz a sorok száma) alapján ezek könnyen összehangolhatók. Első lépésként az elvont jelentéstartalmat kifejező és a *-val/-vel* instrumentalis esetraggal ellátott, birtokos személyjeles főneveket azonosítottuk a korpuszban. Az elvont jelentéstartalom (absztraktumok vö. Paradis 2005) fontos megkülönböztető jegye ugyanis a konstrukciónak, hiszen ez a szempont kizárja a konkrét főnevet szerepeltető mondatokat a keresésből, amelyek általában nem sűrített hasonlatok kidolgozásában vesznek részt (pl.: *„A színpadra belibbenő balett-táncosok öltözőjében kínálta helytel az idős hölgyet.”*) Az elvont jelentéstartalmat a formai oldalról tipikusan

---

<sup>2</sup> Az alábbi linken keresztül elérhető a nyilvános Github repozitórium, ahol az általunk fejlesztett programkód jelenlegi állapota található: <https://github.com/SzemesBotond/s-r-tett-hasonlatok>

az alábbi képzők megléte jelezheti: *-alom/-elem; -ság/-ség; -ás/-és*. Azaz a szóalakra vonatkozó keresés az alábbi végződésekre vonatkozott: *-almával, -elmével, -ságával, -ségével, -ásával, -ésével*, majd ezek közül a találatok közül csak a főnévi, illetve a birtokos paradigmát jelölő tagekkel ellátott szavakat őriztük meg, és ezek képezték a további elemzési metódus bemenetét. A kutatás során ellenőriztük, hogy a szerkezetben milyen gyakorisággal fordulnak elő többes számban az említett absztrakt főnevek, ám azt találtuk, hogy így kapott találatok közül csupán egy, meglehetősen speciális eset vonatkozott a keresett konstrukcióra: *“Száz arany négy krajcárért! kiáltják az ő piaci élelmességükkel, és a tömegben kitör a röhej és vihanolás.”*<sup>3</sup> Ez az eredmény már önmagában a konstrukció részletesebb megismeréséhez vezetett, hiszen kijelenthetjük, hogy prototipikusan egyes számban állnak benne az absztraktumok. Meg kell még jegyeznünk, hogy a konkrét toldalékalakokra való keresést azért részesítettük előnyben az emtsv emMorph moduljával szemben, amely bár lehetővé teszi a képzőkre vonatkozó információk kinyerését, mert jelen esetben egyértelműen meghatározható a keresett főnevek végződése formai úton, így a legpontosabb találatokat ezen keresztül kaphatjuk. Azonban amennyiben hasonulással vagy egyéb formai bizonytalansággal is számolnunk kellene, nem tudnánk megkerülni a nyelvi elemző alkalmazását.

Következő lépésként az azonosított főneveket megelőző szavak szófaja alapján szűkítettük a találatokat. Azokat a találatokat tartottuk meg, amelyeket főnév (*egy balett-táncos méltóságával*) vagy főnév-melléknév szófaji kombináció (*egy balett-táncos üde méltóságával*) előz meg, valamint balra tőle négy token keresési ablakban egy névelő található az elemi mondatban. Meg kell említenünk, hogy így a módszer a birtokos főnév előtti bizonyos jelzős mellérendelői szerkezeteket nem tudja azonosítani (*egy balett-táncos üde, de elszánt méltóságával*), mivel ezen eseteket a komponensek közötti kötőszó előfordulása miatt eleve eldobja a metódus. Ahogy látható, ezen lépés potenciálisan csökkentheti a jó találatok számát, azonban a fals találatok nagyobb számú kiszűrése végett fontos volt megtartanunk. Mindazonáltal kísérleteket folytatunk arra nézve is, hogy ezeket a szerkezeteket az emtsv emChunk moduljával azonosítsuk, amely ugyanis éppen a főnévi csoportok (frázisok) automatikus felismerését teszi lehetővé. Ugyanakkor tapasztalataink szerint ez sem működik hibátlanul az olyan esetekben, amelyeket az említett módszerrel sem tudunk befogni, így további kutatásokra van szükség, hogy emChunk hatékonyságát lemérjük. Biztató azonban, hogy az előző példában (*egy balett-táncos üde, de elszánt méltóságával*) a kifejezéshez tartozó részeket helyesen azonosítja a modul.

Ezt követően a szűkített szólista szótöveivel dolgoztunk tovább, pontosabban azt kívántuk meghatározni, hogy milyen szófajú szavakból képezték a fentebb is listázott képzők az elvont főneveket. Ez azért fontos, mert kutatásunk alapján a melléknévi tő felülreprezentált a keresett konstrukciókban (pl. *méltó-ságával*). Az e-magyar által lemmaként azonosított formák a birtokos és a *-vall/-vel* ragot nélkülöző eseteket, azaz tulajdonképpen magukat az elvont főneveket mutatják (pl. *méltóság*). Ezért ebben a fázisban a képzők eltávolításával hoztuk létre a kívánt tövet, amelyen ismételtlen lefutattuk az e-magyar elemzését. Így tudtuk csak azokra a szavakra korlátozni a keresést, amelyekben melléknévi alapból jönnek létre az elvont birtokos főnevek. Ezek képezik a végső találataink kiindulópontját, ezt a listát azonban kiegészítettük olyan nem melléknévi tövű, esetleg nem az említett képzőkkel képzett főnevekkel, amelyek

<sup>3</sup> Mikszáth Kálmán – Új Zrinyiász (1898)

mégis nagy számban fordulnak elő a konstrukcióban. Ezeket elsősorban Herczeg Gyula szintén erre irányuló szövegei (Herczeg 1958: 314), valamint a tövekre vonatkozó szűkítést megelőző keresések találatai alapján határoztuk meg. A nem melléknévből képzett szavak: *jóérzés, érzés, remegés, derengés, pezsgés, feszültség, ragyogás, kétségbeesés, megjelenés, ragaszkodás, hanghordozás, odaadás, szójárás, tájszólás, lángolás, észjárás, arckifejezés, illetve: nyugalom, izgalom, aggodalom, fájdalom, szánalom, léleknyugalom, unalom, vashűség, buzgalom, értelem, félelem, szorongalom, riadalom, ijedelem; a nem az említett módon képzett főnevek: kedvével, könyörületével, dühével, rajongásával, kitartásával, pompájával, türelmével, fölényével, lendületével.*

Ezáltal létrejött azoknak a szavaknak a listája, amelyek nagy eséllyel szerepelnek a keresett konstrukcióban. Utolsó lépésként azokból a mondatokból, amelyekben ezek a szavak szerepelnek, kiszűrtük azokat az eseteket, amikor a konstrukció közelében olyan szó áll, amelynek kötelező vonzata a *-val/-vel* instrumentális ragos eset, és így nem a sűrített hasonlat meglétét jelzi: *szembe(n), visszaél, helyettesít, összevet, versenyez, küzd, vetekszik, ellenkezik, ellentétben, megeléged, áthidal, rendelkezik, kapcsolatos, tisztába(n), együtt, rendelkezik, összefügg, összetéveszt, összeegyeztet, arány, felruház.*

### 3. A referenciakorpusz<sup>4</sup>

A kereséseket 100 kanonikus magyar regényen végeztük el, ez a korpusz több mondatstilisztikai kutatás alapját képezte már – ezek közül is kiemelendő a tagmondatkapcsolatok kvantitatív elemzését célzó tanulmány (Szemes 2021), amely többek között a hasonlatok gyakorisági vizsgálatát is elvégezte a korpuszban található regényeken. Ezért a későbbiekben a jelen kutatás eredményei jól összevethetők lehetnek ennek a tanulmánynak az eredményeivel, további adalékot szolgáltatva a magyar regény stílustörténetének feltárásához.

A korpusz időbeli kiindulópontja a kiforrott magyar regényirodalom kezdete, azaz az 1830-as évek (Szilágyi és Vadera 2010: 370). A végpontot Nadas Péter *Párhuzamos történetek* című, 2005-ben megjelent regénye jelenti, kizárva a kutatásból a kortárs magyar irodalom elmúlt évtizedeit, mivel nehéz és vitatható lenne olyan szempontot találni, amelynek segítségével csupán pár kanonikus mű a napjaink magyar irodalmából kiválasztható. Kanonikus művekre azért van szükség, mert a 19. század első feléből jóformán csak ezek érhetők el digitalizált formában, ami így meghatározta a korpusz egészének szerkezetét. Ugyanígy az ebből a korszakból elérhető alkotók száma szabta meg azt is, hogy későbbi időszakokból hány szöveg került a kutatásba, mivel arányos időbeli eloszlásra törekedtünk a regények első megjelenését illetően. Ezáltal minden évtizedből legkevesebb 3, legfeljebb 8 szöveg került a korpuszba. Ez összesen 100 regényt és 173 évet, azaz átlagosan 1,7 évenként egy regényt jelent. Szintén az arányosság miatt egy írótól maximum 4, de inkább kevesebb regényt

<sup>4</sup> Az alábbi linken keresztül elérhető a korpuszt képező regények listája és azok bibliográfiai adatai: <https://github.com/SzemesBotond/s-r-tett-hasonlatok/blob/main/A%20reg%C3%A9nyek%20felhaszn%C3%A1lt%20kiad%C3%A1sa%20C3%A9s%20el%C3%A9rhet%C5%91s%C3%A9ge.xlsx>

válogattunk; ez összesen 58 szerzőt eredményezett, mindegyiknek átlagosan 1,7 regényét. Egy szerzőtől abban az esetben szerepel 4 írás, ha azok között jelentős az időbeli távolság, hogy ezáltal vizsgálható legyen az is, hogy az egyes tendenciákat követi-e az adott alkotó életműve, vagy inkább időszaktól független szerzői „ujjlenyomatokról” (Baayen 2002) beszélhetünk.

A szövegek összetételét az irodalomtörténeti konszenzuson túl a Magyar Elektronikus Könyvtár (<https://mek.oszk.hu/>) és a Digitális Irodalmi Akadémia (<https://pim.hu/hu/dia>) adatbázisai határozták meg. Mivel egyik sem érvényesít határozott irodalomtörténeti koncepciót a digitalizáció során, ezért az ELTE BTK Digitális Bölcsészeti Tanszék által egy nemzetközi projekt (*Distant Reading for European Literary History COST Action CA16204* vö. Schöch et al. 2020) keretében összeállított regénykorpuszhoz fordultunk, amely az 1840 és 1920 között megjelent magyar nyelvű prózairodalomból előre meghatározott szempontok szerint válogat (Bajzát és mtsai 2021), valamint az Akadémia Kiadó *Magyar irodalom* című kézikönyvéhez (Gintli szerk. 2010) és egy, a magyar prózastílus történetét tárgyaló korábbi monográfiához (Herczeg 1981). A korpuszt ezenkívül egy-két népszerű szerzővel egészítettük még ki (például Rejtő Jenő, Gárdonyi Géza).

#### 4. A vizsgálat eredményei

A második fejezetben ismertetett eljárást futtattuk a harmadik részben bemutatott referenciakorpuszon. Összesen 561 darab példányt azonosítottunk a módszerrel, azonban, mint ahogyan azt korábban is felvázoltuk, a módszer nem képes a kedvezőtlen találatokat maximálisan kiszűrni. Előnye viszont az, hogy a keresett jelentésszerkezetre vonatkozóan olyan nagy mintát képes biztosítani, amely a további nyelvi elemzésekhez elengedhetetlen, valamint hozzájárul azon törekvésekhez, hogy a magyar irodalom stílustörténete kvantitatívan, mintázatok azonosításával jobban hozzáférhetővé váljon. Az 561 darab példányt manuálisan elemeztük annak érdekében, hogy felmérjük azt, hogy hogyan oszlanak meg a mintában a jelentéssűrítő műveletet végrehajtó szerkezetek és a mintázatra ugyan formailag illeszkedő, de nem az általunk megcélzott adatok. Mivel a kézi elemzés során nem minden esetben volt eldönthető egy-egy példány kategóriába illesztése, ezért mindhárman egymástól függetlenül végeztük a feldolgozást, majd összevetettük egymással az eredményeinket, és a kérdéses eseteket megvitattuk. Mivel törekedtünk arra, hogy a funkcionális kiindulópontot folyamatosan érvényesítsük, ezért nem lehetett mindig igen-nem oppozícióra leszűkíteni a döntéseinket. Úgy döntöttünk, hogy ahelyett, hogy csak a sűrített hasonlatokat címkézzük találatként, a vele szoros hasonlóságot (prototípuselvű kategorizáció vö. Langacker 1987: 16–17, Taylor 1995; Tolcsvai Nagy 2013, 2017b: 36) mutató, jelentéssűrítő szerkezeteket (lásd 1. fejezet) is megtartottuk. Az 561 példányból összesen 43,67%-a volt azon előfordulások száma, amely nem az általunk keresett szerkezet volt (245 db adat), és 316 volt az, amelyet mindhárman elfogadtunk a kézi elemzés során (56,3%).

A jelen dolgozatnak nem célja az eredmények részletes és szisztematikus bemutatása a referenciakorpuszként használt szövegtömeg összes regényének tükrében, ám pár megfigyelésünket és az ezekből fakadó következtetéseinket közöljük. Fontos, hogy a

találatok alacsony száma miatt a számokat ezen a ponton nem arányítjuk a szövegek tokanszámához. A kutatás későbbi fázisában azonban, amikor más típusú sűrített hasonlatok (*-ként*, *-szerű* ragok és a gyanánt névutó) eloszlását is vizsgáljuk, és így több találatot együttesen kezelünk, már mindenképpen relatív gyakorisággal érdemes dolgoznunk.

Herczeg Gyula stílustörténeti koncepciója szerint a keresett szerkezet a 19. század végétől jelenik meg és válik érzékelhető stíluseszközzé a magyar irodalomban (Herczeg 1981:314–16). Eredményeink ezt annyiban igazolják, hogy kiugró értékeket csak 20. században keletkezett regények vesznek fel a keresések és az eredmények kézi ellenőrzése alapján. A kiugró értéket itt úgy értjük, hogy míg regényenként átlagosan 1,92 db ilyen sűrített hasonlat fordul elő a korpuszként alkalmazott szövegtömegben, addig az ettől jelentősen eltérő szövegstruktúrákban legalább 5 ilyen példány található. Kiemelkedik a korpuszból Határ Győző *Héliáne* című műve (32 darab sűrített hasonlat), amely szövegterében amúgy is gyakran dolgozódnak ki hasonlatok (vö.: Szemes 2021) és jellemző rá a figurális tér nagyobb fokú részletezettsége, amelyet a szerkezetek élénk jelzőhasználata és a hasonlatok sűrítettsége tovább fokoz. Szintén jellemzőnek tekinthető a konstrukció Kosztolányi Dezső konkrétságra és pontos megfogalmazásra törekvő prózájára, amely viszont éppen nem a díszítettség, hanem vélhetően a tömörség miatt fordulhat ehhez a szerkezethez. Nádas Péter *Emlékiratok könyve* című művében szintén gyakrabban adathatunk sűrített hasonlatokat (13 darab), mint a referenciakorpusz jelentős hányadában, viszont érdekes lehet azt megjegyeznünk, hogy a *Párhuzamos történetek* című művében már nem mutatható ki hasonló tendencia (3 darab sűrített hasonlat) – ahogyan azt is jeleznünk kell, hogy a közölt adatok az abszolút gyakoriságra vonatkoznak és nem a szövegek hosszával arányos relatív gyakoriságra.

Azon szerkezeteket tekintve, amelyek nem parafrázálhatók az összehasonlító összetett mondatokkal (lásd 1. fejezet), szintén kimutathatók mintázatok (regényenként átlagosan 1,65 példány fordul elő belőlük). Ezek magasabb előfordulásai összefüggést mutatnak bizonyos esetekben a sűrített hasonlatok nagyobb arányú meglétével. Például Szomory Dezső *A párizsi regény* című művében ezen példányok szintén gyakrabban adathatók (14 darab előfordulás), és a sűrített hasonlatok is viszonylag nagyobb arányban fordulnak elő benne (5 darab). De van arra is példa, hogy míg a sűrített hasonlatokkal történő nyelvi konstruálás jellemzőbb az adott regény szövegterében, addig a jelentéssűrítő, módhatározói kapcsolatra parafrázálható szerkesztésmód egyáltalán nem fordult elő a módszerünkkel generáltatott találati listában (pl. Márai Sándor *A gyertyák csonkig égnek* című regényének esetében, vagy pedig Krúdy Gyula *A vörös postakocsi* című prózájának tekintetében).

## 5. A vizsgált konstrukció jelentéshálózatának szerveződése

A funkcionális kiindulópontból végzett elemzés a vizsgált, DET – N<sub>GEN</sub> – N<sub>INS</sub> – V konstrukció poliszém (poliszémia vö. Cuykens és Zawada 2001, Tolcsvai Nagy 2012, 2017b: 268–288, 2021: 113–121) hálózat azon hálózatrészének elemzésére fókuszál, amely közvetlenül összekapcsolódik a sűrített hasonlatként konvencionizálódott funkcióval, s az ezen csomópont körül szerveződő jelentéseket/jelentésárnyalatokat



elemzi az előzőekben bemutatott módszerrel feltárt konkrét nyelvi példányok alapján. A poliszémia tárgyalásakor a Tolcsvai Nagy Gábor (2012: 288) által összeállított tételeket tekintjük kiindulópontnak: a nyelvi jelentés szerkezetességét és fogalmi alapúságát, a dinamikus relációkon alapuló műveleti működést, a kognitív motiváltságot és a jelentésszerkezetek közötti összefüggést, illetve a prototípuselv alapján megképződő hálózatosságot.

Az általunk fókuszba helyezett jelentésszerkezet és az ahhoz kapcsolódó formai megvalósulás szorosan összefügg más konstrukciós jelentésekkel, ezen eltolódások bizonyos esetekben a szemantikai oldalon történnek meg, másfelől lehetnek a formai oldalról is kimutatható elkülönbözések. Ez utóbbira kevésbé tudunk a jelen eljárásban példányokat hozni, mivel a reguláris kifejezés alapján működő mintavételezési módszerünket szigorú megkötésekkel hoztuk létre. A konstrukciótípus jelentéseinek/jelentésárnyalatainak elkülönítésében segítségünkre lehet a már korábban is említett (részleges) parafrazeálhatóság, illetve a birtokos (genitivusi) szerkezettag és a *-val/-vel* instrumentalis esetragos elvont főnév feldolgozása által aktivált fogalmi tartományok (Lakoff 1987) közötti szemantikai távolság mértéke.

(4) Arnótné az öregasszonyok bölcsességével megérzi, hogy a jövevény, ha lopott vagy ölt.<sup>5</sup>

(5) [...], 2-3000 pengőt is elkért egy műtétért, amelyet a szenvedély ihletével s egy finommechanikai műszerész biztonságával és higgadságával vezetett le, a legnehezebb epeműtettel is altatás [...]<sup>6</sup>

A (4)–(5) példányok esetében jól megfigyelhető az általunk fókuszpozícióba helyezett sűrített hasonlatok tipikus, konvencionalizálódott működése. Akár több tagmondatból álló hasonlító szerkezetté is képesek vagyunk őket (részlegesen) parafrazeálni (pl. *olyan bölcsen, mint amilyenek az öregasszonyok*). Ezen parafrazeálhatóság egyfelől eredeztethető abból, hogy a birtokos (genitivusi) szerkezet által aktivált fogalmi tartomány részeként tudjuk feldolgozni a birtokos által aktivált tulajdonságot, és ezen keresztül a két fogalom között létrejövő integrált reláció (*blend* vö. Tolcsvai Nagy 2021: 109–113) könnyedén összeegyeztethető a világról való tudásunkkal, a mentális begyakorlottsága nagyobb. Az ÖREGSÉG fogalmával könnyedén asszociálható a BÖLCSESSÉG, illetve a finommechanikai műszerészek munkavégzéséhez fűzött elvárásaink között szerepel(het) mind a biztonságra, mind pedig a higgadságra való törekvésük. A (4) példamondat feldolgozása azonban elkülönbözést mutat az (5)-től, hiszen a határozott névelő működése révén a feldolgozás során elérhetővé válik számunkra az az információ, hogy az elsődleges figura (*Arnótné*) maga is birtokosok által kidolgozott csoportba tartozik, míg az (5)-ben feltehetőleg az orvosok és a finommechanikai műszerészek precízsege között válik hozzáférhetővé az analógia.

<sup>5</sup> Darvasi László – A könyvmutatványosok legendája (1999)

<sup>6</sup> Déry Tibor – Felelet (1952)

(6) [...], hangja és arcza élénken emlékeztetett a mentoréra az imént. A gyöngéd szív nemességével nyomott el minden szemrehányást, mely a házat ért csapás [...]

A (6) példány esetében képesek vagyunk a több tagmondatból álló hasonlító mintázatához illesztve parafrázálni a birtokos személyjeles kompozitumszerkezetet (“*nemesen, mint a gyöngéd szív*”). Azonban megfigyelhetővé válik az, hogy a konstruálás és a feldolgozás pillanatában jön létre az integrált szemantikai tér a komponensek jelentésstruktúrái között. Ugyan (feltehetőleg) gyengébb asszociatív viszony fennáll a *gyöngéd szív* és a *nemesség* elvontabb jelentésű alszerkezetek között, de ez korántsem olyan könnyen hozzáférhető és feldolgozható, mint az előzőekben idézett példányok esetében ((4)-(5)). A (6) példány továbbá szemlélteti azt, hogy a konstrukció hálózatosan szerveződő jelentései közötti átmenetiséget a kontinuum-elv (Tolcsvai Nagy 2017a: 56) érvényesítésével ragadhatjuk meg. Ahogyan haladunk az egyre absztraktabb integrált relációt kidolgozó kompozitumok felé, úgy csökken a több tagmondatból álló hasonlatra való parafrázálhatóság valószínűsége.

(7) De a háztetőkön is megmozdult a nép: a szabadulás örömeinek dühével szórták az utcán nyüzsgő ellenségre a követ és gerendákat. [...]

A (7) példányban jól kirajzolódik az, hogy azon esetek, amelyeknél a vizsgált kompozitum komponensei maguk is absztrakt entitások, illetve a közöttük létrejövő integrált tér feldolgozhatósága nehezebb, fogalmilag feltűnőbbek, hiszen a birtokos tag (*örömeinek*) és a birtok (*dühével*) eltérő szemantikai tereket aktiválnak. A poétikai stíluspotenciálja a konstrukcióknak vélhetően ezen tulajdonságból fakad. A konstruáló szubjektum ráadásul a fent idézett példány esetében kiaknázza azt, hogy az öröm és a düh egymással látszólagosan ellentétes érzelmi állapot, viszont az általuk aktivált fogalmi tartományokban a FELFOKOZOTT ÉRZELMI INTENZITÁS jelen van, tehát kimutatható részleges érintkezés a szerkezet elemeinek jelentéskomponensei között. Viszont azon példányokat, amelyek a (7)-hez hasonlóan dolgozódnak ki, azokat már nem a sűrített hasonlat címkével láttuk el a manuális elemzés során. A tipikus sűrített hasonlatoktól való viszonylag nagyobb elmozdulást jelöli az, hogy ezen jelentéssűrítő kompozitumok nem parafrázálhatók a több tagmondatból álló hasonlatok mintázatára. A (7)-ben a formai oldalról tekintve további elkülönbözés figyelhető meg, hiszen összetettebb birtokviszony konstruálódik meg (DET – N<sub>GEN</sub> – N<sub>INS</sub> – V), maga a birtokos is egy másik komponens birtokaként képződik meg.

Az ötödik fejezet célja volt, hogy betekintést nyújtson az általunk vizsgált konstrukcionális jelentés szerveződésére hálózatelvűen, kognitív szemantikai kiindulópontot érvényesítve. Megfigyelhettük a vizsgált adatok tükrében, hogy a szemantikai oldalon történő elmozdulás során a kompozitum jelentésárnyalatai variábilisan szerveződnek. Az elmozdulás mögött kimutatható, hogy a komponensszerkezetek absztrakciója, illetve az általuk aktivált szemantikai terek közötti nagyobb távolság és nehezebb összeegyeztethetőségük állnak. Természetesen nem törekedtünk a jelen dolgozatban a teljes jelentéshálózat szisztematikus feltárására, csupán néhány csomópont bemutatására vállalkoztunk.

---

<sup>7</sup> Gárdonyi Géza – A láthatatlan ember (1902)

## 6. Összefoglalás

A tanulmányunk megkísérelte bemutatni a birtokos szerkezetekbe sűrített hasonlatok automatikus, számítógépes azonosítására kidolgozott eljárásunkat, illetve tartalmazza a vizsgált konstrukció poliszém hálózatának funkcionális kognitív szemantikai elemzését. A kutatás során szembesülnünk kellett azzal, hogy a vizsgálat fókuszába helyezett nyelvi mintázat és a vele asszociálódó hasonlító funkció formai elkülönítése a korpuszban korántsem problémamentes folyamat. Az általunk kialakított, reguláris kifejezésen alapuló módszert igyekeztünk a nyelvi anyag kézi elemzéséből származó tapasztalatok nyomán létrehozni. Az így létrejött adatszűrési metódus hatékonysága abban rejlik, hogy a viszonylag kis számban előforduló nyelvi szerkezet a módszerünk segítségével jól detektálhatóvá válik a bemenetet képező szövegtömegben. A jelen vizsgálat referenciakorpuszaként egy olyan 100 kanonikus magyar regényből álló szöveggyűjtemény szolgált, amely egészen az 1830-as évektől kezdve 2005-ig bezárólag tartalmaz műveket. A jelenlegi eredményeink alapján visszacsatolhatunk Herczeg Gyula azon stílustörténeti koncepciójára, miszerint az ilyen, birtokos szerkezetbe sűrített hasonlatok megjelenése a 19. század végére tehető a magyar prózairodalomban. A kimutatásaink alapján ezen szerkezet már korábban adatolható a vizsgált regényekben, viszont kiugró mennyiségbeli érték ténylegesen a 20. századi művek egy részében válik hozzáférhetővé (pl. Határ Győző *Heliáne*).

A funkcionális kognitív nyelvelméleti keretbe illeszkedő poliszémia-elemzésünk arra vállalkozott, hogy a vizsgált konstrukció jelentéshálózatának azon szegmensét elemezze, amely közvetlenül összekapcsolódik a sűrített hasonlatként konvencionalizálódott funkcióval, valamint bemutassa azt, hogy a kontinuum-elv alapján történő jelentésfeltárás gyümölcsöző eredményeket szolgáltathat a jelen esetben.

A bemutatott vizsgálat egy olyan hosszabb távú kutatás alapját képezi, amely a sűrített hasonlító szerkezetek (pl. főnév + *-ként*; főnév + *-szerű*) kvantitatív és kvalitatív vizsgálatára törekszik. Egy ilyen kutatást egyfelől motivál az, hogy a mintázatok nagy mennyiségű adathalmazon való kimutatása a stílustörténeti tudásunkhoz járul hozzá, másfelől pedig a szerkezet működésének használatalapú megközelítését támogatja.

## Hivatkozások

- Baayen, H., van Halteren, H., Neijt, A., Tweddle, F., An experiment in authorship attribution. JADT 2002: 6es Journées internationales d'Analyse statistique des Donn'ees Textuelles. Conference Paper. (2002)
- Bajzát, T., Szemes, B., Szlávich, E. Az ELTE DH Regénykorpusz és lehetőségei. In: Online térben az online térért: Networkshop 30: országos online konferencia. 2021. április 6-9. pp. 63-72. HUNGARNET Egyesület. Budapest (2021)
- Cuyckens, H., Zawada, B. (szerk.) Polysemy in Cognitive Linguistics. John Benjamins, Amsterdam, Philadelphia. (2001)
- Gintli, T. (szerk.) Magyar irodalom. Akadémiai Kiadó. Budapest. (2010)
- Herczeg, Gy. Határozóként álló, -val, -vel ragos utótagú birtokos szerkezetek. Magyar Nyelvőr 84, pp. 311–573. (1958)
- Herczeg, Gy. A régi magyar próza stílusformái. Tankönyvkiadó Vállalat. Budapest. (1981)

- Imrényi, A. Az elemi mondat viszonyhálózata. In: Tolcsvai Nagy, G. (szerk.) *Nyelvtan. A magyar nyelv kézikönyvtára* 4. pp. 663–761. Osiris. Budapest. (2017)
- Indig, B., Sass, B., Simon, E., Mittelholcz, I., Kundráth, P., Vadász, N. emtsv – egy formátum mind felett. In: XV. Magyar Számítógépes Nyelvészeti Konferencia. pp. 235–247. Szegedi Tudományegyetem TTIK, Informatikai Intézet. Szeged. (2019)
- Kugler, Nóra. Az összetett mondat. In: Tolcsvai Nagy, G. (szerk.) *Nyelvtan. A magyar nyelv kézikönyvtára*. 4. pp. 806–895. Osiris. Budapest. (2017)
- Lakoff, G. *Women, Fire, and Dangerous Things*. The University of Chicago Press, Chicago, London (1987)
- Langacker, R. W. *Foundations of Cognitive Grammar. Volume I Theoretical Prerequisites*. Stanford University Press. Stanford–California. (1987)
- Langacker, R. W. *Grammar and conceptualization*. Mouton de Gruyter. Berlin, New York. (1999)
- Langacker, R. W. *Cognitive Grammar: A basic introduction*. Oxford University Press. Oxford. (2008)
- Paradis, C. Ontologies and construals in lexical semantics. *Axiomathes* 15, pp. 541–573. (2005)
- Schöch, C., Eder, M., Rosari, A., Pieter, F., Primorac, A. *Foundations of Distant Reading. Historical Roots, Conceptual Development and Theoretical Assumptions around Computational Approaches to Literary Texts*. (2020) <https://dh2020.hcommons.org/>
- Szilágyi, M., Vaderna G. A klasszikus magyar irodalom (kb. 1750-től kb. 1900-ig). In: Gintli, T. (szerk.) *Magyar irodalom*. pp. 313–637. Akadémiai Kiadó. Budapest. (2010)
- Simon, E., Indig, B., Kalivoda, Á., Mittelholcz, I., Sass, B., Vadász, Noémi. Újabb fejlemények az e-magyar háza táján. In: XVI. Magyar Számítógépes Nyelvészeti Konferencia. pp. 29–42. Szegedi Tudományegyetem TTIK, Informatikai Intézet. Szeged. (2020)
- Szemes, B. Kidolgozott viszonyok. A tagmondatkapcsolatok automatikus azonosításának hasznosíthatósága a stilsztikában és az irodalomtörténet-írásban, *Digitális Bölcsészet* 4 (2021) - megjelenés alatt.
- Taylor, J. R. *Linguistic categorization*. Oxford University Press. Oxford. (1995)
- Tolcsvai Nagy, G. A poliszémia hálózati modellje. *Nyelvtudományi Közlemények* 108, pp. 287–344. (2012)
- Tolcsvai Nagy, G. Bevezetés a kognitív nyelvészetbe. Osiris. Budapest. (2013)
- Tolcsvai Nagy, G. Bevezetés. In: Tolcsvai Nagy G. (szerk.) *Nyelvtan. A magyar nyelv kézikönyvtára* 4. pp. 23–71 Osiris. Budapest. (2017a)
- Tolcsvai Nagy, G. Jelentés. In: Tolcsvai Nagy Gábor (szerk.) *Nyelvtan. A magyar nyelv kézikönyvtára* 4. pp. 207–466. Osiris. Budapest. (2017b)
- Tolcsvai Nagy, G. (szerk.) *Nyelvtan. A magyar nyelv kézikönyvtára* 4. Osiris. Budapest. (2017c)
- Tolcsvai Nagy, G. *Kognitív szemantika*. Eötvös Kiadó. Budapest. (2021)