

Korpusznyelvészeti módszerek a kognitív nyelvészeti kutatásban

Szerkesztők:

**Kugler Nóra, Modrián-Horváth Bernadett
és Simon Gábor**



DiAGram Könyvek 10.

Sorozatszerkesztők:

Laczkó Krisztina, Tátrai Szilárd és Tolcsvai Nagy Gábor



DiAGram Könyvek 10.

Korpusznyelvészeti módszerek a kognitív nyelvészeti kutatásban

Szerkesztette Kugler Nóra, Modrián-Horváth Bernadett és Simon Gábor



A tanulmányok szerzői: Bajzát Tímea Borbála, Ernhöffer Aliz, Horváth Péter, Indig Balázs, Kalivoda Ágnes, Kugler Nóra, Kurucz Réka, Modrián-Horváth Bernadett, Palágyi László, Simon Gábor, Szemes Botond, Szlávich Eszter, Tanos Bálint

A lektori hozzászólások szerzői: Dömötör Andrea, Horváth Péter, Imrényi András, Indig Balázs, Kalivoda Ágnes, Kugler Nóra, Ladányi Mária, Palágyi László, Simon Gábor, Szemes Botond

A tanulmányokat lektorálta: Dömötör Andrea, Horváth Péter, Imrényi András, Indig Balázs, Kalivoda Ágnes, Kugler Nóra, Ladányi Mária, Palágyi László, Simon Gábor, Szemes Botond, Tátrai Szilárd

A kötet az NKFIH 129040, *A magyar nyelv igei konstrukciói. Használatalapú konstrukciós nyelvtani kutatás* keretében készült. A szerzők és a szerkesztők köszönetet mondanak az NKFIH-nak a pályázati támogatásért és az ELTE Eötvös Kiadónak gondos munkájukért.

© Szerzők, 2024

© Szerkesztők, 2024

ISBN 978-963-489-733-0

ISBN 978-963-489-734-7 (pdf)

ISSN 2732-2815



ELTE | BTK



ELTE | EÖTVÖS
KIADÓ
Az ELTE hivatalos könyvkiadója

eltebook.hu

Felelős kiadó: az Eötvös Loránd Tudományegyetem Bölcsészettudományi Kar dékánja

Projektvezető: Csanádi-Egresi Nóra

Tördelőszerkesztő, borítótervező: Balogh Katalin

Nyomdai kivitelezés: Multiszolg Bt.

Tartalomjegyzék

Bevezetés	9
------------------------	----------

Simon Gábor – Modrián-Horváth Bernadett – Kugler Nóra: Korpusznyelvészeti módszerek a kognitív nyelvészeti kutatásban	11
---	----

Szimultán igekötőzés és igeképzés	27
--	-----------

Kalivoda Ágnes – Palágyi László: Szimultán igekötőzés és igeképzés – konstrukcióegyesítés a magyarban	29
Indig Balázs: Hozzászólás a <i>Szimultán igekötőzés és igeképzés – konstrukcióegyesítés a magyarban</i> című tanulmányhoz	59
Ladányi Mária: Hozzászólás a <i>Szimultán igekötőzés és igeképzés – konstrukcióegyesítés a magyarban</i> című tanulmányhoz	65
Kalivoda Ágnes – Palágyi László: Válasz Indig Balázs hozzászólására ..	73

Appozitív határozók	79
----------------------------------	-----------

Indig Balázs – Kurucz Réka – Ernhöffer Aliz – Kugler Nóra: Appozitív határozók korpuszalapú vizsgálata	81
Kalivoda Ágnes – Palágyi László: Hozzászólás az <i>Appozitív határozók korpuszalapú vizsgálata</i> című tanulmányhoz	117

Beférkőző igék kollostrukciós vizsgálata	123
---	------------

Kalivoda Ágnes – Modrián-Horváth Bernadett – Simon Gábor: Beférkőző igék kollostrukciós vizsgálata. Az igekötő és főnévi igenév közé ékelődő igék kolerámaelemzése az MNSZ2-UD korpuszban	125
--	-----

Dömötör Andrea: Hozzászólás a <i>Beférkőző igék kollostrukciós vizsgálata. Az igeikötő és a főnévi igenév közé ékelődő igék kolexéma-elemzése az MNSZ2-UD korpuszban</i> című tanulmányhoz	155
Imrényi András: Hozzászólás a <i>Beférkőző igék kollostrukciós vizsgálata. Az igeikötő és a főnévi igenév közé ékelődő igék kolexéma-elemzése az MNSZ2-UD korpuszban</i> című tanulmányhoz	159
Simon Gábor – Modrián-Horváth Bernadett – Kalivoda Ágnes: Válasz Imrényi András lektori hozzászólására	167

A magmondat szerepű, passzív jelentésű sémák produktivitása 171

Modrián-Horváth Bernadett – Kalivoda Ágnes – Tanos Bálint: A magmondat szerepű, passzív jelentésű sémák produktivitásának kvantitatív vizsgálata	173
Kugler Nóra: Hozzászólás <i>A magmondat szerepű, passzív jelentésű sémák produktivitásának kvantitatív vizsgálata</i> című tanulmányhoz	225
Palágyi László: Hozzászólás <i>A magmondat szerepű, passzív jelentésű sémák produktivitásának kvantitatív vizsgálata</i> című tanulmányhoz	237

A tagmondatkapcsolattól a komplex birtokos szerkezetig 241

Bajzát Tímea Borbála – Szemes Botond – Szilávi Eszter: A tagmondatkapcsolattól a komplex birtokos szerkezetig: hasonlaltípusok azonosítása és vizsgálata	243
Horváth Péter: Hozzászólás <i>A tagmondatkapcsolattól a komplex birtokos szerkezetig: hasonlaltípusok azonosítása és vizsgálata</i> című tanulmányhoz	281
Simon Gábor: Hozzászólás <i>A tagmondatkapcsolattól a komplex birtokos szerkezetig: hasonlaltípusok azonosítása és vizsgálata</i> című tanulmányhoz	283

A lexikai választékosság és szemantikai specifikusság mintázatai ... 291

Horváth Péter: A lexikai választékosság és szemantikai specifikusság mintázatai a magyar kanonikus költészetben	293
Szemes Botond: Hozzászólás <i>A lexikai választékosság és szemantikai specifikusság mintázatai a magyar kanonikus költészetben</i> című tanulmányhoz	333

Bevezetés



SIMON GÁBOR – MODRIÁN-HORVÁTH BERNADETT –
KUGLER NÓRA

Korpusznyelvészeti módszerek a kognitív nyelvészeti kutatásban

Bevezetés

*“Problems should not be an excuse for linguists
to dig deeper into methodological monism:
they are there to be solved.”*
(Arppe et al. 2010: 6)

1. Kognitív nyelvészet és korpusznyelvészet: fegyvertársak vagy jó ismerősök?

A „kognitív korpusznyelvészet” (*cognitive corpus linguistics*) terminus meglehetősen későn, ismereteink szerint 2010-ben jelenik meg először tudományos munkában (Arppe et al. 2010: 2, lásd még Gries 2017: 599). A korábbi évekből egy másik kifejezés (*corpus Cognitive Linguistics*, Glynn 2010: 3) is adatható, amely mintha azt helyezné előtérbe, hogy korpuszon végzett kognitív nyelvészeti kutatások foglalhatók össze e terminus alatt, nem pedig kognitív személelmódú korpusznyelvészeti elemzések. Felmerül tehát a kérdés, egyáltalán milyen viszonyban áll egymással ez a két kutatási terület: természetes és harmonikus szövetségei-e egymásnak, vagy inkább különböző megközelítésmódok, amelyek bizonyos pontokon összekapcsolhatók az eredményeiket tekintve? Ebben a bevezető tanulmányban a kognitív korpusznyelvészet általános bemutatásával kívánjuk keretezni a kötetben közölt, specifikus jelenségekre összpontosító tanulmányokat.

Már csak azért is szükségesnek tűnik ez a keretezés, mert a korpusznyelvészeti alkalmazását a kognitív nyelvészeti kutatásban korántsem övezte a kezdetektől lelkes támogatás. Stefan H. Gries (2006: 3) szerint „[m]eglepő lehet, hogy – a korábban idézett hozzájárulás és a *használatalapúság* terminusának elmúlt évekbeli prominenciája ellenére – viszonylag kevés kivételtől eltekintve a korpuszalapú megközelítések nem élveztek elsőbbséget” a kognitív nyelvészeti kutatásban. Jan Nuyts úgy fogalmaz 2007-es áttekintő fejezetében, hogy a „szisztematikus korpuszalapú elemzések meglehetősen elterjedtek a funkcionális nyelvészetben, de kevéssé azok a kognitív nyelvészetben”, legalábbis ami annak centrumát jelenti: „a korpusznyelvészet nagyon ritka (ha egyáltalán jelen van) a kognitív nyelvészet magjában, de távolról sem kivételes a funkcionális-kognitív nyelvészetben” (Nuyts 2007: 552) – ez utóbbi kapcsán ő is a használatalapúságot meghatározó Barlow és Kemmer, valamint többek között Geeraerts nevét említi. Dylan Glynn pedig komoly ellenállásról, szkeptikusságról ír a korpuszvezérelt eljárásokkal szemben a tágabb kognitív nyelvészeti közösségben (Glynn 2010: 1). A szkepszis és az elutasítás hátterében több tényező is áll, többek között (i) a jelentés objektív mérhetőségének a kérdése; (ii) az azonosítható formákhoz kapcsolódó elméleti kérdések hiánya; illetve (iii) az intuíció szerepének megítélése a nyelvészeti kutatásban. Az első problémakör kapcsán érdemes megemlíteni Chris Sinha kritikáját: „[s]emmilyen kvantitatív módszer [...] nem képes elemezni a jelentés, a szituáció és az intencionalitás összjátékát”, ezért féltő, hogy „a jelentésre irányuló kognitív nyelvészeti elméleti perspektíva [...] addig a pontig gyengül, hogy egy opcionális kiegészítésévé válik a lényegileg adatvezérelt vállalkozásnak” (Sinha 2017: 496, 497). Vagyis a kvantitatív korpuszelemzésben a kognitív nyelvészet könnyen a jelentésképzés folyamatainak marginalizálódását, illetve téves objektiválását láthatja. A második tényezőhöz kapcsolható az a triviálisnak is tűnő érv, hogy bármilyen jelenség pusztá számszerűsítése még nem vezet el a jelenség jobb megértéséhez. Martin Paul Eve (2022: 1–2) szemléletes megfogalmazásában „a ’bálna’ szó megszámlálása a *Moby-Dick*-ben csupán egy dolgot árul el nekünk: milyen gyakran használják a ’bálna’ szót a *Moby-Dick*-ben”.¹ A harmadik tényező esetében pedig utalhatunk arra a „helytelen arroganciára”, amellyel az adatvezérelt elemzések diszkreditálni próbálják az intuícióra (is)

¹ Laura Janda (2017: 513) általánosabban úgy fogalmazza meg kritikai észrevételét, hogy a kvantitatív eljárások nem teszik a kvalitatív nyelvészeti kutatást sem jobbá, sem szükségképpen empirikusabbá vagy tudományosabbá.

támaszkodó kvalitatív vizsgálatokat, figyelmen kívül hagyva, hogy minden kutatás alapvetően intuícióra épül (Arppe et al. 2010: 21). Ahogyan Janda (2017: 513) fogalmaz, mindig lesznek mintázatok a nyelvi adatokban, ezért a nyelvész feladata annak eldöntése, „mit érdemes keresni”.

Ezekkel a kritikus ellenérvekkel természetesen nem nagyon lehet vitatkozni, és bízunk abban, hogy a jelen kötet tanulmányai sem keltik az olvasóban azt a benyomást, hogy szerzőik pusztá számokra redukálnák a konstruálás bonyolult műveleteit, vagy abban a naiv hitben fordulnának a korpuszokhoz (ironikusan szólva „a nyelv elemzésének Szent Gráljához”, hogy ismét Sinlát (2017: 497) idézzük), hogy abban minden egyéni vélekedést meghaladó, objektív válaszokra lelhetnének az őket érdeklő kérdésekre. Mégis az a meggyőződés vezetett bennünket, e kötet szerkesztőit az összeállítás során, hogy a korpusznyelvészet és a kognitív nyelvészet együttműködése a korábbanál nagyobb hatékonysággal tud magyarázatokat kínálni a nyelvleírás, és különösen a kötet középpontjában álló igei konstrukciók jellemzőire, működtetésükre. Ennek alátámasztására a kezdeti kételyek és a szükségszerű kritika után vegyük sorra azokat a tényezőket, amelyek a korpusznyelvészeti kognitív nyelvészet már-már magától értetődő szövetségesévé teszik.

A kognitív nyelvészet sem a nyelvre nem tekint autonóm képességként, sem a nyelven belül nem tételez diszkrét, önmagukba záródó részterületeket (modulokat) – ez a holisztikusság tétele (Gries 2006), amely a lexikogrammatikai szemléletmódhoz vezet, mely szerint nincs éles határ a (mentális) lexikon elemei és a grammatikai szerkezetek között, a nyelvről való tudásunk a szavakat és a konstrukciókat egységesen szervezi (a rule/list fallacy-ról lásd Langacker 1987: 29). Ez a szemléletmód összhangban áll a korpusznyelvészet idiómaelvével (idiom principle), amely szerint a nyelvhasználó nagy számú részlegesen rögzült (semi-preconstructed) szerkezetet tárol, amelyek egységként érhetők el a számára akkor is, ha elemezhetőnek (részekre bonthatónak) tűnnek (Sinclair 1991: 110). Mivel a korpusz nem atomizált szavakat, hanem kifejezés + kontextus mintázatokat tesz azonosíthatóvá, a szó és a szónál nagyobb szerkezet vizsgálatához is kiváló adatforrás, tehát megfelelő kiindulópont a lexikogrammatika kognitív (és konstrukciós) nyelvészeti modelálásához.

Egy másik gyakran hangoztatott érv a használatalapúság tétele, mely szerint (i) a nyelvről való tudásunk a nyelv használatának tapasztalataiból ered, ezért (ii) a nyelv tudományos leírásának a használat adataira kell épülnie (Butler–González–García 2014: 6). Nos, csaknem minden, a kognitív korpusznyelvészet lehetőségeit tárgyaló közlemény felhívja a figyelmet arra,

hogy a kognitív nyelvészet ezen vállalkozását teljes természetességgel támogatja a nyelvhasználati eseményeket adatbázissá szervező korpusznyelvészeti kutatás. Laura Janda (2017: 498, 500) a használatalapúság igényét tartja a kognitív nyelvészetben megfigyelhető „empirikus fordulat” (empirical turn) egyik fő motorjának, amely az előbbit a módszeres korpuszelemzés irányába mozdította el. Dirk Geeraerts és Gitte Kristiansen (2015: 370) ugyancsak a használatalapúságra vezetik vissza a korpusznyelvészet „felémelkedését” a kognitív nyelvészeti kutatásban, a korpusz ugyanis „a nem elicitált, spontán nyelvi viselkedés megőrzőjeként [...] megfelelő empirikus alapot kínál a használatalapú nyelvészet számára”. Antti Arppe és munkatársai is az elsők között említik az autentikus adatokra épülő vizsgálat igényét a kognitív nyelvészetben, melyhez a korpusz jó adatforrás (Arppe et al. 2010: 1). Nick Ellis (2013: 2) pedig azt emeli ki, hogy egy használatalapú modellben a gyakoriság, a tanulás és a nyelv ismerete együtt járnak, az előbbi pedig a korpuszokban válik igazán jól megfigyelhetővé.

Nem csupán arról van azonban szó, hogy a kognitív nyelvészet elméleti előfeltevéseiből következően olyan igényeket támaszt a nyelvészeti magyarázat felé, amelyeket a korpusznyelvészet mint módszertan hatékonyan ki-elégíthet.² Amellett érvelhetünk, hogy a korpusznyelvészeti hagyományban is vannak olyan elvek és tételek, amelyeket célszerű beépíteni a kognitív nyelvészeti kutatásba. Az egyik ilyen klasszikus állítás Firth-től származik: egy szót azokból a társulásokból ismerhetünk meg, amelyekben megjelenik (“you shall know a word by the company it keeps”, Firth 1962: 11). Azaz nem szabad a jelentéseket izolálva, kizárólagosan a szóhoz rendelve kezelni, azok ugyanis a mintázatokhoz is kapcsolódnak, amelyekben a szavak előfordulnak. Ez tulajdonképpen annak a korai felismerése, hogy a kifejezéseink jelentése nem a szavak önálló jelentéseiből adódik össze, hanem a szavak és a környezetük együtteséből emergál (ehhez kapcsolódóan séma és megvalósulás általános viszonyáról lásd Tolcsvai Nagy 2017: 223), más megfogalmazásban a konstrukciónak saját jelentése is van, nem pusztán a szójelentések grammatikai öntőformája (constructional meaning, lásd Goldberg 2006). A korpusznyelvészetben tehát már régóta alapelv az, amit a kognitív és konstrukciós nyelvtanok napjainkban erőteljesen hangsúlyoznak a nyelvről való gondolkodásban.

² Ezt a „módszertani fegyverhordozói” szerepkört tulajdonítja egyébként Dąbrowska és Divjak (2015: 4) a korpuszelemzésnek, akik szerint a kognitív nyelvészet a korpuszokra mint eszközökre, a korpusznyelvészetre pedig mint módszertanra tekint, nem pedig önálló elméletként értelmezi azt.

Egy másik tézis az amerikai strukturalizmus egyik vezető alakjától, Harristól származik, de a korpusznyelvészet is alapelvnek tekinti: a disztribúciós hipotézis szerint a nyelvi elemek funkcionális (szemantikai, pragmatikai) hasonlóságai tükröződnek disztribúciójuk (azaz a nyelvhasználatban megfigyelhető eloszlásuk) hasonlóságaiban (Harris 1970: 785, idézi Gries 2017: 592). Ha tehát egy kifejezés funkcióit kívánjuk feltárni, célszerű a használati mintázatait megfigyelni, mert az ott látható tendenciák összekapcsolhatók szemantikai motivációkkal. Ez az elv motiválhatja egyrészt a disztribúciós szemantikai kutatásokat; másrészt nem nehéz felfedezni benne a diagrammatikus ikonikusság analógiáját, illetve kiterjesztését, amennyiben a mintázatba rendeződés tendenciái ikonikusan mutatják a funkcionális jellemzők érvényesülését.

Ha tehát a korpusznyelvészetet nem csupán módszertani eszköztárnak, hanem a nyelvleírás önálló teoretikus kiindulópontjának tekintjük, amely a nyelv szerkezeteit mindenkor kontextusukban vizsgálja, és a mintázatba rendeződést helyezi a nyelvi rendszer centrumába, akkor még inkább láthatóvá válnak azok a párhuzamok és kapcsolódási pontok, amelyek mentén a kognitív korpusznyelvészet koherens diszciplináris vállalkozásnak bizonyul. Hangsúlyozni kell: ebben a vállalkozásban nem egyszerűen arról van szó, hogy a kognitív nyelvészet kialakítja a maga modelljeit, meghatározza a maga kategóriáit, majd egyszerűen adatokért fordul a korpuszokhoz. Ez az ún. korpuszal szemléltető (corpus-illustrated) kutatási irány kognitív nyelvészet marad minden tekintetben, hiszen csupán a vizsgált jelenség példái származnak a korpuszból. Sokkal termékenyebb annak a felismerése, hogy „mély korrelációk vannak a korpusznyelvészeti és a kognitív nyelvészeti/konstrukcionista elméletalkotás között” (Wulff 2008: 23), ezért a korpuszadatoknak közvetlen magyarázó értékük van a kognitív nyelvészeti kutatás számára.

Melyek a két nyelvleírási irányzat „feltűnő hasonlóságai” (striking similarities, Wulff 2008: 22)? A magas példánygyakoriság például az elsáncolódás/begyakorlódás (entrenchment) mértékének megállapítását segíti, a magas típusgyakoriság a produktivitásról, a szemantikai generalizációról, a grammatikalizációról adhat információkat, egy erős asszociációs mintázat pedig a várhatósággal és az újszerűség alacsony fokával korrelál (Gries 2017: 593). Fordított irányból közelítve: a prototipikalitás vagy a kategorizáció alapszintje összefüggésbe hozható a példánygyakorisági adatokkal és a kontextusok sokféleségével; a recenciahatásra (amivel később találkozunk, az könnyebben felidézhető) magyarázatot adhat a vizsgált nyelvi jelenség szóródása a korpuszban (azaz mennyire egyenletes a megjelenése a korpusz szövegeiben,

vagy mennyire ritka, de tömbösödő az adott nyelvi megoldás); a szaliencia (hozzáférhetőség, feltűnőség) magyarázható a nyelvi szerkezetek statisztikai várhatóságával, probabilitásával; a kontextusból eredő aktiváltság, illetve a forma-funkció párok pedig összekapcsolhatók a konkordanciák visszatérő jellemzőivel, illetve az együttes előfordulás asszociációs mutatóival (Gries 2017: 593–594).

A korpusznyelvészet a maga adattípusaival a kognitív nyelvészeti kategóriák operacionalizálását teszi lehetővé, így módon hozzájárul azok teszteléséhez, igazolásához és/vagy finomításához. Ebben rejlik a kognitív korpusznyelvészet lényege: olyan kutatás, amely „kérdéseket fogalmaz meg az emberi megismerésről olyan módon, hogy a korpuszadatokra épülő statisztikai elemzés válaszokat adhasson rájuk” (Arppe et al. 2010: 2). Belátható, hogy a kognitív nyelvészet mindaddig heurisztikus vállalkozás marad, ameddig nem szembesíti modelljeit, kategóriáit módszeresen feldolgozott nyelvi adatokkal. A korpusznyelvészet pedig mindaddig pusztán mérési módszertan marad, ameddig adatai nem értelmeződnek valamilyen elméleti keretrendszerben. (Ezt nevezi McEnery és Brezina 2022: 48–49 a korpusznyelvészet elméleti aluldetermináltságának: a korpusz nagyon különböző elméleti perspektívák számára kínál adatokat.) Úgy véljük, a kognitív korpusznyelvészet nem egyszerűen két kutatási terület szerencsés fúziója, hanem egymást kiegészítő elemzési módok integrálódása, egyfajta modus vivendi a nyelvleírás számára napjaink digitális bölcsészeti közegében.

Nem egyforma mértékű természetesen a kognitív korpusznyelvészeti kutatásokban sem a korpuszadatok alkalmazása, illetve szerepe. Az imént említett operacionalizálás, tehát meglévő elméleti modellek módszeres összekapcsolása korpuszadatokkal (és egyúttal e modellek korpuszbeli azonosíthatóságának biztosítása) a korpuszalapú elemzések célja, amelyek tehát döntően kvalitatív jellegűek maradnak, ugyanakkor pontos adatokkal támasztják alá megállapításaikat, következtetéseiket (Gries 2006: 5). Arra is van azonban bőven példa (lásd Gries 2017 áttekintését, valamint Glynn 2010 elméleti bevezetőjét), hogy a nyelvész nem valamilyen elmélethez kapcsolódóan végez korpuszelemzést, hanem a vizsgálni kívánt jelenség minden korpuszbeli előfordulását azonosítva, az így nyert minta statisztikai elemzésével kíván eljutni új elméleti belátásokig, esetleg új elméleti modellig. Ezek a korpuszvezérelt kutatások statisztikai mérésekkel, kvantitatív módon (például kollokációelemzéssel, kollostrukciók feltérképezésével vagy korrelációk keresésével) végeznek mintázatkinyerést, majd a mintázatokat a mérési eredmények mentén általánosítják magyarázataikban.

A kvalitatív és kvantitatív, más megközelítésben korpuszalapú és korpuszvezérelt vizsgálatok persze nem határolhatók el élesen: köztes megoldás például, ha az előzetes elméleti modell hipotézisek felállításához vezet, amelyeket aztán a korpusz adatain, statisztikai módszerekkel tesztel a kutató. Ekkor nem egyszerűen operacionalizálást végez, hanem azért használja a korpuszelemzés módszereit, hogy az eredeti elméletet szigorú validálásnak vethesse alá. Köteünkben az olvasó mindegyik megközelítésre lát majd példát a magyar nyelvi géi konstrukcióinak vizsgálatával kapcsolatban.

Az eddigiekben amellet sorakoztattunk fel érveket, hogy a kognitív korpusznyelvészet minden kritikai észrevétel és látszólagos különbözőség mellett is konzekvens, elméletileg megalapozott és ígéretes kutatási terület. Térjünk ki végül azokra az előnyökre, amelyek a korpusznyelvészeti orientációból származhatnak a kognitív nyelvészetben. Mivel a prototipikus korpuszelemzés rendszerszerű és kimerítő jellegű (exhaustive, Gries 2006: 4), minden, a kutatás számára releváns adatra kiterjed, ezért a korábbiaknál szélesebb körű empirikus evidenciák lehetőségét nyújtja a kognitív nyelvészeti vizsgálat számára. Egy másik prototipikus jegye a korpuszon végzett elemzésnek (Gries 2006: 4), hogy a bináris kategóriák mellett a nyelvi jelenségek skalaritását is megragadhatóvá teszi mérhető adatokkal (például asszociációs erősségek formájában), ezért a kognitív nyelvészetben gyakran megjelenő kontinuumelvet empirikusan is alá tudja támasztani (lásd korábbi érvelésünket a holisztikusság kapcsán). Kimerítő jellege és adatvezéreltsége révén a korpuszelemzés olyan megvalósulásait is megfigyelhetővé teheti a nyelvi kategóriáknak, amelyek más módon elkerülnék a figyelmünket (Janda 2017: 511), vagy mert marginálisan találkozunk velük saját nyelvhasználatunkban, vagy mert nem feltételezzük róluk, hogy összefüggésben állhatnak valamilyen általánosabb jelenséggel. Végül az sem elhanyagolható, hogy a korpusznyelvészettel megvalósított módszeres együttműködés a nyelvtechnológia és a természetes nyelvfeldolgozás felé is nyitottá teheti a kognitív nyelvészetet, ez a kollaboráció pedig hosszabb távon növeli mind a számítógéppel támogatott kognitív nyelvészeti kutatások, mind a hétköznapiakra fejlesztett nyelvtechnológiai alkalmazások hatékonyságát (Janda 2017: 512). Ilyen jellegű együttműködésre kötetünk tanulmányai is szép példákat mutatnak.

A kognitív nyelvészet a nyelvet mint a megismerés fontos tényezőjét kutatja, tehát az elme felől közelít a nyelvhez. A korpusznyelvészet a nyelvet mint használati események sorát dokumentálja, és az így nyert adatokon végez méréseket. Magától értetődik, hogy a nyelvhasználat adatai nem reprezentálják közvetlenül az emberi elme folyamatait és struktúráit: miként Dylan

Glynn (2010: 15) fogalmaz, „a korpusznyelvész nem fér hozzá a beszélő elméjéhez”.³ Ráadásul a korpusz sohasem öleli fel a nyelv teljes használatát, hiszen csak törekedhet a reprezentativitásra, de teljesen megvalósítani azt sosem tudja. Mégis feltételezhető, hogy ha egy jelenségnek markáns mintázata van a korpuszban, azt a nyelvhasználók elméje is specifikus módon reprezentálja, tehát a mérhető és a feltételezhető korrelál egymással. Ezt nevezi Tony McEnery és Vaclav Brezina (Popper nyomán) „a valósággal való kvázi-kapcsolatnak”: a korpusz adatai nem tükrözik a nyelvi valóságot vagy a nyelvről való fogalmi tudást, hanem következtetni engednek azokra (McEnery–Brezina 2022: 47–48). Ha ezt elfogadjuk, elkerülhetővé válik a hamis objektivitás csapdája (és az azzal együtt járó redukcionizmusé), ugyanakkor a módszeres adatelemzés stabil alapjára építhető a kognitív nyelvészeti modellalkotás.

2. A kötet koncepciója

A tanulmánykötet esettanulmányok bemutatásával arra fókuszál, hogyan használhatóak a korpusznyelvészeti és nyelvtechnológiai módszerek, a létrehozott adatbázisok és keresőeszközök kognitív nyelvészeti kutatási kérdések megválaszolásában, különösen a magyar igei konstrukciókra irányuló vizsgálatokban.⁴

A szerkesztői koncepciónak megfelelően minden tanulmányban összekapcsolódnak a kognitív nyelvészeti kérdések, problémák, a korpuszelemzés módszerei és a nyelvtechnológiai alkalmazások. A tanulmányokat a két irányultságnak megfelelően választott szakmai lektorok bírálták, minden esetben két nem anonim lektor. A lektorok – nem kötelező jelleggel – véleményüket, meglátásaikat, továbbvezető gondolataikat lektori hozzászólás formájában a kötetben közre is adták. A szerzőknek nemcsak reflektálniuk kellett a bírálatokra az elfogadott javaslatokat a végleges szövegbe beépítve, hanem – ha úgy látták jónak – közlésre szánt választ is megfogalmazhattak

³ Arppe és munkatársai (2010: 2) hasonlóan fogalmaznak: „A megismerés tanulmányozása off-line adatokon keresztül értelemszerűen indirekt, még akkor is, ha ezek az adatok természetesen, reprezentatívak és sokfélék.” Az idézett tanulmányban Gaëtanelle Gilquin pedig még határozottabban érvel: „A puszta szövegekből a kognitív szalienciára irányuló naiv leképezés (mapping) valószínűleg téves. A korpuszok nem léphetnek a kísérletes kutatás helyébe” (Arppe et al. 2010: 8).

⁴ A tanulmánykötet az NKFIH 129040 számú pályázatának (A magyar nyelv igei konstrukciói. Használatalapú konstrukciós nyelvtani kutatás) támogatásával valósult meg. Közvetlen előzménye a pályázat keretében lebonyolított, azonos (*Korpusznyelvészeti módszerek a kognitív nyelvészeti kutatásban*) című, workshop jellegű előadás-sorozat volt (<https://diagram.elte.hu/content/korpusznyelvészeti-módszerek-a-kognitív-nyelvészeti-kutatásban.t.19512>).

(ezzel a lehetőséggel csak kevesen éltek). A lektorok megkapták a bírálatukra adott szerzői válaszokat, megismerték a véglegesített kéziratot, és volt lehetőségük arra, hogy hozzászólásukba ezzel kapcsolatban is beépítsenek észrevételeket. A szerzők pedig megismerve a végleges hozzászólást, ismét megfontolhatták, hogy kívánnak-e a kötet lapjain választ megfogalmazni.

A tanulmánykötet ezzel a megoldással elő kívánja segíteni a különböző nézőpontok közötti eszmecserét, a diszkurzivitást. A megoldás a lektorálási, átdolgozási műveletek transzparenciájához is hozzájárul, ami azért lényeges, mert sokszor az olvasó számára nem válik hozzáférhetővé a szerzők és a lektorok közötti vita, egyezkedés, holott ez is hozzájárul ahhoz, hogy egyre szorosabb együttműködésben, együttes megértésben jelentős komplexitású leírások szülessenek nagymennyiségű nyelvi adat feldolgozására támaszkodva.

A koncepció nem új, előzményként említhetjük például a *Linguistic Discovery* folyóiratnak a szemantikai térképekkel foglalkozó tematikus kötetét (lásd Cysouw–Haspelmath–Malchukov 2010), de megszokottnak sem számít a magyar nyelvészeti szintén. A szerzők és a lektorok közötti diszkusszió közreadása minden alkotó résztvevőt kihívások elé állít, természetesen az olvasót is beleértve. A fejezetszerkezet (tanulmány, lektori hozzászólás(ok), szerzői válasz a lektor(ok)nak) ugyanis nem képezi le az alakulástörténetet, a tanulmány egyszerre kezdeményezője és eredménye is az együtt gondolkodásnak. A lektori javaslatok egy része beépült a tanulmányokba, és ezekre már a tanulmány szövegében is olvashatók reflexiók (lábjegyzetben). A hozzászólásokban pedig olyan megjegyzések is vannak, amelyek a tanulmánynak a lektor által inspirált átdolgozási folyamatához, a folyamat eredmény(esség)éhez fűznek kommentárt. Továbbá – a transzparenciára való törekvés mellett is – a különböző szereplők gondolatai, az eltérő mentális ágensekhez köthető tartalmak szorosan össze is fonódnak, a tanulmány és a hozzászólások emergens módon kapcsolódnak össze. Ebbe a folyamatba hívjuk be alkotó (értelmező, továbbgondoló) és – reményeink szerint – alkalmazó résztvevőként az olvasókat.

3. A kötet tanulmányai⁵

Kalivoda Ágnes és Palágyi László tanulmánya, mely a *Szimultán igekötőzés és igeképzés – konstrukcióegyesítés a magyarban* címet viseli, két passzív

5 Mivel a tanulmányok elkészítésében a Bevezetés szerzői is érintettek, a rövid ismertetések elkészítését úgy osztottuk fel, hogy szerző ne vegyen részt saját vagy általa lektorált munkának a bemutatásában.

állapotváltozást kifejező konstrukció, a *-(s)Ul* és a *-(s)Odik* sémák igekötős megvalósulásait vizsgálja adatvezérelt kvantitatív módszerrel. A szerzők azokat az eseteket tekintik konstrukcióegyesítésnek, amelyekben az igekötőről és a képzőről való tudás nem választható le egymásról, hanem feltételezi egymást. Az általuk bemutatott modell szerint a kétféle képzési séma összekapcsolódását a morfológiai konstrukciók hálózatos szerveződése teszi lehetővé, ezáltal használatalapú magyarázatot adnak egy olyan konstrukcióra, melyet a korábbi szakirodalom csak a potenciális szó fogalmán keresztül tudott leírni. A konstrukcióegyesítés példányokon való vizsgálatához a szerzők az MNSZ2-UD korpusz alapján létrehoztak egy 1,1 millió példányt tartalmazó, szabad hozzáférésű adattáblát, melyben egyéb adatok mellett feltüntetik az igekötő nélküli és igekötős gyakoriságot, az igekötős használat arányát az összes előforduláshoz képest, valamint a konstrukcióban megfigyelhető igekötős használat és az összes igekötős használat arányát. A legnagyobb értékekkel rendelkező példányokat, melyek a konstrukcióegyesítés legjellemzőbb példányai, részletesen bemutatják a tanulmányban, szemantikailag elemezve az igék eseményszerkezetét. Az írás második részében található, szintén adatvezérelt kvantitatív produktivitási mérések árnyalt képet mutatnak az igekötő nélküli, valamint a különféle igekötős konstrukciók használatáról. Az eredmények alapján az igekötővel való együttes használat mindkét vizsgált képzési séma esetében növeli a produktivitást, így például a szakirodalomban eddig nem produktívként számontartott *-(s)Ul* képző/képzési séma esetében kimutatható volt, hogy igekötő nélkül korlátozott produktivitásúnak, igekötővel viszont egyértelműen produktívnek tekinthető.

Indig Balázs, Kurucz Réka, Ernhöffer Aliz és Kugler Nóra *Appozitív határozók korpuszalapú vizsgálata* című esettanulmánya a főnévi alaptagot bővítő, a főnév utáni helyzetben előforduló, azaz appozitív határozók (például *kilábalás a válságból; napilapok nemzeti tulajdonban*) korpuszalapú azonosításával foglalkozik. Ahogy azt hozzászólásukban a lektorok is megjegyzik, „a vizsgált konstrukció korpuszalapú vizsgálata nagy kihívás”. Ugyanakkor épp a megoldás keresése vezetett el ahhoz, hogy a dolgozat fókuszába viszonyítási pontként a *való* jelzősítő szerkezet (például *válságból való kilábalás*) került, nem közvetlenül az appozitív határozós konstrukció. Továbbá az adatok arra is késztették a szerzőket, hogy egy harmadik szerkezetet is bevonjanak a hálózatos leírásba, mégpedig a határozói bővítmény és N alaptag szórendű – *való* jelzősítő nélküli, „szomszédos” – konstrukciót (például *válságból kilábalás*), feltűnővé vált ugyanis ezen szerkezetek nagy száma. A dolgozat megvizsgálja, hogy mely főnévi szóalakpárokkal, páronként milyen gyakorisággal

adatokkal a *való* jelzősítő (például *válságból való kilábalás*), a *való* jelzősítő nélküli (például *válságból kilábalás*) és az appozitív konstrukció (például *kilábalás a [gazdasági] válságból*). Az eredmény eltér az előzmény-szakirodalom alapján megrajzolt képtől.

A *Beférkőző igék kollostrukciós vizsgálatával* foglalkozik Kalivoda Ágnes, Modrián-Horváth Bernadett és Simon Gábor közleménye. Az igekötő + ige + infinitívusz – az ige perspektívájából: beférkőző/beférkőzéses – szerkezet kollostrukciós vizsgálatát a szerzők az MNSZ2-UD korpusz adatain végzik el. A tanulmánynak nem célja a magyar segédige-kategória körülhatárolása, eredményei azonban ebből a szemponttól is értékesek, ugyanis lehetővé teszik az igék kvantitáív jellemzését a beférkőzés (mint a segédigékre lehetségesen jellemző viselkedés), a konstrukcióval való asszociálódás mértéke alapján. Az egyszerű kolecémaelemzés hét lexémát azonosított a beférkőző szerkezettel szignifikánsan asszociálódó igeként: *kell, tud, van/lehet, akar, fog, mer* és *talál*. Ezek mind (pre-)modális és aspektuális-temporális tartományokat nyitnak meg az infinitívusszal kidolgozott folyamat típus értelmezése számára. A beférkőzéses konstrukció komplex magmondatot alkot, melyben az igealak a diskurzus alapjához le is horgonyozza az előhívott folyamatot. A szakirodalomban vagy konszenzusos, vagy legalábbis felvetődött már ezen igéknek a segédigei kategóriába sorolása. Lektor kritika éppen azért fogalmazódott meg a módszerrel kapcsolatban, mert a nyelvészeti intuíció alapján a fenti lista (amely nem tartalmazza például a *szokott, tetszik* igéket) esetleg szűknek tűnhet. A szerzők reflektálnak is a tesztek korlátaira, úgy látják, hogy a szerkezettel gyengébben asszociálódó igék feltárása „a vizsgálatban alkalmazható szignifikanciaszűkítés mellett nem kivitelezhető egzakt módon”, ez további méréseket igényelne (ezekre a szerzői válaszban javaslatok is olvashatók). Összességében a tanulmány (illetve a közleményt és a hozzászólásokat is tartalmazó fejezet) nemcsak a beférkőző igékkel kapcsolatos tudásunkat bővíti, hanem magának az igekötő + ige + infinitívusz sémának és tágabban a hasonló beférkőző konstrukciónak (komplex magmondatoknak) a pontosabb leírásához is hozzájárul. Bár elméleti, kognitív nyelvészeti kiindulópontból a konstrukció vizsgálatához kézenfekvő viszonyítási pont a protoállítás (a semleges állító kijelentő mondat magmondatának) tipikus szórendje, a szerzők rámutatnak arra, hogy az adatvezérelt vizsgálat ilyen előzetes viszonyítási pontot nem tud tételezni, és nem is lehet célja baseline sémát igazolni. A szerzők és a lektorok eszmecseréje ebben a témában gyümölcsözően járult hozzá ahhoz, hogy a „korpuszvezérelt kvantitatív kutatás” és az „introspekción alapuló sugaras hálózatkutatás” viszonya az egyensúly irányába mozduljon el.

A magmondat szerepű, passzív jelentésű sémák produktivitásának kvantitatív vizsgálata című tanulmányban Modrián-Horváth Bernadett, Kalivoda Ágnes és Tanos Bálint négy (eltérő mértékben) passzív jelentésű konstrukció (V-ódik/-ődik, V-(t)atik/-(t)etik, V-va/-ve lesz és V-va/-ve van) produktivitását vizsgálja az MNSZ2, valamint az MNSZ2-UD korpuszok adatain. A szerzők ötvözik az adatok gépi és kézi elemzését, továbbá a produktivitás kvantitatív keretű mérését és a sémák igenemek szerinti besorolásában a kvalitatív módszertant. A szerzők egyik központi módszertani kérdése az volt, hogy összehasonlítható-e a szintetikus és az analitikus konstrukciók globális produktivitása. A (kontrollként a mérésbe beemelt) V-ni fog elemzése (az eredményeknek az alkalmazott keretben való értelmezése) azt jelzi, hogy a globális produktivitás nincs összefüggésben a szintetikus vagy analitikus jelleggel. A vizsgálat eredménye továbbá az is, hogy a produktivitás különböző mutatói alapján (a korpusz egészében) más-más mintázat a kiugró. A V-va/-ve van konstrukció a legnagyobb globális produktivitással rendelkező séma, a legnagyobb típusgyakorisággal a V-ódik/-ődik, míg a legnagyobb potenciális produktivitással a V-va/-ve lesz rendelkezik. A legstabilabban (szinte kizárólagosan) passzív jelentésű séma az alacsony produktivitású V-tatik/-tetik konstrukció, ezt megközelíti a V-va/-ve lesz, túlnyomórészt passzív jelentésű a V-va/-ve van is, a kevésbé stabilan passzív tartományba a V-ódik/-ődik tartozik. A szerzők gondosan elemzik az alkorpuszok szerinti megoszlást is, ennek eredményeként tárul fel az, hogy a spontán hétköznapi nyelvhasználatot leginkább képviselő, „személyes” alkorpuszban a V-va/-ve lesz globális produktivitása kiugróan magas (a séma pedig nagyon markánsan passzív). Ezért a szerzők arra következtetnek, hogy elsősorban a V-va/-ve lesz séma alkalmas arra, hogy magmondatként produktívan megkonstruálja a dinamikus passzív jelentéstartalmat. Az új eredményekhez az is hozzájárul, hogy a szerzők együttesen figyelembe veszik a paradigmatis relációkat és a jelentéviszonyokat. Így például a V-va/-ve lett szinte kizárólagos passzív jelentése nemcsak illeszkedik a dinamikus passzív paradigmába (beletartozik a V-va/-ve lesz sémába), hanem a statikus – dinamikus viszonyban (vö. V-va/-ve volt – V-va/-ve lett) is értelmeződik. A tanulmány a típusgyakoriság megállapításán és a hapaxok vizsgálatán túl a bemeneti igék, valamint a típusok igenem szerinti kategorizációján keresztül eddig nem ismert variabilitást tár fel a sémák megvalósulásában. A dolgozatnak azonban nemcsak a megválaszolt kérdései, elemzési eredményei jelentenek újdonságot, hanem a nyitott kérdései is érdemi hozzájárulások a kognitív korpusznyelvészet vállalkozásához.

Bajzát Tímea Borbála, Szemes Botond és Szilávi Eszter tanulmánya, *A tagmondatkapcsolattól a komplex birtokos szerkezetig: hasonlattípusok azonosítása*

és vizsgálata ötféle hasonlatot kifejező konstrukciótípus kognitív szemantikai jellemzését, ezek automatizált azonosítását és ehhez kapcsolódóan kvantitatív korpuszvizsgálatukat tűzi ki célul. A két tagmondatból álló, prototipikusnak tekintett hasonlító mintázat mellett négy, különböző komplexitású „sűrített” hasonlító mintázat: a *-ként* esetragos, a *-szerű* utótagos, a *gyanánt* névutós és a birtokos szerkezetbe sűrített hasonlító konstrukciók képezik a vizsgálat tárgyát, így a vizsgált konstrukció jelentéshálózatának leírásában érvényesül a funkcionális szemléletre jellemző kontinuum-elv. Kutatói korpuszként az ELTE DH Regénykorpuszának 100 kanonikus regénye szolgált. A szerzők értelmezésében a hasonlatok használatával a megnyilatkozó nem csupán egy entitást egy másik entitáson keresztül tesz hozzáférhetővé, hanem az entitások közötti feldolgozott relációkon keresztül új tudás létrehozása is történik a közös figyelmi jelenetben. A szerzők minden egyes vizsgált konstrukcióra az automatikus felismerhetőséget célzó algoritmusokat dolgoznak ki, melyek különböző lépéseket foglalnak magukba a tokenalapú kereséstől a reguláris kifejezéseken át a gyakorisági listákon, illetve kollokációelemzéseken alapuló stop-szó listák összeállításáig. A gépi elemzés új alkalmazási lehetőséget nyit például a stílustörténeti kutatások számára, de a szerzők röviden bemutatják az egyes hasonlattípusok korrelációvizsgálatának eredményét, valamint a mondatok átlagos hosszúságának korrelációit is az egyes hasonlító konstrukciókkal.

Horváth Péter *A lexikai választékosság és szemantikai specifikusság mintázatai a magyar kanonikus költészetben* című tanulmánya a versek lexikai választékosságának témáját járja körül 24 kanonikus magyar költő művei alapján. A szerző automatizált méréseket dolgoz ki és végez el az ELTE Verskorpuszának versein és referenciakorpuszként az ELTE Regénykorpusz szövegein, a lexikai választékosság mérésének lemmagyakoriság alapú, idegennyelv-elsajátításra kifejlesztett módszerét alkalmazva. A lexikai választékosságot a versek terjedelmével, formai komplexitásával, valamint a vers belső szerkezeti jellemzőivel összefüggésben is megvizsgálja. Az eredmények magasabb lexikai választékosságot mutatnak a közepes hosszúságú (50–199 szót tartalmazó) versek esetében, míg alacsony a mutató a formailag egyszerű versek esetében. A lexikai választékosság terjedelemmel való összefüggéseit a rövidebb (1–49 szavas) versek esetében a fókuszáltabb megfogalmazások, a 200 szónál hosszabb versek esetében a részletesebb leírások magyarázhatják, míg a formailag egyszerű versek alacsonyabb választékossága a versforma és tartalom közti ikonikus leképezésre vezethető vissza. Az egyes verseken belül a szerző egyre csökkenő értékeket állapított meg a versek előrehaladásával, míg a páratlan sorokénál rendre magasabb értékeket talált a versek páros soraiban. A versen belüli egyre csökkenő értékek magyarázatát Horváth

Péter a specifikustól az általános felé haladó tartalom jellegzetességében látja, a versszakokon belüli sorpár alapú fluktuáló mintázatot pedig a szemantikai specifikáltságra vezeti vissza, a két jelentésegység között specifikus-általános absztrakt viszonyt feltételezve. A szerző által alkalmazott vizsgálati módszer segítségével azonosíthatóvá vált a magyar kanonikus költészet kitüntetett verstípusára vonatkozó, a vers lexikai választékosságának és szemantikai specifikusságának makro- és mikroszintű sémája.

Irodalom

- Arppe, Antti – Gilquin, Gaëtanelle – Glynn, Dylan – Hilpert, Martin – Zeschel, Arne 2010. Cognitive Corpus Linguistics: five points of debate on current theory and methodology. *Corpora* 5: 1–27. <https://doi.org/10.3366/cor.2010.0001>
- Butler, Christopher S. – González-García, Francisco 2014. *Exploring Functional-Cognitive Space*. Amsterdam, Philadelphia: John Benjamins. <https://doi.org/10.1075/slcs.157>
- Cysouw, Michael – Haspelmath, Martin – Malchukov, Andrej 2010. Introduction to the special issue „Semantic maps: methods and applications”. *Linguistic Discovery* 8 (1): 1–3. <https://doi.org/10.1349/PS1.1537-0852.A.358>
- Dąbrowska, Ewa – Divjak, Dagmar 2015. Introduction. In: Dąbrowska, Ewa – Divjak, Dagmar (eds.): *Handbook of Cognitive Linguistics*. Berlin, Boston: De Gruyter Mouton. 1–9. <https://doi.org/10.1515/9783110626476-001>
- Ellis, Nick 2013. Corpus study: Cognitive implications. In: Chapelle, Carol A. (ed.): *The encyclopedia of applied linguistics*. H. n.: Blackwell Publishing. <https://doi.org/10.1002/9781405198431.wbeal0261>
- Eve, Martin Paul 2022. *The digital humanities and literary studies*. New York: Oxford University Press. <https://doi.org/10.1093/oso/9780198850489.001.0001>
- Firth, John R. 1962. *Studies in linguistic analysis*. Oxford: Basil Blackwell.
- Geeraerts, Dirk – Kristiansen, Gitte 2015. Variationist linguistics. In: Dąbrowska, Ewa – Divjak, Dagmar (eds.): *Handbook of Cognitive Linguistics*. Berlin, Boston: De Gruyter Mouton. 366–389. <https://doi.org/10.1515/9783110292022-018>
- Glynn, Dylan 2010. Corpus-driven cognitive semantics. Introduction to the field. In: Glynn, Dylan – Fischer, Kerstin (eds.): *Quantitative methods in cognitive semantics: Corpus-driven approaches*. Berlin, New York: De Gruyter Mouton. 1–41. <https://doi.org/10.1515/9783110226423.1>

- Goldberg, Adele E. 2006. *Constructions at work. The nature of generalization in language*. New York: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199268511.001.0001>
- Gries, Stefan Th. 2006. Introduction. In: Gries, Stefan Th. – Stefanowitsch, Anatol (eds.): *Corpora in Cognitive Linguistics. Corpus-based approaches to syntax and lexis*. Berlin, New York: Mouton de Gruyter. 1–17. <https://doi.org/10.1515/9783110197709.1>
- Gries, Stefan Th. 2017. Corpus approaches. In: Dancygier, Barbara (ed.): *The Cambridge handbook of Cognitive Linguistics*. Cambridge: Cambridge University Press. 590–606. <https://doi.org/10.1017/9781316339732>
- Janda, Laura A. 2017. The quantitative turn. In: Dancygier, Barbara (ed.): *The Cambridge handbook of Cognitive Linguistics*. Cambridge: Cambridge University Press. 498–514. <https://doi.org/10.1017/9781316339732>
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar*. Vol 1. Theoretical prerequisites. Stanford: Stanford University Press.
- McEnery, Tony – Brezina, Vaclav 2022. *The Fundamental Principles of Corpus Linguistics*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781107110625>
- Nuyts, Jan 2007. Cognitive linguistics and functional linguistics. In: Geeraerts, Dirk – Cuyckens, Hubert (eds.): *The Oxford handbook of Cognitive Linguistics*. New York: Oxford University Press. 543–565.
- Sinclair, John 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinha, Chris 2017. Opening commentary: Getting the measure of meaning. In: Dancygier, Barbara (ed.): *The Cambridge handbook of Cognitive Linguistics*. Cambridge: Cambridge University Press. 493–497. <https://doi.org/10.1017/9781316339732>
- Tolcsvai Nagy Gábor 2017. Jelentéstan. In: Tolcsvai Nagy Gábor (szerk.): *Nyelvtan*. Budapest: Osiris Kiadó. 207–499.
- Wulff, Stefanie 2008. *Rethinking idiomaticity. A usage-based approach*. London, New York: Continuum.



Simon Gábor

Eötvös Loránd Tudományegyetem

simon.gabor@btk.elte.hu

ORCID: 0000-0001-5233-6313

Modrián-Horváth Bernadett

Szegedi Tudományegyetem

bernadett.modrianhorvath@gmail.com

ORCID: 0000-0003-3834-9163

Kugler Nóra

Eötvös Loránd Tudományegyetem

kugler.nora@btk.elte.hu

ORCID: 0000-0002-8223-2926

Szimultán igekötőzés és igeképzés



KALIVODA ÁGNES – PALÁGYI LÁSZLÓ

Szimultán igekötőzés és igeképzés – konstrukcióegyesítés a magyarban¹

Kivonat

Tanulmányunkban a szimultán igekötőzést és igeképzést mint konstrukcióegyesítést vizsgáljuk adatvezérelt megközelítéssel. Azt feltételezzük, hogy a konstrukcióegyesítés folyamatát a különböző szóképzési konstrukciók hálózatos szerveződése teszi lehetővé, valamint hogy a morfológiai konstrukciók produktivitasában és kidolgozásában a sémáknak és az analógiának egyaránt szerepe van. Kutatásunk alapja az MNSZ2-UD korpusz, amelyből kinyerjük az összes releváns, -(s)*Odik* és -(s)*Ul* képzős derivátumot a kontextusukkal, metaadataikkal és különféle formai jellemzőikkel együtt. Ez összesen 1,1 millió példányt jelent, amelyeket szabadon elérhető adattábla formájában is közzéteszünk. A kutatásunk célja az, hogy feltérképezzük az igekötőket és a passzív állapotváltozást kifejező igeképzőket egyesítő konstrukciók jellemző példányait, valamint a konstrukciók globális produktivitását.

Kulcsszavak: konstrukcióegyesítés, igekötők, igeképzők, magyar nyelv, adatvezérelt megközelítés

¹ Szeretnénk köszönetet mondani a tanulmányunk két bírálójának, Ladányi Máriának és Indig Baláznak a cikk gondos átolvasásáért, a sok hasznos javaslatért. Köszönet illeti Sass Bálintot a módszertani kérdésekben nyújtott segítségéért. Kalivoda Ágnes kutatása a PD 142317 számú projekt keretében, az Innovációs és Technológiai Minisztérium Nemzeti Kutatási Fejlesztési és Innovációs Alapból nyújtott támogatásával, a PD_22 pályázati program finanszírozásában valósult meg. Palágyi László kutatását az NKFIH K 129040 számú pályázata (A magyar nyelv igei konstrukciói. Használatalapú konstrukciós nyelvtani kutatás) támogatta.

1. Bevezetés

Noha az igekötőzésnek számos szemantikai (lásd Szilágyi N. 1996; Szili 2005a, 2005b; Szili 2009; Tolcsvai Nagy 2013), morfológiai (lásd Kiefer–Ladányi 2000) és szintaktikai (többek között É. Kiss 2004) megközelítése ismeretes, abban osztoznak a különböző megközelítések, hogy az igekötős ige szerkezetét alapvetően egy igekötő és egy ige összeépüléseként értelmezik. Azonban az „igekötő+ige” morfológiai felépítés kapcsán valójában nem mindig tudunk redukcionizmussal élni, a szerkezetet nem tudjuk pusztán az igekötő és az ige összeépüléseként értelmezni, ha az igekötőzés nem létező igrén megy végbe. A jelen tanulmány olyan igekötős igrék vizsgálatát célozza, amelyeknek a bázisigéje ún. potenciális szó², más szóval, amelyeknek valójában nincs bázisigéjük. A *bedűhődik*, *feldűhődik* (?*dűhődik*), *elhidegül* (?*hidegül*), *bepipul* (?*pipul*), *elérzékenyül* (?*érzékenyül*), *elmosolyodik* (?*mosolyodik*), *elszomorodik* (?*szomorodik*) igrék arra utalnak, hogy a magyar beszélők tudásának részét képezheti egy igekötőket (*fel*, *el*, *meg*, *be*), melléknévi jelentést (*dűhös*, *hideg* 'idegen, közönyös', *pipa* 'dühös', *érzékeny*) és ahhoz járuló passzív igréképzőket (-(s)*Odik*/(s)*Ul*)³ egységben tartalmazó szerkezeti mintázat. Használatalapú megközelítésben ez a mintázat konstrukcióként, azaz sematikus forma–jelentés párként írható le. Ennek a konstrukciónak a sajátossága, hogy a hálózatosan gyakran összekapcsolódó konstrukciók (igekötőzés, igréképzés) egyesülése (unification) révén áll elő (Booij 2010a: 41–50). Ugyanakkor pusztán kvalitatív elemzéssel, lexikai alapon nem lehetséges a mintázat teljes feltérképezése, a feltételezett konstrukció adatokkal való alátámasztása nyelvtechnológiai kihívásként is értelmezhető. A tanulmány célja az igekötőket és passzív igréképzőket egyesítő konstrukció megvalósulásainak (konstruktumainak) kinyerése

² A potenciális szó fogalmával Ladányi Mária a tanulmányhoz fűzött véleményében részletesen foglalkozik. E helyen annyit jegyezni meg, hogy a fogalom nem csupán a kompetencia és a performancia, hanem a szinkronia és a diakronia dichotomikus elkülönítését is támogatja, mivel eliminálja a nyelvi változást a szóképzést illetően. Véleményünk szerint azonban a szóképzésben a nyelvi működés és változás elválaszthatatlanul van jelen, és egy új szó létrejötte valamilyen termékeny, szabályszerűséget mutató általánosítás alapján tipikus példája a nyelvi változásnak is. Minthogy – használatalapú keretben – a szóképzési sémák a meglévő (és a sémát felidéző) derivátumokról való beszélői tudásra épülnek, egy új derivátum megerősíti (vagy részben módosítja) magát a sémát is séma és megvalósulás dinamikus viszonyrendszerében.

³ A tanulmányban használt fonológiai, morfológiai és korpusznyelvészeti rövidítések jegyzéke az I. számú mellékletben található.

a korpuszból, azok kvalitatív elemzése, valamint a konstrukció produktivitásának kimutatása a hapaxok vizsgálatával. A tanulmány második részében a kutatásnak elsősorban a szóképzésre és az igekötőzésre vonatkozó elméleti keretét ismertetjük. A harmadik részben a kutatás módszertanát mutatjuk be. A negyedik részben azt tárgyaljuk, hogy a kvantitatív kutatás alapján az egyesített konstrukciónak milyen mintázatai vannak az egyes igekötőket és képzőket illetően. Az ötödik részben a konstrukció produktivitásával foglalkozunk, a hatodik részben pedig a tanulmány összeggel zárul.

2. Elméleti háttérfeltevések

A hagyományos morfológiai leírás alapegysége a morféma, a szavak szerkezetességéért a morfémák kombinációja a felelős, az általánosítások (szabályok vagy sémák) a morfémákra vonatkoznak, a szavak konstituenciája a morfémák mint önálló reprezentációval bíró építőelemek révén állandó (rész → egész). Ezzel szemben az igekötőket és igeképzőket egyesítő konstrukció vizsgálatában leírásunk alapegysége a szó(alak) és a morfológiai konstrukció (Booij 2010a). A szóalak egy fonológiai és egy szemantikai struktúra szimbolikus egysége (Langacker 1987), a szóalak morfológiai jellege a forma és a jelentés szó szintű együttes variálódásával áll elő (vö. Haspelmath–Sims 2010: 3), a kötött morfémák a szóalakok egymáshoz való viszonyából általánosított nyelvi egységek, amelyeknek nincs önálló, lexikai elemektől független reprezentációjuk (egész → rész) (Blevins 2006). A morfológiai konstrukció szó szintű sematikus forma–jelentés pár, amely ugyancsak szóalakokból elvont általánosítás eredménye (Booij 2010a). A szó szóalakok hálózata, míg a (forma és jelentés együttes variálódásához hozzájáruló, vagyis morfológiai) szavak szóképzési hálózatot képeznek, amelyek a szóalakhálózatokhoz hasonlóan többirányú kapcsolódások mentén épülnek föl, és a derivátumok morfológiai jellege szintén konstrukciós sémák révén mutatkozik meg.

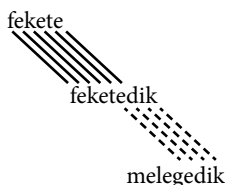
2.1. A szóképzési konstrukciók működése és egyesülése

A konstrukciós szemléletű, szóalapú morfológia a derivátumok struktúrájának elemzését illetően is alternatíváját jelenti a morfémaalapú leírásoknak. Minthogy a szavak derivációs hálózatot képeznek, a szóképzési művelet mindig túlmutat az alapszó és a derivátum viszonyán. Eszerint például az *erősít* derivátum a $[[N-s]_{ADJ-ít}]_V$ általánosítás eredményeként éppúgy leírható

(lásd *erő* → *erős* → *erősít*), mint a [ADJ-ít]v általánosítás eredményeként (*erős* → *erősít*) attól függően, hogy a derivátumot a derivációs hálózat mely részére vetítjük rá. A deriváció hálózatos jellege magyarázatul szolgál a morfémaalapú disztribúciós elemzések azon problémájára, hogy „sokszor nem könnyű eldönteni, hogy két képző egymásutánjáról vagy egy összetett képzőről van-e szó” (Keszler 2000: 315), illetve arra, hogy egy szóképzési művelet hogyan mehet végbe nem létező szavakon (lásd *tanú* → **tanús* → *tanúsít*). Utóbbi jelenség halmozódása annak is az indikátora, hogy két (vagy több) szóképzési séma az egyesülés (unification) irányában alakul (Booij 2010a: 41–50).

A szóképzési konstrukciók forrásalapú (source-oriented) és kimenetalapú (product-oriented) általánosításokra épülnek (Bybee 2001). A forrásalapú általánosítás információt tartalmaz az alapszó jellegéről, illetve az alapszó és a derivátum közötti formai és szemantikai leképezés módjáról. A kimenetalapú általánosítás viszont csupán azt rögzíti, hogy a szóképzési művelet eredményének milyen formai és szemantikai sémát kell megvalósítania. A forrásalapú általánosítás tehát különböző sémák hálózatos, asszociatív viszonyára épül, például meghatározhatja a konstrukció fonológiai és/vagy szemantikai alkalmazási tartományát/hatókörét, formai és jelentésbeli kötöttségeit (lásd Ladányi 2017: 550–554). A szóképzési konstrukcióban a forrásalapú és kimenetalapú általánosítások különböző mértékben játszanak szerepet. Mind alaki, mind szemantikai kidolgozás szempontjából kimenetalapú sémát képviselnek például a *battyog* és *billeg* típusú, [(C)(C)VC:Og] alaki mintázatot kidolgozó onomatopoeitikus igék (Kalivoda 2021), amelyeknek gyakran nincs is alapszavuk, vagyis a jelentésük és alakjuk pusztán kimenetalapú általánosítások felől értelmezhető. Ebben az esetben a hangzásséma hangsúlyozottan globális (Szilágyi N. 2013) abban az értelemben, hogy a szerkezet részletezőbb jelentésű alkotórészeiről való tudásunk (például *batty*-, *bill*-) sem választható el a derivátumtól, csupán a teljes szerkezetben érvényesül. A forrásalapú és kimenetalapú általánosítások egyaránt fontosak lehetnek a konstrukcióról való tudásunkat illetően. A tanulmányban tárgyalt, passzív igeképző konstrukciót képviselő *sárgul*, *barnul*, *szürkül*, *kékül*, *zöldül*, *őszül* igék szemantikai kidolgozása során fontos szerepet játszik az a forrásalapú általánosítás, hogy a derivátum a névszói komponensek (például színnevek) által felidézett állapotba kerülést jelöli. Az ige alaki kidolgozása során ugyanakkor az a kimenetalapú általánosítás is fontos szerepet játszik, hogy a derivátum két vagy három szótagos, illetve *-CUL* szótagra végződik. Az egy szótagos melléknevek és két szótagos, mássalhangzóra végződő melléknevek tehát egyszerűen kiegészülnek az *-UL* végződéssel

(*kék* → *kékül*, *zöld* → *zöldül*, *halvány* → *halványul*), a két szótagos, magánhangzóra végződő melléknemek viszont úgy dolgozzák ki a konstrukciót, hogy az alapszó második szótagjának magja helyett a toldalék *u* vagy *ü* hangja szerepel, így illeszkedve a mintázat -*CUI* végződéséhez: *sárga* → *sárgul*, *szürke* → *szürkül*. Az adott színt tehát a *sárg-*, *barn-*, *szürk-* *kék-*, *zöld-* tövek egyaránt felidézik, a sémát a két szótagos, magánhangzóra végződő és az egy szótagos, mássalhangzóra végződő melléknemek különbözőképpen, de hasonló eredménnyel dolgozzák ki: *sár\$gul* ~ *zöl\$dül* ~ *ké\$kül* ~ *ö\$szül* (a \$ jel szótaghatárt jelöl). A magyar színnevek tehát a következő kimenetalapú általánosítást dolgozzák ki ebben a konstrukcióban: [(C)V(C)\$CUI]. A fentebb tárgyalt, passzív állapotváltozást kifejező igei konstrukció forrásalapú általánosítása részben átfedésben van a kimenetalapú általánosítása szerint (egyes szám harmadik személyben, jelen időben és kijelentő módban ikes) legalább négy szótagú, felső szótagszám-határt pedig nem tartalmazó [Adj-Od(ik)]v sémával. A három szótagú, vagy két szótagú, de mássalhangzóra végződő szavak nem mindig illeszkednek a [(C)V(C)\$CUI] sémába (*meleg* → **melegül*, *beteg* → **betegül*, *fekete* → ?*feketül*), kidolgozhatják viszont ugyanebben a jelentésben a szótagstruktúrát illetően nem specifikált [ADJ-Od(ik)]v sémát, például *melegedik*. Mind az [ADJ-Od(ik)]v, mind az [ADJ-UI]v szóképzési séma jelentése: 'az alapszóban megadott tulajdonságúvá válik'. A folyamat nem kötődik cselekvő ágenshez, szereplője az átélő, az experiens, például *öreg-ed(ik)*, *sötét-ed(ik)*, *meleg-ed(ik)*, *vékony-ed(ik)*. A kimenetalapú általánosítás az [ADJ-Od(ik)]v sémánál is a tő „csonkulásához” vezet az alapszóhoz képest, az -ú, -ü végű névszók a szóvégi magánhangzó nélkül tesznek eleget annak az általános követelménynek, hogy a derivátum alapalakjának -COd(ik) végződésűnek kell lennie: (*savanyú* →) *sava\$nyodik* ~ *mele\$gedik* (← *meleg*). Ugyanakkor például a *fekete* alapszó szóvégi -e magánhangzója átfedésben van a konstrukció hangzássémájával, ezáltal a *feketedik* harmadik -e- magánhangzója a részletező jelentésű alapszót és a képzőt is felidézi attól függően, hogy a hálózat mely részeire vetítjük rá, vagyis többszörösen motivált, ahogy az 1. ábrán is látható.



1. ábra. A *feketedik* ige harmadik *-e*- elemének többszörös motiváltsága

Az [ADJ-*Od(ik)*]v konstrukció az *-s* melléknévképzős konstrukcióval egyesített formában is termékeny. Az egyesülés (unification) kiindulópontja, hogy a derivációs hálózatban az *-s* melléknévképzős séma és a passzív állapotváltozást kifejező igei konstrukció paradigmaticusan összekapcsolódik. Az [ADJ-*Od(ik)*]v konstrukciót ugyanis számos *-s* képzős melléknév kidolgozza: például *erős-odik*, *rozsdás-odik*, *szőrös-odik*, *poros-odik*, *felhős-odik*, *házas-odik*. Az egyesülés hátterében feltételezhetően a két képző tömbösödése (chunking) áll, amely folyamat számos *-(V)sOdik* végű igében végbe-mehetett. Az *-s* gyakori tövégi előfordulása révén egyesült az igeképzővel és sematizálódott a konstrukcióban, amelyet ezáltal az alapszó transzparensabb módon dolgozhat ki: *egyszerű-södik* (vö. nincs sem **egyszer-edik*, sem **egyszerű-s* szó a magyarban, azaz a konstrukcióegyesítés nélkül az ige felépítése nem magyarázható meg). A konstrukcióegyesítés leírható úgy is, mint egy többszörös forrásalapú általánosítás kimenetalapúvá válása. A többszörös forrásalapú általánosítás információt tartalmaz a leképezési lánc módjáról. Az *erő* – *erő-s* – *erő-s-odik* hálózatrész leírható a [[N-(V)s]Adj-*Od(ik)*]v többszörös forrásalapú általánosítás eredményeként. A kimenetalapú általánosítás viszont csupán azt rögzíti, hogy az ige formai oldalon *-sOd(ik)* végződésű, szemantikai oldalon pedig a szó névszói, részletező jelentésével asszociált állapotba kerülést jelzi: [állapot-*sOd(ik)*]v. A konstrukcióegyesítésben tehát fontos szerepet játszik a konstrukciók közeledése a derivációs hálózatban (networking) többszörös forrásalapú általánosítások mentén (lásd [[N-(V)s]Adj-*Od(ik)*]v), ez vezet a különböző konstrukciók egyesüléséhez egy olyan kimenetalapú általánosítás jegyében, amelyet számos, eredetileg különbözőképpen létrejövő derivátum képvisel. Pusztán ebben a(z) alapszó alakját háttérbe helyező) megközelítésben állítható párhuzamba például a *karcsúso-dik* (**karcsús-*) és az *erősödik* szó alakja (lásd még **korszerűs-*, *korszerű-södik*, **észszerűs-*, *észszerűsödik*, **nagykorús-*, *nagykorúsodik*, **egyértelműs-*, *egyértelmű-södik*). A számítógépes nyelvészeti eszköztár lehetőséget nyújt

a sémához csak kimenetalapúan kapcsolódó, vagyis vélhetően egyesített konstrukciót képviselő szavak feltérképezésére.

2.2. Az igekötős konstrukció

Minthogy az igekötő és az ige szemantikai egysége nem, illetve nem feltétlenül jár együtt fonológiai egységességgel (az igekötős ige nem fonológiai szó), az igekötős igék a szó és a szónál nagyobb (szintaktikai) szerkezetek között helyezhetők el a lexiko-grammatikai kontinuumon. Az igekötők az esetragokhoz hasonlóan számos szintaktikai konstrukcióhoz hozzájárulnak, amelyekben a jelentésük erősen konstrukciófüggő. Például az ún. okozott mozgás konstrukcióban az igekötők a tárgyeset, a delativus és (opcionálisan) az instrumentalis mintázatában a morfológiai esetekhez hasonló funkciót töltenek be, jelentésük a teljes konstrukció felől értelmezhető. A *lerúg* ige igekötője például nem feltétlenül a rúgás irányát, hanem a (rúgással) okozott mozgás irányát jelöli a *lerúgta a virágot az erkélyről (a labdával)*, vö. *ledobta a labdát az erkélyről* (Palágyi 2022).

Az igekötős igei konstrukciók feltűnően dinamikusak abban az értelemben, hogy némely jelentésük – lásd irányjelentés, aspektuális jelentés – kifejezetten magas általánossági szinten kezelhető, míg más jelentéseik kidolgozásában nyilvánvaló a már meglévő példányok és az analógia szerepe. Az igekötős ige prototipikus kompozitumszerkezete aszimmetrikus: a sematikusabb szemantikai szerkezetű komponens, az igekötőt kidolgozza egy részletezőbb jelentésű komponens, az ige (Tolcsvai Nagy 2013). Azonban az igekötős sémák gyakran a vártnál nagyobb mértékben módosítják az igék jelentését. A példányok szerepét mutatja, hogy az új igekötős igék jelentése pusztán az igekötő + ige kapcsolódásból nem mindig vezethető le, a megértéséhez szükség van analógiás csoportként a hasonló jelentésű példányok aktiválására is: *jár : repül : autózik : utazik ~ bejárta : beutazta : beutózta : berepülte* (+ accusativus, például *Európát*). Ezek a gyakran metaforikus jelentések jól elkülönülnek egymástól, a *le* igekötő például hozzájárul a tranzitív 'felülmúl' (*legyőz*) és az intranszitiv, egy szereplőt profiláló 'passzíválódik/passzívál' jelentéshez is (*leáll*) (Palágyi 2022). Jellemző, hogy az igen specifikus konstrukciós jelentések sokszor viszonylag kevés, ugyanakkor viszonylag gyakori igekötős igére épülnek, mégis produktívak. A következő igék például a [*le-V*] konstrukció 'felülmúl' sematikus jelentését dolgozzák ki: *lefutja/leússza az ellenfelét, ledumálja/leérveli a vitapartnerét*.

Az igekötős igei konstrukciók tehát jól mutatják, hogy a séma mint általánosítás és az analógia mint példány szerepének megkülönböztetése mögött valójában csupán szemléleti különbség rejlik. A mi megközelítésünk szerint a kidolgozás sémaalapúsága (szabályszerűsége) és analógiás (egyedi) jellege pusztán a konstrukciót képviselő mintázatok eltérő méretéből adódó, látszólagos megkülönböztetés (vö. Booij 2010b). Az analógiának is központi motívuma az (akár egyetlen példány valamely tulajdonságán alapuló) általánosító művelet, miképpen az általánosítások (sémák) mind alakilag, mind szemantikailag lexikai nyálábok sematizációi, amelyekkel dinamikus séma-megvalósulás viszonyt tartanak fenn a sematizáció után is (Bybee 2010), vagyis nem a példányaitól független reprezentációk. A sémák alapjául szolgáló lexikai nyálábok feltérképezése korpuszalapon válik lehetségessé.

3. A kutatás módszere

A kutatásunk során adatvezérelt, kvantitatív megközelítést alkalmazunk. A munkánk reprodukálhatósága érdekében elérhetővé tettünk a https://github.com/kagnes/odik_ul repozitóriumban minden olyan scriptet és szöveges segédfájlt, amelyet az adatgyűjtéshez és adatelemzéshez készítettünk. Nyilvánosan elérhetővé tettük az adatgyűjtés eredményeként létrejött adattáblát is, erről bővebben lásd a 3.4. alfejezetet.

3.1. A felhasznált korpusz

A kutatásunk alapja az MNSZ2-UD korpusz (Kalivoda et al. 2024). Ez a megjelenés előtt álló korpusz az 1,04 milliárd szavas, 20–21. századi szövegeket tartalmazó Magyar Nemzeti Szövegtár 2 (Oravecz et al. 2014) legújabb változata. Számos korpusztisztítási lépésen esett át, amelyek közül a legfontosabbak a paragrafus-szintű duplumszűrés, a nem magyar nyelvű bekezdések automatikus azonosítása és törlése, valamint az 500 tokennél hosszabb mondatok kiszűrése. Az utóbbiak jellemzően nem valódi mondatok, hanem kódrészletek, vagy hibásan feldolgozott weboldalak. Ezáltal a szöveg minősége jobb, de a korpusz mérete kisebb az MNSZ2-höz képest (776,9 millió szövegszó).

Az MNSZ2-UD kapcsán fontos még kiemelni, hogy új elemzési szintekkel gazdagodott: az emMorph (Novák et al. 2016) által létrehozott, részletes morfológiai elemzéssel, valamint függőségi elemzéssel, amely a Universal

Dependencies formalizmusát követi (de Marneffe et al. 2021). A kutatásunk számára alapvető fontosságú ez a két új annotációtípus. A részletes morfológiai elemzés abban segít, hogy könnyen megtaláljuk a *-(s)Odik* és *-(s)Ul* képzős igéket, és pontosan azonosítani tudjuk a szóképzés bázisául szolgáló névszókat. A függőségi elemzés pedig abban nyújt segítséget, hogy az igétől elvált igekötőt a megfelelő igetőhöz tudjuk kapcsolni, valamint könnyen meg tudjuk állapítani az ige vonzatkeretét is.

3.2. Az adatgyűjtés folyamata

A korpuszból lekérdeztünk minden olyan mondatot, amely legalább egy *-(s)Odik* vagy *-(s)Ul* képzős finit igét illetve infinitívuszt tartalmaz. Ezeket a részletes morfológiai elemzésben a `[_AdjVbz_Ntr/V]` címke keresésével lehetett azonosítani. A címke feloldása: olyan igeképző, amely melléknévből intranszitiv igét képez. Az (1) példa a *pirosodik* igeen keresztül illusztrálja azt, hogy a keresett címke hogyan jelenik meg a részletes morfológiai elemzésben.

- (1) `piros[/Adj|col]odik[_AdjVbz_Ntr/V][Prs.NDef.3Sg]`⁴

Azt tapasztaltuk, hogy a mai magyar nyelvre kifejlesztett `emMorph` elemző bizonyos alakulatokban már nem azonosítja ezeket a képzőket, ilyen például az *elgonoszul* ige (ebben tehát a *gonosz* tő nem jelenik meg). Ezért végeztük egy lemma-alapú keresést is, amelyben csak azt figyeltük, hogy az adott ige lemmája az *-ul*, *-ül*, *-odik*, *-edik*, *-ődik* betűsorok valamelyikére végződik-e. Ezeket a lemmákat külön is elemeztük az `emMorph` középmagyar nyelvállapotra finomhangolt verziójával. Az elemzés eredményét minden – összesen 5839 darab – lemma esetében átnéztük, és a hamis pozitív találatokat kiszűrtük. Ilyen volt például a *kutyul*, amelynek esetében nem arról van szó, hogy valami kutyává válik. Természetesen előfordulnak olyan találatok, amelyek erősen kontextusfüggők, ilyen például a *vadul* szó. Ez lehet hamis pozitív találat, amennyiben határozói funkcióban, 'vad módon' jelentésben áll (például *vadul forgatja a szemét*), és lehet számunkra releváns találat is, '(lassanként) vaddá válik' jelentésben (például *nehogy állatokat táplálván*

⁴ A példában látható `emMorph` címkek feloldása: `[/Adj|col]` = szint jelölő melléknév; `[_AdjVbz_Ntr/V]` = olyan igeképző, amely melléknévből intranszitiv igét képez; `[Prs.NDef.3Sg]` = tárgyatlan ragozás, egyes szám harmadik személy

és nevelvén maguk is állatokká vaduljanak). Minden egyes előfordulás kézi átnézésére nem vállalkoztunk, mivel 6,2 millió mondatból álló anyaggal volt dolgunk; az ilyen kétértelmű eseteket egységesen megőriztük. Ezáltal hibás adatok is bekerülnek a vizsgálatainkba, de ez ekkora mennyiségű szöveg automatikus feldolgozása során elkerülhetetlen.

Ezután kinyertük a mondatokból minden egyes *-(s)Odik* vagy *-(s)Ul* képzős ige összes olyan morfológiai és szintaktikai jellemzőjét, amelyet lényegesnek tartottunk a kutatásunk szempontjából. Ezek a következők: lemma, tehát az inflexiós toldalékok nélküli szóalak; igekötő;⁵ információ arról, hogy az igekötő prefixumként fordul elő vagy elválik az igéjétől; igelemma (igekötő nélkül); az igeképzés bázisául szolgáló névszó; igeképző; vonzatok és bővítmények. Ami az igekötőket illeti, a prefixumként megjelenőket a részletes morfológiai elemzés segítségével tudtuk azonosítani, az elváltakat pedig a függőségi elemzés *compound:preverb*⁶ relációja alapján találtuk meg. A vonzatokat és bővítményeket úgy állapítottuk meg, hogy az ige dependenseit kerestük, ezek lemmáját és esetragját nyertük ki. A névutókat is esetragként kezeltük. A (2) példában összefoglalva láthatók az eddig bemutatott jellemzők, azokkal a változónevekkel, amelyeket később az adattáblában is használunk.

(2) Legerősebben két kép *elevenedett meg* az emlékeimben.

→ form:	elevenedett
→ lemma:	meg+elevenedik
→ prev:	meg
→ prevtype:	SEP (azaz: <i>separated</i> 'igétől elvált')
→ verb:	elevenedik
→ root:	eleven
→ suffix:	-(s)Odik
→ argframe_cases:	Ess Ine Nom
→ argframe_long:	erős[Ess] kép[Nom] emlék[Ine]

⁵ Egymás alakvariánsainak tekintettük és azonos alakra normalizáltuk a következő igekötőket: *belé* → *bele*, *bé* → *be*, *föl* → *fel*, *föül* → *felül*, *fönn* → *fenn*, *odább* → *odébb*, *reá* → *rá*, *teli* → *tele*.

⁶ A magyar igekötőre nehéz megfelelő angol terminust találni, a nemzetközi szakirodalomban (*verbal*) *prefix*, *preverb*, *coverb* vagy *verbal particle* néven hivatkoznak rá. A UD formalizmus a *preverb* ill. *prev* címkéket használja, ezért mi is ezeknél maradtunk az adattábla készítése során. Ez az elnevezés arra utal, hogy az igekötő leggyakoribb, tipikus pozíciója az ige előtt van, de természetesen más pozíciókban is előfordulnak igekötők.

Az igeképzés bemenetül szolgáló névszókat megvizsgáltuk word2vec szóbeágyazáson alapuló módszerrel is (Siklósi–Novák 2016).⁷ Az eljárás lényege az, hogy a szavakat egy valós vektortér pontjaiként képezzük le, és ebben a térben a hasonló szavak egymáshoz közel, az eltérőek egymástól távol rendeződnek el. A hasonlóság a választott modelltől függően szintaktikai vagy szemantikai jellegű lehet. Mi az utóbbira fókuszáltunk: arra voltunk kíváncsiak, hogy kirajzolódnak-e szemantikailag hasonló csoportok az igeképzések bemenetül szolgáló szavak között. Minden találat esetében megjelenítjük annak a klaszternek – csoportnak – a számkódját, amelybe az adott névszó a szóbeágyazás során került. A (3) példásor a 3-as klaszter tagjait illusztrálja. Amint látjuk, ide különféle nemzetiségnevekből képzett szavak kerültek.

- (3) a. Nagypját még Virágnak hívták, az ír körülmények között ez *angolosodott* Bloomra.
 b. A bányászok kezdetben németek voltak, akik később *elhorvátosodtak*.
 c. [...] a második világháború után részben újra *visszaszlovákosodtak*.
 d. E város régi német lakosai a 16-ik században *tótosodtak el* egészen.
 → w2v_cluster: 3

Az adatgyűjtés során megőriztük a *-(s)Odik* és *-(s)Ul* képzős igék mondat-szintű kontextusát is, valamint a dokumentumszintű metaadatokat: a publikálás évét, a stílusregisztert és a fájlnévet. Ezeket mutatja be a (4) példa.

- (4) Jónak nevezte az előterjesztésben, hogy *egységesülnek* a közteher-alapok.
 → doc_year: 2008
 → doc_style: press
 → doc_id: 2.press_hu_origo_uzletinegyed_007_2008.vert

⁷ Indig Balázs jelzi a lektori hozzászólásában, hogy a választott szóbeágyazási módszer nem járul hozzá a tanulmányunkban bemutatott eredményekhez. Bár végül valóban nem használtuk fel az eredmények értelmezésében, egy korábbi, hasonló szellemben készült publikáció (Kalivoda 2021) meggyőzően mutatja be a módszer hasznosságát, hozzánk hasonlóan csak a bázisszavakra alkalmazva. Az adattáblában meghagytuk ezt az információt, mert biztosak vagyunk benne, hogy jövőbeli kutatások még hasznát tudják venni.

- left_context: Jónak nevezte az előterjesztésben, hogy
- kwic:⁸ egységesülnek
- right_context: a közteheralapok.

3.3. Módszer az egyesített konstrukciók azonosítására

Miután a korpuszban szereplő összes, *-(s)Odik* és *-(s)Ul* képzős igéhez előálltak a fentebb bemutatott adatok, lehetővé vált annak a meghatározása, hogy ezek milyen valószínűséggel képviselnek egyesített konstrukciót. Arra számítottunk, hogy ha az igekötős alak jóval gyakoribb, mint az igekötő nélküli, az az egyesített konstrukció indikátora lehet. Ezen a ponton felmerül az a fontos módszertani kérdés, hogy pontosan mit értünk „jóval gyakoribb” alatt. Ennek a megválaszolására saját módszert dolgoztunk ki a korpuszadatok ismeretében, amely három számértéken alapszik:

1. Az igelemma tokengyakorisága, az igekötős és igekötő nélküli eseteket egybeszámolva (freqsum). Erre azért van szükség, hogy az adatritkaságból adódó problémákat elkerüljük. Például az adattáblában a *metálosodik* ige nem szerepel önmagában, csak a *meg* igekötővel együtt (*megmetálosodott*), de mivel ez az egyetlen előfordulása a korpuszban, nem indokolt az egyesített konstrukció jellemző példányaként figyelembe vennünk.
2. Az igelemma igekötős előfordulásainak száma osztva az összes előfordulásának számával (prev_vs_all). Ez az érték azt mutatja meg, hogy az adott ige jellemzőbben áll-e igekötővel, mint igekötő nélkül, azonban arról, hogy egy- vagy többféle igekötővel áll-e, még nem teszünk pontos állítást. Például a korpuszban gyakran, 13747-szer előforduló *bizonyosodik* mindig valamilyen igekötővel jelenik meg, de változó, hogy melyik ez az igekötő (például *be-*, *meg-*, *rá-*, ... *-bizonyosodik*).
3. Az adott igekötő+igelemma előfordulásainak száma osztva az igelemma összes igekötős előfordulásának számával (actprev_vs_allprev). Ezen az értéken keresztül azt láthatjuk, hogy az igelemma mennyire „preferál” egy bizonyos igekötőt. Például a *bizonyosodik* igelemma az esetek 79%-ában a *be* igekötővel áll.

⁸ A korpusznyelvészethen használt kwic feloldása *keyword-in-context*, azaz a keresett szó vagy kifejezés.

Az olyan igék esetében van a leginkább okunk egyesített konstrukciót feltételezni, amelyeknél mindhárom felsorolt érték magas: az ige kellően gyakori ahhoz, hogy megalapozott állítást lehessen tenni róla, nagyon gyakran – vagy kizárólagosan – igekötővel áll, ezen belül pedig főként egy konkrét igekötővel szokott előfordulni. Ilyen igéket mutatunk be részletesen a 4. fejezetben.

3.4. Az elkészült adattábla

Az adatgyűjtési folyamat eredményeként egy olyan adattáblát kaptunk, amely a *-(s)Odik* és *-(s)Ul* képzős finit igék és infinitívusok összes, MNSZ2-UD korpuszban szereplő előfordulását tartalmazza. Ez összesen több mint 1,1 millió találatot jelent. Az adattábla formátuma TSV (tab separated values): egy egyszerű táblázat, amelyben az első sor a fejléc, a további sorok az egyes korpuszbeli előfordulásoknak felelnek meg, az oszlopok pedig a vizsgált jellemzőknek. Az oszlopok részletes ismertetése a II. számú mellékletben található. Az erőforrás nyílt hozzáférésű, a <https://zenodo.org/record/7607661> linken érhető el.

4. Az egyesített konstrukciók jellemző példányai

Az 1. táblázatban szereplő igekötős igék esetében a legvalószínűbb, hogy az egyesített konstrukciót képviselik, azaz esetükben az igekötőről és a képzőről való tudás nem választható szét, hanem feltételezi egymást. A táblázat azokat az igéket tartalmazza – igekötő, bázisszó, képző bontásban –, amelyek legalább 500-szor fordultak elő, az esetek legalább 90%-ában az ige igekötővel áll, illetve az esetek legalább 90%-ában egy meghatározott igekötővel fordul elő. Ladányi Mária javaslatára először ‘adott igekötő – adott képző’ párok szerint rendeztük az adatsort, hogy a tipikus mintázatok azonnal kivehetők legyenek, majd ezen belül az ‘akt. IK %’ értékek csökkenő sorrendjében.

Az általánosságban elmondható, hogy ezen igék a folyamatot a végpont felől figyeltetik meg, minthogy az igekötő általában végpontot helyez a folyamatra (Tolcsvai Nagy 2013). A táblázatban szereplő igék tanúsága szerint ennek számos szemantikai motívuma lehet.

1. táblázat. Azok az igekötős igék, amelyek a legnagyobb valószínűséggel képviselnek egyesített konstrukciót

igekötő	bázisszó	képző	ált. IK %	akt. IK %	gyakoriság
el	külön	-(s)Ul	1,00	1,00	1823
el	híres	-(s)Ul	0,99	1,00	657
el	törpe	-(s)Ul	0,94	0,99	1953
el	homályos	-(s)Ul	0,95	0,98	533
el	hideg	-(s)Ul	0,97	0,97	675
el	sötét	-(s)Ul	0,91	0,96	1511
el	bizonytalan	-(s)Odik	0,99	1,00	2772
el	néptelen	-(s)Odik	0,99	1,00	711
el	értéktelen	-(s)Odik	0,92	1,00	589
el	adós	-(s)Odik	0,99	0,99	2170
el	hatalmas	-(s)Odik	0,96	0,99	1091
el	komor	-(s)Odik	0,93	0,98	552
el	mérges	-(s)Odik	0,94	0,95	1541
meg	valós	-(s)Ul	0,99	1,00	42 513
meg	sebes	-(s)Ul	0,99	1,00	11 892
meg	semmis	-(s)Ul	0,99	1,00	6397
meg	testes	-(s)Ul	0,96	0,99	1450
meg	özvegy	-(s)Ul	0,99	0,94	688
meg	hibás	-(s)Odik	1,00	1,00	1919
meg	honos	-(s)Odik	0,98	0,99	1031
be	népes	-(s)Ul	0,97	0,97	1181
be	lázás	-(s)Odik	0,99	1,00	558
ki	egész	-(s)Ul	1,00	1,00	7918
fel	bősz	-(s)Ul	0,92	0,97	652

A jellemző példányok közül néhány pontszerűen (vagyis nem skalárisan) feldolgozott jelentésszerkezetű eseményt jelöl. A *meghibásodik*, *megsebesül*, *megözvegyül* igék alapvetően két, egymástól lényegileg különböző állapot pontszerű egymásba váltását jelölik (például a *meghal* igéhez hasonlóan), vagyis a meghibásodás, megsebesülés és megözvegyülés feldolgozása

magával vonja a végpont (valami/valaki hibás, sebesült, özvegy stb.) megfigyeltetését. Előbbi igékkel a skalaritás bizonyos mértékű (fogalmi) hiányának szempontjából rokoníthatók a *kiegészül*, *elkülönül*, *megsemmisül*, *megvalósul*, *megtestesül*, *meghonosodik* igék (a középfokú melléklevelek létrehozásának kérdésessége a skalaritás fogalmának hiányára utal, vö. *?egészebb*, *?különebb*, *?semmisebb*, *?valósabb* 'létezőbb' értelemben, *?testesebb* 'testté váltabb' értelemben, *?honosabb*). A *megsemmisül* és *megvalósul* párhoz hasonlóan antonímiát látunk továbbá az eseményszerkezetüket tekintve hosszabb időintervallumot, nem pontszerű eseményt jelölő *benépesül* és *elnéptelenedik* igéknél. A szemantikai motívumok kapcsán érdemes hangsúlyoznunk, hogy a jellemző példányok között nem egyszerűen olyan igéket találunk, amelyeknek a fogalmi tartalma motiválja az esemény végpont felőli megfigyeltetését, hanem az egyesített konstrukciót képviselő igék a fogalmi tartalmat – esetleg más alternatíváktól eltérően – a végpont felől figyeltetik meg. Jó példa erre az *elsötétül* ige, amelynek ismeretes a hasonló fogalmi tartalmú, de imperfektív jelentésszerkezetet (is) jelölő, alapvetően skalárisabb természetű, nagyobb időintervallumot átfogni képes *(be)sötétedik* alternatívája is. A jellemző példányok lexikalizálódásának magasabb fokát mutatja továbbá az *elhidegül* ige, amely alternatíváival szemben *(hideg)edik*, *(hideg)szik* ma már kifejezetten kiterjesztett jelentésben használatos *(elhideg)ülnek egymástól*.

5. Produktivitás és hapaxok

Noha a hapaxok (lásd a már említett *megmetálosodik* igét) nem tekinthetők az egyesített konstrukció tipikus példányainak, nyilvánvaló indikátorai az egyesített konstrukció produktivitásának (Baayen 1989). Ezért érdemesnek tartottuk néhány kvantitatív produktivitás-vizsgálat elvégzését is a korpuszból kinyert adattáblán. A továbbiakban rövid áttekintést adunk a produktivitás számszerűsítésének a lehetőségeiről, majd rátérünk az eredményeinkre. Ezek közül elsőként az igekötős és igekötő nélküli, *-(s)Odik* illetve *-(s)Ul* képzős igékre vonatkozó vizsgálat, majd a konkrét igekötők szerint elvégzett vizsgálat eredményeit ismertetjük.

5.1. Áttekintő a kvantitatív produktivitásról

Egy produktív szóalkotási minta alapján elvileg tetszőleges számú, szemantikailag transzparens szó hozható létre egy adott szemantikai tartományban (Kiefer–Ladányi 2000: 149). A nyelvi rendszer szintjén a produktivitás egy lehetőség: a minden további nélkül létrehozható szavak és szóalakok nem biztos, hogy ténylegesen létrejönnek, és még kevésbé valószínű, hogy benne lesznek egy korpuszban. Emiatt több elméleti morfológus (például Dressler 2003) nem tartja helyesnek a kompetenciaszintű lehetőség performanciaszintű valószínűsége alapjánuló vizsgálatát.⁹ Az is külön problémát jelent, hogy a korpuszokból szemantikai információt csak közvetetten tudunk kinyerni, a produktivitás pedig szorosan összefügg a szemantikával. Így a produktivitás mérésére kidolgozott, kvantitatív módszerek szemantikai információt expliciten nem vesznek figyelembe; ezt leginkább különféle gyakoriságokon keresztül próbálják megragadni. Mi azonban úgy gondoljuk, hogy az utóbbi irány, a korpuszadaton alapuló, kvantitatív vizsgálat olyan tendenciákat mutathat ki, amelyek a kompetenciát kutatóknak is relevánsak lehetnek, és amelyekre introspekción alapú módszerrel nem lehetne szert tenni. A használatalapú megközelítések szerint ugyanis a kompetenciaszintű lehetőséget alapvetően a performancia, a nyelvi tapasztalatok határozzák meg, vagyis nincs köztük dichotomikus különbség.

A morfológiai produktivitás kvantitatív meghatározása Baayen (1989) nevéhez köthető. Baayen több típust és ennek megfelelően több mérési módszert definiál, amelyek a produktitásnak különféle aspektusait igyekeznek megragadni.¹⁰ Ezeket az elmúlt több mint harminc évben számos korpusznyelvészeti kutatásban alkalmazták, így az évek során sokat változtak, finomodtak az eredeti módszerek. Napjainkban a legszélesebb körben alkalmazott módszer a **globális produktivitás** (Global Productivity, G) meghatározása (Baayen–Lieber 1991). Ezt négy érték ismeretében állapíthatjuk meg. Az alábbiakban röviden ismertetjük ezeket, zárójelben megadva a leggyakrabban használt angol elnevezésüket, valamint a nemzetközileg sztenderd jelöléseket, amelyekkel a cikkünkben hivatkozni fogunk rájuk:

⁹ A Dressler nevéhez köthető természetes morfológia kvalitatív elemzése a produktivitást illetően azt vizsgálja, hogy egy szóképzési séma milyen akadályok leküzdésére alkalmas (például jövevényszavakon, betűszavakon való alkalmazás), és szintén alkalmas a produktivitás fokozatosságának megragadására. E kvalitatív produktitásfogalomról részletesebben lásd Ladányi (2007: 29–40, 47–50), valamint Ladányi Mária cikkünkről írt bírálatát.

¹⁰ Ezekről bővebben lásd Kalivoda (2019) és Modrián–Horváth–Kalicoda–Tanos (ebben a kötetben). Az utóbbi tanulmány részletesen tárgyal olyan általános módszertani megfontolásokat is, amelyekre a kvantitatív fókuszú kutatásokban tekintettel kell lennünk.

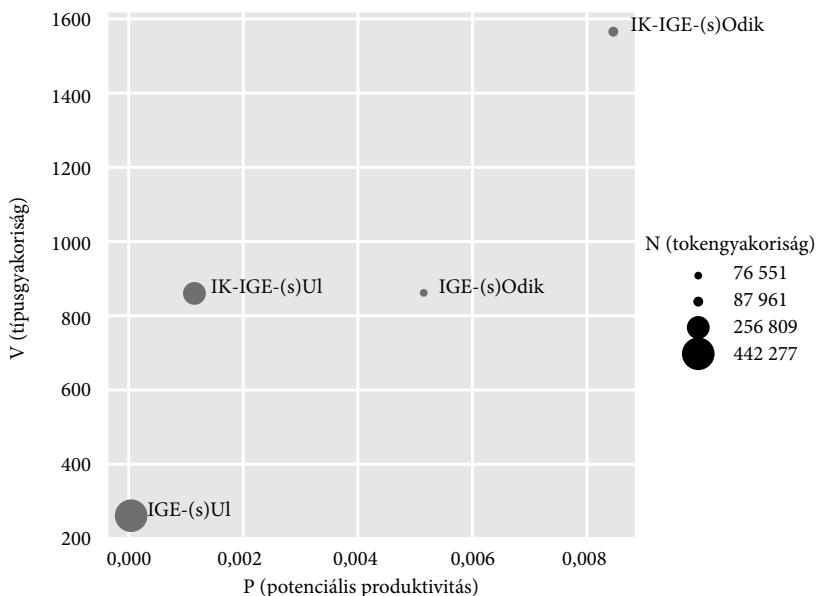
1. **Tokengyakoriság** (token frequency, N): azt mutatja meg, hogy az adott affixum hány példányban adatolható a korpuszban.
2. **Típusgyakoriság** (type frequency, V): a szótípusok – esetünkben igeképzések – száma, amelyekben az affixum előfordul.
3. **Hapaxok száma** (hapax legomena, n1): a korpuszban egyszer előforduló szótípusok száma, amelyek az affixumot tartalmazzák. A hapaxok jellemzően produktív szóalkotás eredményei, így a morfológiai szerkezetük és a jelentésük is kompozicionális – legalábbis nagyobb eséllyel az, mint a gyakori szavak esetében.
4. **Potenciális produktivitás** (Narrow Productivity, P): egy származtatott érték, amelyet úgy kapunk, hogy a tokengyakoriságot elosztjuk a hapaxok számával. Arról ad jóságot, hogy az affixumnak várhatóan mekkora szerepe lesz a szóalkotásban a közeljövőben.

A globális produktivitás sajátossága az, hogy nem egyetlen számértékkel fejezhető ki, hanem vizuálisan megjelenítve nyer értelmet. Úgy kapjuk meg, hogy a potenciális produktivitást és a típusgyakoriságot egy-egy dimenzióknak tekintjük – az előbbi jellemzően a koordináta-rendszer X-tengelyén, az utóbbit az Y-tengelyén vesszük fel –, és ebben a két dimenziós térben helyezzük el a vizsgált affixumokat a két számértéküknek megfelelően. Minél magasabb értékei vannak egy affixumnak az egyes tengelyeken, annál produktívabbnak tekinthető. Még szemléletesebb képet kaphatunk a produktív értékekről akkor, ha az affixumok tokengyakoriságát az adatpontok nagyságával szemléltetjük.

5.2. Igekötő nélküli, valamint igekötős -(s)Odik és -(s)Ul szerkezetek

Elsőként azt vizsgáltuk meg, hogy hogyan alakulnak a globális produktív értékek a -(s)Odik és -(s)Ul képzős igék esetében akkor, ha igekötőkkel, illetve ha igekötők nélkül állnak. A mérésünkhöz csak a 2000-ben vagy annál később publikált szövegeket vettük figyelembe.¹¹ Az eredményt a 2. ábrán mutatjuk be, az egyes adatpontok részletes értékei pedig a III/A. mellékletben láthatók.

¹¹ Mindkét bírálónk jelzi, hogy ezt a döntést szükséges lenne indokolni. A döntést megelőző gondolatmenetünk az volt, hogy kvantitatív produktív meghatározásánál célszerű minél rövidebb időintervallumot választani, mivel ez egy szinkron korpuszvizsgálatra kidolgozott módszer. Az általunk használt korpusz 1937 és 2014 közötti szövegeket tartalmaz, 77 év összevonása azonban sokat torzíthatott volna az eredményen. Azzal, hogy – önkényesen, de nem ok nélkül – a 2000-es évet jelöltük ki határvonalnak, a vizsgálatot a 21. század elejére szűkítettük le.



2. ábra. Globális produktivás értékek az igekötő nélküli illetve igekötős, -(s)Odik, illetve -(s)Ul képzős igék esetében

A 2. ábráról leolvasható, és a korábbi szakirodalom alapján (lásd Kiefer–Ladányi 2000: 154–155; Ladányi 2007: 245–247; Ladányi 2017: 608) számítani is lehetett rá, hogy jelentős a különbség a -(s)Ul és -(s)Odik képzők produktivitásában: a -(s)Ul képző önmagában – igekötő nélkül – szinte egyáltalán nem produktív, a -(s)Odik viszont igen. Ugyanakkor megfigyelhetjük azt is, hogy a -(s)Ul tokengyakorisága sokszorosa a -(s)Odik képzőének. Ez főként olyan igéknek köszönhető, amelyek jelentése nem vagy alig transzparens, a képzőt a beszélők többsége feltételezhetően nem is érzékeli ezekben a szavakban. Az öt leggyakoribb -(s)Ul képzős ige a *készül*, *épül*, *bizonyul*, *részesül* és *bővül*. Az is látható, hogy a tokengyakoriság közel fordítottan arányos a produktivitással – ez alátámasztja Bybee (1985) tézisé, miszerint a tokengyakoriság az adott lexikai egységet erősíti, és nem (vagy legalábbis kevésbé) járul hozzá a mintázat produktivitásához. Az igekötőzés mindkét képző esetében a globális produktivás jelentős növekedését eredményezi. A típusgyakoriságok azonos mértékben növekednek, a potenciális produktivás pedig a -(s)Odik esetében jobban, a -(s)Ul esetében valamivel kevésbé. Noha a -(s)Ul képző

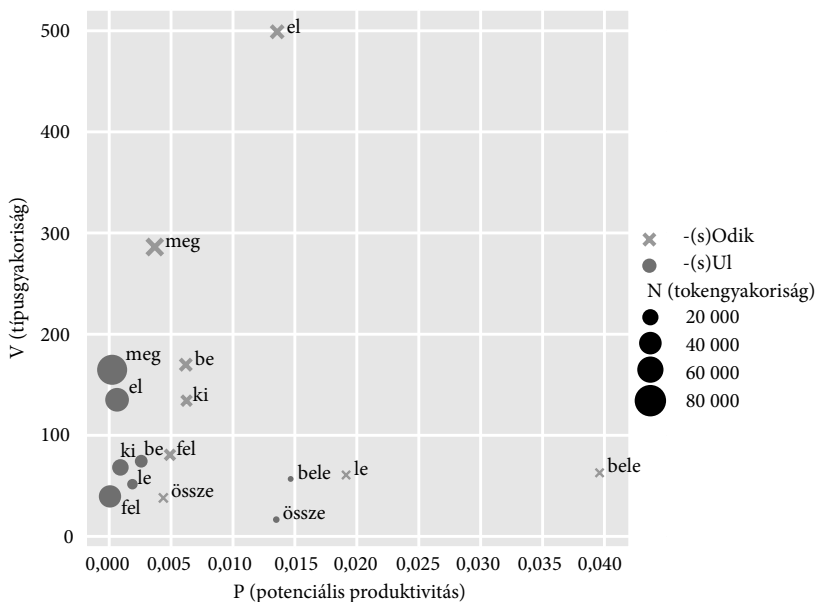
önmagában kevésbé produktív, a táblázatból az is látható, hogy viszonylag termékeny az igekötővel egyesített konstrukció komponenseként. Az (5a) igekötő nélküli, az (5b) igekötős igékre hoz néhány példát.

- (5) a. zöldeskékül, szállóigésül, iparosul, tartósul, férgesül
b. felnemesül, hozzáaljasul, elmémesül, belerészegül, behomályosul, bedemagógul, elunalmasul, felelevenül, beletéteszesül, begyávul, hozzáferdül, rálelkesül, szétpuhul

Már a bevezetőben említett, előzetesen megfigyelt példányok esetében is feltűnő volt az, hogy a 2.1 részben tárgyalt szótagszám az *-Ul* igeképzőt illetően kimenetelapú általánosításként nem működik. Az *érezékenyül* igének például a szótagszáma is arra utal, hogy nem vezethető vissza egy (akár csak potenciális szóként felfogott) *érezékenyül* igére. Az *érezékenyül* ige ugyanis nem lehet potenciális szó sem, mivel nem felel meg az *-Ul* képzős séma szótagszáma vonatkozó általánosításának. A bázisszó tehát az egyesített konstrukció esetében két szótagosnál hosszabb is lehet (lásd még Kalivoda 2021: 60–61), a hapaxok közül ebbe a sorba illeszkedik (5b) adatai közül a *bedemagógul*, *elunalmasul* és a *felelevenül* (vö. *felelevenedik*) ige.

5.3. Igekötők közötti különbségek

Érdemesnek tartottuk úgy is elvégezni a produktivitás-mérést, hogy az igekötőket külön-külön is vizsgáltuk. Arra számítottunk, hogy ezáltal látszódnia fog, melyik igekötő milyen mértékben járul hozzá egy-egy szerkezet produktivásához. Olyan igekötős, *-(s)Odik* illetve *-(s)Ul* képzős szerkezeteket vizsgáltunk, amelyekben az adott képzővel az adott igekötő legalább ötszázszor előfordult – tokengyakoriságot nézve –, így a vizsgált igekötők a *meg*, *el*, *fel*, *le*, *ki*, *be*, *bele* és *össze* lettek. Az előző méréshez hasonlóan itt is a 2000-ben vagy később publikált szövegeket vettük figyelembe. A 3. ábra mutatja be a mérés eredményét, az egyes adatpontok részletes értékei a III/B. mellékletben láthatók.



3. ábra. Globális produktivás értékek különböző igekötőkkel ellátott, -s)Odik, illetve -s)Ul képzős igék esetében

A 3. ábra alapján a vizsgált minimálpárok globális produktivitása úgy alakul, hogy a -s)Odik képzővel együtt álló igekötő egyértelműen produktívabb, mint a -s)Ul képzővel álló. Ez alól csak az össze igekötő kivétel: az össze...-s)Ul potenciális produktivitása magasabb, mint az össze...-s)Odik szerkezeté, de típusgyakoriságát illetően kicsivel elmarad tőle. Ennek azonban pusztán az adatrítkaság az oka: az előbbi konstrukció kilenc példánnyal fordul elő a korpuszban, az utóbbi hússzal, és a két alacsony gyakoriság közötti kontraszt nagyobbak tűnik, mint a magas gyakoriságok közötti. Emellett megfigyelhetjük azt is, hogy az el...-s)Odik és a bele...-s)Odik konstrukciók globális produktivitása kimagasló, viszont más-más tengelyen. Az előbbinek a típusgyakorisága magas, az utóbbinak pedig a potenciális produktivitása. Mivel a típusgyakoriság egy olyan érték, amelyet közvetlenül a korpuszból ki tudunk nyerni – szemben a potenciális produktivitással, amely származtatott érték –, megállapíthatjuk, hogy az el...-s)Odik konstrukció a legproduktívabb a méréseink alapján. Ez a megállapítás összhangban van azzal, hogy a jellemző példányok

(lásd 4. rész, 1. táblázat) között az *el* igekötős igék abszolút többségben, az *el...-(s)Odik* mintázatot képviselő igék pedig relatív többségben vannak.

6. Összegzés

A tanulmányban egy egyedi példányokon alapuló megfigyelést, a szimultán igekötőzést és igeképzést vizsgáltuk kvantitatív alapon, adatvezérelten. A szimultán igekötőzést és igeképzést konstrukcióegyesítésként közelítettük meg, és azt feltételeztük, hogy ezt a folyamatot a különböző szóképzési konstrukciók hálózatos szerveződése teszi lehetővé. További elméleti háttérfeltevésünk, hogy a morfológiai konstrukciók produktivitásában és kidolgozásában a különböző általánosításoknak és a gyakori példányoknak, vagyis a sémáknak és az analógiának egyaránt szerepe van. A korpuszalapú kutatás az igekötőket és passzív állapotváltozást kifejező igeképzőket egyesítő konstrukciók jellemző példányainak és a konstrukciók számszerűsített, globális produktivitásának feltérképezésére irányult. A kutatás alapját az MNSZ2-UD korpuszban található összes *-(s)Odik* vagy *-(s)Ul* képzős finit ige és infinitívusz képezte.

Az egyesített konstrukció jellemző példányainak tekintettük azokat az igéket, amelyeknek az esetében az igekötős ige jóval gyakoribb, mint az igekötő nélküli, illetve az ige kifejezetten egy adott igekötővel fordul elő nagy arányban. A jellemző példányokat három számérték segítségével határoztuk meg. Az egyesített konstrukciót képviselő jellemző példányok esetében megfigyelhető az ige által jelölt állapotváltozás pontszerűsége, vagy legalábbis az állapotváltozás skalaritásának relatív hiánya, ami fogalmi magyarázatul szolgál arra, hogy miért tipikus az adott folyamat igekötővel való, vagyis végpont felőli megfigyeltetése. A jellemző példányok között továbbá több antonimát is találtunk, a pozitív és negatív irányú változásokat kódoló igekötős igék eszerint sok esetben implikálhatják egymást. A tanulmányban megvizsgáltuk az egyesített konstrukciók globális produktivitását is. A vizsgálat egyik fő eredménye, hogy azt a szakirodalomból ismert megállapítást, hogy a *-(s)Ul* képző ma már nem produktív, az adatok ismeretében korrigálnunk kell: az igekötő nélküli igével álló *-(s)Ul* képző nagyon kevésbé, az igekötős igével álló viszont egyértelműen produktív. Összességében mind a *-(s)Ul*, mind a *-(s)Odik* képző produktívabb az igekötőkkel egyesített konstrukciók komponenseként, mint önmagukban.

Források

- MNSZ2=Magyar Nemzeti Szövegtár v2.0.5. <http://clara.nytud.hu/mnsz2-dev/index.html>. (Lásd Oravecz–Várad–Sass 2014.)
- MNSZ2-UD = Magyar Nemzeti Szövegtár v2.0.5 UD treebank. (Lásd Kalivoda et al. 2024.)
- Verbal derivational suffixes in Hungarian: -(s)Odik and -(s)Ul. <https://zenodo.org/records/7607661>. (Lásd a jelen cikket; Kalivoda–Palágyi 2024.)

Irodalom

- Baayen, Harald 1989. *A Corpus-Based Approach to Morphological Productivity (Statistical Analysis and Psycholinguistic Interpretation)*. Doktori értekezés. Amsterdam: Centrum voor Wiskunde en Informatica.
- Baayen, Harald – Lieber, Rochelle 1991. Productivity and English derivation: a corpus-based study. *Linguistics* 29 (5): 801–844. <https://doi.org/10.1515/ling.1991.29.5.801>
- Booij, Geert 2010a. *Construction Morphology*. Oxford: Oxford University Press.
- Booij, Geert 2010b. Compound construction: schemas or analogy? A construction morphology perspective. In: Scalise, Sergio – Vogel, Irene (eds.): *Cross-disciplinary issues in compounding*. Amsterdam: John Benjamins. 93–108. <https://doi.org/10.1075/cilt.311.09boo>
- Blevins, James P. 2006. Word-based morphology. *Journal of Linguistics* 42: 531–573. <https://doi.org/10.1017/S0022226706004191>
- Bybee, Joan 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins. <https://doi.org/10.1075/tsl.9>
- Bybee, Joan 2001. *Phonology and language use*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511612886>
- Bybee, Joan 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511750526>
- de Marneffe, Marie-Catherine – Manning, Christopher D. – Nivre, Joakim – Zeman, Daniel 2021. Universal Dependencies. *Computational Linguistics* 47 (2): 255–308. https://doi.org/10.1162/coli_a_00402
- Dressler, Wolfgang U. 2003. Degrees of grammatical productivity in inflectional morphology. *Rivista di Linguistica (Italian Journal of Linguistics)* 15 (1): 31–62.

- É. Kiss Katalin 2004. Egy igeekötőelmélet vázlata. *Magyar Nyelv* 100: 15–43.
- Haspelmath, Martin – Sims, Andrea D. 2010. *Understanding morphology*. London: Hodder Education.
- Kalivoda Ágnes 2019. Véges erőforrás végtelen sok igeekötős ígére. In: Berend Gábor – Gosztolya Gábor – Vincze Veronika (szerk.): *XV. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2019)*. Szeged: Szegedi Tudományegyetem, TTIK, Informatikai Intézet. 331–344.
- Kalivoda Ágnes 2021. Az igeekötők produktív kapcsolódási mintái. *Argumentum* 17: 56–82. <https://doi.org/10.34103/ARGUMENTUM/2021/4>
- Kalivoda, Ágnes – Malouf, Robert – Ackerman, Farrell – Sass, Bálint 2024. *Building a dependency treebank from the Hungarian Gigaword Corpus*. Kézirat, megjelenés előtt.
- Keszler Borbála 2000. A szóalkotás módjai. In: Keszler Borbála (szerk.): *Magyar grammatika*. Budapest: Nemzeti Tankönyvkiadó.
- Kiefer Ferenc – Ladányi Mária 2000. A szóképzés. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 3. Morfológia*. Budapest: Akadémiai Kiadó. 137–164.
- Ladányi Mária 2007. *Produktivitás és analógia a szóképzésben: elvek és esetek*. Budapest: Tinta Könyvkiadó.
- Ladányi Mária 2017. Alaktan. In: Tolcsvai Nagy Gábor (szerk.): *Nyelvtan*. Budapest: Osiris Kiadó. 503–660.
- Langacker, Ronald W. 1987. *Foundation of Cognitive Grammar (Vol. 1). Theoretical prerequisites*. Stanford: Stanford University Press.
- Modrián-Horváth Bernadett – Kalivoda Ágnes – Tanos Bálint 2024. A magmondatszerepű, passzív jelentésű sémák produktivitásának kvantitatív vizsgálata. In: Kugler Nóra – Modrián-Horváth Bernadett – Simon Gábor (szerk.): *Korpusznyelvészeti módszerek a kognitív nyelvészeti kutatásban*. Budapest: ELTE BTK – ELTE Eötvös Kiadó. 153–203.
- Novák, Attila – Siklósi, Borbála – Oravecz, Csaba 2016. A New Integrated Open-source Morphological Analyzer for Hungarian. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Portorož, Slovenia: European Language Resources Association (ELRA). 1315–1322.
- Oravecz, Csaba – Várad, Tamás – Sass, Bálint 2014. The Hungarian Gigaword Corpus. In: *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*. Reykjavík: European Language Resources Association (ELRA). 1719–1723.

- Palágyi László 2022. A metafora, az analógia és a szemantikai megfelelések szerepe a *le* igekötős konstrukciók jelentésében. In: Tátrai Szilárd – Tolcsvai Nagy Gábor (szerk.): *A magyar mondat és kontextuális környezete*. Budapest: ELTE BTK – ELTE Eötvös Kiadó. 15–48.
- Siklósi Borbála – Novák Attila 2016. Beágyazási modellek alkalmazása lexikai kategorizációs feladatokra. In: *XII. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi Tudományegyetem, TTIK, Informatikai Intézet. 3–14.
- Szilágyi N. Sándor 1996. *Hogyan teremtsünk világot?* Kolozsvár: Erdélyi Tankönyvtanács.
- Szilágyi N. Sándor 2013. A mi (rend)szertelen nyelvünk. In: Kádár Edit – Szilágyi N. Sándor (szerk.): *Analógia és modern nyelvelírás*. Kolozsvár: Erdélyi Múzeum-Egyesület. 235–263
- Szili Katalin 2005a. A *be* igekötő jelentésváltozásai I. *Magyar Nyelvőr* 129: 151–164.
- Szili Katalin 2005b. A *be* igekötő jelentésváltozásai II. *Magyar Nyelvőr* 129: 282–299.
- Szili Katalin 2009. A *fel*, *le* és egyéb igekötős igék formai-szemantikai viszonyának kérdéséhez. *Magyar Nyelv* 105: 175–188.
- Tolcsvai Nagy Gábor 2013. Az igekötő + ige szerkezet szemantikája. *Nyelvtudományi Közlemények* 109: 187–226.



Simultaneous use of preverbs and verb formation suffixes – construction unification in Hungarian

This paper presents a data-driven approach to explore the simultaneous application of Hungarian preverbs and verb formation suffixes. We consider this phenomenon construction unification and assume that the process is made possible by the network organization of different word formation constructions. Our further assumption is that schemas and analogy play a role in the productivity and development of morphological constructions. The text source used in our research is the HGC-UD corpus, from which we extract all relevant derivatives, their contexts, metadata, and various formal features. We convert the result of this process (more than 1.1 million instances) into a freely accessible dataset. Our research aims to extract the

typical exemplars of the constructions unifying preverbs and verb derivational suffixes expressing a passive change of state, as well as to determine and analyze the global productivity of these constructions.

Keywords: construction unification, preverbs, verb formation suffixes, Hungarian grammar, data-driven approach

Kalivoda Ágnes

HUN-REN Nyelvtudományi Kutatóközpont

kalivoda.agnes@nytud.hun-ren.hu

ORCID: 0000-0003-2520-5523

Palágyi László

Eötvös Loránd Tudományegyetem

palagyilaszlo@btk.elte.hu

ORCID: 0000-0002-1437-4746

I. számú melléklet: Rövidítésjegyzék

Fonológiai jelölések:

C	mássalhangzó
C:	hosszú mássalhangzó
V	o, ö, e vagy a
O	o, ö vagy e
U	u vagy ü

Szófaji jelölések:

N	főnév
ADJ	melléknév
V	ige

A cikkben használt emMorph címkék:

[/Adj col]	színt jelölő melléknév
[_AdjVbz_Ntr/V]	olyan igeképző, amely melléknévből intranszitiv igét képez
[Prs.NDef.3Sg]	tárgyatlan ragozás, egyes szám harmadik személy

A produktivitásmérésben használt nemzetközi jelölések:

G	globális produktivitás
N	tokengyakoriság
V	típusgyakoriság
n1	hapaxok száma
P	potenciális produktivitás

Egyéb jelölések:

?	a szerzők intuíciója alapján kérdéses grammatikalitású szó
*	a szerzők intuíciója alapján agrammatikus szó
(...)	nem mindig megjelenő elem
[...]	séma
\$	szótaghatár

Amint látható, a V és az N jelölés túlterhelt, mert mindig igyekeztünk az adott szakterületen (fonológiában, morfológiában illetve korpusznyelvészetben) sztemerddé vált jelölést alkalmazni. Abban bízunk, hogy a kontextus egyértelművé teszi ennek a két rövidítésnek a feloldását a konkrét esetekben.

II. számú melléklet: Az adattábla oszlopai

Oszlopnév	Leírás
id	A korpusztalálát egyedi azonosítója.
form	A szóalak kisbetűsítve. Ez azt jelenti, hogy ha az adott szó a korpuszban mondatkezdő helyzetben, nagybetűvel állt, itt akkor is kisbetűvel szerepel. Az eredeti írásmód a <i>kwic</i> cellában lesz látható.
lemma	A lemma (inflexiós morféimák nélküli szóalak) kisbetűsítve. Igekötős igék esetében az igekötő és az ige határán + jel áll. A lemmában akkor is feltüntettük az igekötőt, ha az elvált volt.
prev	Az igehez tartozó igekötő kisbetűsítve.
prevtype	A cella értéke PFX, ha az igekötő tapad az igehez, és SEP, ha az igekötő elvált.
verb	Az igelemma kisbetűsítve, minden esetben igekötő nélkül.
root	Az igeképzés bázisul szolgáló névszó kisbetűsítve.
suffix	A képző: <i>-ul/ül/sul/sül</i> esetében <i>-(s)Ul</i> , <i>-odik/edik/ödik/sodik/sedik/södik</i> esetében <i>-(s)Odik</i> .
w2v_cluster	Annak a klaszternek az azonosító száma, amelybe a bázisszó – a <i>root</i> mező tartalma – került a szóbeágyazás során.
argframe_cases	Az igeinek az adott mondatban megjelenő vonzatai, amelyeket esetragokkal jelölünk.
argframe_long	Az igeinek az adott mondatban megjelenő vonzatai, amelyeket lemma+esetrag kombinációkkal jelölünk.
doc_year	A dokumentum keletkezésének vagy publikálásának az éve. Amennyiben ismeretlen az évszám, a cella értéke 0.
doc_style	A dokumentum stílusrétege.
doc_id	A dokumentum egyedi azonosítója.
left_context	A korpusztalálátot megelőző mondatszakasz.
kwic	A korpusztalálát.
right_context	A korpusztalálátot követő mondatszakasz.
freqsum	Az igelemma tokengyakorisága, az igekötős és igekötő nélküli eseteket egybeszámolva.

Oszlopnév	Leírás
prev_vs_all	Az igelemma igekötős előfordulásainak száma osztva az összes előfordulásának számával.
actprev_vs_allprev	Az adott igekötő+igelemma előfordulásainak száma osztva az igelemma összes igekötős előfordulásának számával.

A hiányzó cellaértékek jelölésére alulvonást (_) használunk.

III. számú melléklet: A produktivitásmérésekhez kapcsolódó adatok

A. Igekötő nélküli ill. igekötős -(s)Odik/-(s)Ul:

szerkezet	N	V	n1	P
IGE-(s)Odik	76.551	867	398	0,0052
IGE-(s)Ul	442.277	265	60	0,0001
IK-IGE-(s)Odik	87.961	1.570	751	0,0085
IK-IGE-(s)Ul	256.809	864	307	0,0012

B. Igekötők közötti különbségek:

igekötő	képző	N	V	n1	P
be	-(s)Odik	12.251	169	77	0,0063
be	-(s)Ul	9.210	76	25	0,0027
bele	-(s)Odik	780	63	31	0,0397
bele	-(s)Ul	1.558	57	23	0,0148
el	-(s)Odik	18.160	496	248	0,0137
el	-(s)Ul	66.099	136	44	0,0007
fel	-(s)Odik	6.603	81	34	0,0051
fel	-(s)Ul	48.460	40	11	0,0002
ki	-(s)Odik	8.669	133	55	0,0063
ki	-(s)Ul	21.212	67	22	0,0001
le	-(s)Odik	1.670	61	32	0,0192
le	-(s)Ul	7.505	52	14	0,0019
meg	-(s)Odik	30.587	285	117	0,0038
meg	-(s)Ul	97.539	164	37	0,0004
össze	-(s)Odik	4.496	38	20	0,0044
össze	-(s)Ul	663	17	9	0,0136

INDIG BALÁZS

Hozzászólás a *Szimultán igekötőzés és igeképzés – konstrukcióegyesítés a magyarban* című tanulmányhoz

1. Általános benyomásaim

A mű a magyar nyelv igekötős igéinek, valamint azok felépítésének vizsgálatára vállalkozik. A szerzők az állításaikat a hagyományos leíró nyelvészeti irodalom áttekintése és egy számítógép segítségével korpuszon végzett vizsgálat eredményeként teszik meg. Habár a teljes igeképzés leíró és számítógépes nyelvészeti szempontból történő alapos vizsgálatára nincs egy cikkben lehetőség, a szerzők által választott *-(s)Ul* és *-(s)Odik* képzők vizsgálata részletes betekintést enged a módszer részleteibe, mely alapján a szóképzés több példája is hasonlóan kidolgozhatóvá válik.

A cikk az interdiszciplináris jellegéből adódóan jobban hajlik a leíró nyelvészeti irány kibontására, ami több helyen a számítógépes szempontok kidolgozásának rovására történik meg. Véleményem szerint érdemes lenne a nyelvelméleti részt – mely dicséretesen sok példát tartalmaz – strukturált ábrák formájában, esetleg táblázatban megjeleníteni (például 2.2. fejezet és hasonló részek), hogy a számítógépes nyelvészet felől érkező olvasó is könnyebben magáénak érezze az érvelést. A hivatkozott művek részletesebb bemutatása lehetővé tenné, hogy kinek-kinek a gyengébb oldalát képező tudományterület ismeretének hiányosságai ellenére jól érthető és követhető maradjon a leírás.

A mérés kifejezetten jól van leírva és újszerű eredményeket mutat. Az ábrák segítik a megértést, viszont néhol azonban egységesítésre, bővítésre szorulnak. Habár felfedezni vélek egynémely módszertani limitációt, melyeket az alábbiakban kifejték, összességében úgy látom, hogy nagyon alapos cikk született, mely az alább javasolt változtatások figyelembevételével javasolt a közlésre.

2. Nyelvészeti meglátások

Tudományos háttéremből fakadóan, több esetben az az érzésem, hogy túl mély előismereteket igényel a cikk megértése, melyek rövid magyarázatokkal enyhíthetők lennének. Például a „sokszor nem könnyű eldönteni, hogy két képző egymásutánjáról vagy egy összetett képzőről van-e szó” (Keszler 2000: 315) idézet esetén kifejtve, hogy pontosan milyen módon próbálták meg eldönteni a kategorizációt. Esetleg röviden felvonultatva, hogy a számítógépes nyelvészeti eszköztár milyen kiaknázatlan lehetőségeket tartogat (például kollokációs metrikák) a kérdés további vizsgálatára, mintegy kontrasztot képezve. Specifikusan azt is érdemes megmagyarázni, hogy miért szükséges szigorúan szétválasztani vagy egyben hagyni „két képzőt”, mely többnyire együtt fordul elő, de potenciálisan van bizonyíték, hogy valamilyen formában külön is működnek az adott környezetben belül (például régebbi nyelvállapot). Ha eltekintünk ettől a szigorú feltételtől, mennyivel lesz rosszabb a magyarázat a jelenségre?

A „Jellemző, hogy az igen specifikus konstrukciós jelentések sokszor viszonylag szűk számú, ugyanakkor viszonylag gyakori igekötős igére épülnek, mégis produktívak.” mondatnál az olvasónak önkénytelenül is az jut eszébe, hogy az innovatív nyelvhasználat kifejezetten táptalaja az ilyen jellegű produktivitásnak (például *be- kap, nyom, tol, csap egy szendvicset a szünetben*). Kérdés, hogy a különféle stílusregiszterekben ugyanúgy viselkedik-e az említett konstrukció? A korpusz által felölelt évek diakroniájában tekintve tapasztalható-e változás?

3. A felhasznált számítógépes eszköztár

Dicséretes, hogy a nyelvész körökben méltán használt Magyar Nemzeti Szövegtár tovább fejlődik és még meg se jelenik az új verzió, de már mérések készülnek rajta. Véleményem szerint kicsit bővebb kifejtést érdemelne, hogy milyen speciális tulajdonságai teszik kifejezetten alkalmassá erre a specifikus mérésre a nevezett korpuszt, mik ezek közül az elengedhetetlenül szükségesek és melyek csak extrák. A cikket olvasva felötlik az emberben, hogy vajon mennyire jó az, ha minden mérést egy korpuszon végzünk el? Nem lenne-e célszerű a számítógépes nyelvészeti kánonban megszokott módon egy korpuszt/korpuszrészletet a hipotézis felállítására míg egy másik tőle független adathalmazt a hipotézis validálására használni feltéve, ha nem jár

aránytalanul sok többletmunkával? Így elkerülhető lenne, hogy abba a szerencsétlenségbe essünk, hogy a nyelv helyett csak az adott korpusz részletes tulajdonságait ismerjük meg. Különös tekintettel arra, amikor kvantitatív módszert használva a szavak gyakoriságára vagy épp ritkaságára hivatkozunk. Az ezen hibák elkerülésére, illetve fenn nem állására vonatkozó érvekkel javasolom kiegészíteni a cikket.

Ami a felhasznált eszközöket illeti, nem világos számomra, hogy a középmagyar emMorph mennyivel adott hozzá többet a vizsgálathoz, mint ha például reguláris kifejezések segítségével végezték volna a szerzők a szavak osztályozását. Hiszen, mint írják, a végén úgyis kézzel átnézték az eredményt. Dicséretesnek tartom, hogy kiemelik, hogy az emMorph különböző változatainak erősségeit és gyengeségeit, de úgy érzem, hogy a szerzők indokolatlanul ragaszkodnak ehhez az eszközhöz, amikor hasonló hatékonysággal egyszerűbben is meg lehet oldani ugyanazt a feladatot. Továbbá kódok a cikkhez mellékelése által a szerzők potenciálisan segíthetnek másokat az állítások ellenőrzésében és továbbfejlesztésében, szemben a jelenlegi felállással, amely korlátozza a későbbi kutatókat. Ugyanennél a résznél meglátásom szerint érdemes lenne ismertetni a konkrét számokat. A „jóval több” és „Minden egyes előfordulás kézi átnézésére nem vállalkoztunk, mivel milliós nagyságrendű találati listával volt dolgunk; az ilyen kétértelmű eseteket egységesen megőriztük. Ezáltal hibás adatok is bekerülnek a vizsgálatainkba, de ez ekkora mennyiségű szöveg automatikus feldolgozása során elkerülhetetlen, és számottevően nem befolyásolja az eredményeinket.” jellegű kvalitatív jelzők elveszik az élet az amúgy kiváló számítógépes vizsgálatnak. Pontosan milyen arányban került az anyag átnézésre kézzel? Mennyire reprezentatív? Milyen hibák maradhattak és pontosan miképpen befolyásolják a későbbieket és ezek a tényezők miért nem fontosak? Ekkora anyag esetében fontosnak tartanám jobb példák kiválasztását. A terminális állapotváltozásként értelmezett *vadul* ('vaddá válik') szó véleményem szerint nem jelenhet meg ebben az értelemben igekötő nélkül („elvadul”), mely megkülönbözteti a hallgató számára a „vad módon” jelentéstől, ami a folyamatossága miatt csak igekötő nélkül létezhet („vadul”). Amennyiben arra akar példa lenni, hogy a két alaknak a jelentés szintjén milyen a viszonya egymással, bővebb magyarázat szükséges. A metrika leírásánál is érdemes olyan példát választani, amely nem a leírás fordítottját bizonyítja, azaz, „adattáblában a *metálosodik* ige nem szerepel önmagában, csak a megikézővel együtt (*megmetálosodott*), de mivel ez az egyetlen előfordulása a korpuszban, nem indokolt az egyesített konstrukció jellemző példányaként

figyelembe vennünk.” vö. „a *megmetálosodott* ige egyszer fordul elő, ezért csak azt állíthatjuk, hogy nincs arra bizonyítékunk, hogy igekötő nélküli változata is létezik, de ez nem kizárható, így semmilyen szempontból nem tudjuk figyelembe venni statisztikailag.” Formálisan a metrika számítógépes és matematikai értelemben is egy olyan függvényt jelöl, aminek tetszőleges a bemenete és egy szám a kimenete. A cikkben definiált „metrika” viszont három számból áll, ezzel pedig három külön metrikáról és azok összefüggéséről van szó. Ha egyként akarjuk értelmezni őket, akkor valahogyan súlyozandóak, hogy egy számot alkossanak a végén. Ez nem elvárt, viszont a jelenleg használt fogalmak megkövetelnék. Érdemes lenne ezen kívül egy számszerű „elvárt”, indikatív küszöbértéket megadni már itt objektivitás miatt (vö. „mindhárom felsorolt érték magas”) és mint minden ilyen paraméterbeállítást validálni egy másik korpuszon.

4. A szóbeágyazások körkörössége

Az ismertetett szóbeágyazás alapú módszer szerepe nem derül ki a cikkből, csak annyi, hogy mint kiegészítő információ hozzá lett adva a keletkezett adattáblához. Miért is jó ez? Véleményem szerint ilyen formán és körkörös érvelésre ad lehetőséget a felhasználása. A cikkben nincs kifejtve, hogy a szóbeágyazás működése folytán az adott szó előfordulásainak egyes kontextusaiból veszi azt az információt, mely segítségével algoritmikusan, a skatulya elv alapján csoportokba szorítjuk a szavakat. A csoportosítás mikéntjét legjobban a bemeneti szavak módosításával lehet elérni, ezzel előállítva például szintaktikus és szemantikus csoportosításokat (lásd a hivatkozott cikk). Ilyen célszerű módosítás lehet például a szavak szétvágása az igekötők és a képzők mentén, mely azirányban befolyásolja a rendszert, hogy a függetlennek vélt intuíciónak megfelelőbben viselkedjen körkörösé téve az érvelést, miszerint „a vektortér alapú csoportosítás visszaigazolja az intuíciónkat”. Úgy vélem, hogy a tanulmányban bemutatott eredményekhez nem járul hozzá a választott szóbeágyazási módszer alkalmazása.

Nagyobb távlatban viszont érdemes foglalkozni a szóbeágyazások felhasználásával. Ha úgy fogjuk fel a szóbeágyazást, hogy nálunk sokkal nagyobb szöveget sokkal jobban és több szempont szerint végignézzve, az intuíciónkat mintegy felerősítve jobb, több és részletesebb példákat mond, akkor a kimenete kritikai éllel vizsgálhatóvá válik: Hol érez rá jól? Hol általánosít túl? Mit nem vesz észre? Ha csak az első szempontot, a pozitív visszacsatolást

tekintjük, akkor a saját intuíciónk csapdájába kerülünk. Arról nem is beszélve, hogy mivel ezek a modellek a legtöbb elérhető korpuszon tanítva lettek, a validálás is sokkal nehezebbé válik. Meglátásom szerint ugyanezt a megerősítést vagy cáfolatot sokkal direktebb módon meg lehet szerezni a körösség elkerülésével. Ha a konstrukcióként felfogott igekötős szerkezet példányaikat célszerűen általánosítjuk (például a mozaik n -gram és mozaik szózsák modellel (Indig–Bajzát 2023) és a cikkben javasoltakkal (részben) összhangban az igekötő és a képzők mentén szétvágjuk, viszont az egyszerűség kedvéért a teljes képzősorozatot egyként tekintve abban a formában ahogy a példamondatban szerepelt, akkor definiálhatóak olyan slotok, például „az igeképzés bemenetül szolgáló névszó” helye, mely kicserélhető hasonlóan működő szinonimákra (lásd *be- kap, nyom, tol, csap, talicskázik egy szendvicset*). A környezet általánosításának mértékétől és irányától függően pedig közvetlenebbül és pontosabban szabályozhatóvá válik, hogy mi engedhető meg a slotban. Ha ebben az irányban tovább gondolkodunk, akkor más slotok is definiálhatóak az hipotetikus igei szerkezeten belül: például a vonzatkeret vonzatok eseteivel (lásd Mazsola, Sass Bálint 2009), amivel megvizsgálható, hogy egy adott „az igeképzés bemenetül szolgáló névszó” milyen „környezetben” képes megjelenni: azaz milyen vonzatkeret, igekötő és képzősor kíséretében. Ezzel nem csak a névszók válnak osztályozhatóvá, hanem teljes igei paradigmákat tudnánk azonosítani. Természetesen ez egy, a jelen cikktől független kutatás keretében valósítható meg.

5. Összefoglalás

A cikk a lehetetlenre vállalkozik: korpuszalapú, számítógépes vizsgálattal rést kíván ütni azon a gordiuszi csomón, amely a magyar igekötős igék, azon belül az igévé képzett névszók leíró nyelvi elméletéből származtatott konstrukciók erdejébe visz minden érdeklődő kutatót. Mint minden kezdeti lépés tartalmaz bizonytalanságokat, ugyanakkor figyelemre méltó alapos vizsgálatokat és jó meglátásokat is. Közlésre javaslom a cikket kiegészítve a limitációkkal, hogy a későbbiekben táptalajául szolgálhasson a jövőbeli, remélhetőleg mind jobban adatorientáltabb nyelvészeti kutatásoknak.

Irodalom

Indig Balázs – Bajzát Tímea Borbála 2023. Bags and Mosaics: Semi-automatic Identification of Auxiliary Verbal Constructions for Agglutinative Languages. In: Vetulani, Zygmunt – Paroubek, Patrick (eds.): *Human language technologies as a challenge for computer science and linguistics*. Poznań: Adam Mickiewicz University. 111–116.

Keszler Borbála 2000. A szóalkotás módjai. In: Keszler Borbála (szerk.): *Magyar grammatika*. Budapest: Nemzeti Tankönyvkiadó.

Sass Bálint 2009. „Mazsola” – eszköz a magyar igék bővítményszerkezetének vizsgálatára. In: Váradi Tamás (szerk.): *Válogatás az I. Alkalmazott Nyelvészeti Doktorandusz Konferencia előadásaiból*. Budapest: MTA Nyelvtudományi Intézet. 117–119.

Indig Balázs

Eötvös Loránd Tudományegyetem

indig.balazs@btk.elte.hu

ORCID: 0000-0001-8090-3661

LADÁNYI MÁRIA

**Hozzászólás a *Szimultán igekötőzés és igeképzés –
konstrukcióegyesítés a magyarban* című tanulmányhoz**

1. Általános értékelés

A tanulmány többszörösen is fontos hozzájárulás a magyar morfológiai leíráshoz. Először is azért, mert a szerzők egy már jó ideje megfigyelt, de eddig csupán a potenciális szó fogalmának segítségével leírt jelenséget állítanak a figyelem középpontjába, és a konstrukcióegyesítés fogalmának bevezetésével – használatalapú elméleti keretben – magyarázatot kínálnak fel rá. Másrészt azért, mert ezt a felvetett általánosítást az elméleti háttérrel harmonizáló módon, empirikus alapokon, korpuszvizsgálattal, az összes releváns adat tekintetbe vételével, adatvezérelten fejtik ki, illetve igazolják. A szerzők így meggyőző módon kapcsolják össze azt az elméleti megközelítést (a nyelv hálózatos szerveződése, használatalapúsága, a sémák és az analógia szerepe), amelyet Palágyi László eddigi saját munkáiban (például Palágyi 2021, 2022a, 2022b) az MNSZ2 korpuszából vett kisebb mintákon alkalmazott és igazolt, valamint azt a nyelvtechnológiai módszert (az MNSZ2 összes releváns adatának gépi feldolgozása és az adatok kvantitatív jellemzőiből kirajzolódó összefüggések leírása), amelyet Kalivoda Ágnes a korábbiakban elsősorban az igekötős igék, illetve az igekötők vizsgálatában érvényesített (például Kalivoda 2021a, 2021b, 2022) – igazolva ezáltal a kétféle módszertan összekapcsolásának lehetőségét és hasznosságát.

A szerzők a kutatás módszertanát alaposan megtervezték, és a vizsgálatot gondosan kivitelezték. A levont következtetésekkel egyetértek. A tanulmány gondolatmenete jól követhető, a megfogalmazásmód az esetek legnagyobb részében teljes mértékben érthető. Azokat a szöveghelyeket, amelyeknek az esetében véleményem szerint világosabb fogalmazásra lenne szükség, a kéziratban jelöltem.

Kisebbségi megjegyzéseimet, javítási javaslataimat is a tanulmány szövegébe írtam bele, ezen a helyen nem ismételtem meg őket, annál inkább, mivel ezek közül némelyeket a szerzők a tanulmány szövegének végső megformálásakor talán tekintetbe is vesznek. Néhány jelenséggel/fogalommal kapcsolatban vannak olyan megjegyzéseim, amelyeket e helyt is szeretnék megfogalmazni. Ezek – saját korábbi kutatásaimat is tekintetbe véve talán nem meglepő módon – a potenciális szó (és a nem létező szó), az analógia, valamint a produktivitás fogalma.

2. Potenciális szó és nem létező szó

A tanulmány 2.1. alpontjában szerepel a következő mondat:

A deriváció hálózatos jellege magyarázatul szolgál arra, hogy „sokszor nem könnyű eldönteni, hogy két képző egymásutánjáról vagy egy összetett képzőről van-e szó” (Keszler 2000: 315), illetve arra, hogy egy szóképzési művelet hogyan mehet végbe nem létező (vagy potenciális) szavakon (lásd *tanú* → **tanús* → *tanúsít*).

A „nem létező (vagy potenciális) szavakon” megfogalmazás azt sugallja, hogy a nem létező szó és a potenciális szó szinonim terminusok, vagyis a két fogalom teljes mértékben fedi egymást, ugyanazt a jelenséget ragadja meg. Véleményem szerint ez nem szerencsés, a két fogalom azonosítása mindkét oldalról nézve problematikus. Egyrészt a potenciális szó fogalmát valóban szokás használni abban az esetben is, amikor a több (vagy összetett) képzőt tartalmazó derivátum létrehozásában két lépcsőt tételeznek, és az első képzővel létrehozott derivátum nem létező szó (a szerzők által közölt példa is egy ilyen esetet mutat be), ugyanakkor a potenciális szó fogalma véleményem szerint nem korlátozódhat a szóképzés nem létező köztes szavaira, a fogalom tartalma ennél tágabb.

Másrészt azokban a szabályelvű keretekben sem tekintik az összes nem létező szót egyben potenciális szónak is, amelyekben az utóbbi fogalommal operálnak. Potenciális szónak ebben a keretben is csak olyan szót lehet tekinteni, amely megfelel a nyelvi rendszer szabályainak, ezek alapján elvileg létrehozható, de valamilyen okból nem létezik, nem attesztálható.

Annak az oka, hogy egy derivátum nem jön létre, lehet például az, hogy már van egy olyan konvencionális, illetve a közösségi norma szerint el-

fogadott, létező szó, amelyet a beszélők használnak, ezért a szabályos derivátumot nem hozzák létre. (Kiefer–Ladányi (2000:157) szerint ez az ún. „lexikai akadályozás elve”, amely nem abszolút érvényű.) Például potenciális szónak tekinthető a %*harapgat* derivátum, amely teljes mértékben megfelel a nyelv szabályainak, de nincs rá találat az MNSZ2-ben, mivel a beszélők nem ezt a gyakorító igeképzés általános mintázatát követő derivátumot, hanem a normatív *harapdál*-t használják helyette (utóbbira 848 találat van az MNSZ2-ben). Ezekben a norma által szabályozott esetekben a potenciális derivátum alkalmilag mégis létrejöhet, például a gyermeknyelvben vagy a kreatív nyelvhasználat során, mert ezekben az esetekben a norma korlátozó szerepe nem érvényesül (ilyen például a gyermeknyelvben a *píllógó* ’szempilla’ vagy a *festékez* ’fest’ – ez a jelenség a pszicholingvisztikában az ún. „alkotó hibázás” vagy „okos hiba”, vö. Pléh 2000: 952, 990).

Emellett annak, hogy egy adott derivátum nem létezik, akár az is lehet az oka, hogy (a mintázat produktivitása ellenére, ill. annak ellenére, hogy a norma sem korlátozza a használatát) egyelőre egy beszélő sem hozta létre – vagyis bármikor aktualizálódhat a nyelvhasználatban (vö. Ladányi 2007: 67–68, 74–75). Például a *punkoskodik* derivátumra ugyan csak 2 találatot adott ki az MNSZ2 keresőfelülete (a *punk* alapszó előfordulási száma 5966) – de mivel van ilyen szó a korpuszban, a *punkoskodik* már nem potenciális, hanem aktuális szónak tekinthető. Ugyanakkor a *steampunkoskodik* előfordulására egyelőre nincs egy adat sem, sem az MNSZ2-ben, sem a Google keresőben; tehát a *steampunkoskodik* még potenciális szónak számít. (A *steampunk*-ra 381, a *steam punk*-ra pedig 2 találat van az MNSZ2-ben.)

A tanulmányban használnál tágabb a nem létező szó fogalma is. A nem létező szavak nem azonosíthatók teljes mértékben a potenciális szavakkal, mert vannak olyan nem létező szavak is, amelyek azért nem jönnek létre, mert (a potenciális szavakkal ellentétben) ellentmondanak a nyelv fonológiai vagy morfológiai szabályszerűségeinek. A fonológiára nézve funkcionális keretben lásd például Markó hangtani fejezetének egy részét (2017: 188–191) a mássalhangzó-fonémák lehetséges sorbarendezi mintázatairól, formális keretben pedig lásd például Törkenczy szótagról szóló fejezetében az általános szabályokat a fonológiai rosszformáltságról az elágazó szótagkezdetekben (1994: 283). A morfológiára nézve lásd például az angol igék beillesztését magyarba: a **print* ’printel/nyomtat’ vagy a **szken* ’szkennel/beolvas’ szó nem létezik magyar igeként, mert igék nem illeszthetők be a magyarba igeképző nélkül (vö. Ladányi 2007: 48, 2017: 607).

Visszatérve a kiindulóponttra: a nyelv potenciális szavai, annak ellenére, hogy megfelelnek a nyelv szabályszerű mintázatainak, különböző okokból aktuálisan valóban nem léteznek (bár vannak olyan potenciális szavak, amelyek – akár a normát is áthágva – aktualizálódhatnak). Azok a nem létező szavak azonban, amelyek nem felelnek meg a nyelv fonológiai vagy morfológiai szabályszerűségeinek, nem azonosíthatók a potenciális szavakkal.

Mindent összevéve, a szerzőknek ezen a szöveghelyen érdemes lenne valamilyen értelmezést adniuk arra nézve, hogy a potenciális szó és a nem létező szó terminusnak mi a viszonya egymáshoz, illetve az adott helyen milyen értelemben használják őket. Egyúttal érdemes lenne megcserélni a sorrendet, és a nyelv potenciális (nem létező) szavairól beszélni, mert ez az adott összefüggésben magyarázat nélkül is jobb, mint fordítva.

3. Séma és analógia

Örömmel olvastam a tanulmányban a séma és az analógia viszonyáról írtakat. Nagyon fontosnak tartom az analógia fogalmának rehabilitálását, hiszen ezt a – történeti nyelvészetben a nyelv működése szempontjából a szabályszerűséghez kötött és központinak tartott – fogalmat a 20. századi formális megközelítésekben szinte kiiktatták, és csak az egyes elszigetelt nyelvi egységek közötti viszonyokat megragadó, ún. felszíni analógia fogalmát tartották meg. A nyelv használatalapú és hálózatos megközelítése az 1990-es, 2000-es évekre (legalábbis a funkcionális megközelítésekben) az analógia fogalmát visszahozta a nyelvészeti köztudatba. (Ennek alapján az analógia fontosságát és nyelvi rendszerbeli szerepét én magam is igyekeztem elismertetni, vö. Ladányi 2007.) A szerzők azonban nem egyszerűen visszaemelik az analógiát a nyelvi leírásba azzal, hogy „a morfológiai konstrukciók produktivitásában és kidolgozásában a sémának és az analógiának egyaránt szerepe van” (kivonat, lásd 29. oldal), hanem részletesebb kifejtésükben megmutatják a séma és az analógia alapvetőbb és dinamikus összefüggését, a mennyiségi (látszólagos) különbségek mellett lényegi összefüggéseket, illetve azonosságukat (a kiemelések az enyéim):

Az igekötős igei konstrukciók tehát jól mutatják, hogy **a séma mint általánosítás és az analógia mint példány szerepének megkülönböztetése mögött valóban csupán szemléleti különbség rejtőzik.** A mi megközelítésünk szerint a kidolgozás sémaalapúsága (szabályszerűsége) és analógiás (egyedi) jellege

pusztán a konstrukciót képviselő mintázatok eltérő méretéből adódó, látszólagos megkülönböztetés (vö. Booij 2010b). Az analógiának is központi motívuma az (akár egyetlen példány valamely tulajdonságán alapuló) általánosító művelet, miképpen az általánosítások (sémák) mind alakilag, mind szemantikailag lexikai nyalábok sematizációi, amelyekkel dinamikus séma-megvalósulás viszonyt tartanak fenn a sematizáció után is (Bybee 2010), vagyis **nem a példányaiktól független reprezentációk**. A sémák alapjául szolgáló lexikai nyalábok feltérképezése korpuszalapon válik lehetségessé. (2.2. alpont)

A fentiekkel teljes mértékben egyet tudok érteni, és úgy vélem, hogy séma és analógia viszonyának e felfogása komoly elméleti jelentőséggel bír általában és a jelen tanulmány szempontjából is: megalapozza a használat-alapú elméleti megközelítés és a korpuszalapú, adatvezérelt vizsgálatok összekapcsolását.

4. Produktivitás

A szerzők a tanulmányban a vizsgálat jellegénél fogva a produktivitás kvantitatív felfogására építenek, mégpedig egy olyan összetett megközelítésmódot alapul véve, amelyben a produktivitás három mérhető (tokengyakoriság, típusgyakoriság, hapaxok száma) és egy kiszámítható formáját (potenciális produktivitás) egyszerre veszik tekintetbe, és ezek alapján a produktivitás mértékét egy koordináta-rendszer kétdimenziós terében (x tengely: potenciális produktivitás, y tengely: típusgyakoriság) láthatóvá is teszik. Ez az eljárás az olvasó számára nagyon plasztikusan mutatja meg az összefüggéseket, és segíti a szöveges értelmezések megértését. A szerzők a vizsgált, különböző mintázatokat követő igék tokengyakoriságát a koordináta-rendszerben elhelyezett pontok nagyságának segítségével érzékeltetik, ami szintén nagyon látványos, de ezzel kapcsolatban meg kell jegyezni, hogy nagyon kicsi tokengyakoriság esetén a pontok olyan picikék, hogy félő: nyomtatásban ezek esetleg nem fognak látszódni. Ezért – ha lehetséges – javaslom a tokengyakoriságot érzékeltető pontok arányos megnövelését.

Mielőtt azonban a szerzők bevezetnék a kvantitatív produktivitásfogalmat, amellyel dolgoznak, egy általános produktivitásdefinícióból (Kiefer–Ladányi 2000: 149) indulnak ki, majd hivatkoznak Dressler 2003-as munkájára, miszerint Dressler „nem tartja helyesnek a kompetenciaszintű lehetőség performanciaszintű valószínűségeen alapuló vizsgálatát” (5.1. alpont).

Ez utóbbi megállapítás igaz, és a gondolatmenet szempontjából fontos is. Ugyanakkor – míg a kvantitatív produktivásfogalomnak (érthető és szükséges módon) egy, az utóbbi 30 év során kialakult, összetett, finomított változata jelenik meg a tanulmányban – addig a szerzők a kvalitatív produktivásfogalomra csak a legáltalánosabb, fokozatiságra nem utaló, igen–nem típusú változatában, leegyszerűsítve utalnak, lehetséges árnyalásáról még egy lábjegyzetben sem szólnak. A szövegből emiatt úgy tűnik, mintha a kvalitatív produktivásfogalom meghatározásában a fogalom árnyalása (összetettségének és fokozati jellegének feltárása) fel sem vetődött volna. Ezt nem tartom helyesnek, mivel éppen a szövegben más összefüggésben hivatkozott Dressler dolgozta ki a produktivitás mértékének megállapítására vonatkozó kritériumrendszer természetes nyelvelméleti keretben, először az inflexiók (Dressler 1997), később a derivációs morfológiára nézve (Dressler–Ladányi 2000), majd az idegen képzős derivátumokkal kapcsolatban a kritériumrendszer tovább módosult (Ladányi 2007: 179–181). Emellett a kvalitatív megközelítésen belül a produktivitás összetettsége is szóba került (például az Osiris Nyelvtan alaktani fejezetében is (vö. Ladányi 2017: 545–564), amelyben a sémaerősségen, illetve a típusok számán túl az alapbeállítás, azaz a sémát korlátozó tényezők hiánya, illetve az alkalmazási tartomány növekedése is jelentőséggel bír a produktivitás alakulásában).

Természetesen nem kérem számon a szerzőktől a fentieknek akár csak vázlatos ismertetését sem, hiszen a tanulmányban nem építenek a kvalitatív produktivásfogalomra. Véleményem szerint azonban – mivel hivatkoznak rá, emellett Dressler nevét is megemlítik – a kvalitatív produktivásfogalom összetettségének és fokozatiságának felvetődéséről legalább egy lábjegyzetben, nagyon röviden említést kellene tenniük.

5. Egyéb megjegyzések

A tanulmányban felhasznált adatok forrása az MNSZ2 legújabb változata, az MNSZ2-UD korpusz, amely az MNSZ2-höz képest „számos korpusztisztítási lépésen” esett át, és „új elemzési szintekkel gazdagodott” (2.2. alpont). Ezt az új korpuszváltozatot forrásként is jó lenne feltüntetni az irodalomjegyzékben, annak ellenére, hogy még nem jelent meg (jelenleg az új korpuszváltozatról szóló cikk szerepel csak az irodalomjegyzékben).

A vizsgálat során az MNSZ2-UD korpuszból létrehozott, hatalmas adatmennyiséget (1,1 millió szót), az összes releváns adatot tartalmazó adattábla

nemcsak a tanulmányban bemutatott vizsgálatnak az alapja, hanem – ahogyan arra maguk a szerzők is rámutatnak – kiindulópontja lehet más vizsgálatoknak is, ezért javaslom azt is, hogy ennek az elérhetőségét külön is emeljék ki az irodalomjegyzékben.

A tanulmány végéről hiányoltam egy rövidítésjegyzéket, mivel az I. számú mellékletben csak az adattáblában szereplő rövidítések feloldása van megadva. Különösen a glosszákból előforduló rövidítések pontos feloldására lenne szükség, vagy ha a szerzők a tanulmányban nem akarnak teljes rövidítésjegyzéket szerepeltetni, akkor az első rövidítésnél jelezni kellene, illetve az irodalomjegyzékben is meg kellene adni annak/azoknak a konvenció(k)nak az elérhetőségét, amely(ek) a rövidítések forrása(i). Ilyenek lehetnek például a lipcsei glosszázási szabályok – bár nekem úgy tűnt, hogy a glosszákból szereplő rövidítéseknek nem mindegyike szerepel azokban sem, ezért az ott elő nem forduló rövidítések jegyzékére mindenképpen szükség lenne. (Emellett, a jobb olvashatóság érdekében, bizonyos ritkább rövidítéseket talán akkor is érdemes lenne helyileg feloldani, ha egyébként van rövidítésjegyzék, és az adott rövidítés abban szerepel is.)

Irodalom

- Dressler, Wolfgang U. 1997. On productivity and potentiality in inflectional morphology. *CLASNET Working Paper 7*. Univ. de Montréal.
- Dressler, Wolfgang U. 2003. Degrees of grammatical productivity in inflectional morphology. *Rivista di Linguistica* 15: 31–62.
- Dressler, Wolfgang U. – Ladányi, Mária 2000. Productivity in Word Formation (WF): a morphological approach. *Acta Linguistica Hungarica* 47: 103–144. <https://doi.org/10.1023/A:1014010530824>
- Kalivoda Ágnes 2021a. Igekötős szerkezetek a magyarban. Doktori (PhD) értekezés. Pázmány Péter Katolikus Egyetem Bölcsész- és Társadalomtudományi Kar. Kézirat.
- Kalivoda Ágnes 2021b. Az igekötők produktív kapcsolódási mintái. *Argumentum* 17: 56–82. <https://doi.org/10.34103/ARGUMENTUM/2021/4>
- Kalivoda Ágnes 2022. PrevDistro: An open-access dataset of Hungarian preverb constructions. *Acta Linguistica Academica* 69: 549–563. <https://doi.org/10.1556/2062.2022.00578>

- Kiefer Ferenc – Ladányi Mária 2000. A szóképzés. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 3. Morfológia*. Budapest: Akadémiai Kiadó. 137–164.
- Ladányi Mária 2007. *Produktivitás és analógia a szóképzésben: elvek és esetek*. Budapest: Tinta Könyvkiadó.
- Ladányi Mária 2017. Alaktan. In: Tolcsvai Nagy Gábor (szerk.) 2017. *Nyelvtan*. Budapest: Osiris Kiadó. 503–660.
- Markó Alexandra 2017. Hangtan. In: Tolcsvai Nagy Gábor (szerk.) 2017. *Nyelvtan*. Budapest: Osiris Kiadó. 75–203.
- Palágyi László 2021. A magyar nomen actionis és metonimikus kiterjesztései. *Argumentum* 17: 374–398. <https://doi.org/10.34103/ARGUMENTUM/2021/19>
- Palágyi László 2022a. *Morfológia az egésztől a részekig: A magyar főnév esete*. Budapest: Szépirodalmi Figyelő Alapítvány.
- Palágyi László 2022b. A metafora, az analógia és a szemantikai megfelelések szerepe a le igekötős konstrukciók jelentésében. In: Tátrai Szilárd – Tolcsvai Nagy Gábor (szerk.): *A magyar mondat és kontextuális környezete*. Budapest: ELTE BTK – ELTE Eötvös Kiadó.
- Pléh Csaba 2000. A magyar morfológia pszicholingvisztikai aspektusai. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 3. Morfológia*. Budapest: Akadémiai Kiadó. 951–1020.
- Törkenczy Miklós 1994. A szótag. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 2. Fonológia*. Budapest: Akadémiai Kiadó. 273–392.

Ladányi Mária

Eötvös Loránd Tudományegyetem, Budapest

ladanyi.maria@btk.elte.hu

ORCID: 0000-0001-9893-8504

KALIVODA ÁGNES – PALÁGYI LÁSZLÓ

Válasz Indig Balázs hozzászólására

Mindenekelőtt szeretnénk megköszönni Indig Balázs minden részletre kiterjedő, konstruktív hozzászólását, valamint azokat a kisebb módosítási javaslatokat, amelyeket informálisan küldött a cikkünkhöz. Bár nem ismerjük a kötethez készült összes lektori véleményt, valószínűsítjük, hogy Indig Balázs hozzászólása képviseli a leghatározottabban a számítógépes nyelvészeti szempontokat, amelyeket egy ilyen interdiszciplináris kutatásnál érdemes szem előtt tartani. Válaszunkban idézzük a bírálókat azokat a részeit, amelyekkel kapcsolatban hosszabb mondanivalónk van: indokoljuk, hogy miért nem, vagy miért nem úgy végeztük el a javítást, ahogy a bíráló kérte. A bíráló többi részével a tanulmány végső változatában foglalkoztunk, beépítettük a javaslatokat.

Tudományos háttéremből fakadóan, több esetben az az érzésem, hogy túl mély előismereteket igényel a cikk megértése, melyek rövid magyarázatokkal enyhíthetők lennének. Például a „sokszor nem könnyű eldönteni, hogy két képző egymásutánjáról vagy egy összetett képzőről van-e szó” (Keszler 2000: 315) idézet esetén kifejtve, hogy pontosan milyen módon próbálták meg eldönteni a kategorizációt. Esetleg röviden felvonultatva, hogy a számítógépes nyelvészeti eszköztár milyen kiaknázatlan lehetőségeket tartogat (például kollokációs metrikák) a kérdés további vizsgálatára, mintegy kontrasztot képezve. Specifikusan azt is érdemes megmagyarázni, hogy miért szükséges szigorúan szétválasztani vagy egyben hagyni „két képzőt”, mely többnyire együtt fordul elő, de potenciálisan van bizonyíték, hogy valamilyen formában külön is működnek az adott környezetben belül (például régebbi nyelvállapot). Ha eltekintünk ettől a szigorú feltételtől, mennyivel lesz rosszabb a magyarázat a jelenségre?

A leírást igyekeztünk világosabbá tenni, és készítettünk egy rövidítésjegyzéket is, de ettől függetlenül nem valószínű, hogy nyelvészeti előképzettség nélküli olvasó számára ez egy jól követhető cikk lesz. Jó lenne persze

a kályhától indulni, de a teljes közérthetőség elérését nem tartjuk reális célnak. Az általunk bemutatott felfogást részletesebben szembeállítottuk a hagyományos leírással a 2. fejezetben. Ami a kollokációs metrikákat illeti, ezekbe nem szeretnénk belemenni, mert nagyon nagy falat, de érdekes kísérleteknek adna teret, ebben egyetértünk. A régebbi nyelvéllapok szintén nagyon nagy falatot jelentenek, és ebben a tanulmányban nem szeretnénk ilyen irányba (is) elmenni. Viszont annyit jeleznénk, hogy mi is elkezdtünk korábban gondolkodni erről, és a 21st International Morphology Meetingen lesz erről egy előadásunk (2024 augusztusában).

A „Jellemző, hogy az igen specifikus konstrukciós jelentések sokszor viszonylag szűk számú, ugyanakkor viszonylag gyakori igekötős igére épülnek, mégis produktívak.” mondatnál az olvasónak önkénytelenül is az jut eszébe, hogy az innovatív nyelvhasználat kifejezetten táptalaja az ilyen jellegű produktivitásnak (például be- kap, nyom, tol, csap egy szendvicset a szünetben). Kérdés, hogy a különféle stílusregiszterekben ugyanúgy viselkedik-e az említett konstrukció? A korpusz által felölelt évek diakróniájában tekintve tapasztalható-e változás?

A produktivitás és az innovatív nyelvhasználat szerintünk ugyanaz. Persze feltűnőbb a kreativitás, ha sokak számára nem magától értetődő analógiát alkalmaz a beszélő. Érdekesnek tartjuk a felvetést, mert valószínűsíthető, hogy a hagyományos megoldások és az újítás nem ugyanolyan súllyal esnek latba minden regiszterben, és ezt akár lehetne korpuszszárazon is vizsgálni, de szerintünk ez messzire vezet ettől a tanulmánytól.

A metrika leírásánál is érdemes olyan példát választani, amely nem a leírás fordítottját bizonyítja, azaz, „adattáblában a *metálosodik* ige nem szerepel önmagában, csak a *meg* igekötővel együtt (*megmetálosodott*), de mivel ez az egyetlen előfordulása a korpuszban, nem indokolt az egyesített konstrukció jellemző példányaként figyelembe vennünk.” vö. „a *megmetálosodott* ige egyszer fordul elő, ezért csak azt állíthatjuk, hogy nincs arra bizonyítékunk, hogy igekötő nélküli változata is létezik, de ez nem kizárható, így semmilyen szempontból nem tudjuk figyelembe venni statisztikailag.”

Mi itt nem látjuk a problémát. A *megmetálosodott* ige pont azt mutatja, hogy fontos a *freqsum* érték, mert ha csak a másik kettőt használnánk, akkor ez az ige is egyesített konstrukció példányaként jelenne meg. De ez nem lenne

szerencsés, mivel valójában nincs róla elég adatunk, tehát nem tudunk róla nyilatkozni.

A cikket olvasva felötlik az emberben, hogy vajon mennyire jó az, ha minden mérést egy korpuszon végzünk el? Nem lenne-e célszerű a számítógépes nyelvészeti kánonban megszokott módon egy korpuszt/korpuszrészletet a hipotézis felállítására míg egy másik tőle független adathalmazt a hipotézis validálására használni feltéve, ha nem jár aránytalanul sok többletmunkával? Így elkerülhető lenne, hogy abba a szerencsétlenségbe essünk, hogy a nyelv helyett csak az adott korpusz részletes tulajdonságait ismerjük meg. Különös tekintettel arra, amikor kvantitatív módszert használva a szavak gyakoriságára vagy épp ritkaságára hivatkozunk. Az ezen hibák elkerülésére, illetve fenn nem állására vonatkozó érvekkel javaslom kiegészíteni a cikket.

Ez a megjegyzés is messze vezet a cikktől, nem gondoljuk, hogy ilyen típusú kiegészítésre szorulna a cikk. Abban egyetértünk, hogy az mindig pozitívum, ha egy mérés több korpuszon is el van végezve, de ez – ahogy a bíráló is sejteti a hozzászólásában – az esetek nagy részében aránytalanul sok többletmunkával jár. A fent leírt eljárás egyébként a nyelvtechnológiában megszokott – ahol valóban elég nagy tétje és ipari felhasználása is lehet egy-egy ilyen munkának –, de a korpusznyelvészetben nem az (egyébként a validálási lépést maga a bíráló is kihagyja több publikációjában). A *számítógépes nyelvészet* elnevezés kicsit egy kalap alá veszi ezt a két területet, de a mi munkánk egyértelműen korpusznyelvészeti, nem pedig nyelvtechnológiai.

Ami a felhasznált eszközöket illeti, nem világos számomra, hogy a középmagyar emMorph mennyivel adott hozzá többet a vizsgálathoz, mint ha például reguláris kifejezések segítségével végezték volna a szerzők a szavak osztályozását. Hiszen, mint írják a végén úgyis kézzel átnézték az eredményt. Dicséretesnek tartom, hogy kiemelik, hogy az emMorph különböző változatainak erősségeit és gyengeségeit, de úgy érzem, hogy a szerzők indokolatlanul ragaszkodnak ehhez az eszközhöz, amikor hasonló hatékonysággal egyszerűbben is meg lehet oldani ugyanazt a feladatot.

Ezzel nem értünk egyet. Ha reguláris kifejezésekkel próbálnánk megoldani ezt a keresést, rengeteg fals pozitív találatot kapnánk. A *-(s)Odik* képzőre úgy lehetne keresni, hogy *s[?oöe]dik\$* (a lemmában, case insensitive módon). Ha kipróbáljuk ezt a keresést, többek között ilyen szavakat kapunk találatként:

közlekedik, álmodik, küszködik stb. Ezek egyikében sem a minket érdeklő képző szerepel. A *-(s)Ul* esetében is hasonlóan rossz a helyzet a *s?[uü]l\$* reguláris kifejezéssel: *fordul, kerül, merül* stb. – ezek mind fals pozitív találatok. Tehát morfológiai elemző nélkül sokkal kevésbé hatékonyan lehet csak megoldani ezt a feladatot, mivel sokkal több hamis találatot kellene kézzel kiszűrni. Nem szólva arról, hogy a bázisszavakat milyen nehéz lenne morfológiai elemző nélkül megállapítani.

Ugyanennél a résznél meglátásom szerint érdemes lenne ismertetni a konkrét számokat. A „jóval több” és „Minden egyes előfordulás kézi átnézésére nem vállalkoztunk, mivel milliós nagyságrendű találati listával volt dolgunk; az ilyen kétértelmű eseteket egységesen megőriztük. Ezáltal hibás adatok is bekerülnek a vizsgálatainkba, de ez ekkora mennyiségű szöveg automatikus feldolgozása során elkerülhetetlen, és számottevően nem befolyásolja az eredményeinket.” jellegű kvalitatív jelzők elveszik az élet az amúgy kiváló számítógépes vizsgálatnak. Pontosan milyen arányban került az anyag átnézésre kézzel? Mennyire reprezentatív? Milyen hibák maradhattak és pontosan miképpen befolyásolják a későbbieket és ezek a tényezők miért nem fontosak?

A kvalitatív jelzőket konkrét számokra cseréltük a végső verzióban, ahol csak tudtuk, és a kézi átnézéshez kapcsolódó listákat is publikussá tettük Githubon a kódokkal együtt (lásd https://github.com/kagnes/odik_ul). A „számottevően nem befolyásolja az eredményeinket” állítást kivettük a cikkből, mert biztosak ebben tényleg nem lehetünk, csak annyit tudunk mondani, hogy a kapott eredmények az intuíciónkkal összhangban vannak. A reprezentativitásról nem tudunk konkrét számok formájában nyilatkozni. Azt, hogy milyen jellegű hibák maradhattak – kétértelműségből adódók –, tárgyaljuk a cikkben, az arra történő utalást, hogy ezek nem annyira fontosak, a végső verzióból kivettük. Szeretnénk megjegyezni, hogy a bíráló ebben a bekezdésben szerintünk irreális elvárásokat fogalmaz meg a tanulmánnyal kapcsolatban. A kézi átnézés, annotálás önmagában sok munka – esetünkben nagyjából 10 munkaóra volt –, a reprezentativitás vizsgálata, a hibatípusok részletes elemzése, sőt a várható hatásai elemzése indokolatlanul sok munkával járna. A kutatás háttéranyagait publikussá tettük Githubon, így aki ilyen szinten akar elmélyedni a témában, azelőtt most már nyitva áll a lehetőség.

Ekkora anyag esetében fontosnak tartanám jobb példák kiválasztását. A terminális állapotváltózásként értelmezett *vadul* ('vaddá válik') szó véleményem szerint nem jelenhet meg ebben az értelemben igekötő nélkül („elvadul”), mely megkülönbözteti a hallgató számára a „vad módon” jelentéstől, ami a folyamatossága miatt csak igekötő nélkül létezhet („vadul”).

Szerintünk a *vadul* jó és szemléletes példa, inkább csak a jelentését írtuk körül pontatlanul. A szövegben ezt pontosítottuk, és hoztunk kontextussal együtt példákat: „Ez lehet hamis pozitív találat, amennyiben határozói funkcióban, 'vad módon' jelentésben áll (például *vadul forgatja a szemét*), és lehet számunkra releváns találat is, '(lassanként) vaddá válik' jelentésben (például *nehogy állatokat táplálván és nevelvén maguk is állatokká vaduljanak*).” Az első esetben hibásan van igének elemezve a határozószó, ezért jelent meg nálunk találatként, és ezért okoz itt problémát.

Formálisan a metrika számítógépes és matematikai értelemben is egy olyan függvényt jelöl, aminek tetszőleges a bemenete és egy szám a kimenete. A cikkben definiált „metrika” viszont három számból áll, ezzel pedig három külön metrikáról és azok összefüggéséről van szó. Ha egyként akarjuk értelmezni őket, akkor valahogyan súlyozandóak, hogy egy számot alkossanak a végén. Ez nem elvárt, viszont a jelenleg használt fogalmak megkövetelnék. Érdeemes lenne ezen kívül egy számszerű „elvárt”, indikatív küszöbértéket megadni már itt objektivitás miatt (vö. „mindhárom felsorolt érték magas”) és mint minden ilyen paraméterbeállítást validálni egy másik korpuszon.

A *metrika* szót valóban pongyolán használtuk. Mivel formális értelemben elfogadható, egy számértékből álló metrikát nem hoztunk létre, a cikk javított verziójában csak számérték(ek)ről, és nem metrikáról írunk. Az indikatív küszöbértékünk – amelyet nem így nevezünk, de a 4. fejezetben kiderül – az, hogy a *freqsum* minimum 500, a *prev_vs_all* és *actprev_vs_allprev* pedig minimum 90%. Ez természetesen nem vihető át egy az egyben másik korpuszra (az MNSZ2-UD-ben a minimum 500-as *freqsum*, azaz tokengyakoriság más korpusz esetén túl alacsony vagy épp túl magas küszöbérték lehet). Értjük a bíráló aggályát ezzel kapcsolatban, de mi ezen a ponton egyelőre nem jutottunk előrébb, és azt gondoljuk, hogy bár nem tudjuk megfelelően bizonyítani a három számérték használhatóságát általánosabban – nem korpusz-specifikusan –, maga az ötlet mindenképp publikálásra érdemes egy

ilyen workshop-jellegű kötetben. A három számérték összevonása is hasznos lenne, de egyelőre nem tudunk olyan megoldással szolgálni, ami ne járna jelentős információvesztéssel.

Az ismertetett szóbeágyazás alapú módszer szerepe nem derül ki a cikkből, csak annyi, hogy mint kiegészítő információ hozzá lett adva a keletkezett adattáblához. Miért is jó ez? Véleményem szerint ilyen formán és körkörös érvelésre ad lehetőséget a felhasználása. A cikkben nincs kifejtve, hogy a szóbeágyazás működése folytán az adott szó előfordulásainak egyes kontextusaiból veszi azt az információt, mely segítségével algoritmikusan, a skatulya elv alapján csoportokba szorítjuk a szavakat. A csoportosítás mikéntjét legjobban a bemeneti szavak módosításával lehet elérni, ezzel előállítva például szintaktikus és szemantikus csoportosításokat (lásd a hivatkozott cikk). Ilyen célszerű módosítás lehet például a szavak szétvágása az igezőtők és a képzők mentén, mely az irányban befolyásolja a rendszert, hogy a függetlennek vélt intuíciónak megfelelőbben viselkedjen körkörösé téve az érvelést, miszerint „a vektortér alapú csoportosítás visszaigazolja az intuíciónkat”. Úgy vélem, hogy a tanulmányban bemutatott eredményekhez nem járul hozzá a választott szóbeágyazási módszer alkalmazása.

Köszönjük a szóbeágyazással kapcsolatos meglátásokat, a válaszunkat beemeltük a cikkbe. A 7. lábjegyzetben térünk ki erre, hivatkozva a lektori hozzászólásra.

Appozitív határozók



INDIG BALÁZS – KURUCZ RÉKA – ERNHÖFFER ALIZ –
KUGLER NÓRA

Appozitív határozók korpuszalapú vizsgálata

Kivonat

A tanulmány a főnévi alaptagot bővítő, a főnév utáni helyzetben előforduló (N1.NOM N2.CASE), azaz appozitív határozók (például *visszaélés a hatalommal; napilapok nemzeti tulajdonban*) korpuszalapú azonosításával foglalkozik. Az N.CASE a konstrukció leírásában főnévi esetalakot szimbolizál a nominativusi és az accusativusi esetalak kizárásával (OBL : MODE megkülönböztetés nélkül). A dolgozat kognitív nyelvészeti keretben bemutatja az appozitív határozós szerkezet problematikáját, és megfogalmazza azt a kutatási kérdést, amely az adatvételezést életre hívta. Ezt követően ismerteti az MNSZ2-ben (v2.0.5) végzett adatkeresési, lekérdezési folyamatot az első próbálkozásoktól addig a kontextusalapú keresésig, amellyel értékelhető eredményekhez jutottunk. A sikeres lekérdezést az N2.CASE *való* N1.NOM konstrukció adataira alapoztuk, és ez ki is jelölte a jelen vizsgálat határait. A korpusvizsgálat így azokra a potenciális N1.NOM N2.CASE appozitív határozós szerkezetekre szűkült, amelyeknek alternatívája az N2.CASE *való* N1.NOM konstrukcióban történő megvalósulás. A dolgozat tehát az N2.CASE *való* N1.NOM konstrukció adataiból kiindulva azzal kapcsolatban közöl eredményeket, hogy az appozitív határozós szerkezet konstruálási módként milyen mértékben tekinthető a *való* szerkezet alternatívájának. Ezzel összefüggésben, szintén alternatív konstrukcióként számításba veszi az N2.CASE N1.NOM, tehát szomszédos, határozói bővítmény és N alaptag szórendű konstrukciót is (például *válságból kilábalás*). A tanulmány csak kezdeményezi, de nem

végzi el az N2.CASE *l[eé]vő* N1.NOM konstrukció¹ hasonló elemzését, reflektál azonban az ezzel kapcsolatos előzetes megfigyelésekre.

Kulcsszavak: jelzői értékű határozó, *való* segédszó, összetett jelző, függőségi viszony (dependenciaviszony), szórend

1. Bevezetés és a kognitív nyelvészeti kérdésfeltevés

1.1. Az appozitív határozóról

A magyar függőségi szakirodalomban problematikus a főnév után álló, a jelölt entitás körülményét részletező appozitív határozó megítélése (lásd (1)–(2)).

- (1) Másrészt a divat: *tartozás egy csoporthoz*, egyben az önazonosság-tudat erősítésére szolgál. (#106017260)²
- (2) nincsenek *napilapok nemzeti tulajdonban* (#363282000)

Mivel az appozitív határozó szemantikailag a főnévvel áll szoros kapcsolatban, a főnév és nem az ige (vagy más lexikális kifejezéssel alkotott magmondattal, illetve igenévi alaptag)³ bővítményeként kezeljük. A magyarban az utóbbi, az igétől, igei jellegű kifejezéstől való „függés” a tipikus megoldása a határozók bekapcsolódásának a mondat viszonyhálózatába (Kugler–Szabó 2024, vö. Imrényi 2017).⁴

Az (1) példában a *divat* értelmezője az appozitív határozós szerkezet (*tartozás egy csoporthoz*), a főnévi alaptag csak a határozói vonzattal együtt alkalmas a divatjelenség pontosítására, részletezőbb kifejtésére. Az appozitív határozó (N.ALL) az (1)-ben a *tartozás* ’folyamat’ jelentésű deverbális főnév vonzata.

¹ A *l[eé]vő* forma azt jelöli, hogy mind a *levő*, mind a *lévő* alak figyelembe lett véve.

² A dolgozatban ezzel a formátummal az MNSZ2-ből vett adatokat jelöljük, a példa mellett zárójelben a szóazonosító van megadva. A dőlt kiemelés a szerzőktől származik.

³ Lásd Imrényi 2017.

⁴ Antal László is amellett érvelt, hogy a határozó az igével van intim viszonyban (Antal 1975: 44). De épp azért, mert az általunk appozitív határozónak nevezett kifejezés nem az ige körülményét dolgozza ki, Antal új terminust hozott létre a nominális alaptaghoz kapcsolódó „adverbialis bővítmény” (sic!) kategorizálására: *kiegészítő* (Antal 1985: 44).

A (2)-ben a napilapoknak az állapotukon keresztüli specifikációját végzi el a *nemzeti tulajdonban* appozitív szabad határozói lánc, és ezzel a bővítmény hozzájárul a *napilapok* főnévvel jelölt dolgok értelmezéséhez. A *nincs* igével kifejezett tagadás a teljes szerkezetet a hatókörébe vonja, tehát a *nemzeti tulajdonban* állapot jellegű körülménnyel részletezett *napilapok* létezésének tagadását jelöli.⁵

Az appozitív határozói konstrukció tehát nem a jelenet egészéhez tesz hozzá valamit a körülmény jellegű viszonyoknak a részletezésével, hanem a főnévvel jelölt dolog specifikálásához, alkategorizálásához, sikeres referenciális azonosításához járul hozzá (vö. Radden–Dirven 2007: 143–145; Roskó 2020, 2023). Ezek a műveletek pedig alapvetően a jelzős szerkezeteket jellemzik, és a főnév (típus) nominálissá (példánnyá) alakításához kapcsolódnak a nyelvi kifejezésben (Langacker 2008: 264). A határozói kifejezések ebben a konstrukcióban követik a nominálist.⁶

A jelen tanulmány Kugler és Szabó (2024) alapján az appozitív határozó terminust alkalmazza a vizsgálat fókuszába helyezett konstrukcióban előforduló, a nominális körülményét részletező kifejezésekre. Az *appozitív* csak a főnevet követő⁷ helyzetre reflektál, nem implikálja egy eredetileg elől álló kifejezés hátravetését. A terminusválasztás ugyanakkor szándékoltan jelzi, hogy az appozitív határozói konstrukciós séma asszociatív viszonyban áll nemcsak az ígét bővítő határozói szerkezetekkel (vö. *tartozik* N.ALL), hanem a körülmény kidolgozásával specifikáló attributív szerkezettel is (vö. *nemzeti tulajdonban lévő napilapok*; *nemzeti tulajdonú napilapok*), amelyben a bővítmény megelőzi a főnevet. Az asszociatív viszonyok hálózatába beletartozik az értelmezős szerkezet is,⁸ az *appozitív* komponens tehát az értelmezővel való művelati kapcsolatra is reflektál. Mindkét szerkezet egy a főnév által már megnyitott tartomány részletezéséhez, nagyobb fokú elaborációjához járul hozzá (Langacker 2012).

⁵ Az állapotnak a közvetlen tagadása a *nem nemzeti tulajdonban* [vannak] lenne, a dolog bizonyos állapotban levésének (integrált) negációja pedig *nincsenek nemzeti tulajdonban napilapok* volna.

⁶ A nemzetközi szakirodalom a kifejezésre legáltalánosabban a *postnominal modifier* terminust alkalmazza (lásd például Radden–Dirven 2007: 145). Antal (1985: 47) szerint ez a szórendi elrendeződés jellemző, de nem kizárólagos, és a szomszédosság sem feltétel.

⁷ A követő helyzetben a szomszédosságra vonatkozó értelmezésünk azt megengedi, hogy a bővítmény saját bővítményei az appozitív határozó elé és így a nominális és az appozitív határozó közé kerüljenek.

⁸ A jelen munka az értelmezős szerkezet tagjainak kapcsolatát Imrényi (2017: 756–758) javaslatát követve kontextualizáló viszonyoknak tartja.

Mivel a specifikáló és kidolgozó részletezés, valamint az alkategorizálás művelete az attributív szerkezetekben tipikus (vö. [vmely] *csoporthoz való tartozás; nemzeti tulajdonban lévő napilapok; nemzeti tulajdonú napilapok*), a hagyományos függőségi elemzés *jelzői értékű határozó*-nak nevezte ezt a mondatrészt (lásd Hámori 1954).⁹ A strukturális szemléletű leírások szintén vizsgálják az attributív szerkezettel való kifejezés lehetőségeit (Rákosi 2021: 349, 353). Mivel a magyarban a főnév bővítményei tipikusan az alaptag előtt állnak, az appozitív határozó sajátos, szórendileg az értelmezővel rokonítható specifikáló, részletező szerkezet. A főnév utáni és az ige előtti helyzetben a határozói viszony többértelmű lehet, ez a szórendi elrendezés ugyanakkor erőteljesen kezdeményezi az ígéhez kötődést, ezért a többértelműség csak akkor áll fenn, ha a főnévvel nagy konceptuális egységben is feldolgozódhat az adott 'körülmény' jelentés (lásd (3)).

(3) Értesült a botrányról *a ligetben*. (Balogh 2000: 136)

A *ligetben* szabad határozó a (3)-ban értelmezhető igei bővítményként ('a ligetben értesült a botrányról') és a főnévhez kapcsolódó határozóként is ('a ligetbeli / ligetben történt botrányról értesült'). Az ambiguitásnak ezt az esetét már korábban is tárgyalta a szakirodalom (többek között Hámori 1954; B. Lőrinczy 1962; Komlósy 1992: 463; Balogh 2000; lásd részletesen Kugler–Szabó 2024).

Kugler és Szabó (2024) szerint a többértelműség akkor áll fenn, ha a szerkezet mind az igei, mind a főnévi kapcsolat esetében értelmezhető, és mindkét értelmezés hasonlóan életképes, hasonló valószínűséggel hozható létre a befogadás folyamatában. Ezzel kapcsolatban Kugler és Szabó (2024) azt a kutatási kérdést fogalmazta meg, hogy mitől függhet a határozó, illetve a mondat többféle lehetséges értelme közötti választás, illetve ha mindkét lehetséges értelmezést megalkotja a mondat befogadója,¹⁰ akkor van-e olyan, amelyik

⁹ Az egyéb megnevezések is általában annak a szándéknak felelnek meg, hogy a jelző és a határozó közötti átmenetiséget érzékeltessék (lásd Hámori 1954 és ennek kognitív függőségi nyelvtani reflexiójához Kugler–Szabó 2024). A 20. század közepéig a jelzői főkategória tűnik dominánsnak (vö. Rác 1968: 293), Hámori (1954) után azonban inkább a határozó alkategóriájának tartják az appozitív határozót (vö. B. Lőrinczy 1962: 169; Balogh 2000: 135–136; Keszler 2000: 431).

¹⁰ A tanulmány és a tervezett vizsgálat is az ambiguózus szerkezetek értelmezésére fókuszál, az értelmezői pozícióhoz kapcsolódó műveletekre igyekszik következtetni. Ezek a szerkezetek a létrehozás szempontjából alapvetően nem többértelműek. Köszönjük Horváth Péternek, hogy felhívta a nézőpont explicitálásának szükségességére a figyelmünket.

könnyebben, ezért rövidebb idő alatt születik meg; továbbá az utóbbi esetben megfigyelhető-e jellemző eltolódás valamelyik értelmezés felé, tehát lehetséges-e, hogy nem csak egyéni tényezőktől függ a kézenfekvőbb értelmezés. Az empirikus vizsgálathoz azonban nem állt Kugler és Szabó rendelkezésére a használatalapúság koncepciójának megfelelő nyelvi adat, ezért kezdeményezték az erre irányuló korpuszvizsgálatot. A jelen vizsgálat ehhez a célhoz kapcsolódik: az appozitív határozók tényleges nyelvi adataihoz kíván jutni korpuszalapúan, az adatvételezés során a humán beavatkozást a szükséges legkisebb mértékre korlátozva.

1.2. Az N2.CASE való N1 konstrukció és az appozitív határozó összefüggéséről

Az appozitív határozói szerkezetben a függőségi viszony azonosíthatósága, az összetartozás felismerhetősége különböző mértékű lehet. A szokatlan szórend ellenére is egyértelműen azonosítható a függőségi viszony akkor, ha az alaptag olyan képzett főnévi kifejezés, amelynek a fogalmi szerkezetében centrális a határozói bővítményként megkonstruált szereplő (például *tartozás egy csoporthoz*, lásd (1), vö. *tartozik [valahova / egy csoporthoz]*).¹¹ Bár az (1) példában aligha kérdőjelezné meg bárki is azt, hogy a *valahova / egy csoporthoz* határozói bővítmény vonzat, és ez összefügg azzal, hogy az összetartozási viszony landmarkja centrális az ige (*tartozik*) és a deverbális főnév (*tartozás*) fogalmi szerkezetében,¹² a vonzatosságnak, illetve a vonzat (komplementum) és a szabad határozó megkülönböztetésnek a jelen vizsgálat nem tulajdonít meghatározó jelentőséget. Egyrészt azért, mert ez a megkülönböztetés nem végezhető el éles határok mentén (vö. Kálmán 2007), másrészt nem oldható meg nyelvtchnológiai eszközökkel (sem a bináris, sem a skaláris felfogás mellett).

¹¹ A lektori véleményre reflektálva jelezzük, hogy a deverbális alaptag csak megkönnyíti a viszony azonosítását, nem állítjuk azt, hogy ez volna az egyetlen motivációja a viszony felismerésének. A dolgozat további része tárgyalja is a nem deverbális alaptagokat, és a konstrukciók hálózatos viszonyában, a szemantikai kapcsolatok erősségében jelentőséget tulajdonít a gyakoriságnak.

¹² A FrameNet *belong* (Frame: Membership; COD: be a member of) keretbe tartozóként határozza meg a GROUP és a MEMBER elemeket. A magyarban a poliszém *tartozik/tartozás* alaptag jelentésszerkezetei konstrukciósan jól elkülönülnek (vö. *tartozik vki vkinek vmivel: tartozik vki vkihez/vmihez*), a határozói bővítmények tehát a jelentés egyértelműsítésében is szerephez jutnak.

Kohezív a kapcsolat a komponensek között akkor is, ha az alaptag 'dolog' jelentésén keresztül, metonimizációval¹³ magyarázható a viszony megkonstruálódása a landmarkként szolgáló entitással (például az út [visz/vezet] *valahova*, lásd (4)).

(4) Mutasd meg neki az utat a kávéházba (#8837331)

Valamivel kevésbé kohezív a kapcsolat, és többet teljesít az entitásokra vonatkozó enciklopédikus tudás, ha a dolgok közötti reláció nagyon általános, sematizált (például 'létezés'). A (2) példában (*napilapok nemzeti tulajdonban*) a viszony úgy explikálható, hogy 'a napilapok a nemzet tulajdonában / nemzeti tulajdonban vannak/állnak'. A létezését/aktualizációját pedig eleve feltételezzük azoknak a dolgoknak/eseményeknek, amelyekkel kapcsolatban információt osztunk meg, de azt is tudjuk, hogy a napilapoknak van kiadójuk, lehet tulajdonosuk stb.¹⁴

A leíró nyelvtanok általában hivatkoznak arra, hogy az appozitív határozós szerkezet jelzős szerkezeté alakítható, sőt ezt a szempontot próbává is teszik (például Keszler 2000: 431). A jelzővé alakítás próbáját transzformációként bírálta Antal (1985: 31–32), de a jelzős szerkezettel vont párhuzamot a viszonyok összevetésében ő is elfogadta (1985: 44). Rákosi (2021: 349, 353) szintén vizsgálja az attributív szerkezettel való kifejezés lehetőségeit.

Az appozitív határozó esetalakját megtartó jelzős szerkezetekben tipikusan *való* (5a) vagy *levő* (6a) kifejezések alkalmazásával valósul meg a jelzősítés, de más participiumi kifejezés is előfordulhat elől álló határozó esetén (például *a csoporthoz történő csatlakozás*). Az *-i* melléknévképzővel is lehetséges az esetrag megtartása, de ez nem tipikus, és nem is gyakori megoldás (7).

¹³ A DOLOG AZ ESEMÉNY HELYETT, lásd a (4)-ben az ÚT AZ ÚTON VALÓ HALADÁS HELYETT fogalmi metonímia (Kövecses 2005: 149) általános alkalmazásán keresztül tudunk magyarázatot adni a nem deverbális főnév appozitív (a (4)-ben lativusi hely-) határozói bővítményére. Ezeknek a viszonyoknak a metonimikussága nem feltétlenül reflektált, lehet nagymértékben begyakorlott. A metonimizációnak ez a begyakorlottsága teszi a szerzőket szkeptikussá a 'tény, cselekvés' és az 'eredmény' főnévi ontológiáknak a *való* szerkezettel történő kidolgozáson alapuló próbájával kapcsolatban is (Szabolcsi–Laczkó 1992: 238–239).

¹⁴ A hasonló parafrázisok mutatják azt, hogy az appozitív határozók esetében problematikus a viszony önálló állításként való megjelenítése (vö. Komlósy 1992: 447), mert nem világos, hogy milyen kifejezés lehet még része az állításnak, milyen formulálás felel meg meg annak, hogy a szerkezetek ugyanazt a viszonyt konstruálják meg. Például a *darabokra* predikatív határozó esetében a *Kati darabokra törte a játékeit*. mondathoz rendelhető állítás Komlósi (1992: 448)-ban: *A játékek darabokban vannak*.

- (5) a. A kisebbségi csoporthoz való tartozás kérdésében nyilatkozatra senki sem kötelezhető. (#50246894)
b. *A tartozás kisebbségi csoporthoz van.
- (6) a. Ez egy magántulajdonban levő jacht! (#2115301)
b. A jacht magántulajdonban van.
- (7) a. „[...] jobb ember nem használ, mondjuk, villát a szőlőszemek fürtrőli leszedéséhez.” (Nádasdy 2003: 97)
b. *A leszedés a fürtről van.

Az (5a), (7a) típusú jelzős szerkezetekben a *való* kifejezéssel, ill. az *-i* melléknévképzővel jelölt műveletet Lengyel (2000: 254–255) „melléknévesítés”-ként említi,¹⁵ de a jelzősítés megfélelőbbnek tűnik.¹⁶ Érveink között szerepel, hogy

- a művelet alapvetően szintaktikai szóképzés abban az értelemben, hogy a kifejezés szintaktikai szerepének, azaz a függőségi viszonyhálózatba való bekapcsolódás lehetséges módjának a megváltoztatására irányul;
- ennek megfelelően a tipikus, szóalapú képzéstől eltérően nem vagy nem csak szóalapú, hanem szerkezetalapú is lehet (például az *elnyomottak mellett való kiállását* (#689154325));
- és ezért közel áll más, tipikusan jelzősítő szóképzési művelethez:¹⁷ az *-i* (például *megköszönte Románia NATO melletti kiállását*

¹⁵ A melléknévesítés, illetve jelzősítés korábban is felbukkant már a szakirodalomban (lásd ehhez M. Korchmáros 1974: 313).

¹⁶ Ezt a jelenséget A. Jászó Anna (1991) ómagyar adatokra vonatkozóan is megfigyelte és leírta: „A *való* igenévnék a kései ómagyarban igen gyakori képzőszerű szerepe, melynek révén a határozó alakú jelzőt formailag is jelzővé változtatja” (A. Jászó 1991: 330). A jelzővé alakító funkciót a korai és a kései ómagyarban egyaránt adatolja, és létigei participiumi alakok közül (*való : levő ~ lévő*) egyértelműen inkább *való*-ra tartja jellemzőnek (A. Jászó 1992: 426).

¹⁷ Természetesen más, szószerkezetekhez kapcsolódó képző is létezik, de a szerzők az érvelésben csak a jelzősítő funkción alapuló kapcsolatokat kívánták bemutatni. Jellemző azonban például a „főnevesítő” -*As* képző esetében is (Kiefer 1998: 230) a szintaktikai kapcsolódás megváltoztatására irányuló szóképzési művelet. Ugyanakkor az appozitív határozós szerkezet esetében a Kiefer (1998)-ban olvasható leírás sem érvényesül teljesen mértékben, mivel a szerkezetalapú képzés esetében a képzőnek az utolsó hangzó elemhez kellene kapcsolódnia (például *csoporthoz tartozik*: (*csoporthoz tartozás*), míg az (1)-ben *tartozás* egy *csoporthoz* megvalósulás van), vagyis a képzés művelete után kellene a határozónak az alaptagot követő pozícióba kerülnie. A jelen megközelítés ezt elveti, mert nem vezeti le az appozitív határozós konstrukciót más szerkezetekből, hanem konstrukciós sémaként kezeli (Langacker 2008: 167; Tolcsvai Nagy 2017: 211). Ez azonban nem zárja ki azt, hogy a konstrukciók viszonyban álljanak egymással az alternatív nyelvi konceptualizáció eseteiként (Langacker 2008: 30, 43; Tolcsvai Nagy 2017: 213).

(#162243887)), az $-Ű \sim -jŰ$, valamint az $-(V)s$ képző viselkedéséhez (lásd Kiefer 1998: 230),¹⁸

- a művelet eredménye tipikusan nem egy a melléknév sémájába illeszkedő derivátum,¹⁹ szófajilag nem is feltétlenül melléknév. Például az ún. névutómelléknév (például *melletti*) esetében a képzés a névutós szerkezet ($N\text{ mellett}$) jelzősítését végzi el.

Mint a (7) példán is látható, az $-i$ képző jelzősítő funkciója annyira produktív, hogy a képzési művelet ragos főnévi „alapszóval” is lehetséges.

Lengyel (1996, 2000: 254–255) egy olyan megkülönböztetést is bevezet, amelyben az (5a) *csoporthoz való* szerkezete összetett jelző, amelyben a *való* kopula.²⁰ A kopula a *csoporthoz* határozót jelzősíti, ezért a függőségi viszonyhálózatban azzal egyetlen csomópontot (összetett mondatrészt) alkot. A *való* tehát ebben a konstrukcióban Lengyel (1996, 2000) szerint kopula, nem igenév, ezért igévé nem lehet alakítani (lásd 5b). Ezzel szemben a (6a) *magántulajdonban levő* kifejezése szintagma, abban a *levő* nem kopula, hanem igei alapszavú participium, amelynek saját bővítménye a *magántulajdonban* állapothatározó (lásd 6b).²¹

Mivel általánosan elfogadott szempont a magyar nyelvészeti szakirodalomban, hogy az appozitív határozók jelzős szerkezetté alakíthatók, továbbá a lexikális megoldások közül a *való* a legerőteljesebben grammatizálódott jelzősítő „eszköz” (vö. A. Jászó 1991: 330), a dolgozat szerzői erre építik korpusz-

¹⁸ Bár a származékok képzőnként eltérő valószínűséggel predikatív használatúak is lehetnek, az attributív használat jellemzően gyakori, az $-Ű \sim -jŰ$ képző és a *való* segédszó esetében pedig ez a tipikus.

¹⁹ Lásd Palágyi 2022a: 98.

²⁰ Pontosabban (kopulatív) segédigenév (Lengyel 2000: 252).

²¹ Hasonlóan járt el M. Korchmáros (1974) is. Bár a *való* és a *levő* kifejezést egyaránt jelzősítőnek tartja, a *levő*-t a létige igenévi alakjaként határozza meg, míg a *való* igenévi használatát a mai magyarban már kiveszetteknek vagy archaikusnak ítéli. A jelen dolgozat azt nem vizsgálja, hogy elkülönül-e, és ha igen, milyen mértékben különül el a *való* és a *levő* használata a hasonló, jelzősítő szerkezetekben, mivel ez nem lehetséges nyelvtchnológiai eszközökkel, az igévé alakíthatóság ellenőrzése kutatói introspekciót igényel, és sokszor bizonytalan (lásd M. Korchmáros 1974: 317–318). Az adatok kézi elemzésénél azonban nem tévesztjük szem elől ezt a szempontot sem, és hangsúlyozzuk azt is, hogy a szerkezetben lévő főnévi kifejezés ’esemény’ vagy ’dolgozatosság’ jelentése is hatással van arra, hogy milyen mértékben alkot egységet a szerkezet, ill. milyen mértékben tűnik a függőségi viszony alapján felbonthatónak.

alapú vizsgálatukat,²² és az N2.CASE *való* N1.NOM²³ konstrukció adataiból kiindulva vizsgálják korpuszalapúan a potenciális N1.NOM N2.CASE appozitív határozós szerkezeteket. A kutatás fő kérdése az, hogy az appozitív határozós szerkezet konstruálási módként milyen mértékben tekinthető az N2.CASE *való* N1.NOM szerkezet alternatívájának.

Az appozitív határozó a hagyományos leíró megközelítésben egy inkább kivételes és sajátos jelenség (Hámori 1954; Keszler 2000), mivel a főnév által jelölt entitást határozó és nem a nominális szerkezetben megszokott jelző részletezi. Ha ez a határozói kidolgozás előfordul, akkor a határozó a főnévi alaptagot követi (a klasszikus példával *A puszta télen*, id. Keszler 2000: 431), és nem megelőzi (**a télen puszta*), mivel a főnév előtti helyzetben a jelző felel meg a várható szintaktikai megoldásnak (*a téli puszta*; vö. Keszler 2000: 431).²⁴ A korpuszadatok alapján azt is vizsgálat tárgyává tesszük, hogy az N2.CASE *való* N1.NOM és az appozitív megfelelők hálózatában van-e helye az N2.CASE N1.NOM, tehát határozói bővítmény és N alaptag szórendű (tehát szomszédos, segédszó nélkül, közvetlen függőségi viszonyt képviselő) konstrukciónak.

1.3. Részösszefoglalás: a kutatás célja és fő kérdései

A jelen kutatás minimális korpusznyelvészeti célja az, hogy az appozitív határozók tényleges nyelvi adataihoz jusson a kijelölt korpuszban (MNSZ2), és így megbízható adatokat szolgáltatson az appozitív határozók empirikus vizsgálata számára (lásd Kugler–Szabó 2024).

Az optimális cél az, hogy lehetőleg minden olyan nominativusi főnévi alaptagú appozitív határozói adathoz hozzájussunk a kijelölt korpuszban (MNSZ2), amelynek van *való* konstrukciós alternatívája (N2.CASE *való* N1.NOM). Az optimális cél teljesülése esetén a következő kérdésekre keressük a választ:

1. Mely főnévi szóalakpárokkal, páronként milyen gyakorisággal adathozható az N2.CASE *való* N1.NOM konstrukció?

²² Lényeges azonban, hogy csak a korpuszvizsgálat tervezését építik erre a szerzők, nem a magyarázatukat. Az utóbbi hálózatos viszonyt feltételez a konstrukciók között, és kiterjedtebben is kezeli az érintett szerkezetek körét.

²³ A dolgozat módszertani részében fogjuk indokolni, hogy miért szűkítettük a vizsgálatot nominativusi főnévi alaptagra.

²⁴ Ismét utalunk a szórenddel kapcsolatban Antal (1985: 47) ettől eltérő vélekedésére.

2. Az N2.CASE *való* N1.NOM konstrukcióban adatolt főnévi szóalakpárok milyen gyakorisággal fordulnak elő N1.NOM N2.CASE appozitív határozói szerkezetben?
3. Ugyanezek a főnévi szóalakpárok milyen arányban fordulnak elő N2.CASE N1.NOM konstrukcióban?²⁵

Az 1–3) kérdésre nyelvtechnológiai módszerekkel fogunk választ adni. Az adatok szűrőpróbaszerű ellenőrzésével és elemzésével megkísérülünk képet alkotni az alkalmazott módszertan megbízhatóságáról is.

A továbbiakban bemutatjuk az alkalmazott módszertant (2.), az MNSZ2-ben (v2.0.5) végzett adatkeresési, lekérdezési folyamatot az első próbálkozásoktól (2.1.) a részeredményekkel járó, kontextusalapú keresésig (2.2.), majd ismertetjük az eredményeket (3.).

2. A vizsgálat módszertana

A 2.1. vázolja azt az előzményt (Kugler–Szabó 2024), amelynek a nyomán a szerzők (Ernhöffer, Kugler és Kurucz; még nyelvtechnológus támogatása nélkül) keresést végeztek az MNSZ2-ben, és röviden leírja a mintavételezési próbálkozások tapasztalatát. A 2.2. pontban olvasható vizsgálat Indig Balázs közreműködésével zajlott, ő végezte az N2.CASE *való* N1.NOM lekérdezéstől a korpuszadatok kezelését és a kutatási célnak megfelelő feldolgozását.

2.1. Előzmények és mintavételezési próbák

Kugler és Szabó (2024) a többértelműséget kívánta vizsgálni, és ehhez alaptagként a *bizalom* nominativusi főnévi szóalakkkal, illetve egy másik lekérdezésben appozitív határozóként a *gyerekekkel* instrumentalisi főnévi szóalakokkal²⁶ indított keresést a valószínűsített appozitív szerkezetekre olyan beállítással, hogy a *bizalom* alaptagot inessivusi bővítmény kövesse, a másik lekérdezésben a *gyerekekkel* bővítményt pedig a nominativusi alaptag megelőzze 1–2 ablakban. Az alaptag nominativusi esetalakjának a rögzítése az

²⁵ A *való* konstrukciónak az esetében figyelt fel Indig Balázs arra, hogy a határozó akár a jelzősítő kifejezés nélkül, közvetlenül is kapcsolódhatna a nominális profilját meghatározó főnévi szóalakhoz (például *kábítószerral való visszaélés : kábítószerral visszaélés [gyanújával]*).

²⁶ Ezt a keresést Rákosi (2003) inspirálta; általános szempont volt, hogy a határozó vonzat és szabad határozó egyaránt lehessen.

alapértelmezésre támaszkodik. Az alapértelmezés szerinti főnévi szóalak egybeesik az alapalakkal, tehát az alaki jelölletlenség reflektál a könnyű hozzáférésre.²⁷ Ezt a szempontot a jelen vizsgálat is érvényesíti.

Kugler és Szabó (2024) általános, mindkét mintára érvényes megállapítása az volt, hogy az appozitív határozói konstrukció azonosítása csak olyan szőrendben vetődött fel, amikor a határozó szomszédos volt a nominálissal.

Az első korpuszalapú keresések támaszkodtak Kugler és Szabó (2024) javaslatára, N2.INS és N2.INE appozitív határozóra irányultak, a részletes keresésben beállítottuk, hogy az appozitív határozó nem lehet birtokos személyjeles. A szépirodalmi alkorpuszt kizártuk.²⁸ Szűrést végeztünk L1–2 ablakban N.NOM szóalakra.²⁹ Az appozitív határozó jellemzően névelős főnévi szóalak, ez indokolta az ablak méretét. A találatok száma így is feldolgozhatatlan méretű volt, ezért az újabb lekérdezésekben beállítottuk, hogy L1 pozícióban névelő álljon (1. táblázat). A szűrési beállítás adatvesztést okozott, mert az L1 helyzetű főnevek és a jelzős appozitív szerkezetek (például *a kezében ébenfekete mappa vérszín betűkkel*) nem kerültek be a konkordanciába.

1. táblázat. Részletes keresés kétszeres szűréssel N.INS és N.INE határozókkal

L2	L1	nodus	a találatok száma
N.NOM	ART	N.INS	81 180
N.NOM	ART	N.INE	184 712

²⁷ Könnyű hozzáférésen azt értjük, hogy az alanyi viszonyban tipikusan az ige által jelölt folyamat elsődleges szereplője reprezentálódik. Ez a magyar főnévi szóalakok gyakorisági viszonyai szerint jellemzően a nagy gyakorisággal is együtt jár (más esetalakok viszonyában). Vö. Ladányi 2017: 504, 521; Tolcsvai Nagy 2017: 320; Palágyi 2022a: 100.

²⁸ A lektori hozzászólás rámutat arra, hogy ez a lépés magyarázatot kíván. A szépirodalmi alkorpusz kizárásával is követtük Kugler és Szabó (2024) eljárását, és mindkét vizsgálatban azért, mert tapasztalatunk szerint a szépirodalmi alkorpuszból származó adatok (a poétikai funkcióval összefüggésben) sok esetben megnehezítik a vizsgált szerkezet megbízható azonosítását, elemzését. Ha sikerül kifejleszteni az appozitív határozós szerkezetre jól működő lekérdezést, akkor a konstrukció diskurzustípusokkal, illetve nyelvhasználati szinterekkel való összefüggése is kutathatóvá válik, és feltétlenül érdemes lesz kiterjeszteni a vizsgálat hatókörét is.

²⁹ Ez azt jelenti, hogy a határozót (nodust) megelőző 1. vagy 2. token N.NOM szóalak.

Mivel a névelő helyén (L1) egyéb kifejezések is előfordultak (például *mind(en)*, *holmi*, *némi*) az ezekre indított negatív szűréssel dolgoztunk a konkordancia további megtisztításán.³⁰ A találatokba betekintve elvetettük a mintavételezést, mert nagyon kevés volt a releváns adat, a határozó ugyanis többnyire nem a nominálishoz tartozott. Az appozitív határozói előfordulások jellemzően igealak nélküli, felsorolásszerű szövegrészekben, címekben bukkantak fel (lásd (8)–(9)).

- (8) d) a szabad kapcsolattartás az anyaországgal és más közösségekkel (#49326355)
 (9) 9. cikk (az anyanyelv használata az oktatásban) (#49681521)

A kudarcra ítélt N1.NOM N2.CASE lekérdezések után Kurucz Réka azt javasolta, hogy Sass Bálint kontextusalapú keresés elvét kövessük (Sass 2019, 2022: 611–612), és induljunk ki az N2.CASE *való* N1.NOM konstrukció megvalósulásából. Ez a javaslat gyümölcsözőnek bizonyult. Először megtartottuk a szűkítést instrumentalisi és inessivusi főnévi szóalakokra, de elvégeztük a keresést mind a *való*, mind a *l[eé]vő* jelzősítőkkel (2. táblázat).

2. táblázat. Keresés *való* és *l[eé]vő* jelzősítőkkel N.INS, N.INE határozókkal

L1	nodus	R1	a találatok száma
N.INS	<i>való</i>	N.NOM	25 912
N.INE	<i>való</i>	N.NOM	8492
N.INS	<i>l[eé]vő</i>	N.NOM	422
N.INE	<i>l[eé]vő</i>	N.NOM	52 825

A *való* illetve *levő/lévő* jelzősítőket nóduszként kijelölő, L1-en N.INS vagy N.INE szóalakokra, R1-en pedig nominálisra szűrő lekérdezéssel nyert konkordanciák értékelése során többszintű gyakorisági listákkal dolgoztunk. Ezeket úgy kértük le, hogy az appozitív határozói konstrukciónak megfelelő sorrendben az 1. szinten az R1-en található N1.NOM, a 2. szinten az L1-en található N2.CASE szóalakok összesített gyakorisági adataival kiadják azokat a szópárokat, amelyek így együtt a legnagyobb gyakorisággal fordulnak elő az adott konkordanciában.

³⁰ A konkordanciák exportálásához nyújtott segítséget köszönjük Sass Bálintnak.

A N2.CASE *való* N1.NOM konstrukciók határozóit tekintve a N2.INS *való* N1.NOM esetében a társas állapotok jelölése a legjellemzőbb, a partner N2-beli kidolgozásával, például *barátokkal való kapcsolattartás* (695), *szervezetekkel való együttműködés* (76), ezenkívül körülményként dolgozódnak ki olyan entitások is, amelyekre az N1 nominálisban megadott folyamat, esemény irányul, például *gyerekekkel való foglalkozás* (67), *bőrrel való érintkezés* (42), *érzelmeikkel való játék* (9), és ilyenek a *visszaélés* változatos adatai is, például *kábítószerral való visszaélés* (813), *hatalommal való visszaélés* (598).

A N2.INE *való* N1.NOM konstrukció határozóinak legtipikusabb képviselőit a 'megvalósulási színtér' jelentés fogja egybe, ez változatosan konstruálódhat meg, többek között humán szervezeteken vagy eseményeken keresztül is, például *archívumokban való kutatás* (45), *bűnszervezetben való részvétel* (60), *kupákban való szereplés* (32), *versenyben való helytállás* (11), *öngyilkosságban való közreműködés* (15), legjellemzőbben ilyenek a *részvétel* nagy típus- és példánygyakorisággal rendelkező adatai, például *játékokban való részvétel* (156), *eljárásban való részvétel* (47). Egyes N2.INE határozók a folyamat platformját, médiumát, például *levélben való szavazás* (46) vagy manifesztálódását adják meg, például *nemzetben való gondolkodás* (20).

Az N2.INS *való* N1.NOM és az N2.INE *való* N1.NOM konstrukció leggyakoribb szópárjainak listáiban jellemző bizonyos N1.NOM szóalakok ismétlődése különböző határozók mellett, például az INS esetében *visszaélés*: *kábítószerral* (813), *hatalommal* (598), *adatokkal* (251); *összhang*: *törvénnyel* (37), *jogszabályokkal* (21), *egyezménnyel* (19); *kapcsolat*: *Istennel* (76), *országokkal* (28), *természettel* (19); az INE esetében *részvétel*: *játékokban* (156), *programban* (127), *képzésben* (77); *szereplés*: *kupákban* (32), *klubban* (21), *osztályban* (15); *hit*: *Istenben* (53), *életben* (8), *összeesküvés-elméletekben* (7). Ez az adott folyamat típus nagyfokú begyakorlottságáról tanúskodik, amelyhez kapcsolódóan az N2.CASE a jelzőkre jellemző alkategorizálással dolgozza ki a megegyező N1.NOM által egymással összefüggő típusokban ugyanannak az eseménynek mint kategóriának a különböző alkategóriáit.³¹

Mindkét gyakorisági listában nagy, az N2.INS *való* N1.NOM adataiban kiemelkedő az N1.NOM komponens esetében az -*As* képzős konstrukció aránya. Ez a megfigyelés felvetette annak a szükségességét, hogy vizsgáljuk, pontosan milyen az -*As* képzős és az egyéb N1 alanyesetű névszók aránya, valamint

³¹ Ezek a gyakorisági adatok azonban nemcsak a vizsgált konstrukciót, hanem részint a korpuszt is jellemzik: bizonyos kifejezések nagy gyakoriságú adatolhatósága az MNSZ műfaji sajátosságaiból adódik.

hogy milyen változatosság figyelhető meg a nem -*Ás* képzős N1-ek esetében. Ezt az R1 pozícióra végzett pozitív illetve negatív szűréssel kérdeztük le, mindkét esetben a [word="*: [ái]s. * [ái]s [oe]k"]³² CQL-el.

3. táblázat. Az -*Ás* képzős és nem -*Ás* képzős N1.NOM előfordulások arányai a N2.INS *való* N1.NOM és N2.INE *való* N1.NOM lekérdezésekben

konstrukció	találat	%	konstrukció	találat	%
N2.INS <i>való</i> N1.NOM	25 912	100,00	N2.INE <i>való</i> N1.NOM	8492	100,00
N2.INS <i>való</i> N1.NOM- <i>Ás</i>	20 391	78,69	N2.INE <i>való</i> N1.NOM- <i>Ás</i>	4394	51,74
N2.INS <i>való</i> N1. NOM-non <i>Ás</i>	5521	21,31	N2.INE <i>való</i> N1.NOM-non <i>Ás</i>	4098	48,26

Az -*Ás* képzőnek szentelt kitüntetett figyelmet a nagy gyakoriság mellett az alapozza meg, hogy deverbális konstrukcióként az igei alapra jellemző körülményviszonyok a névszóban is leképeződhetnek, ami konstrukciós alapon magyarázza a névszóknál atipikus körülményviszonyok kidolgozódását. Erre számos példát látunk mind az INS (például *visszaélés*, *gazdálkodás*, *szembenézés*, *együttműködés*), mind az INE (például *kételkedés*, *gondolkodás*, *tobzódás*) -*Ás* képzős főnevekre illeszkedő konkordanciáinak adataiban.

Az N2.INS mellett N1.NOM-ként az -*Ás* képzős konstrukcióra nem illeszkedő, egyéb alanyesetű szóalakokat listázó („non*Ás*”) konkordanciában más deverbális konstrukciók is adathozhatóak, például *kapcsolat*, *konzultáció*, *győzelem*, a deverbális konstrukciók mellett viszont nagy arányban jelennek meg olyan főként ’esemény’ jelentésű főnevek (például *vita*, *játék*, *munka*, *harc*, *konfliktus*, *verseny*), amelyeknek szintén kidolgozódhat saját, körülményjellegű bővítményük. Az N2.INE szerkezeteinek „non*Ás*” konkordanciáját a szintén deverbális *részvétel* N1.NOM szóalak uralma jellemzi. Kisebb arányban ugyan, de itt is megjelennek deverbális szóképzési sémába nem illeszkedő ’esemény’ jelentésű főnevek (például *ima*, *munka*). A nem (vagy nem nyilvánvalóan) deverbális adatok (mind az INS mind az INE esetében)

³² A lekérdezés így a többes számú -*Ás* képzős megvalósulásokat együttesen kezeli az egyes számakkal. Köszönjük lektoraink javaslatát az elegánsabb CQL-formulálásra (lásd a lektori hozzászólásban), az adott vizsgálati szakaszban a részfejezet szerzői (Ernhöffer Aliz, Kurucz Réka és Kugler Nóra) ezzel dolgoztak, ezért csak ezt tartjuk etikusnak közölni.

azért is relevánsak az appozitív határozói konstrukció szempontjából, mert ezeknek az N1 adatai jól tudnak működni appozitív konstruálással is (például *tűzzel való játék ~ játék a tűzzel*). Ezeknek a lekérdezéseknek a tanulságait jelen vizsgálat a három konstrukció összevetésében (3.1., 3.2.) még nem használta föl, további vizsgálatok és különösen egy empirikus kísérlet megtervezése számára azonban számottevőek lehetnek.

A N2.CASE *való* N1.NOM konstrukció N2.INS-re és N2.INE-re koncentrálnál próbakeresései és a leggyakoribb szópárokat kiemelő többszintű gyakorisági listák túlnyomórészt olyan adatokat eredményeztek, amelyeknek lehetőségeik az appozitív megvalósulásaiak is, bár ezek nem mind tűnnek megszokottnak.³³ A lekérdezések tapasztalatai arról tanúskodnak, hogy az N2.CASE *való* N1.NOM alternatív konstrukcióként alkalmas kiinduló keresés egy az appozitív határozós konstrukció elemzését célzó vizsgálat számára.

Az N2.INS *[[eé]vő* N1.NOM, N2.INE *[[eé]vő* N1.NOM konstrukciók esetében jellemzően működött a létigés N.INS *van* (például *a lánnnyal van; tekintettel van*), illetve az N.INE *van* parafrázis (például *előzetesben van; válságban van*). Tehát ez a konstrukció valóban nem annyira absztrakt, mint a *való* mintázata. Ugyanakkor a határozók az N2.INE *[[eé]vő* N1.NOM konstrukcióban nemcsak, sőt nem is elsősorban 'valahol levést' jelölnek (például *közelében, kézben/kezeiben*), hanem inkább állapotokat.³⁴ Az 1000 fölötti példánygyakoriságú N2.INE szóalakok a gyakoriság csökkenő sorrendjében: *tulajdonában* (3329), *tulajdonban* (3040), *helyzetben* (2747), *érvényben* (2503), *folyamatban* (2361), *hivatalban* (1883), *állapotban* (1237).

Az instrumentalisi határozók részben társas állapotokat jelölnek (például *lánnnyal* (10a), *istennel*), de még ezeknek az esetében sem egyértelmű, hogy természetes volna az appozitív határozó (10b), az értelmezős szerkezet (10c) inkább valószínűsíthető alternatíva.

³³ Például *a platínával való összevetés ~ ?összevetés a platínával; közveszéllyel való fenyegetés ~ ?fenyegetés közveszéllyel*.

³⁴ Köszönjük a lektori hozzászólásnak azt az értékes szempontját, hogy az inessivus eset poliszémiaja, a BENNE LEVÉS viszony absztrahálódása és az állapotok tartályszerű leképezése magyarázza az adatolt szemantikai viszonyokat és azok összefüggését, továbbá az inessivusi esetalak gyakoriságát is. Ugyanakkor az inessivus erősen absztrakt viszonyokat is konceptualizál (vö. *kérdésekben/ügyekben való döntés*), és egyelőre nem rendelkezünk megbízható ismeretekkel a típusjelentések megvalósulási arányaival kapcsolatban. Továbbá például a RAJTA LEVÉS viszony és a FELÜLET séma esetében is megfigyelhető hasonló kiterjesztés a téri viszonyból az állapotokra (például nemcsak *éber állapotban*, hanem *éberen van* is lehetséges). Ettől függetlenül intuitívnak tartjuk a szemantikai motivációt az inessivusi alak gyakoriságával kapcsolatban.

- (10) a. többször meg is lopta az éppen a lánnyal *lévő* kuncsaftokat (#604236477)
 b. ??meglopta a kuncsaftokat a lánnyal
 c. meglopta a kuncsaftokat, a lánnyal *lévőket*

Az N2.INS $[l[e\acute{e}]v\acute{o}]$ N1.NOM esetében tipikus (és nagyobb példánygyakoriságú) az elvontabb viszony (például *hatással* (181), *tekintettel* (50), lásd (11)).³⁵ Ezeknek az esetében még kevésbé várható az appozitív konstruálás.

- (11) [...] az életmű csúcscsaira tekintettel *lévő* Mátyás[-könyvet] (#25663346)

A fenti példák azt a régi szakirodalmi megállapítást is jól illusztrálják, hogy a $[l[e\acute{e}]v\acute{o}]$ -konstrukció megvalósulásai nem volnának természetesek az ige-név nélkül (**az éppen a lánnyal kuncsaftok*, **az életmű csúcscsaira tekintettel Mátyás-könyv*; vö. M. Korchmáros 1974).³⁶

A próbakérdés megerősítette, hogy a *való* konstrukció lényegesen gyakoribb, erősebben absztrahálódott a *levő/lévő* adataihoz képest. Az erősebben absztrahálódott N2.INS a *való*, míg a jellemzően állapotot megkonstruáló N2.INE a $[l[e\acute{e}]v\acute{o}]$ -konstrukcióiban volt lényegesen gyakoribb. A *való* konstrukció határozóinak az esetében várható inkább az, hogy azok appozitív szerkezetben is megjelennek,³⁷ illetve a *való* konstrukcióinak az esetében a határozó akár a jelzősítő kifejezés nélkül, közvetlenül is kapcsolódhat a nominális profilját meghatározó főnévi szóalakhoz (például *kábítószerral való visszaélés* : *kábítószerral visszaélés [gyanújával]*). Indig Balázs segítségével

³⁵ A típusok száma alapján nem elhanyagolható a mérték kifejezése se (például *méterrel* (6), *számmal* (3), *októval* (1)).

³⁶ Visszaautalunk a korábbi lábjegyzetre a szempont bizonytalanságával kapcsolatban (Korchmáros 1974: 317–318), és megemlítjük, hogy például a *felelőse tulajdonképpen mégsem csak a nikotin, hanem a füstben lévő szénmonoxid, amely a vérbe kerül* (#388206125) adat esetében elvégezhető az igévé alakítás (vö. *a szénmonoxid a füstben van*), de ez nem felel meg jól a szerkezet specifikáló értelmének ('a füstben lévő, a füsttel belélegzett szénmonoxid'). Lásd még az *energiaszektorban lévő fejlesztések forrását szolgálja* (= *a fejlesztések az energiaszektorban vannak*). Az N2.INS konstrukciókban még több ilyen bizonytalanságot tapasztaltunk. Másfelől a *való* szerkezeteiben is lehetségesnek mutatkozott az igei szerkezettel történő megkonstruálódás is, és nem csak a nyilvánvalóan igenévének tartható találatok esetében (például [vmelyik] *ajtónál való felszállás* – *a felszállás az első ajtónál van; vmire való tekintettel* – *vki tekintettel van vmire*).

³⁷ A. Jászó (1992) a kései ómagyarban is a *való* dominanciáját látta a jelzősítésben („A kései ómagyar korban kettőjük közül a *való* volt az uralkodó” A. Jászó 1992: 426).

olyan lekérdezést végeztünk, amely tetszőleges (de nem nominativusi vagy accusativusi) főnévi szóalakot is megenged az n2 esetében (lásd a következő, 2.2. alpontban).

2.2. Kontextusalapú keresés és adatfeldolgozás

A kontextusalapú kereséshez és a szerkezetek páronként való összehasonlításához az eredeti lekérdezés CQL alakjából indultunk ki:³⁸

```
[msd="^FN.+ $" & msd!="^+.NOM$" & msd!="^+.ACC$"] [word="^való$"]
[msd="^*.MN.+NOM$" | msd="^SZN.+NOM$"]{0,2} [msd="^FN.+NOM$"]
```

Az eredményben az első és utolsó szó rendre megfelel az n2 és n1 pároknak. Az egyes párok vizsgálatához viszont páronként két lekérdezést szükséges elvégezni, amelyeket automatizáltan hajtottunk végre a gyakori elemektől a ritkábbak felé haladva, a lekérdezések nagy száma miatt csak a 10-nél gyakoribb eseteket figyelembe véve.

Az alábbi két CQL lekérdezéssel vizsgáltuk az egyes szó párok sorrendjének előfordulásait:

```
[word="n2"] [msd="^DET.+ $" ]{0,1} [(msd="^*.MN.+NOM$" & lemma!="
való") | msd="^SZN.+NOM$"]{0,2} [lemma="n1" & msd="^+.NOM$"]
```

```
[word="n1"] [msd="^DET.+ $" ]{0,1} [(msd="^*.MN.+NOM$" & lemma="
való") | msd="^SZN.+NOM$"]{0,2} [word="n2"]
```

Például *Unióhoz (való) csatlakozásunk* vs. *csatlakozás az Unióhoz, kábítószerral (való) visszaélés* vs. *visszaélés kábítószerral*.

Az összes n2 – n1 párra egyenként lefuttatva mindkét lekérdezést és az eredményeket kiexportálva az egyes gyakoriságpárokat táblázatba gyűjtöttük, hogy összehasonlíthatóvá váljanak. Lektori javaslatra a gyakoriságokat két változatban is megadjuk: az elsőben a közvetlenül a korpuszból nyert gyakoriságokat, a másodikban a többször előforduló mondatok szűrésével nyert gyakoriságokat vizsgáljuk. A duplikátumszűrt adat így tartalmaz

³⁸ Lektori javaslatra kiegészítettük a melléknévi címkék definícióját, hogy tartalmazza a felsőfokú és igekötővel kezdődő melléknéveket is. Ez három olyan új esetet eredményezett, melyeknek 9-nél nagyobb a gyakorisága. A dolgozat csak a javított megoldást közli, a kézirat első változatának keresőkifejezése a lektori hozzászólásban megtalálható.

olyan egyszer szereplő kifejezéseket, melyek mondat szinten ismétlődnek a korpuszban, ezt a tulajdonságot nem vizsgáltuk az első adathalmazban. A kapott mondatok kézi ellenőrzésére az adatmennyiség miatt nem volt lehetőség. Az exportált adatok segítségével viszont szükség esetén szűrőpróbaszerű ellenőrzések és további vizsgálatok végezhetők.

3. Eredmények

A 3.1. pontban áttekintő módon mutatjuk be azokat a számszerűsíthető eredményeket, amelyekhez kézi adatellenőrzés és elemzés nélkül jutottunk. Ezt követően a 3.2. alponthoz megvizsgáljuk a legalább 30 előfordulással rendelkező N2.CASE *való* N1.NOM típusokat, és választ keresünk arra, hogy ezek milyen viszonyban vannak az appozitív (N1.NOM N2.CASE) és az N2.CASE N1.NOM konstrukcióval. Az adatok által meghatározott lehetőségeken belül bizonyos mértékű diszkusszió az eredményeket bemutató alpontokban is helyet kapott.

3.1. Gyakorisági adatok

A 2.2. alponthoz bemutatott eljárással 1628 típust, a javított CQL-lel és duplumszűréssel, limit nélkül pedig 1685 típust, azaz bizonyos N2 és N1 főnévi szóalakpárt tartalmazó N2.CASE *való* N1.NOM konstrukciót sikerült azonosítani (lásd alább, az 5. táblázatban).

A típusok elkülönítéséhez szükséges néhány technikai megjegyzést fűzni. A típusok szóalakokat képviselnek (például *hatalommal való visszaélés* (609);³⁹ *hatalommal való visszaélések* (12)), a párokat nem lemmatizáltuk. A kereső és a gépi feldolgozás sajátosságaiból adódóan megkülönböztettük azonos morfémaszerkezetű főnévi szóalak esetében is az írásmódban jelentkező különbségeket, tehát például külön típusként tartjuk számon a *hatalommal való visszaélés* (609) és a *Hatalommal való visszaélés* (35) kifejezéseket.⁴⁰ A nyilvánvalóan hibás találatokat (például *ördögtől való a* (59)) ebben a szakaszban még nem távolítottuk el kézzel.

³⁹ A típusok megadásában a kifejezés mellett zárójelek között látható szám az előfordulások száma.

⁴⁰ A lektori hozzászólás problematizálja ezt a döntést. A célunk az volt, hogy ne veszítsük el a köznév : tulajdonnév megkülönböztetést, de ne legyen szükség elemzői döntésre azzal kapcsolatban, hogy az írásmód különbsége a kutatás szempontjából releváns-e vagy nem (megkülönböztet-e típust vagy nem).

Az N2 főnév esetalakja (msd kódja) alapján a *való* konstrukció határozói viszonyainak a jelölésében a 4. táblázatban látható gyakorisági adatokat kaptuk. Megjegyezzük, hogy az irányhármasság szerinti „lativusi” irányt⁴¹ kifejező esetragok gyakorisága szempontjából némi eltolódás lehetséges a ’vmilyen célra rendelt, szánt, illő’ (például *szemétre való műsora* (272); *Kenyérhez való liszt* (25)), illetve a ’vminek megfelelő mennyiségű’ jelentésű (például [2] *télre való hó* (31); [egy] *kötetre való novella* (27)), az igenévi és a melléknévi használat bizonytalan határmezsgyéjén elhelyezhető, nem egyértelműen segédszói használatoknak a konkordanciába való bekerülése miatt.

4. táblázat. Az N2.CASE különböző esetalakjainak gyakorisági viszonyai az N2.CASE *való* N1.NOM konstrukcióban

N2 esetalakja	előfordulások száma		példa
	duplum- szűrés nélkül	duplum- szűréssel	
INS (instrumentalis)	29 320	11 968	<i>kábítószerral való visszaélés</i>
SUB (sublativus)	23 448	9320	<i>pártállásra való tekintet</i>
ALL (allativus)	18 690	12 170	<i>unióhoz való csatlakozás</i>
INE (inessivus)	10 610	4464	<i>programban való részvétel</i>
ABL (ablativus)	8074	2790	<i>házaszálytól való eltérés</i>
SUP (superessivus)	7341	2675	<i>térképen való megjelenítése; telefonon való érdeklődésem</i>
ILL (illativus)	6485	1937	[vmilyen] <i>pontba való szál- lítás</i>
ELA (elativus)	5938	2109	<i>válságból való kilábalás</i>
DEL (delativus)	4356	1118	<i>hitről való tanúságtétel</i>
DAT (dativus)	3682	1647	<i>kornak való megfelelés</i>
CAU (causalis)	1409	385	[a] <i>rendszerért való aggo- dalom</i>

⁴¹ A lativusi (vö. *hova?*) irány itt az ablativustól (vö. *honnan?*) és a locativustól (vö. *hol?*) való megkülönböztetést szolgáló megnevezés. Ide sorolhatnánk lativusi eredete miatt a factivust is, de mint látjuk, ennek megvalósulása a konstrukcióban nem gyakori.

N2 esetalakja	előfordulások száma		példa
	duplum- szűrés nélkül	duplum- szűréssel	
FOR (formalis)	393	292	<i>hatóanyagként</i> való felvétele [vminek]
ADE (adessivus)	298	170	[vmelyik] <i>ajtónál</i> való fel- szállás
FAC (factivus)	191	10	<i>tisztté</i> való kinevezése
TER (terminativus)	163	55	[vminek a] <i>végéig</i> való meg- hosszabbítása [vminek]
TEM (temporalis)	12		<i>leszálláskor</i> való kinyitása [vminek]; [vminek a] <i>kezdetekor</i> való leszokás
ESS (essivus) + <i>képpen</i>	4 1		[vminek a] <i>fedezetül</i> való felajánlása [vminek]; <i>csataképpen</i> való meglopás

A duplumszűrés nem volt alapvető hatással a különböző N2.CASE esetalakok arányára és gyakoriságuk sorrendjére, de az látható a 4. táblázatból, hogy az allativusi főnévi N2 szóalak kismértékben gyakoribb duplumellenőrzés után, mint az instrumentalisi esetalak (mind a példány, mind a típusgyakoriság szempontjából). A duplumok szempontjából szűrt adatsorban nem volt adat olyan N2 szóalakszerkezetre, amelynek ne lettek volna példányai a nem szűrt mintában is. Az allativus és a sublativus nagy részesedésére (ahogy fentebb már jeleztük) hatással lehet az, hogy a vizsgálat szempontjából nem releváns, de a kézi ellenőrzés hiánya miatt a mintába bekerülő, valamilyen elvárásnak/célnak megfelelő; valamihez illő⁴² jelentésű *való* melléknév adatai is ezeknek a kosntrukcióknak a gyakoriságát növelik.

Az instrumentalisi kimagasló gyakorisága (leggyakoribb vagy második leggyakoribb adatszűréstől függően) visszaigazolja Kugler és Szabó (2024) javaslatát, bár annak nem a gyakoriság volt az elsődleges szempontja.⁴² Egyértelműen jelzi a *való* erős grammatikalizálódottságát is, mivel az N2.INS adatok

⁴² A másik keresett alak az inessivus volt, ami szintén gyakori esetalak, a 4. a fenti gyakorisági listában.

tipikusan absztrakt viszonyokat (nem eszközt vagy társas állapotot jelölnek; ehhez lásd még a főnévi párok gyakorisági listáját is a 7. táblázatban).

A folyamatok fizikai körülményeit kidolgozó határozók esetében (lásd a 4. táblázatban például a temporalist és a terminativust) az látható, hogy a *való* és a *levő/lévő* különbsége nem elsősorban abban ragadható meg, hogy a *levő/lévő* inkább igenév, ezért a szerkezeteinek inkább lehetséges a létigét tartalmazó parafrázisa (például *előzetesben lévő* [vki] ~ [vki] *előzetesben van*), hiszen a hasonló szerkezetekben a *való*-konstrukciónak is lehet igei szerkezet az alternatívája (például *leszálláskor való kinyitása* [vminek] ~ [vminek] *a kinyitása / kinyitás leszálláskor van*).⁴³ A lényeges különbség tehát az, hogy a [leé]vő konstrukció tipikusan a szereplők állapotát, statikus szituációt dolgoz ki, míg a *való* konstrukció a folyamatokat, történéseket részletezi a körülmény elaborálásával.⁴⁴

3.2. A háromféle konstrukció megvalósulási arányai

Ebben a fejezetpontban az N2.CASE *való* N1.NOM, az N1.NOM N2.CASE (appozitív határozós) és az N2.CASE N1.NOM konstrukció gyakorisági viszonyaival foglalkozunk.

A duplumszűrés nélküli és a szűrt adatok kézi ellenőrzés nélküli áttekintő adatait az 5. táblázat mutatja be. Kézi ellenőrzés nélkül a gyakorisági adatok

⁴³ A tanulmány korábbi változatában a *történik* igével alkotott parafrázissal illusztráltuk a folyamatjelentést, de a *történik* jelentésszerkezete több szempontból (például temporalitás) is eltér a *van* igétől. Köszönjük, hogy lektoraink hozzászólásukban felhívták a figyelmünket erre a problémára, és értékes megjegyzést fűztek ehhez (lásd Kalivoda és Palágyi hozzászólásában). A tanulmány az eredeti változatában is több ponton kifejtette tette, hogy a *való* erősebben absztrahálódott a vizsgált szerkezetekben, mint a *levő/lévő*, és ennek egyik szimptomája az, hogy kevésbé várható, és jellemzőbben nem lehetséges a szerkezet létigés parafrázisa. Ugyanakkor az is megfigyelésünk volt, hogy bizonyos appozitív szerkezetek nem a *való* vagy a *levő/lévő* jelzősítő konstrukcióval, hanem a *történik* ige igenévi alakjával parafrázálhatók (például *lövöldözés az egyetlen – az egyetlen történt lövöldözés*). A jelen dolgozat azonban azt még nem tudta kidolgozni, hogyan kapcsolódnak ezek a dinamikusabb eseménykonstruálással összefüggő szerkezetek a vizsgált konstrukcióhálózathoz.

⁴⁴ Ahogy a korábbi szakirodalomból is láttuk (például M. Korchmáros 1974, Lengyel 2000), az erősebben absztrahálódott konstrukciókban fel sem vetődik a *való* igenévként történő elemzése, igévé *való* átalakítása (ha mégis, akkor is csak ellenpróbaként). Nagyon lényeges a dolgozat lektorainak az a kiegészítése (lásd a lektori hozzászólásban), hogy ugyanakkor a *való* konstrukció sem mindig dinamikus, ahogy azt a dolgozatban közölt adatok is felmutatják (például a *tartozás, kapcsolat* N1 esetében, vö. 7. táblázat), a *való* konstrukció is reprezentálhat statikus (körülmény jellegű) szituációt. A szerzők köszönik, hogy a lektori hozzászólás kiemeli a konstrukcióról alkotott képnek ezt a részletét, és így hozzájárul az összetett jobb érzékeltetéséhez.

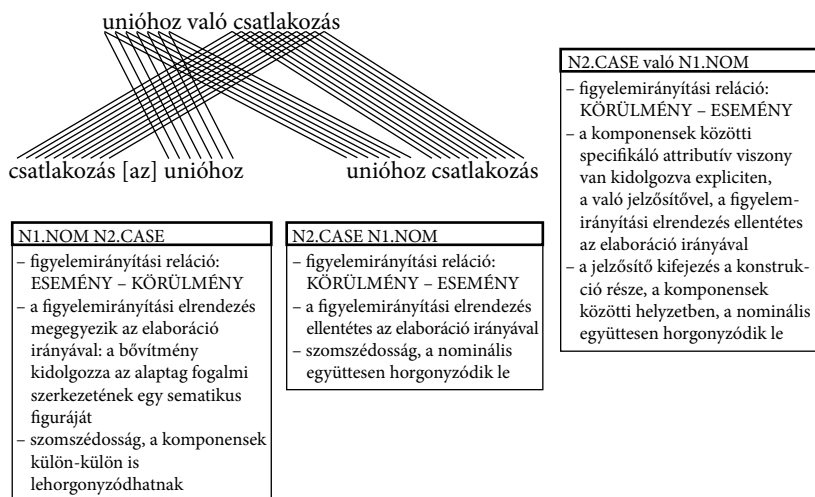
mindkét minta esetében erősen torzítanak, ugyanis mind a *való* melléknéves (például *szemétre való műsora*), mind a hibásan nominativusi alakként jelölt, de esetragos, névutós szerkezettel kapcsolatban álló kifejezések (például N2.SUB *való tekintet[el]*; N2.SUB *tekintettel*) és a nem főnévi kifejezések (például határozott névelő a *mire való a* kifejezésben) adatait tartalmazza. Az N2.CASE *való* N1.NOM típusainak a száma azért tér el a kétféle mintában, mert az első lekérdezéskor csak a minimum 10 előfordulással rendelkező kifejezéseket vettük figyelembe, a későbbinél nem volt ilyen limitáció. Duplumszűrés után az N2.CASE N1.NOM konstrukció típusgyakorisága kevésbé értékelhető áthatónak, de az N1.NOM N2.CASE adatokhoz (amelyek között az appozitív határozós konstrukció megvalósulásaira számítunk) viszonyított aránya így is nagy, a *való*-konstrukció típusai 39,52%-ban *való* nélkül, N2.CASE N1.NOM alakban is adatolhatók, míg az N1.NOM N2.CASE szórendű kifejezéspár típusainak az aránya 24,21%.

5. táblázat. A konstrukciók gyakorisági viszonyai a *való*-konstrukcióhoz viszonyítva

Konstrukció	Duplumszűrés nélkül				Duplumszűréssel			
	típus		példány		típus		példány	
	db	%	db	%	db	%	db	%
N2.CASE <i>való</i> N1.NOM	1628	100,00	50 392	100,00	1685	100,00	31 548	100,00
N1.NOM N2.CASE	392	24,07	10 263	20,36	408	24,21	8389	26,59
N2.CASE N1.NOM	1578	96,92	42 105	83,55	666	39,52	31 322	99,28

Az alábbi (1.) ábrában foglaljuk össze a konstrukciók viszonyát a formai és a szemantikai pólust egyaránt figyelembe véve.⁴⁵

⁴⁵ A leírásban a főnévi kifejezések szomszédosságát úgy értjük, hogy a főnevek közé legfeljebb az adott főnévi láncához tartozó kifejezés (például a főnév jelzői bővítménye) kerülhet.



1. ábra. A három konstrukció kapcsolatrendszere

Az N2.CASE *való* N1.NOM konstrukció mintájának összesen 1628 típusából kiválasztottuk a legalább 30 előfordulással rendelkezőket (377 típus).⁴⁶ Két nyilvánvalóan hibás pár (*mire való a* (63), *ördögtől való a* (59)) kizárásával az elemzett típusok száma 375. A 375 típus összesen 30772 előfordulást képvisel. A három konstrukció gyakorisági viszonyait a 6. táblázat mutatja be.

6. táblázat. A három konstrukció duplumszűrés nélküli gyakorisági viszonyai a legalább 30 előfordulással rendelkezők *való*-konstrukciókhoz viszonyítva

konstrukció	típus		példány	
	db	%	db	%
N2.CASE <i>való</i> N1.NOM	375	100,00	30 772	100,00
N1.NOM N2.CASE (appozitív határozós)	115	30,66	1095	3,55
N2.CASE N1.NOM	196	52,26	9973	32,40

⁴⁶ Ezt a vizsgálatot a duplumszűrés nélküli adatvételezés legalább 30 előfordulással rendelkező kifejezéseinek végeztük el. A lektori javaslat alapján elvégeztük a duplumszűrést, de az elemző részt nem változtattuk meg, mivel a szerkezetek arányai nem változtak meg, inkább az a jellemző, hogy a duplumszűrés után alacsonyabb lett az adatolt szerkezetek előfordulásának a száma.

Meglepő adat, hogy az appozitív határozós szerkezetnél mind a típusok, mind a példányok számát tekintve gyakoribb a főnévi alaptagot megelőző, jelzősítő nélküli, tehát az alaptaggal szomszédos bővítmény. Erdemes azonban óvatosan bánni a számszerűsíthető adatokkal.

Ha megvizsgáljuk az N2.CASE N1.NOM konstrukció kiugró (100 fölötti) előfordulással rendelkező adatait, akkor a 7. táblázatban látható kifejezés-párokat kapjuk.

7. táblázat. A leggyakoribb N2.CASE N1.NOM típusok és előfordulásuk száma

Duplumszűrés nélkül			Duplumszűréssel		
N2	N1	db	N2	N1	db
vele	kapcsolat	3256	vele	kapcsolat	2354
befolyással	üzérkedés	614	velük	kapcsolat	1032
kábítószerral	visszaélés	528	lakáshoz	jutás	837
közügyektől	eltiltás	403	kábítószerral	visszaélés	464
nemzethez	tartozás	291	szabályokkal	összhang	427
joggal	visszaélés	272	befolyással	üzérkedés	425
közveszéllyel	fenyegetés	261	joggal	összhang	330
törvénnyel	összhang	260	nekem	dolog	305
életre	nevelés	250	jogszabályokkal	összhang	287
házzsabálytól	eltérés	208	közügyektől	eltiltás	284
lőszerral	visszaélés	201	szülőföldön	maradás	255
lőfegyverrel	visszaélés	192	nemzethez	tartozás	240
adattal	visszaélés	192	közveszéllyel	fenyegetés	237
levélben	szavazás	184	törvénnyel	összhang	232
csoporthoz	tartozás	172	adattal	visszaélés	229
közösséghez	tartozás	169	munkához	jutás	226
vitára	alkalmasság	156	életre	nevelés	218
nyugdíjra	jogosultság	108	nyelveken	szólás	203
			társadalomra	veszélyesség	201
			erre	utalás	189
			velünk	kapcsolat	188
			kisebbséghez	tartozás	184

Duplumszűréssel		
n2	n1	db
házzsabálytól	eltérés	182
lőfegyverrel	visszaélés	173
lőszerrel	visszaélés	170
arra	tekintet	167
arra	hivatkozás	163
életmódra	nevelés	151
csoporthoz	tartozás	149
joggal	visszaélés	148
információhoz	jutás	148
vitára	alkalmasság	142
közösséghez	tartozás	136
gyűlöletre	uszítás	128
járművezetéstől	eltiltás	126
nekem	hely	126
törvényekkel	összhang	119
bírósághoz	fordulás	109
emberekkel	kapcsolat	102

A 7. táblázat szerkezetei többnyire jogi-közigazgatási terminusok jellemző alaptagokkal (például *visszaélés*, *fenyegetés*, *üzerkedés*), amelyek a cselekmény fő típusát, a bővítmények pedig az alkategóriáit jelenítik meg. Akadnak azonban köznapibb kifejezések is, például az *életre nevelés*.

Igen jelentősen torzított azonban az arányokon a *vele kapcsolat* találat, amely nem a *vele való kapcsolat* jelzősítő kifejezés nélküli „párja”, hanem vagy a (12) példához hasonló valóban határozós szerkezetet képviseli (erre azonban csak 3 adat volt), vagy valamilyen hiba folytán a *kapcsolat* nem nominativusi alakjaival megalkotott kifejezés (például *a vele kapcsolatba kerülő*; *vele kapcsolatban álló/lévő*), amelyben nincs közvetlen függőségi viszony a főnévi, illetve főnévi névmási komponensek között, legnagyobb arányban pedig névutós szerkezet (13). Grammatikalizálódottsága miatt épp az utóbbi dobta meg ilyen jelentősen a találatok számát (nincs más kifejezéspár többezres találatsszámmal).

- (12) A vészjelzés leadása óta nincs vele kapcsolat. (#621629523)
 (13) A néphit vele kapcsolatban azt tartja, hogy [...] (#2916690)

Hasonlóan nem illeszkedik a keresett szerkezetre a *törvénnyel összhang* sem, mivel a találatok nem nominativusi alakban tartalmazzák az *összhang* lemmát (14), amely egy funkciógés (vagy könnyűgés) szerkezet (például *összhangban áll, összhangba hoz*) részeként bővül N2.INS komponenssel. Az azonban kétségtelen, hogy az *összhangban* szóalak közvetlenül is bővül a *törvénnel* szóalakkal úgy, hogy a bővítmény megelőzi az alaptagot (vö. „az adatvédelmi törvénnel összhangban szorgalmazni fogja...” #50132000).

- (14) A javaslat az adatvédelmi törvénnel összhangban áll (#375932538)

Bár az N2.CASE N1.NOM konstrukció gyakori megvalósulásai azt mutatják, hogy a mintázat a jogi-közigazgatási-politikai nyelvhasználati színterekre inkább jellemző, és gyakran terminusértékű, az *életre nevelés* (250), *konszenzusra törekvés* (59), *biztonságra törekvés* (57) típusok jelzik, hogy ennél lényegesen kiterjedtebb lehet a konstrukció alkalmazása, és erősen motiválhatja a begyakorlott használat (lásd még az *Európai Unióhoz / unióhoz / Unióhoz csatlakozás* (68) példát is ehhez).

A 8. táblázat bemutatja azt a képet, amely az adatvételezés alapján a három konstrukció viszonyáról (a *való*-szerkezet viszonyítási pontjából) kirajzolódik. A táblázatban + jel jelöli, ha van, – jel, ha nincs adat. Ha a manuális elemzés nélküli, számszerűsíthető adatokat a fentiek alapján különös óvatossággal kezeljük, akkor a leginkább megbízható összefüggés az, amikor egyik *való* nélküli szerkezetre sincs adat (lásd az utolsó sort). Minden más esetben szükségesnek látszik az adatok kézi ellenőrzése.

8. táblázat. A három konstrukció típusainak (főnévi szóalappárjainak) megvalósulási viszonya

	N1.NOM N2.CASE (appozitív)	N2.CASE N1.NOM (balra bővülő, szomszédos)	típusok száma (db)	típusok aránya (%)
N2.CASE <i>való</i> N1.NOM	+	+	82	22
	–	+	114	30
	+	–	34	9
	–	–	145	39

A típusok (n=375) megvalósulási viszonyai alapján az N2.CASE *való* N1.NOM konstrukció 39%-ban nem alternatívája a másik két konstrukciónak az adott körülményviszonyok megkonstruálásában. A jelzősítő kifejezést nem tartalmazó szerkezetek közül lényegesen több típussal adathozható az elől álló és az alaptaggal szomszédos határozói bővítmény (52%), mint az appozitív (31%). Ha mindkét konstrukció előfordulhat az adott főnévi szóalakpárral (22%), akkor is jelentős gyakorisági eltérés lehet a megvalósulások között (például *kábítószerral visszaélés* (528), *visszaélés* [...] *kábítószerral* (236); [*vmilyen*] *szervezetekkel együttműködés* (40), *együttműködés* [...] *szervezetekkel* (3)), tipikusan az elől álló határozós szerkezet javára (vö. a példánygyakoriságokat bemutató 6. táblázat adataival).

Ettől nem független az sem, hogy több olyan típus van, amelyben csak az elől álló határozó adathozható, az appozitív nem (30%), míg viszonylag csekély (9%) a csak appozitív szerkezetben megjelenő szóalakpár.

Az N2.CASE N1.NOM konstrukciónak a jelen vizsgálattal adatolt gyakorisága teljesen új magyarázatát kívánja meg a konstrukciók viszonyának. A határozónak ezt a szórendi pozícióját (jelzősítő nélkül, az alaptaggal szomszédosan) korábban nem valószínűsítette a szakirodalom. Ugyanakkor az N2.CASE *való* N1.NOM már az ómagyar korban konvencionálódott konstrukció begyakorlott szórendi mintázata nyilvánvalóan előkészítette a talajt a gyakori használatú, egységgé váló kifejezések rövidülésének. Ezt figyelhettük meg a jogi-politikai-közigazgatási szövegek terminológiájában (lásd (15), 7. táblázat).

- (15) Ott kezdeném, ha korrupciós bűncselekmény, vesztegetés, *befolyással üzérkedés* gyanúja merül fel, akkor [...] (#53300043)

A kategóriák (15)-beli formulálása számára nyilvánvalóan kedvezőbb a *befolyással üzérkedés*, mint a *befolyással való üzérkedés*, különösen mivel az egész szerkezet gyakran egy nagyobb lánc része (a (15)-ben a *gyanúja* birtokos jelzős szerkezetének része, de lásd még *vétsége miatt*, *vádjával* stb.).

A gyakorisági adatok alapján tehát erősebb kapcsolat feltételezhető a *való* konstrukció és az elől álló határozós szerkezet (N2.CASE N1.NOM) között, mint a szakirodalom által leírt összefüggésben: a jelzősítő (N2.CASE *való* N1.NOM) és az appozitív határozós szerkesztésmód egymást kiegészítő konstruálási stratégiái között. Az elől álló határozói bővítmény konstrukciója ugyanakkor nem korlátozódik a szakkifejezésekre. Egyrészt a gyakori *való* konstrukciók bármely szintéren redukálódhatnak, másrészt a hétköznapi nyelvi tevékenység részévé váló szakkifejezések (lásd N2.ALL *csatlakozás*, például *unióhoz/Unióhoz*

csatlakozás; N2.SUB *alkalmasság*, például *általános vitára alkalmasság*; *munkára alkalmasság*; *válságból kilábalás* (261), *követelményeknek megfelelés* (40)) is erősíti a mintázat termékenységet (például *multtal szembenezés* (21)⁴⁷).

A *való* konstrukció mellett csak appozitív szerkezetben előforduló szó-párok kisebb részben téves találatok (például a *spórolás szempontjából* (2) névutós szerkezet, nem a jelzősített kifejezés alternatív megkonstruálása), vagy egységként viselkedő, rögzült szerkezetek, ilyen például az „Emészt engem a buzgóság házadért” evangéliumi idézet 16-szor ismétlődő előfordulása. Jellemző, a 34-ből 15 típusra érvényes, hogy az alaptag birtokos személyjeles (például [*vkiknek a*] *viszonya egymáshoz* (40); [*vkiknek a*] *joga az élethez* (13)). Ezek a birtokos szerkezet tipikus szórendjét érvényesítik, amely leképezi a referenciapont-szerkezet referenciapont – cél elrendezését (Tolcsvai Nagy 2005, 2017: 250–251). Mivel kisebb valószínűséggel várható, hogy a birtokos jelző eltávolodjon a jelzett szótól vagy kövesse azt, ahogy az elől álló határozó esetében történne (vö. [*vkiknek*] *egymáshoz viszonya v. egymáshoz viszonya* [*vkiknek*]; [*vkinek*] *az élethez joga v. az élethez joga* [*vkinek*]), nem meglepő, hogy ilyenkor a begyakorlott szórenddel kompatibilis appozitív szerkezetet részesítik előnyben a beszélők.⁴⁸

Kiemelkedő gyakoriságú a *félelem attól* (104) (összevetve az *attól félelem* 0 előfordulásával), amelyben az appozitív határozó erősen sematizált jelentésű, utalószói szerepű névmási bővítmény. A sematizáltság részben magyarázza is az előfordulások nagy számát (a részletezőbb jelentésű főnévi kifejezésekhez képest). Lényegesebb azonban, hogy az utalószó az appozitív szerkezetben a főmondat végére kerül,⁴⁹ és így szomszédos a mellékmondatdal, amelyhez a mentális elérést biztosítja (lásd (16), vö. Kugler 2022: 218).

⁴⁷ A ném. *Vergangenheitsbewältigung* motivációját itt nem tárgyaljuk. Bár kétségtelen, hogy a kifejezések sok szempontból mutatják az összetételek jellemzőit, morfológiai jelöltségük miatt elég stabil a különírásuk, továbbá épp az appozitív alternatíva jelzi, hogy a főnévi szóalakok külön is lehorgonyozódhatnak epiztemikusan (vö. „Ennek következtében Ausztriában elmaradt az a *katartikus szembenezés a náci multtal*” (#129321662)).

⁴⁸ Ennek a morfoszintaktikai szempontnak a relevanciáját mutatja, hogy az N2.ALL *jog* konstrukciónak több elől álló határozói megvalósulása is van, ha a *jog* nem része birtokos szerkezetnek, például *jog* [...] *munkához* (6). Hasonlóan, a *részvétel* N1 is előfordul többféle elől álló határozós szerkezetben (például *eljárásban* (3) / *versenyen* (2) *részvétel*), de birtokos személyjeles alakokkal csak appozitív szerkezetben van adata (*részvétele* [...] *forradalomban* (1)).

⁴⁹ Egy előfordulás, amelyben kijelölő jelző („a félelem attól a pillanattól, amikor” (#10650950)), a tagmondat végére kerülés (természetesen) nem teljesül.

- (16) Régebben jogos volt a félelem attól, hogy adósságszpirálba kerül a főváros [...] (#176484684)

4. Összefoglalás

A vizsgálat egyértelműen teljesítette azt a minimális célt, hogy korpuszalapúan appozitív határozói konstrukciók tényleges adataihoz jusson, és ezzel támogassa az appozitív határozós szerkezetek empirikus kutatását (lásd 1.1., 1.3. alpont, vö. Kugler–Szabó 2024). A többértelműség empirikus ellenőrzése és megfigyelése szempontjából ugyanakkor inkább az N2.CASE *l[eé]vő* N1.NOM, és nem az N2.CASE *való* N1.NOM konstrukcióból való kiindulás tűnik kedvezőnek. Az adatokba való betekintés alapján ugyanis a 2.2. alpontban leírt módon kinyert szerkezetek erősen kohezívek, a határozó sok esetben az alaptaggal jelölt entitás alkategorizálását végzi el (például a *visszaélés a hatalommal* a visszaélésként kategorizált cselekmények egyik típusát jelöli), az inkább elaboráló kifejezések is jellemzően centrálisak, magkomponensek az alaptag fogalmi kerete szempontjából (például *szembenezés* N2.INS, *hűség* N2.ALL). Ebből az következik, hogy kevésbé várható esetükben a (3) példához hasonló többértelműség, míg például a *faluban levő fiatalság* (#3799187) szerkezettel párhuzamba állítható *fiatalság* N2.INE esetében találunk adatot erre is (17).

- (17) mivel tölti az idejét a mai kínai fiatalság a városokban. (#287068013)

A (17) egy lehetséges értelmezése az appozíciónak megfelelő ’városi, városokban élő fiatalság’ időtöltésére vonatkozik, míg az N.INE igei bővítményként kezelve a ’városokban történő időtöltés’-ről szól, bárhol lakjanak is a fiatalok.

Járulékos megfigyelésünk, hogy a *való* és a *levő/lévő* közötti lényeges különbség az, hogy a *l[eé]vő* konstrukció tipikusan egy statikus szituációt, a szereplők állapotát dolgozza ki, míg a *való* konstrukció a folyamatokat, történéseket részletezi a szereplő jellegű körülmény elaborálásával. Ez nem magyarázható önmagában a *való* szemantikájával,⁵⁰ hanem a létige participiumi alakjainak konstrukcióhoz kötött, több évszázados funkcionális elkülönülésével (vö. A. Jászó 1991, 1992) és jelentős részben a konstrukció

⁵⁰ Hiszen a *van* : *lesz* alapszó esetében inkább az utóbbi a dinamikus, és az előbbi a statikus (Modrián–Horváth 2024).

profilját meghatározó főnévi alaptagnak a 'folyamat' jelentésével függ össze.⁵¹

A dolgozat megvizsgálta, hogy mely főnévi szóalakpárokkal, páronként milyen gyakorisággal adatolható az N2.CASE *való* N1.NOM konstrukció. Ennek eredménye eltér a szakirodalom alapján kirajzolódó képtől. A típusok megvalósulási viszonyai alapján az N2.CASE *való* N1.NOM konstrukció 39%-ban nem alternatívája sem az appozitív konstrukciónak, sem az N2.CASE N1.NOM szerkezetnek az adott körülményviszonyok megkonstruálásában. Tehát nem tartható az az elképzelés, hogy az appozitív határozónak a *való*-konstrukció egyfajta elkerülési módja, vagy fordítva (vö. M. Korchmáros 1974). Az is váratlan adat, hogy az elől álló határozós, jelzősítő kifejezést nem tartalmazó N2.CASE N1.NOM konstrukció lényegesen több típussal adatolható (52%), mint az appozitív (31%), továbbá az N2.CASE N1.NOM az előfordulások gyakorisága alapján is vezet a közvetlen határozói bővítményt tartalmazó két szerkezetet összevetve (lásd 5. táblázat).

A gyakorisági adatok alapján erősebb asszociatív kapcsolat állapítható meg a *való* konstrukció és az elől álló határozós szerkezet (N2.CASE N1.NOM) között, mint a jelzősítő (N2.CASE *való* N1.NOM) és az appozitív határozós szerkesztésmód között. Bár az elől álló határozói bővítmény nagyon gyakori egység státuszú kifejezésekben (főleg terminusokban), ezek a hétköznapi nyelvi tevékenység részévé is válhatnak (például az *Unióhoz csatlakozás*) mintákkal erősítve a konstrukció termékenységét. Másrészt a gyakori *való* konstrukciók egység státusz nélkül is megszokottá teszik a határozói bővítményt követő főnévi alaptag szórendet, a *való* nélküli szerkezet bármely szintéren előfordulhat (például *életre nevelés*). A *való* nélküli szerkezet, amely szórendjével leképezi a nominálisra jellemző kidolgozási-specifikálási irányt (lásd a szórenddel jelölt jelzői viszonyokat) konstrukciónak sematizálódhat, konstrukcionálódhat. Az előzménykonstrukciónak csak a konstrukcionálódás folyamata szempontjából van jelentősége. A vizsgált szerkezetek konstrukciós sémák, amelyeket séma – megvalósulás viszonyban értelmeztünk.⁵² Az adatok egyértelművé is teszik, hogy az *életre nevelés* kifejezés az N2.CASE N1.NOM séma megvalósulása, nem másik kifejezésből redukált vagy levezetett szerkezet. Míg az *életre való nevelés*-nek 45 előfordulása van az MNSZ2-ben (*életre szóló/kiható* találatok mellett), addig az

⁵¹ Ehhez az alaptagok ontológiáját szükséges volna vizsgálni, de a 2.1. pontban bemutatott mintákra érvényes az, hogy a *[[e]vő* konstrukció esetében inkább jellemző a 'dolog' profil.

⁵² A kifejezések tehát nincsenek levezetési viszonyban.

életre nevelés (tipikusan jelzővel bővülve, például *családi/egészséges életre nevelés*) 136 példánnyal adatolható. A magyarázatot az -*Ás* képzős főneveknek abban a sajátosságában tudjuk megragadni, hogy képesek részlegesen illeszkedni az ige morfoszintaktikai sémájába (Palágyi 2022a: 202kk., 226), és így közvetlenül bővíülhetnek a folyamat szereplőjét megjelenítő határozóval (de természetesen birtokos jelzővel is). Ugyanakkor a bővítményt a főnévi alaptagokra jellemző, a jelzős szerkezetben tipikus szórendben is felvehetik, sőt adataink szerint ez történik gyakrabban.⁵³ Az -*Ás* képzős főneveknek ez a vegyült karaktere magyarázatot tud adni arra is, hogy bár az appozitív határozó az -*Ás* képzős főnevek esetében lényegesen gyakoribb, mint tőszavak mellett, a határozói bővítménynek az alaptagot követő szórendi pozíciója mégis ritkább, mint a típusos, megelőző helyzet.

A 2.2. pontban bemutatott eljárással nagymennyiségű adathoz jutottunk, ezeket az adatokat nem dolgoztuk fel teljeskörűen, első lépésben a konstrukciók összevetésére törekedtünk. Jövőbeli terveink között szerepel az adatok mozaik n-gramos (trigramos) feldolgozása (Bajzát–Indig 2023; Indig–Bajzát 2023) és beható elemzése.

Források

MNSZ2 = Magyar Nemzeti Szövegtár v2.0.5 (<http://corpus.nytud.hu/mnsz/>)
(Lásd Oravecz–Váradi–Sass 2014.)
FrameNet. <https://berkeleyfn.framenetbr.ufjf.br/about> (Lásd Ruppenhofer et al. 2016.)

Irodalom

- A. Jászó Anna 1991. Az igenevek. In: Benkő Loránd (főszerk.): *A magyar nyelv történeti nyelvtana I. kötet. A korai ómagyar kor és előzményei*. Budapest: Akadémiai Kiadó. 319–352.
A. Jászó Anna 1992. Az igenevek. In: Benkő Loránd (főszerk.): *A magyar nyelv történeti nyelvtana II/1. kötet. A kései ómagyar kor. Morfematika*. Budapest: Akadémiai Kiadó. 411–454.

⁵³ Lásd részletesen a cselekvésnévi szerkezeteket Palágyi (2022a: 224–229, 2022b) munkáiban.

- Antal László 1975. Két karcolat a leíró nyelvészet köréből. *Magyar Nyelv* 71: 40–46.
- Antal László 1985. *A hatodik mondatrész*. Budapest: Magvető Kiadó.
- Bajzát Tímea – Indig Balázs 2023. A (pre)modális szemantikai térrel összekapcsolódó főnévi igeneves szerkezetek vizsgálata című kutatásban alkalmazott kódok. Repozitórium. <https://github.com/bajzattimi/Research-of-infinitive-structures-related-to-the-modal-semantic-domain> (2023-07-17)
- Balogh Judit 2000. A főnév. In: Keszler Borbála (szerk.): *Magyar grammatika*. Budapest: Nemzeti Tankönyvkiadó. 127–141.
- B. Lőrinczy Éva 1962. A határozó. In: Tompa József (szerk.): *A mai magyar nyelv rendszere. Leíró nyelvtan II*. Budapest: Akadémiai Kiadó. 162–259.
- Hámori Antónia 1954. A jelzői értékű hátravetett határozó használatának kérdéséhez. *Magyar Nyelv* 50: 419–431.
- Imrényi András 2017. Az elemi mondat viszonyhálózata. In: Tolcsvai Nagy Gábor (szerk.): *Nyelvtan*. Budapest: Osiris Kiadó. 664–760.
- Indig, Balázs 2017. Mosaic n-grams: Avoiding combinatorial explosion in corpus pattern mining for agglutinative languages. In: Vetulani, Zygmunt – Paroubek, Patrick – Kubis, Marek (eds.): *Human language technologies as a challenge for computer science and linguistics*. Poznań: Adam Mickiewicz University. <https://real.mtak.hu/73335/> (2023-07-17)
- Indig, Balázs – Bajzát, Tímea 2023. Bags and mosaics: Semi-automatic identification of auxiliary verbal constructions for agglutinative languages. In: Vetulani, Zygmunt – Paroubek, Patrick – Kubis, Marek (eds.): *Human language technologies as a challenge for computer science and linguistics*. Poznań: Adam Mickiewicz University. 111–116.
- Kálmán László 2007. Miért nem vonzanak a régensek? In: Kálmán László (szerk.): *KB 120: A titkos kötet. Nyelvészeti tanulmányok Bánréti Zoltán és Komlós András tiszteletére*. Budapest: Tinta Könyvkiadó. 229–246.
- Keszler Borbála 2000. A határozó. In: Keszler Borbála (szerk.): *Magyar grammatika*. Budapest: Nemzeti Tankönyvkiadó. 423–443.
- Kiefer Ferenc 1998. Alaktan. In: É. Kiss Katalin – Kiefer Ferenc – Siptár Péter: *Új magyar nyelvtan*. Budapest: Osiris Kiadó. 187–289.
- Komlósi András 1992. Régensek és vonzatok. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 1. Mondattan*. Budapest: Akadémiai Kiadó. 299–528.
- Kövecses Zoltán 2005. *A metafora. Gyakorlati bevezetés a kognitív metafora-elméletbe*. Budapest: Typotex.

- Kugler Nóra 2022. Milyen kérdésekhez és válaszlehetőségekhez vezet az összetett mondat leírásában a kontextus módszeres figyelembevétele? In: Tátrai Szilárd – Tolcsvai Nagy Gábor (szerk.): *A magyar mondat és kontextuális környezete*. Budapest: ELTE BTK – ELTE Eötvös Kiadó. 213–243.
- Kugler Nóra – Szabó Dorina Szabina 2024. A függőségi viszony azonosítását befolyásoló szórendi jelenségek kísérletes vizsgálatának megalapozása. In: Laczkó Krisztina – Tátrai Szilárd – Tolcsvai Nagy Gábor (szerk.): *A magyar nyelv igei konstrukciói. Használatalapú függőségi vizsgálatok*. Budapest: ELTE. 255–282.
- Ladányi Mária 2017. Alaktan. In: Tolcsvai Nagy Gábor (szerk.): *Nyelvtan. A magyar nyelv kézikönyvtára* 4. Budapest: Osiris Kiadó. 503–660.
- Langacker, Ronald W. 2008. *Cognitive Grammar: A basic introduction*. Walter de Gruyter, Berlin. <https://doi.org/10.1093/acprof:oso/9780195331967.001.0001>
- Langacker, Ronald W. 2012. Substrate, system, and expression: Aspects of the functional organization of English finite clauses. In: Brdar, Mario – Raffaelli, Ida – Žic Fuchs, Milena (eds.): *Cognitive linguistics between universality and variation*. Newcastle upon Tyne: Cambridge Scholars Publishing. 3–52.
- Lengyel Klára 1996. Van a magyarban összetett jelző! *Magyar Nyelvőr* 120: 195–209.
- Lengyel Klára 2000. A segédigék és származékaik. In: Keszler Borbála (szerk.): *Magyar grammatika*. Budapest: Nemzeti Tankönyvkiadó. 252–258.
- M. Korchmáros Valéria 1974. A való és a levő igenevek használatáról. In: Imre Samu – Szathmári István – Szűts László (szerk.): *Jelentés- és stilisztika*. Nyelvtudományi Értekezések 83. Budapest: Akadémiai Kiadó. 313–319.
- Modrián-Horváth Bernadett 2024. Vázlat a magyar passzív jelentésű konstrukciók vizsgálatához a használatalapú nyelvtan szemszögéből. In: Tátrai Szilárd – Tolcsvai Nagy Gábor (szerk.): *A magyar nyelv igei konstrukciói. Használatalapú függőségi vizsgálatok*. (Kézirat, megjelenés alatt.)
- Nádasy Ádám 2003. *Ízlések és szabályok. Írások nyelvről, nyelvészetről 1990–2002*. Budapest: Magvető.
- Oravecz, Csaba – Várad, Tamás – Sass, Bálint 2014. The Hungarian Gigaword Corpus. In: *Proceedings of LREC 2014*.
- Palágyi László 2022a. *Morfológia az egésztől a részekig. A magyar főnév esete*. Budapest: Szépirodalmi Figyelő Alapítvány.

- Palágyi, László 2022b. Possessing a process. A cognitive semantic approach to the action nominal construction in Hungarian. *SLH [Studia Linguistica Hungarica]* 34: 118–138. <https://ojs.elte.hu/slh/article/view/5407/4378> (2023-07-17) <https://doi.org/10.54888/slh.2022.34.118.138>
- Rácz Endre 1968. Mondattan. In: Bencédy József – Fábián Pál – Rácz Endre – Velcsov Mártonné: *A mai magyar nyelv*. Budapest: Tankönyvkiadó. 205–458.
- Radden, Günter – Dirven, René 2007. *Cognitive English Grammar*. Amsterdam, Philadelphia: John Benjamins. <https://doi.org/10.1075/clip.2>
- Rákosi, György 2003. Comitative arguments in Hungarian. In: Heeren, Willemijn – Papangeli, Dimitra – Vlachou, Evangelia (eds.): *Uil-OTS Yearbook 2003*. Utrecht: Utrecht Institute of Linguistics OTS. 47–57.
- Rákosi, György 2021. PPs used as arguments. In: É. Kiss, Katalin – Hegedűs, Veronika (eds.): *Syntax of Hungarian: Postpositions and postpositional phrases*. Amsterdam: Amsterdam University Press. 285–354. <https://doi.org/10.2307/j.ctv1kmj874.8>
- Roskó Mira 2020. Az értelmező mint eltagolt szerkezettypus. *Ösvények 2020*: 117–124.
- Roskó Mira Rózsa 2023. Főnévi jelzők. *Ösvények* [megjelenés alatt].
- Ruppenhofer, Josef – Ellsworth, Michael – Petruck, Miriam R. L – Johnson, Christopher R. – Baker, Collin F. – Scheffczyk, Jan 2016. *FrameNet II: Extended Theory and Practice*. (https://framenet.icsi.berkeley.edu/fndrupal/the_book) (2022.10.30.)
- Sass Bálint 2019. Keresés korpuszban. Oktatási segédanyag, PDF.
- Sass, Bálint 2022. Principles of corpus querying: A discussion note. *Acta Linguistica Academica* 69 (4): 599–614. <https://doi.org/10.1556/2062.2022.00581>
- Szabolcsi Anna – Laczkó Tibor 1992. A főnévi csoport szerkezete. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 1. Mondattan*. Budapest: Akadémiai Kiadó. 179–298.
- Tolcsvai Nagy Gábor 2005. A magyar birtokos szerkezet jelentéstana, kognitív keretben. In: Kertész András – Pelyvás Péter (szerk.): *Általános Nyelvészeti Tanulmányok XXI. Tanulmányok a kognitív szemantika köréből*. Budapest: Akadémiai Kiadó. 43–70.
- Tolcsvai Nagy Gábor 2017. Jelentésan. In: Tolcsvai Nagy Gábor (szerk.): *Nyelvtan*. Budapest: Osiris Kiadó. 207–499.



A corpus-based study of Hungarian appositive adverbial constructions

The study deals with the corpus-based identification of appositive adverbials (e.g. *visszaélés a hatalommal* 'misuse.NOM power.INS, misuse of power'; *napilapok nemzeti tulajdonban* 'newspaper.PL.NOM national ownership.INE, newspapers in national ownership, nationally owned newspapers') that occur in postnominal position (N1.NOM N2.CASE). In the description of the construction, N.CASE symbolizes a noun case form with the exclusion of nominative and accusative case forms (without OBL : MODE distinction). The paper presents the problem of the appositive adverbial construction in a cognitive linguistic framework and formulates the research question that gave rise to the data collection. It then describes the data retrieval and query process in Hungarian Gigaword Corpus (HGC version 2.0.5, Hungarian acronym: MNSZ2) from the first attempts to the context-based search that yielded valuable results. The successful query was based on the data from the N2.CASE *való* ('be.PTCP') N1.NOM construction, and this defined the boundaries of the present study. The corpus analysis was thus limited to potential N1.NOM N2.CASE appositive adverbial constructions, for which the alternative is to implement N2.CASE *való* N1.NOM construction. The paper thus uses the data from the N2.CASE *való* N1.NOM construction to report results on the extent to which the appositive adverbial structure (N1.NOM N2.CASE) can be considered an alternative to the N2.CASE *való* N1.NOM construction. In this context, it also considers the N2.CASE N1.NOM, i.e. the adjacent adverbial construction (not including *való* attributive marker), as an alternative construction. The study only initiates but does not carry out a similarly detailed analysis of the N2.CASE *levő/lévő* ('be.PTCP') N1.NOM construction, but reflects on preliminary observations on this.

Keywords: „attributive adverbial”, appositive adverbials / postnominal adverbial modifiers, *való* ('be.PTCP') in complex attributive construction, dependency grammar/analysis, word order

Indig Balázs

Eötvös Loránd Tudományegyetem

indig.balazs@btk.elte.hu

ORCID: 0000-0001-8090-3661

Kurucz Réka

Eötvös Loránd Tudományegyetem

kuruczreka@student.elte.hu

Ernhöffer Aliz

Eötvös Loránd Tudományegyetem

lizalizer@student.elte.hu

Kugler Nóra

Eötvös Loránd Tudományegyetem

kugler.nora@btk.elte.hu

ORCID: 0000-0002-8223-2926

KALIVODA ÁGNES – PALÁGYI LÁSZLÓ

Hozzászólás az *Appozitív határozók korpuszalapú vizsgálata* című tanulmányhoz

A tanulmány a magyar nyelv leírása szempontjából kifejezetten problematikus és összetett szintaktikai viszonynak, az appozitív határozóknak a vizsgálatára vállalkozik. A korábbi vizsgálatoktól eltérően adatközpontúan közelíti meg a jelenséget, és kvantitatív szempontokat is érvényesít. Korpuszalapon feltérképezi a konstrukció kategorizáló funkcióját és jól begyakorlott példányait, illetve más konstrukciókkal való összefüggését a konstrukciók hálózatában. Utóbbi eredmény a további kutatási irányok meghatározása szempontjából is jelentőséggel bír. Fontosnak tartjuk kiemelni, hogy a vizsgált konstrukció korpuszalapú vizsgálata nagy kihívás. Pusztán szóalakokra és morfoszintaktikai címkékre támaszkodva nehéz pontosan és egyúttal nagy fedéssel is megragadni; a munkát esetleg megkönnyítő, szintaktikai elemzést tartalmazó, nagyméretű és nyilvánosan is elérhető korpusz pedig nem állt a szerzők rendelkezésére.

Későbbi levelezésünk során arra is felhívták a szerzők a figyelmünket, hogy a helyzetük gépi szintaktikai elemzővel sem lett volna könnyebb, sőt szándékosan tartózkodtak ennek a használatától. Ahogy Kugler Nóra jelezte: „[...] a magyarban nagyon összetett folyamat vezet el ahhoz a döntéshez, hogy egy határozó az igéhez vagy esetleg egy főnévhez kapcsolódik-e. Például *A lövöldözésről az egyetemen mindenki értesült.* mondat esetében (prozódiai egyértelműsítés nélkül) olyan többértelműség áll fenn, amelyet egyelőre nem tudna kezelni egy elemző, tehát még a nominálissal való szomszédosság [az igétől való nagyobb távolság] esetén sem egyértelmű, hogy a határozó a főnévi lánc része-e.”

A szerzők alaposan átgondolt, jó módszertant követtek, a kutatásuk reprodukálható. Az eredmények nyelvészeti és nyelvtechnológiai szempontból is érdekesek és előremutatók. A bírálatunkban elsősorban a módszertani leírás kisebb hézagaira mutattunk rá. Nagy örömmel vettük, hogy a szerzők mindent javítottak, vagy gondosan reflektáltak rá a tanulmányban.

1. Kisebb módszertani kérdések, amelyeket a szerzők maradéktalanul megválasztak

1. A 2.1. fejezetben szerepel a következő mondat: „A szépirodalmi al-korpuszt kizártuk.” Ez nem egy magától értetődő lépés, szükséges volna megindokolni, hogy miért döntöttek így a szerzők. Feltételezzük, hogy a sok vers, illetve a regények esetében a szándékosan régies vagy mes-terkelt nyelvhasználat lehet a döntés oka, de jó lenne ezt explicitté tenni.
2. Rögtön ezután következő mondat: „Szűrést végeztünk L1–2 ablakban”. Itt arra hívnánk fel a figyelmet, hogy ezek keresőfelület-specifikus fogalmak, az MNSZ2 esetében a NoSketch Engine-ben megszokott terminusok. A közérthetőség érdekében érdemes lenne a legelső említésüknél nagyon röviden leírni, hogy mit jelentenek.
3. A 2.2. fejezetben a CQL-ek alapján az látszik, hogy csak az elemzés elején keresik a szerzők az MN címkét ($msd="^MN.+NOM\$"$), így kimaradnak például a felsőfokú melléknevek (pl. a *legszebb* címkéje FF.MN.NOM). Ez elkerülte a szerzők figyelmét, vagy valamilyen okból szándékosan csinálták így? (Esetleg egyáltalán nem várható felsőfokú melléknév ebben a szerkezetben?)
4. Szintén a 2.2. fejezetben szerepel, hogy „Duplikátumszűrés nem történt.”. Amennyiben a szerzők folytatják a kutatást az MNSZ2 korpuszon, javasoljuk, hogy végezzenek duplumszűrés: sajnos a duplumok aránya meglehetősen magas ebben a korpuszban, a személyes alkorpuszban a 20%-ot is meghaladja. Ha a keresőfelületen keresztül dolgozik az ember, nem igazán tudja ezt a problémát orvosolni, de az itt bemutatott módszerbe az automatikus duplumszűrés jól beiktatható (az adatok le-kérése után és az Indig-Bajzat (2023) módszer alkalmazása előtt).
5. A dolgozat szerzői a jelzői szerepű *lévő* és *való* kapcsán szemantikai, használatalapú különbséget állapítanak meg: „A lényeges különbség tehát az, hogy a *l[eé]vő* konstrukció tipikusan a szereplők állapotát, statikus szituációt dolgoz ki, míg a *való* konstrukció a folyamatokat, történéseket részletezi a körülmény elaborálásával” (12. oldal). Ezt a megkülönböztetést a korpuszadatok alátámasztják, de ez esetben megemlíthető, a bevezetőben szereplő *kisebbségi csoporthoz való tartozás* mint állapotváltozást nem magába foglaló, statikus helyzetet jelölő állapot, egyáltalán nem számít reprezentatív példánynak. Azt azonban jól példázza, hogy noha a *való* inkább dinamikus jelentésű főnevek jel-zője, használata (erősebb grammatikalizációja folytán) szélesebb körű,

mint a statikus állapotot megkonstruáló *lévő* konstrukcióé (lásd még a nem állapotváltozást, hanem állapotot jelentő főnevet a 4. táblázat rendszerért való aggodalom példájában).

2. Egyéb észrevételek, amelyekre a szerzők hivatkoznak és válaszolnak a tanulmányban

6. A szerzők a következőt írják az adatfeldolgozásról: „Az összes n_2 – n_1 párra lefuttatva mindkét lekérdezést és az eredményeket kiexportálva az Indig–Bajzát (2023) közleményben definiált módszer szerint (a részletes leíráshoz lásd Bajzát–Indig (é.n.)) az egyes gyakoriságpárokat táblázatba gyűjtöttük, hogy összehasonlíthatóvá váljanak.” Alapvetően jó megoldás, hogy ha már született egy külön cikk a módszerről, akkor itt nem kell belebonyolódni a részletes ismertetésébe, de a mostani leírás nagyon kevés, szinte semmi nem derül ki belőle. A cikk kulcsszavai között szerepel a *függőségi elemzés*, így – legalább is nyelvtechnológus szemmel – arra számítanánk, hogy a találatokon a szerzők függőségi elemzést is lefutttattak, és ez a módszer egyik kulcspontja. Tudomásunk szerint azonban nem függőségi elemzésre épül a hivatkozott cikk módszere, hanem n -gram szekvenciákra (amik szavak, lemmák és morfoszintaktikai címkék változatos kombinációi), ezt tehát jó lenne két-három mondatban világosabbá tenni.
7. A 3.1. fejezetben szerepel ez a mondat: „A kereső és a gépi feldolgozás sajátosságaiból adódóan megkülönböztettük azonos morfémaszerkezetű főnévi szóalak esetében is az írásmódban jelentkező különbségeket, tehát például külön típusként tartjuk számon a *hatalommal való visszaélés* (609) és a *Hatalommal való visszaélés* (35) kifejezéseket.” Nem világos, hogy miért döntöttek így a szerzők, ezt jobban meg kellene indokolni. A táblázatokban főként köznevek szerepelnek, így nem tűnik célszerűnek külön tartani a kis- és nagybetűs verziókat. A keresőben valóban kissé körülményes case-insensitive keresést indítani, de a gépi feldolgozás szakaszában ez egy egyszerűen megoldható helyzet. Esetleg vártak a szerzők nagy számban tulajdonneveket is?
8. Az 1.2 fejezetben szerepel ez a mondat: „Az appozitív határozói szerkezetben a szokatlan szórend ellenére is jól magyarázható a függőségi viszony létrejötte akkor, ha az alaptag olyan képzett főnévi kifejezés, amely az alapszavától »örökli« a határozói bővítményt”. Mint azt

a 2.1-ben az előzménynek számító korpuszvizsgálat eredményei is alátámasztják, a főnévnek nem szükséges deverbálisnak lennie, az ‘esemény’ jelentésű főnév az igei derivátumának határozóival is asszociálódhat, (*verseny/versenyez harc/harcol a győzelemért*). Sőt, önmagában a szemantikai viszony is elégségesnek látszik a megfelelő vonzat “megörökléséhez”, lásd az *[elkezdődött a] munka a regényen* kifejezést, amelyet feltehetően sokkal inkább a *dolgozik* ige superessivus vonzata motivál (*dolgozik a regényen*), mint a konstrukciót (szemantikai okokból) szintén kidolgozni képes *munkálkodik* igéé (*munkálkodik a regényen*).

9. A 2.1-es fejezetben olvasható, hogy a határozók az N2.INE *l[eé]vő* N1. NOM konstrukcióban nem elsősorban ‘valahol levést’ jelölnek, hanem állapotokat. Nem egészen világos, hogy ez konstrukcióspecifikus jellemző-e, hiszen ebben a bekezdésben a nem igenévi parafrázis is szóba kerül (*előzetesben van* stb.). Ugyanitt megemlíthető lenne, hogy az állapotok tartályszerű leképezése meglehetősen univerzális fogalmi metaforaként ismert a kognitív nyelvészetben, és a magyar inessivus eset hozzájárulása ennek szimbolizációjához korántsem megkérdőjelezhető.
10. A dolgozat szerzői elvetik a jelzői szerepű *lévő* és *való* kapcsán a szófaji/kategoriális különbséget. Az egyik érvük szerint “a *való*-konstrukciónak is lehet igei szerkezet az alternatívája” (12. oldal), és erre példaként *a leszálláskor való kinyitása [vminek] – [vminek] a kinyitása leszálláskor történik* pár szolgál. A *történik* ige megítélésünk szerint annyiban bonyolítja a helyzetet, hogy a *való*-val (főleg dinamikus jelentésű főnév esetében) gyakran egyébként is helyettesíthető a *történő* (*első ajtón történő felszálláskor*). Formális (leíró) szempontból pedig az igei alternatíva nem helyettesítheti az igévé (létigévé) való átalakíthatóság kritériumát, vagyis mondható, hogy a jelzői szerepű *történő* igenév, az ezt parafrázáló *történik* pedig ige, míg a *való*-nak nincs igei párja. Mindenesetre a *való* konstrukciófüggő kopulaként való megközelítése mellett szóló érv, hogy absztraktabb a jelentése (erősebben grammatikalizálódott) és a *lévő*-vel szemben gyakran elhagyható a szerkezetből.
11. A CQL kifejezések mind helyesek, de pár esetben redundánsak, van elegánsabb verziójuk is:

- $[word="*[áé]s.[*][áé]s[oe]k"] \rightarrow [word="*[áé]s([oe]k)?"]$
- $\& msd!="^+.NOM\$" \& msd!="^+.ACC\$" \rightarrow \& msd!="^+. (NOM|ACC)\$"$
- $[msd="^MN.+NOM\$" \mid msd="^SZN.+NOM\$"]\{0,2\} \rightarrow [msd="^ (M|SZ) N.+NOM\$"]\{0,2\}$

Köszönjük, hogy a szerzők minden tekintetben figyelembe vették a bírálattunkat, és gratulálunk a nagyon igényesen megírt, érdekes tanulmányhoz!

Kalivoda Ágnes

HUN-REN Nyelvtudományi Kutatóközpont

kalivoda.agnes@nytud.hun-ren.hu

ORCID: 0000-0003-2520-5523

Palágyi László

Eötvös Loránd Tudományegyetem

palagyilaszlo@btk.elte.hu

ORCID: 0000-0002-1437-4746

Beférkőző igék kollostrukciós vizsgálata



KALIVODA ÁGNES – MODRIÁN-HORVÁTH BERNADETT –
SIMON GÁBOR

**Beférkőző igék kollostrukciós vizsgálata.
Az igekötő és főnévi igenév közé ékelődő igék
kolxémaelemzése az MNSZ2-UD korpuszban¹**

Kivonat

Jelen tanulmány az igekötő és főnévi igenév közé való beférkőzés kollostrukciós vizsgálatát mutatja be. A kolxémaelemzés az egyes lexémák korpuszbeli gyakoriságához viszonyítva állapítja meg a vizsgált konstrukcióval való asszociáltság mértékét, és hozzájárulhat a konstrukció alapjelentésének leírásához. A beférkőzést a szakirodalom a segédigei funkcionálás formailag jól megragadható aspektusaként azonosítja. Vizsgálatunkkal azt kívántuk feltárni, hogy a vizsgált korpusz igéi közül melyek mutatják jellemzően ezt a viselkedést, mivel ez indikátora lehet ezen igék segédigei jellegének.

A korpuszvizsgálat összesen hét statisztikailag szignifikáns mértékben vonzott igét állapított meg, melyek jelentése modális és temporális jellegű. A hét ige szinte teljes egyezést mutat a központi segédigéknek Tolcsvai Nagy (2009) által szemantikai és grammatikalizációs alapon kiemelt listájával. A vonzott lexémák köréből kizárt igék között is vannak olyanok, amelyek

¹ Szeretnénk köszönetet mondani a cikkünk két bírálójának, Dömötör Andreának és Imrényi Andrásnak a tanulmány alapos átolvasásáért és az ahhoz fűzött sok hasznos javaslatért. Kalivoda Ágnes kutatása az 142317 számú OTKA PD projekt keretében, az Innovációs és Technológiai Minisztérium Nemzeti Kutatási Fejlesztési és Innovációs Alapból nyújtott támogatásával, a PD 22 pályázati program finanszírozásában valósult meg. A dolgozat elkészítését Modrián-Horváth Bernadett és Simon Gábor esetében az NKFIH 129040 számú pályázata (A magyar nyelv igei konstrukciói. Használatalapú konstrukciós nyelvtani kutatás) támogatta.

adatai gyakori beférkőző használatot mutatnak, ezek további vizsgálatokat igényelnek.

Kulcsszavak: beférkőző pozíció, segédige, kollostrukció, egyszerű kolexéma-elemzés

1. Beférkőzés és segédigeség

Jelen tanulmány a segédigeséggel igen gyakran asszociált konstrukció, az igekötő és főnévi igenév közé való beférkőzés kollostrukciós vizsgálatát mutatja be. Az 'igekötő – ige – infinitívusz' szerkezetet számos magyar nyelvészeti tanulmány a segédigeség egyik lehetséges, vagy talán legmegbízhatóbb kritériumának tekinti, azonban jelen tanulmányban az alább kifejtett elméleti és metodológiai megfontolások alapján a konstrukciós nyelvészet értelmében vett önálló konstrukciós mintázatként, forma-jelentés párként kezeljük.² Korpuszalapú vizsgálatunk a mintázat formai oldalából indul ki, és nagy mennyiségű szövegadat elemzésével keresi az azzal legerősebben asszociált lexémákat, amelyek jelentése a konstrukciós jelentésre enged következtetni. Ezt az eljárást, az egyszerű kolexémaelemzést – a disztinktív és a kovariációs kolexémaelemzés mellett, amelyekkel együtt alkotja a kollostrukciós elemzések családját –, a konstrukciós nyelvészet keretében fejlesztették ki (bővebb ismertetését lásd a 2. pontban).

1.1. Mi a beférkőzés?

A beférkőző és keretes szerkezetek jelenléte – az összetartozó elemek megszakítása egy harmadik elem által az A1XA2 séma alapján, ahol A1A2 alkotja a keretet, X a beférkőző elem – a (morfo)szintaktikai szerveződés fontos mintája a magyar nyelvben is. Jelen van a ragozástól (például *leg-jo-bb*) és szóképzéstől kezdve (például az *el-néptelen-edik* típusú, igekötővel

² A konstrukció terminust – összhangban a konstrukciós nyelvtannal (vö. Goldberg 2006) – tanult forma-jelentés párként értjük itt, amely nem vezethető le egészében valamelyik alkotóeleméből. Mivel a konstrukció fogalma többféleképpen is értelmezhető a különböző konstrukciós nyelvtanok kiindulópontjából, továbbá egy nyelv konstrukcióinak bemutatása empirikus vizsgálatok eredményeként tehető meg (lásd Simon megj. e.), ebben a tanulmányban előfeltételezzük, hogy a beférkőző mintázat konstrukcióként működik, de nem célunk ennek a konstrukciónak sem az empirikus megalapozása, sem a kimerítő leírása.

együttjáró igeképzés, vö. Palágyi–Kalivoda 2024) az analitikus állítmány default megvalósulásán (*meg fogjuk látni, be lesz zárva*) és a kéttagú partikulákon át (*még ... is/sem*) a tagmondatok konstruálásáig (*hogy ... -e, még ha ... is*), vö. Modrián–Horváth (2020). Az említett szerkezetek mindegyike rendelkezik egy „rögzített”, kevésbé variábilis, valamint egy sematikusabb taggal; az infinitívuszos beférkőző szerkezetben a kevésbé variábilis tag a beférkőző ige. A tag értelemben vett absztrakt konstrukciós mintázat általános funkciója a figyelem fenntartása, valamint a konceptuális egység ikonikus megvalósítása lehet.

Tanulmányunk az igekötő és infinitívusz közé beférkőző ragozott igékkel foglalkozik. Beférkőzés alatt a szakirodalomban (vö. Kálmán C. et al. 1989) leginkább az igekötő és infinitívusz közé beékelődő igék megjelenését szokták érteni, tágabb értelemben a frazeológiai szókapcsolatokba való beférkőzést is megemlítik (például *húzni fogja a lóbőrt*, vö. Kálmán–Rádai 1996). E kétféle keret (igekötős ige, frazeológiai szókapcsolat) magmondatot alkot Imrényi (2017) terminológiája szerint. Fontos kiemelni, hogy mindkét típus számos besorolási nehézséget vet fel átmeneti jellegű konstrukciók esetén, amelyeket az igekötők grammatikizációja (vö. Kalivoda 2021; lásd az Imrényi 2013: 292 lapról származó 1a-b példákat) és a frazeológiai szókapcsolatok fokozatos lexikizációja (vö. Modrián–Horváth 2020, lásd 2a-b) eredményez.

- (1) a. János *el fog utazni* Párizsba. (Imrényi 2013: 292)
b. János részt szeretne venni a kiállításon. (Imrényi 2013: 292)
- (2) a. *húzni fogja a lóbőrt* (Kálmán–Rádai 1996: 86)
b. pár napig *húzni fogja* a hátsó lábát (#83135508)³

A keretet az általunk vizsgált esetben tehát egy infinitívuszos konstrukció adja, vagyis a beférkőzés tulajdonképpen egy komplex (igekötős igenévi) konstrukció és egy finit igealak összekapcsolódásával kialakuló sajátos konstrukciós mintázat. A kolecémavizsgálat során technikai okokból az infinitívuszos keretek megállapításánál csak az igekötőként annotált első tagokat vettük figyelembe, de a jelenség lehetséges tágabb értelmezésére reflektálunk az 5. pontban is.⁴

³ A tanulmányban az MNSZ2-ből származó példákat a szó sorszámával adjuk meg, hogy biztosítsuk a visszakereshetőségüket. Azok a példák, amelyeknél nincs forrás megadva, a szerzők kitalált példái.

⁴ Imrényi András bírálatában javasolja a „puszta főnévi igemódosítók” figyelembe vételét a vizsgálathoz. E javaslat megvalósítása egyrészt jelentős technikai nehézségekbe ütközik,

Az infinitívuszos konstrukcióba beférkőző elemek első ránézésre sokféléknek tűnhetnek, azaz nemcsak igék, hanem más elemek is be tudnak ékelődni az igekötő és az infinitívusz közé, azonban ennek korpuszvizsgálatok szerint korlátai vannak. Egyrészt az igekötő és infinitívusz egymástól való eltávolodása az esetek 97,9% százalékában nem több, mint három szó (vö. Kalivoda 2021: 62, még részletesebb adatokért Kalivoda–Prószéky 2023: 239). Másrészt az egyéb beférkőző szavak mellett jellemzően található egy ragozott ige (például *kell*, *akar*, vö. 3a-b), ritkábban más modális kifejezés (például *szokás*, *kénytelen*).

- (3) a. *Oda még ma menni.
b. Oda akar még ma menni.
- (4) a. *Oda volna menni.
b. Oda kellett volna menni.

A fennmaradó néhány esetben pedig a ragozott igekötős ige sorrendjét is megszakítani képes *sem*, *nem*, *is* szavak fordulnak elő a fent említett infinitívuszos keretben.⁵ Ez alapján indokoltnak találtuk a vizsgálat beférkőző igékre való szűkítését.

1.2. Miért nem azonos a segédigeség a beférkőzéssel?

A beférkőzés tulajdonságát szinte minden, a magyar segédigékkel foglalkozó tanulmány megemlíti és összefüggésbe hozza a segédigei státusszal: Kálmán C. et al. (1989) úttörő tanulmánya, amely az utóbbi évtizedek legbefolyásosabb műve a témában, a segédigeség fő kritériumának tekinti a csüggés (infinitívusz melletti enklitikus helyzet, például *menni fog*) és beférkőzés szórendi tulajdonságait; Imrényi (2013) a beférkőző használatot segédigei használatnak tekinti. Ezzel kapcsolatban mégis jelentős elméleti kíváncságnak tűnik mind

másrészt azonban az említett főnévi elemek vélhetőleg éppúgy megjelennek a szignifikánsan vonzott, mint a taszítottként megjelölt lexémákkal alkotott konstrukciókban, ezért abból indulunk ki, hogy bevonásuk érdekében nem változtatná meg, legfeljebb csak árnyalná a beférkőző igékről alkotott képet.

⁵ A fenti állítást felülvizsgáltuk az MNSZ2 korpuszból származó 500 találatos mintán. Keresési feltételek: főnévi igenév igekötő nélkül, pozitív szűrés (-3, -2) szó távolságban igekötőre, negatív szűrés (-3, -1) szó távolságban tetszőleges írásjelre, negatív szűrés (-3, -1) szó távolságban igére. A kapott találatok a fentiekén túl csak hibás annotáció (például *viszont* kötőszó, *át* névutó vagy *túl* határozószó igekötőként annotálása), elütés, hibás lemmatizáció eredményei.

a beférkőzés, mind más, kritériumként használt morfoszintaktikai jellegzetesség megfelelő elméleti alapokra helyezése, mivel a vizsgált kritériumok kiválasztása sokszor leginkább intuitívnak tűnik (lásd 1.3.).

A segédigesség fogalmának megragadásához a legmegfelelőbb elméleti keretnek a grammatikalizációelmélet látszik (vö. Heine 1993), mely az auxiliárizáció, azaz segédigévé válás diakrón folyamatának általános és speciális (egyes nyelvekre szabott) leírásával nyújt fogódzkodót a segédigésülés mértékének megítéléséhez (a magyar nyelvre való alkalmazáshoz lásd Tolcsvai Nagy 2009, Modrián-Horváth 2009).⁶ A beférkőzés tulajdonsága ez alapján tekinthető egyfajta grammatikalizációs markernek, amely a klitizáció speciális eseteként a kötöttség, vagyis szintagmatikai kohézió (boundedness, Függungsenge, vö. Diewald 1999, Dér 2008) ikonikus⁷ kifejezőeszköze. A grammatikalizációelmélet azt a folyamatot írja le, hogyan lesznek az eredetileg (inkább) lexikális kifejezőeszközökből (inkább) grammatikai kifejezőeszközök, amely folyamat során fokozatos szemantikai és formális autonómiavesztés figyelhető meg. Ez az autonómiavesztés magyarázhat számos, a segédigésülésnek tulajdonított sajátosságot, mint például a saját vonzatok elvesztése vagy a ragozás lekopása (például a *kell*, *lehet*, *tetszik* főigeként teljes paradigmával rendelkező igék hiányos paradigmája infinitívuszi vonzat mellett).

A „segédige” terminust a jelen dolgozatban is a „jelentős mértékben grammatikalizálódott ige” jelentésben használjuk, amely magába foglal több igei jelentéstípust, így a ragozási paradigmában szerephez jutó *van* és *lesz* igéket (például *be van zárva*, *be lesz zárva*), valamint a modális jelentésű *akar*, *tud*, *kell* igéket egyaránt. Az első típusban szereplő (segéd)igék az episztemikus lehorgonyzásban játszanak szerepet, míg az infinitívusszal álló (segéd)igék ezenfelül modális, illetve temporális jelentéstartalommal is hozzájárulnak az elemi jelenet konstruálásához. Jelentésükből adódóan az említett típusú igék alkalmasak a grammatikalizáció ösvényén való elinduláshoz, de a jelen tanulmány a grammatikalizáció folyamatának sem szemantikai, sem formai aspektusaival nem foglalkozik részletesen. A tanulmány (i) a segédigéket

⁶ Bajzát (2020) megemlíti, hogy „[a *tud*] segédige beférkőző szórendi pozícióra való törekvése vélhetően követte a jelentésének absztrahálódását” (Bajzát 2020: 10, tartalmi kiegészítés a szerzőktől), ezzel – némileg kifejtetlenül – szintén a grammatikalizáció kontextusába helyezi ezt a változást.

⁷ Haiman (1985: 129-130) „clitic climbing”-nek nevezi és az auxiliárizációval, segédigésüléssel együtt fellépő jelenségnek tartja a magyar igeragoknak, valamint igekötőknek a (segéd)igén való megjelenését; a segédigésülést a konceptuális fúzió jelének tekinti.

nyitott kategóriaként kezeli, amelynek belső szerveződése is van; (ii) a tanulmány és a vizsgálat nem kívánja sem meghatározni, sem egyértelműsíteni a segédigék kategóriájának meghatározását. Ugyanakkor bízunk abban, hogy a beférkőzés mint mintázatba rendeződés korpuszalapú feltérképezése hozzájárulhat a segédigék nyitott, prototípuselvű kategóriaként történő értelmezéséhez.

A fentiekből látható, hogy a beférkőzés tulajdonsága a grammatikalizáció-elmélet alapjára helyezve tekinthető a segédigésülés egyik, de korántsem egyetlen sajátosságának. A többi jellemző tulajdonság elméleti és empirikus vizsgálata további, lehetőleg korpuszalapú vizsgálatokat kíván.

1.3. A segédigeség kritériumalapú meghatározásáról: morfoszintaktikai sokszínűség és metodológiai opportunizmus

A segédigéket Bolinger megállapítása szerint morfológiai sokszínűség jellemzi (Bolinger 1980: 297, az angol segédigék sokféleségét demonstráló táblázatát idézi Heine 1993: 72). A grammatikalizálódott igék funkciójuktól, eredeti jelentésüktől és számos más tényezőtől függően nagyban különbözhetnek egymástól ragozás és preferált konstrukciók tekintetében, így például a szakirodalom szerint beférköző *akar* ige jelentéséből adódóan többnyire ágens jellegű alannyal társul, probléma nélkül nominalizálható és állhat infinitívuszban (lásd a 4-5 példákat).

- (4) amelynek közelében az élni *akarás* már csupán pókhálóból van szöve (#1071610)
- (5) mindent lehet, csak *akarni* kell (#1223248)

E sokszínűség miatt a szakirodalomban, vélhetően a potenciális segédigék számát csökkentendő, a segédigeség kritériumainak kiválasztásában egyfajta apriorizmust és metodológiai opportunizmust (a terminushoz vö. Croft 2001: 30kk., 41kk.) figyelhetünk meg, a fent említettek szerint leginkább anélkül, hogy a segédige terminus „segéd-” előtagja reflektálva lenne, és a kritériumok hierarchiája elméleti alapozást nyerne (ez alól két kivételt említettünk az 1.1. pontban).

A kritériumok megalapozott felállításához először is szükség van a segédige terminus definíciójára. Segédige alatt a magyar és nemzetközi szakirodalomban formailag, szemantikailag (például temporális, aspektuális és modális, vö. Heine 1993) és funkcionálisan (helyenként kopulák) sokféle igét

szoktak érteni, amelyek fő közös vonása a magas fokú grammatikalizáltság (lásd Heine 1993). Az igék segédige voltát igazoló kritériumokkal kapcsolatban azonban nagy bizonytalanságot tapasztalunk. Érdemes közelebbről megnézni, hogy milyen forrásokból származnak a kritériumok, felállításukkor találhatunk-e elméleti alapozást.

A tesztelt tulajdonságok egy része leíró megfigyeléseken alapul, amelyek feltűnő morfoszintaktikai különbségeket fednek fel a különböző igék használatában (kifejezetten a magyar nyelvre specifikus például a beférkőzés). Más részük a nemzetközi, tipológiai szakirodalomban található segédigei jellegzetességeket igyekeznek a magyarra átültetni (például Kenesei 2000, de vö. Tolcsvai Nagy 2009 tudománytörténeti kiigazítását „Heine listájával” kapcsolatban), amennyiben a kritériumok megfelelnek a szerzők elvárásainak; ezt a jelenséget nevezi Croft (2001: 30) tipológiai metodológiai opportunizmusnak („cross-linguistic methodological opportunism”).

A kritériumok kiválasztásánál számos munkában kifejezésre jut az e döntést joggal övező bizonytalanság, esetleg önkényesség. E tekintetben ikonikus Lengyel (2000b: 61) összefoglalása: „Úgy tapasztaltam, hogy a magyarban nincsen olyan alapvető vagy fő rendező szempont, mellyel az analitikus igealakokat egészen pontosan el tudnánk választani a főnévi igenévi bővítményű igéktől.” Továbbá Lengyel (2000b: 62): „Összesen 16 szempontot próbáltam ki, jórészt kíváncsiságból, mert nem voltam meggyőződve az összes szempont hasznosságáról. [...] Voltak olyan szempontok, melyek használhatatlanok voltak [...]; volt olyan, mely a várakozásom ellenére még a fog segédigeségét is bizonytalanította.” Ezt kissé kielezve így kommentálhatnánk: az a priori kritériumok sokszor használhatatlannak bizonyulnak, és elvetésre kerülnek, ha a szintén a priori segédigének tekintett igék státuszát kétségbe vonják (ez Croft szavaival élve a nyelven belüli metodológiai opportunizmus esete, „language-internal methodological opportunism”, vö. Croft 2001: 41kk.).

Látható, hogy az a priori kritériumokkal való operálás nem vezet el a segédigeség általános meghatározásához, miközben feltételezhető, hogy a kategóriának vannak kitüntetett vagy centrális (prototipikus) példányai, valamint az egyes kritériumok önmagukban is vizsgálhatók függetlenül attól, hogy a segédigeség meghatározásában mindenképpen kitüntetett szerepet tulajdonítanánk nekik.

1.4. A vizsgálati módszer választásának indoklása

A „segédigékkel” foglalkozó magyar szakirodalom túlnyomó részére – a „szigorú formális-disztribúciós kritériumokkal” (vö. Kenesei 2000) dolgozókra is, amilyen M. Korchmáros (1997) vagy Kálmán C. et al. (1989) – az introspekciós vizsgálati módszer jellemző. Ezzel az a legnagyobb probléma, hogy a grammatikalizációs kontinuum miatt sokszor fokozati jellegű különbségek vannak az egyes igék között, amelyek kategóriába sorolása egyetlen beszélő explicit módon hozzáférhető nyelvi tudása alapján kétséges. Az alkalmazott tesztek éppen a határesetek megítélésekor nem tudnak megbízható eredményt adni, akkor is, ha az egyes szerzőknek határozott véleményük van az adott ige hovatartozásáról: a csoportok elkülönülését nem lehet empirikus (jelen esetben korpusznyelvészeti) módszerrel igazolni. A Kenesei (2000) által elsőként vizsgált kritérium, a nominalizálhatóság esetében egy rövid, az MNSZ2-ben (v2.0.5, Oravecz–Váradi–Sass 2014) végzett felülvizsgálat azt mutatja, hogy a nominalizálható (a) csoportba sorolt (Kenesei 2000: 163) *óhajt* és *bír* esetében 1-1, a *szándékozik* esetében 2, a *kezd és kíván*⁸ igékből 3-3 és a *mer* igéből 7 valódi találatot kaptunk csupán (a *tud* és *akar* igékből számos nominalizáció található), míg a nem nominalizálható csoportban a *kell* igéből két nominalizációt; ez utóbbi ige gyakorisága miatt természetesen kevésbé esik latba a talált két nomen actionis, azonban az első csoport igéi közötti szórványos (1-2 előfordulással rendelkező) találatok nem támasztják alá ezek külön csoportba sorolását. Hasonlóan kevésbé verifikálható korpusznyelvészeti módszerrel (az MNSZ2 alapján) például a határozói igenév + *lenni* konstrukció segédigékkel való előfordulása (Kenesei 2000: 177), mivel az MNSZ2 korpuszban nincs találat a határozói igenévtől 2 szó távolságban *lenni* szóra (pozitív szűrés -2, +2 szavas ablakon belül).

A segédigék különböző tulajdonságaival kapcsolatban már születtek korpuszalapú vizsgálatok (Bajzát 2020, 2022; Kalivoda–Prószéky 2024, Modrián–Horváth 2009, 2020). Ezek közül némelyek a kontextus szűkítésével pontosabb eredményt kaphattak egy-egy jelenségre vonatkozóan, így Bajzát (2020) a *tud* ige vizsgálata során külön elemzi az igemódosítót tartalmazó, valamint egyéb kontextusokat, Kalivoda–Prószéky (2024) kizárja a vizsgálatból a vélhetően nem semleges szórendű (nem kijelentő modalitású és bal

⁸ További 3 példában egy Széchenyi-idézet bukkan fel: „minden időben volt és lesz is panasz, de az igen különbözik a javítani kívánástól” (#398520887; vö. még #386528105, #398530246)

perifériával rendelkező) mondatokat (vö. 5.2.). A jelen kutatás figyelembe vesz minden kontextustípusban való előfordulást, mivel nincs abszolút „semleges” mondat, minden megnyilatkozás információszerkezetét befolyásolja az adott setting (alap). A nagyméretű korpuszok a teljes nyelvhasználatot kívánják reprezentálni úgy, ahogy az a nyelvhasználók elméjében is érvényesítheti gyakorisági hatásait.⁹ Ez alapján feltételezhető, hogy a korpuszban ritkább a *be óhajt menni* konstrukció – annak ellenére, hogy az óhajt igét rendszerint a beférkőző igék között tartják számon – a mai nyelvhasználók számára is kevésbé természetes, és a szerzők intuíciója alapján is jelöltebbnek hat, mint a *be akar menni* konstrukció.

Kutatásunkban azt vizsgáltuk, hogy beférkőzés jelenségét (tehát az ige-kötős főnévi igenévi alakok folytonosságának megszakítását, illetve az ige-kötő és az igenév közé kerülést) mely igék mutatják statisztikailag szignifikáns módon. Azt kívántuk feltárni, hogy a vizsgált korpusz igéi közül melyek mutatják jellemzően ezt a viselkedést, mely a segédigei funkcionálás formailag jól megragadható aspektusa, ezáltal indikátora lehet ezen igék segédigei jellegének. Fontosnak tartjuk megjegyezni, hogy az itt bemutatott vizsgálat eredményeiből nem következik automatikusan a beférkőzést nagy arányban megvalósító igék segédigeként való osztályozása. Az viszont egyértelműen bizonyítható, hogy bizonyos igék preferálják ennek a mintázatnak a kialakítását főnévi igenévi alakokkal együttes előfordulásban, míg mások nem mutatnak ilyen hajlandóságot a korpusz adatai alapján.

2. A kollostrukciós elemzés

Ahhoz, hogy megállapítsuk a beférkőző szerkezettel való asszociáltság mértékét, olyan eljárásra van szükség, amely nem csupán azt mutatja meg, hogy melyik igék gyakoriak a vizsgált mintázatban, hanem azt is, hogy az adott ige korpuszbeli eloszlása alapján mennyire várható a vizsgált konstrukcióban való megjelenése. Ennek kiderítésére a kollostrukciós elemzés módszerét választottuk, azon belül is az egyszerű koléxémaelemzést. A módszerről általában lásd Stefanowitsch–Gries (2003), Gries (2015), Stefanowitsch (2013),

⁹ Természetesen nem gondoljuk, hogy a korpusz – bármilyen terjedelmű is – mindazt a nyelvhasználatot fel tudná mutatni, amely alakítja a beszélők nyelvi tudását. Éppen ezért helyesebb „kvázi-realitásról” beszélni a korpusz által reprezentált nyelvet tekintve, vagyis arról, hogy a reprezentatívnak szánt korpusz „kvázi-kontaktusban” áll a mindenkori nyelvi valósággal (lásd McEnery–Berzina 2022: 48).

Gries (2019), Stefanowitsch (2020: 270–275). Magyar nyelvre adaptálására lásd Simon (2022a, b).

A kollostrukciós elemzés arra a kérdésre keresi a választ, hogy mely szavak mely konstrukciókkal fordulnak elő együtt szignifikáns gyakorisággal (Stefanowitsch–Gries 2003: 213). Adottnak tekint egy bizonyos konstrukciót, és azt méri lexémák egy célcsoportjának korpuszbeli adataira támaszkodva, hogy mely lexémákat vonzza (attract) a konstrukció, illetve melyeket taszítja (repel). A vonzás/taszítás ténye azzal magyarázható, hogy az adott lexéma gyakrabban/ritkábban fordul elő a konstrukcióban annál, mint az várható lenne. A vonzott lexémák a konstrukció kolexémái, a kolexéma és a konstrukció együttesen adja a kollostrukciót a leírásban. Kutatásunkban a beférkőzést konstrukcióként értelmezzük, 'igekötő¹⁰ + finit igealak + főnévi igenévi alapalak' mintázatban, ez lesz az a kollostrukció, amelynek a jellemző igei kolexémáit keressük az elemzéssel. A kollostrukciós elemzés tehát a kollokációelemzés kiterjesztése, ugyanakkor sajátos változata is, amennyiben nem két szó együttes előfordulását teszi megfigyelhetővé, hanem egy szó és egy konstrukció összekapcsolódását. Mivel a kollostrukciós elemzés elméleti hátterét a konstrukciós nyelvtanok adják, és a lexikogrammatika kontinuumában a szó és a szónál nagyobb kifejezések társulására fókuszál (Stefanowitsch 2013), ma már önálló korpuszelemzési eljárásként tekinthetünk rá.

1. táblázat. A beférkőző mintázat igei komponenseinek kontingenciatáblája

	beférkőző konstrukció	nem beférkőző konstrukció
az ige előfordulásai az adott konstrukcióban	a	b
más ige előfordulásai az adott konstrukcióban	c	d

¹⁰ Az igekötők kategóriája folyamatosan bővül új elemekkel a grammatikalizáció révén, ezért vannak olyan „igekötőszerűen viselkedő elemek” (Imrényi András lektori javaslata) is a nyelvhasználatban, amelyek igekötői státusa külön vizsgálatot érdemel. Mivel ebben az elemzésben nem volt lehetőségünk ezt a problémát is megnyugtatóan rendezni (egyúttal az igekötők kurrens kategóriájával előállni), a mintavétel során nyelvtchnológiai eszközre támaszkodtunk az igekötők azonosításában, vállalva, hogy így csak a kategória centrális elemeire terjed majd ki az elemzés. Mindenképpen megfontolandó a jövőben az igekötők kategóriájának a lehető legpontosabb körülírása, amely vélhetően az elemzés eredményeit is még pontosabbá teszi majd.

A kollostrukciós elemzés alapvetően egy ún. kontingenciátábla (lásd 1. táblázat) adataira épül: a táblázat tartalmazza az adott ige korpuszbeli gyakoriságát a vizsgált főnévi igenévi beférkőző mintázatban (a), az adott ige összes előfordulását más (nem beférkőző) konstrukciókban (b), más igék gyakoriságát az adott konstrukcióban (c), illetve más igék előfordulásait más konstrukciókban (d). Mivel az itt bemutatott kutatás során nem általában bármilyen beférkőző mintázatban vizsgáltuk az igék megjelenését, hanem specifikusan az igekötővel rendelkező főnévi igenevek esetében, a teljes alapsokaságot ezeknek a főnévi igeneves adatoknak a gyakorisága jelentette a vizsgálatunkban, ezt tekintettük tehát a korpuszbeli (vizsgált) konstrukciók számának, és ebben az alapsokaságban értelmeztük a beférkőzés kialakulását igék mentén.

A tábla megmutatja az adott ige és a konstrukció együttes előfordulásait (a), de kalkulálható belőle a várható együttes előfordulások mértéke is az ige korpuszbeli gyakorisága alapján, azt feltételezve, hogy az ige adott konstrukcióban való gyakorisága azonos más konstrukciókban való gyakoriságával¹¹ (az ige összes előfordulásának és a konstrukció összes előfordulásának szorzata elosztva az összes ige és az összes konstrukció előfordulásainak összegével, lásd Levshina 2015: 228). Az asszociálódás alapelve, hogy a tényleges gyakoriság nagyobb, mint a várható gyakoriság. Természetesen annak a pusztá megfigyelése, hogy a tényleges és a várható gyakoriság eltér, még nem jelenti, hogy az asszociálódás ténye szignifikánsan igazolható is. A beférkőzés kapcsán például a mérések azt mutatják, hogy a *talál* ige (például *meg találja kérdezni*) esetében a gyakoriság és a várható gyakoriság között kicsi az eltérés az előbbi javára ($404 > 338,90$), mégis erősen vonzott kollexémának mutatja az elvégzett elemzés (lásd később). A *kelletik* ige esetében viszont (például *fel kelletik jegyezni*), ahol hasonlóan kisléptékű az eltérés ($10 > 6,20$), a vonzás ténye nem szignifikáns (lásd később). Vagyis a kétféle gyakoriság közötti különbség pusztá megfigyelése még nem jelenti, hogy az adott igét a konstrukció vonzott vagy taszított kollexémájában tekinthetjük-e.

Éppen ezért az asszociálódásnak statisztikai tesztelési lehetőségei vannak, ezek jól ismertek a kollokációelemzésben (lásd Evert 2009). Ám a számítások egy része normál eloszlást és homogén varianciát feltételez az alapsokaságra

¹¹ A feltételezés, amely a várható gyakoriság számításának háttérében áll, egyúttal azt is jelenti, hogy a vizsgált konstrukciónak nincs semmilyen hatása az ige korpuszbeli disztribúciójára. A statisztikai mérés szempontjából ez az a nullhipotézis, amelynek a tarthatóságát a teszt vizsgálja: a kapott p-értékek annak a valószínűségét fogják megmutatni, hibázunk-e, ha elvetjük ezt a nullhipotézist, és az alternatív hipotézisnek (miszerint a beférkőzés konstrukciója hatással van az ige korpuszban megfigyelhető viselkedésére) adunk igazat.

(például t -érték vagy z -érték), más részük érzékeny az alacsony előfordulási adatokra, ezeknek kiemelt erősséget tulajdonít (például a χ^2 -próba vagy a log-likelihood teszt, lásd Stefanowitsch–Gries 2003: 217–218). Mivel a korpuszbeli eloszlás általában nem normál jellegű, és mivel sok esetben igen alacsony gyakorisággal jelennek meg egyes szavak a vizsgált konstrukcióban, a potenciális koleszmák feltérképezéséhez a Fisher–Yates Egzakt tesztet – a továbbiakban FYE – javasolják a módszer kidolgozói (Stefanowitsch–Gries 2003: 218). Az FYE előnye, hogy nem érzékeny az alacsony gyakoriságokra, és nincsen semmilyen előfeltevése a vizsgálni kívánt alapsokaság méretéről és jellegéről. Mint minden kontingenciateszt, nullhipotézise, hogy nincs semmilyen asszociáció a konstrukció és a szó között, eredménye pedig egy p -érték, amely minden vizsgált szó esetében segít eldönteni a nullhipotézis fenntarthatóságát vagy elvethetőségét. Kiemelt előnye a FYE-tesztnek, hogy a p -érték interpretálható egyúttal az asszociáció erősségeként is (Stefanowitsch 2013), tehát minél kisebb értéket kapunk a teszt eredményeként, (i) annál inkább valószínű, hogy az adott szó asszociálódik a konstrukcióval, és (ii) annál erősebbnek tekinthető ez az asszociáció.¹² Az értelmezés további megkönnyítése céljából a kapott p -értékek tízes alapú logaritmikus transzformációját is célszerű elvégezni, ez ugyanis a nagyon alacsony értékeket (tehát a nagyon erősnek tekinthető asszociációkat) magas értékekké alakítja, és a -végtelen – +végtelen skálán helyezi el az adatokat (Gries–Hampe–Schönefeld 2005, Levshina 2015: 232–233).

3. Adatok és az elemzés menete

A kutatásunk során igyekeztünk minőségi és mennyiségi szempontokat egyaránt figyelembe venni: a nyelvtechnológia adta lehetőségeket kihasználva több millió adatponttal dolgoztunk, de a munkafolyamat több pontján adat-tisztító lépéseket iktattunk be – gépi és kézi módszereket egyaránt –, hogy

¹² A kollostrukciós elemzés immár két évtizedes történetében több kritika is felmerült mind a FYE-teszt alkalmazhatóságával, mind a p -érték erősségként való értelmezésével szemben. Schmid és Küchenhoff (2013) például elutasítja az utóbbit, ám Gries (2015) szerint, noha a p -érték valóban az asszociáció szignifikáns jellegét mutatja, nem pedig a hatás erősségét, korrelál azzal, ezért alkalmazható a koleszmák sorrendezésére is. Gries és Durrant (2020: 144) szerint ráadásul a FYE-teszt p -értéke erősen korrelál a log-likelihood-teszt G^2 értékével. Megjegyzendő, hogy legújabbban a módszer egyik létrehozója, Stefan Th. Gries már egy multidimenziós változatot tart alkalmazhatónak (Gries 2019), ezzel a fejlesztéssel azonban ebben a tanulmányban nem foglalkozunk.

minél tisztább legyen az elemzendő anyag. Elsőként bemutatjuk a felhasznált korpuszt, majd az adatgyűjtési folyamat főbb lépéseit és az adatelemzési módszert ismertetjük.¹³

3.1. A felhasznált korpusz

A kutatásunk alapja az MNSZ2-UD korpusz (Kalivoda et al. 2024). Ez a megjelenés előtt álló korpusz az 1,04 milliárd szavas, 20–21. századi szövegeket tartalmazó Magyar Nemzeti Szövegtár 2 legújabb változata. Számos korpusztisztítási lépésen esett át, amelyek közül a legfontosabbak a paragrafus-szintű duplumszűrés, a nem magyar nyelvű bekezdések automatikus azonosítása és törlése, valamint az 500 tokennél hosszabb mondatok kiszűrése. Az utóbbiak jellemzően nem valódi mondatok, hanem kódrészletek, vagy hibásan feldolgozott weboldalak. Ezáltal a szöveg minősége jobb, de a korpusz mérete kisebb az MNSZ2-höz képest (776,9 millió szövegszó). A korpusz új elemzési szintekkel is gazdagodott: az emMorph (Novák et al. 2016) által létrehozott, részletes morfológiai elemzéssel, valamint a Stanzával (Qi et al. 2020) készített függőségi elemzéssel, amely a Universal Dependencies formalizmusát követi (de Marneffe et al. 2021).

A kutatásunkhoz főként arra volt szükség, hogy a lehető legpontosabban tudjuk azonosítani azt a főnévi igenevet, amelyhez egy-egy elvált igekötő tartozik. Ezért úgy döntöttünk, hogy a korpusz szövegeit végigelemezzük az emPreverb eszközzel is (Pethő et al. 2022).¹⁴ Ez a szabályalapú algoritmus kifejezetten a magyar igék és igekötők összekapcsolására készült, és ezzel érhető el a legpontosabb annotáció Pethő et al. (2022: 86–87) kiértékelése alapján.

¹³ A felhasznált korpusz és az adatgyűjtési folyamat jelentős része megegyezik Modrián-Horváth – Kalivoda – Tanos (ebben a kötetben) munkájával. Ennek az az oka, hogy az ott vizsgált passzív jelentésű konstrukciók között is van olyan, amelyik beférkőző (például *el van döntve*), tehát itt és az említett tanulmányban technikailag ugyanolyan keresést igénylő szerkezeteket vizsgálunk.

¹⁴ Dömötör Andrea felveti a bírálatában azt a kérdést, hogy miért a morfoszintaktikai címkekre támaszkodó emPreverb eszköz mellett döntöttünk, ha a korpuszban függőségi elemzés is rendelkezésre állt, amely szintaktikai relációk vizsgálatára triviális választás lehetne. A kérdés nagyon jó, és valóban magyarázatra szorul a döntésünk. Ahogy a bíráló is gyanítja, a függőségi annotáció minősége sajnos ingadozó, és az infinitívuszi szerkezetek tekintetében csak a nagyon gyakori esetekre megbízható, lásd még Kalivoda–Prószék (2023: 225–226). Ezért a konkrét probléma megoldására az emPreverb annotációja bizonyult hasznosabbnak.

3.2. Az adatgyűjtés és -tisztítás folyamata

A korpuszból lekérdeztünk minden olyan szóalakot, amely infinitívuszi toldalékkal szerepelt, ügyelve arra, hogy a magyarban a főnévi igenév személyragos is lehet (-*nom/nem/nöm*, -*nod/ned/nöd* stb.). Minden szóalak esetében megvizsgáltuk, hogy tartozik-e hozzá prefixált vagy elvált igekötő. Az utóbbit az emPreverb annotáció segítségével találtuk meg. Ennek a lépésnek az volt a célja, hogy az igekötős igelemmák egyforma reprezentációt kapjanak, függetlenül az igekötő pozíciójától (például az *imádja végignézni* és a *végig kellett néznie* esetében a főnévi igenév lemmája egységesen *végig#néz* lett). Egy más alakvariánsainak tekintettük és azonos alakra normalizáltuk a következő igekötőket: *belé* → *bele*, *bé* → *be*, *föl* → *fel*, *föül* → *felül*, *fönn* → *fenn*, *odább* → *odébb*, *reá* → *rá*, *teli* → *tele*.

Így kinyertük a korpuszból a főnévi igenevek összes példányát az azokat tartalmazó teljes mondatokkal együtt. Ezt az adathalmazt tovább szűrtük a beférkőző szerkezetekre, amelyekből kigyűjtöttük a beférkőzőként adathozható lemmák (*kell*, *szabad*, *akar* stb.) listáját. Minden lemmához kétféle gyakorisági értéket rendeltünk a korpusz alapján: egyrészt azt, hogy hányszor fordult elő beférkőző szerkezetben, másrészt azt, hogy általában véve hányszor fordult elő főnévi igenévvel. Ezáltal kizártuk a főnévi vonzattal más jelentésben/olvasatban előforduló példányokat (például *levest mer*, *bírja a búráját*). A *szerelem* lemmához tartozó találatokat két részre bontottuk: *szerelem* illetve *szerelemne* lemmák találataira. Ezt a döntést az indokolta, hogy mind az intuíciónk, mind a szakirodalomból ismert korábbi megfigyelések alapján meglehetősen különböző eloszlási mintákra számítottunk a *szerelem*, illetve a *szerelemne* esetében.

Annak érdekében, hogy minél kevesebb hamis pozitív találatunk legyen, a potenciális segédigétől legfeljebb három szónyi távolságra kerestünk infinitívuszi vonzatot, a két elem közötti további finit igéket, kötőszavakat és írásjeleket kizárva. Szintén a minél tisztább adathalmaz érdekében döntöttünk amellett, hogy csak azokat az adatsorokat őrizzük meg további elemzésre, amelyekben a vizsgált lemma beférkőzőként minimum kétszer, összesen minimum ötször fordult elő. Az ennél ritkább lemmák elemzése hibás, vagy félregépeltek voltak. Az adatokat ezután manuálisan ellenőriztük, és kizártuk a teljes egészében hibás annotáció miatt nyert találatokat (például *hív*, *tér*, *érkezik*, *utazik*, *ül* mint beférkőző ige),¹⁵ valamint az

¹⁵ Például *meg is hívja vacsorázni* (#777015923), *amíg vissza nem térnek dolgozni* (#423490340), *de még meg sem érkezünk enni* (#228810794).

egy-egy jó találattal rendelkező *vágyik* és *kéretik* igéket (ez utóbbiak az előzetes vizsgálatok szerint a taszított kategóriába tartoztak)¹⁶, és egyesítettük azokat a lemmákat, melyek gyakori gépelési hiba (például *akad>akar*; *kellett, kall, kél>kell*) vagy hibás annotáció miatt (például *mered>mer*, *tessen>tetszik*, *szoktat>szok(ik)*) külön lemmaként jelentek meg az adatsorban. Az így előállt adattáblán végeztük el a kollostrukciós elemzést, amelyet a következő fejezetben ismertetünk.

3.3. Az adatok elemzése

A beférkőzésre való hajlandóság statisztikai tesztelését az RStudio (v.4.1.0, R Core Team 2021) statisztikai elemző szoftverrel végeztük, a *collostructions* package (v.0.1.0) *collex()* függvényével (Flach 2017), az FYE-teszt opciót választva. A függvény automatikusan elvégzi a logaritmikus transzformációt is, kiértékeli a vizsgált szavak asszociálódásának statisztikai szignifikanciáját, és sorba rendezi azokat a transzformált p-érték mentén.

4. Eredmények

Az alábbiakban mind a 42, a beférkőzés fenti kritériumainak megfelelő ige asszociációs adatait közöljük a mérés eredményeként. A táblázat első oszlopa a potenciális kollexémákat tartalmazza, a második ezen igék korpuszbeli gyakoriságát, majd a konstrukcióbeli gyakoriságot mutatja a harmadik oszlop. A negyedik a várható gyakoriságok adatait ismerteti. Az elemzés szempontjából talán a legszemléletesebb az ötödik oszlop, amely a kollexéma vonzott (*attr*), illetve taszított (*rep*) státusát ismerteti. A legfontosabb pedig a hatodik és a hetedik oszlop: ezek mutatják a logaritmikusan transzformált p-értéket (az asszociáció erősségeként, ahol *Inf* jelzi a végtelen erősséget, *-Inf* pedig a negatív végtelen értéket), valamint az asszociáció szignifikáns jellegét (a csillagok száma arányos a szignifikanciaküszöb szigorodásával,¹⁷ „ns” jelöli a nem szignifikáns eredményt).

¹⁶ A találatok: A *zínfláció* mértékét pedig le *kéretik* vonni belőle (#1056477830); *meg vágyta írni az - enciklopédikusnak tervelt - Nagy Kaka Könyvet* (#299765553).

¹⁷ **** = $p < 0,00001$, *** = $p < 0,0001$, ** = $p < 0,001$, * = $p < 0,01$, = $p < 0,05$

2. táblázat. A beférkőzés vonzott és taszított kolexémái a korpusz adatai alapján

Kolexéma	Gyakoriság a korpuszban	Gyakoriság a konstrukcióban	Várható gyakoriság	Státus	Az asszociáció erőssége	Szignifikancia
kell	1 960 680	506 794	225 325,00	attr	Inf	****
tud	943 760	162 624	108 458,60	attr	Inf	****
van	962 391	139 980	110 599,80	attr	Inf	****
akar	460 166	90 546	52 883,10	attr	Inf	****
fog	588 318	87 554	67 610,60	attr	Inf	****
mer	43 105	6806	4953,70	attr	156,97068	****
talál	2949	404	338,90	attr	3,86904	***
kelletik	54	10	6,20	attr	1,06649	ns
merészkedik	71	4	8,20	rep	-1,10782	ns
kívánczik	86	4	9,90	rep	-1,60466	*
kegyeskedik	232	2	26,70	rep	-9,61656	****
számít	261	2	30,00	rep	-11,05485	****
méltóztatik	959	49	110,20	rep	-11,27681	****
merészel	1793	108	206,10	rep	-14,68721	****
akarózik	560	4	64,40	rep	-23,60525	****
iparkodik	764	3	87,80	rep	-35,28484	****
bátorkodik	823	4	94,60	rep	-36,88772	****
tartozik	832	2	95,60	rep	-40,33980	****
próbálkozik	918	5	105,50	rep	-40,35736	****
gondol	1667	2	191,60	rep	-84,01679	****
bír	17 222	1159	1979,20	rep	-97,98818	****
óhajt	4142	92	476,00	rep	-110,87301	****
szeretne	129 449	11 915	14 876,50	rep	-157,84642	****
szokik	80 017	6854	9195,70	rep	-161,83583	****
szíveskedik	3167	2	364,00	rep	-163,01145	****
szándékozik	7151	167	821,80	rep	-184,55379	****
tetszik	22 918	1323	2633,80	rep	-193,75257	****
kényszerül	6688	37	768,60	rep	-289,14550	****
kíván	139 341	10 445	16 013,30	rep	-Inf	****

Kollexéma	Gyakoriság a korpuszban	Gyakoriság a konstrukcióban	Várható gyakoriság	Státusz	Az asszociáció erőssége	Szignifikancia
próbál	139 250	628	16 002,90	rep	-Inf	****
szabad	48 418	502	5564,30	rep	-Inf	****
sikerül	162 244	101	18 645,40	rep	-Inf	****
szeret	9653	98	1109,30	rep	-Inf	****
hagy	17 566	83	2018,70	rep	-Inf	****
enged	11 451	57	1316,00	rep	-Inf	****
illik	9344	47	1073,80	rep	-Inf	****
készül	10 704	42	1230,10	rep	-Inf	****
lesz	80 117	19	9207,20	rep	-Inf	****
tervez	7900	18	907,90	rep	-Inf	****
látszik	14 491	12	1665,30	rep	-Inf	****
vél	10 487	12	1205,20	rep	-Inf	****
segít	18 987	5	2182,00	rep	-Inf	****

A táblázat értelmezéséhez mindenekelőtt fontos kiemelni, hogy csupán az első nyolc ige azonosítható vonzott kollexémaként, sorrendjük szerint egyre gyengébben asszociálódnak a konstrukcióval. Egyszerűbben fogalmazva, a *kell* szeret a leginkább beférkőzni igekötő és infinitívusz közé, a *talál* esetében a legcsekélyebb a beférkőzésre való hajlandóság (ám még mindig erősen szignifikáns ez a tulajdonsága), a *kelletik*-ről pedig nem állapítható meg, hogy szignifikáns kollexémája a beférkőzésnek, noha az adatok az ige asszociáltságát mutatják. A *van* lexéma egyaránt tartalmazza a *lehet*, *van* és *volna* töveket is.¹⁸

¹⁸ Az adatelemzés utolsó fázisában észleltük, hogy ebben az esetben is érdemes lett volna újramegmentizálni az eredeti adatokat, és ezt a három tövet egyúttal külön lemmaként is kezelni, ugyanúgy, ahogy megkülönböztettük a *szeret* és *szeretne* lemmákat is. A szakirodalom és az intuíciónk alapján arra számítottunk, hogy beférkőzőként a *lehet* alak jelenik meg dominánsan. Ezt az adatkezelési megfontolást a kutatás jövőbeli szakaszában érvényesíteni fogjuk. Itt jegyeznénk meg azt is, hogy amikor a *volna* szó igei komplexumot alkotott más főigékkel, akkor az adott találatot a főige találataihoz számoltuk (például *kellott volna* → *kell*, *akartam volna* → *akar*). A *volna* lemma csak akkor jelent meg az adataink között, amikor a főige hibásan volt írva, így nem kapott a korpuszban automatikus elemzést (például *szerett volna*, a *szeretett volna* helyett).

A táblázat nagyobbik része a taszított kolerémákat mutatja, vagyis azokat, amelyek megjelennek ugyan beférkőző szerkezetben, de meglehetősen ritkán; a beférkőzés inkább atipikus tulajdonságuk. Sorrendjük növekvő aszerint, mennyire erős a taszítottság mértéke, és mennyire szignifikáns ez az eredmény. A *merészkedik* ugyan kimutathatóan nem preferálja a beférkőzést, ám ez az eredmény nem tekinthető statisztikailag szignifikánsnak. A többi ige esetében viszont kvantitatív módon is igazolható, hogy nem jellemző rájuk a beférkőzés. Mindez azonban korántsem jelenti, hogy ne jelennének meg beférkőzésben: a *szeretne* és a *kíván* például 10 000 adatpont fölötti gyakoriságot mutat ebben a mintázatban, ez azonban elenyésző a teljes korpuszbeli (azaz más infinitívuszi vonzatú konstrukciókhoz kapcsolódó) gyakoriságukhoz képest. Ezzel összevetve számos ige (például *akarózik*, *tartozik*, *gondol*, *szíveskedik*, *segít*) megjelenése beférkőzésben legfeljebb 5 előfordulással adatolható. Fontos tehát ismét megjegyezni, hogy a konstrukcióban való előfordulások száma nem, vagy nem feltétlenül korrelál a konstrukcióval való asszociálódás mértékével. Az is lényeges, hogy a teszt lehetővé teszi az asszociálódás és a taszítottság közötti egyértelmű határ megállapítását, de ez a határ mesterséges, az elemzés eredményeként felállítható „vonal”, hiszen a vizsgált igék egy kontinuumon rendezhetők el a konstrukcióval való asszociálódás szempontjából.

5. Az eredmények értelmezése és értékelése

5.1. A vizsgálat eredményeinek értékelése

5.1.1. Vonzott és taszított lexémák és a konstrukciós jelentés

Az egyszerű kolerémaelemzés lényegi vonása, hogy nem az abszolút gyakoriságot veszi figyelembe annak megítélésekor, hogy egy adott lexéma mennyire asszociálódik a vizsgált konstrukcióval. Könnyű belátni, hogy egy eleve gyakori igének eleve nagyobb az esélye erre: például a *kell* csaknem kétmillió adatával (1960680 találat) sokkal nagyobb valószínűséggel vesz részt beférkőzésben, mint a korpuszban mindössze 86 előfordulással adatolható *kíváncozik*. Felvethető, hogy az adott ige beférkőző adatait arányosítani lehet az ige összes előfordulásához, és az így kapott arányszámmal jellemezhető a beférkőzési hajlandóság. Például a *mer* esetében ez az arány 15,79% (6806/43105), míg a *szabad* esetében (igeként annotálva) 1,04% (502/48418), tehát a hasonló gyakoriságú igék közül az egyik nagy arányban produkál

beférkőzést, a másik jóval kisebb arányban. Ám például a *szeretne* esetében ez az arány 9,20% (11915/129449), vagyis megközelíti a tipikusan beférkőző *mer* arányát, mégis a statisztikai mérés azt mutatja, hogy a *szeretne* nem asszociálódik a beférkőzés jelenségével (ami – mint látjuk – nem jelenti azt, hogy ne fordulna elő gyakran ilyen mintázatban).

A vizsgált konstrukcióval erősen asszociált lexémák feltárásának az az egyik célja, hogy a talált lexémák jelentése alapján következtethessünk a konstrukció jelentésére, mivel az asszociáció erőssége jó indikátora a szemantikai kompatibilitásnak (vö. Stefanowitsch–Gries 2003: 229; a szerzők kiemelik azt is, hogy erősen poliszém konstrukciók megfelelő leírásához részletesebb elemzésre van szükség). Stefanowitsch–Gries (2003) például az angol ditranzitív konstrukció vizsgálatakor azt találták, hogy a legerősebben vonzott lexémák (álljon itt példaként az első öt: *give* 'ad', *tell* 'mond', *send* 'küld', *ask* 'kér, kérdez', *show* 'mutat') 'transzfer' jelentéssel rendelkeznek. A beférkőző igei szerkezet kolerémavizsgálata kizárólag a következő igéket találta szignifikánsan beférkőzőnek: *kell*, *tud*, *akar*, *fog*, *mer* és *talál*, legutóbbit csak alacsonyabb szignifikanciaküszöbvel. A legerősebben vonzott kolerémák a *kell*, *tud*, *akar* és *fog*, amelyek végtelen asszociációs erősséget mutatnak (azaz a pusztán véletlen korreláció valószínűsége annyira alacsony, hogy ennek tízes alapú logaritmus a végtelenhez közelít). Mind a négy ilyen lexéma modális jelentésű, illetve a *fog* esetében komplex temporális jelentésről beszélhetünk. Ez alapján arra következtethetünk, hogy a beférkőző konstrukció az adott infinitívuszi alakban megjelenő ige¹⁹ modalizálására szolgál. Az alacsonyabb szignifikanciaküszöb mellett szignifikánsan beférkőző *mer* és a nem szignifikánsan vonzott *talál* lexémák jelentését szintén tekinthetjük modálisnak (megjegyezve, hogy a *talál* használata szinte kizárólag a szépirodalmi alkorpuszhoz kötődik). A *mer* ige megfelelő olvasatának prototipikus jelentése feltűnően komplex, szerephez jut benne egy (társadalmi-)normatív elvárásrendszer (ez a deontikus modalitásra jellemző), valamint a trajektorhoz köthető volitív jelentéskomponens.

Feltűnő a vonzott kolerémák alacsony száma és a kisszámú vonzott koleréma asszociációs erőssége csakúgy, mint a taszított lexémák taszítottságának erőssége: a vizsgált konstrukció valóban csak néhány lexémával fordul elő szignifikáns relatív gyakorisággal. Ezenfelül az összes felmerült

¹⁹ Nem kívánunk állást foglalni az infinitívuszi alakok szófajiségével kapcsolatban; amennyiben a beférkőző igét segédigének tekintjük, az infinitívuszi alakot a ragozási paradigma részeként kezeljük.

lexéma száma is rendkívül alacsony. A többi – nulla beférkőző előfordulással rendelkező – ige taszítottságának mértéke nem állapítható meg további számítások nélkül (vö. Stefanowitsch 2006): ehhez szintén a fent ismertett kontingenciátábla adataira, valamint a 0 előfordulású igék gyakorisági adataira van szükség. Ez alapján egy beférkőző előfordulás nélküli gyakori igének (például **be megy főzni az ebédet*) az infinitívusszal megjelenő lexéma korpuszbeli gyakorisága alapján jóval magasabb lesz a taszítottsága, mint egy alacsony gyakoriságú lexémának.

Az ún. taszított lexémák adatainak értelmezésénél az adott vizsgálati kontextusban különösen nagy körültekintésre van szükség. Az eredmények értelmezéséhez fontos szem előtt tartani, hogy az egyszerű kolexémaelemzés nullhipotézise a lexémák random megjelenése a konstrukciókban. Azonban egyrészt a legtöbb ige nem rendelkezik beférkőző használattal, másrészt a beférkőzőként is használható igék asszociálódnak más szórendi konstrukciókkal is; ez megnehezíti a beférkőzőnek tekinthető igék elválasztását a más taszított igéktől. Ha azonban figyelembe vesszük a megfigyelt értékeknek az elvárt értékektől való eltérését, árnyaltabb képet kapunk. A 2. táblázat adatai szerint például a *próbál* és a *kíván* lexémák elvárt gyakorisága egyaránt 16 000 körüli, de a *próbál* megfigyelt értékének mintegy tizenhatszorosa a *kíván* igére vonatkozó megfigyelt érték, noha a szignifikanciasztek alapján mindkét esetben taszított lexémáról van szó.

Azoknak a lexémáknak az esetében, amelyeknél a megfigyelt adatok megközelítik az elvárt adatokat, joggal állíthatjuk, hogy beférkőző, de a konstrukcióval nem szignifikánsan asszociált igékről van szó. Ilyenek különösen a nagyobb gyakoriságú *kelletik*, *merészel*, *bír*, *óhajt*, *szeretne*, *szokik/szokott*, *szándékozik*, *tetszik*, *kíván*. Ezek között akadnak a magyarországi sztenderd köznyelvben kevésbé használatos igék is (például *kelletik*, lásd a (6) példát).

(6) néha át kellett hívni a Kokker gyári őrt (#35339445)

Releváns meglátásokkal gazdagíthatnak tehát a táblázatban megjelenő igék: a „beférkőző” igeként számontartott igéknek a (köznyelvben) nem beférkőző igéktől való elkülönülése fokozati jellegű, míg a szignifikánsan vonzott lexémák, beférkőző igék kategóriájának „centrumának” megállapítása a FYE teszttel egyértelmű eredményekkel szolgált.

Érdeemes szemügyre venni a táblázat vége felé szereplő igéknek a beférkőző konstrukcióban való megjelenését is. Az ’infinitívusz + ige’ kapcsolat a két folyamat típusjelentésű szó valamilyen mértékű konceptuális

integrációját fejezi ki. Az aktuálisan beférkőzőként való használat nagyobb integrációra utalhat (Imrényi 2013 itt segédigei használatról beszél), míg a nem beférkőző használat nagyobb konceptuális autonómiára (vö. a (7-8) példamondat különbségeit). A példákban egy kétcselekvős igét, az *enged* igét látjuk; ezt az ige típust a szakirodalom jellemzően kizárja a segédigék vizsgálati köréből (vö. például Lengyel 2000a, b; Kálmán C. et al. 1989 kontrollcsoportként tartja meg őket), holott funkcionálisan igenem-változást tudnak a típusba tartozó igék kieszközölni, amely például a műveltető képzővel analóg funkció.

- (7) Voltam a Computer Centre-ben most es az egyik srac mutatott olyan WWW oldalt, amiről ha *le engedsz tolni* egy cookie-t, akkor lecsereli az autoexec.bat-odat, meg a config.sys-edet valami olyanra, hogy kerdes nelkul format c: /u >null, vagy hasonlo.²⁰
- (8) El kellene döntened, hogy *engedsz letölteni* fájlokat vagy sem. Mert egyszerre engedni, meg nem engedni, az nem megy.²¹

5.1.2. Szórendi változatok, amelyek kihatással lehetnek a konstrukció gyakoriságára

A kollostrukciós vizsgálatok tárgya a szakirodalomban többnyire nem szórendi, hanem egyéb (morfo)szintaktikai szerkezet, nem számít egyedinek azonban a szórendnek a vizsgálatba történő bevonása sem. A magyar beférkőző igei konstrukciók az eredmények alapján a többi igétől statisztikailag jelentősen elkülönülő kategóriát alkotnak, ami indokolja a szórendi szempontnak a vizsgálatba történő bevonását. Nem hagyhatjuk ugyanakkor figyelmen kívül, hogy a magyar nyelv egyes speciális szórendi konstrukciói (így különösen a felülíró szerkezetek, vö. Imrényi 2017, illetve a különböző mondatmodalitásokkal korpuszvizsgálatok szerint erőteljesen korreláló szórendi változatok, vö. Péteri 2015) kizárják a beférkőző szórendi megvalósulást. A következő példák három felülíró szerkezetet tartalmaznak: egy tagadósós (9), egy kérdősós (10) és egy azonosító operátort (11), amelyeket közvetlenül követ a ragozott ige, így a beférkőző használat nem valósulhat meg még igeikötős főnévi igenév esetében sem.

- (9) *Nem óhajtom megismételni.* (#9258600)

²⁰ <http://www.graph.hix.hu/arch/?page=issue&issueid=50101> [2024. 02. 07.]

²¹ <https://prog.hu/tudastar/15688/php-download-manager> [2024. 02. 07.]

- (10) kérdezze meg az urakat, *mikor óhajtják befejezni* a közelharcot.
(#2117866)
- (11) De én *itt óhajtom meginni*. (#9180090)

Mindez a kollostrukciós vizsgálati nézőpontból annyit jelent, hogy bizonyos igék erőteljesebben asszociálódhatnak a megfelelő (felülíró, felszólító stb.) szerkezetekkel, mint a semleges mondatra jellemző protoállítás szórendi konstrukciójával; ez az egyik oka annak, hogy a „taszított”, de az elvárt gyakorisághoz közeli megfigyelt értékeket mutató lexémákat nem zárhatjuk ki a beférkőző igék közül, csupán megállapíthatjuk, hogy kevésbé szoros asszociációt mutatnak a vizsgált szerkezettel.

A fenti kontextusokon túl említést érdemelnek azok a szerkezetek is, amelyek – az igekötő és az infinitívusz által alkotott kerethez hasonlóan – mintegy körbefogják a ragozott igéket, de a vizsgálatban használt technikai eszközök nem annotálták őket összetartozó egységekként. Ilyen például a határozószó és infinitívusz közé való beférkőzés (12), a lexikalizálódó igekötők esete (13), vagy az egyéb összetartozó konstrukciók közé való beférkőzés (14), ahol nem az infinitívuszon „csügg” a ragozott ige, hanem annak bővítményén.

- (12) *körbe kezdtem lődözni* (#16308551)
- (13) Ez csak oktatófilm, *szénné kell egyszerűsíteni*, hogy mindenki értse
(#1139942000)
- (14) Ekkor már az is *fonáknak kezdett tűnni*, hogy a gyerekek csak az
anyjukat szólították maminak (#234525)

5.2. Az eredmények összevetése a magyar szakirodalom eredményeivel

Amint a 2. pontban kifejtettük, a segédigék listáját a legkülönbélebb kritériumok és kritériumkombinációk mentén állították fel a korábbi kutatásokban. A jelen vizsgálat eredményeivel a szakirodalomban szereplő listák közül a legteljesebb átfedést Tolcsvai Nagy (2009) szemantikai és grammatikalizációs kritériumokra épülő elemzése mutatta, ami annyiban meglepő, hogy a hivatkozott tanulmány deklaráltan nem a formális-szintaktikai kritériumokra épít, míg a kollostrukciós vizsgálat korpuszelemzésre épült és statisztikai eszközökkel dolgozott. A hivatkozott tanulmányban szereplő („nem teljes” lista, lásd Tolcsvai Nagy 2009: 385) központi segédigék a következők: *tud, bír, akar, mer, fog, kell, lehet*. Ezek közül a jelen kollostrukciós vizsgálat

csupán a *bír* esetében nem állapított meg szignifikáns asszociáltságot, és a *talál* esetében volt szignifikáns a vonzottság mértéke alacsonyabb szignifikanciaküszöb mellett.

A segédigeség több tulajdonságát korpuszalapon vizsgáló korábbi tanulmány (Modrián-Horváth 2009) gyakran beférkőzőként ítélte meg a következő igéket a beférkőző használat százalékos megoszlása alapján: *kell, lehet, tetszik, mer, akar, tud, fog, talál, szokott, kíván, szeretne, méltóztatik*. A listában aláhúzással jelöltük azokat a lexémákat, amelyek a Modrián-Horváth (2009) mellékletében szereplő táblázat alapján 10% feletti beférkőző használatlaltal fordultak elő; ezek közül csupán a *tetszik* nem szerepel a jelen vizsgálat alapján szignifikánsan vonzott lexémák listáján.

A korpusznyelvészeti módszerrel dolgozó tanulmányok közül szintén hasonló eredményekre jutott a beférkőző tulajdonság tekintetében Kalivoda-Prószéky (2024). Ez a tanulmány egy, a jelen vizsgálatától teljesen eltérő módszerrel, hierarchikus agglomeratív klaszterezéssel vizsgálta az infinitívuszt vonzó igéket, Kálmán C. et al. (1989) elemzési szempontjait alkalmazva egy modern nagyméretű korpuszon. Az ezzel a tanulmánnyal való összevetés érdekes felismerése, hogy ott a fókuszos és tagadó (Imrényi 2017 terminológiájában felülíró), kérdő, felkiáltó és komplex mondatot tartalmazó kontextusok kiszűrése után is csak négy ige tűnt ki a beférkőző tulajdonság alapján: *akar, fog, kell* és *kíván*. Ez arra utalhat, hogy az említett kontextusokban való megjelenés nem befolyásolja számottevően az eredményeket a szignifikáns kölcsönös kollostrukciós vonzottság tekintetében.

A beférkőzés lehetőségének megítélése nem tartozott a jelen vizsgálat kiinduló kérdésfeltevései közé, mégis feltételezhetjük, hogy azon igék esetében, amelyeknél az adatsorban szereplő „megfigyelt gyakoriság” nem maradt el számottevően az „elvárt gyakoriság” értékétől, a beférkőzést alapvetően megengedő igékről van szó. Ezek az igék, az *akar, bír, fog, kell, kelletik, kíván, lehet, mer, merészel, óhajt, szándékozik, szeretne, szokik/szokott, talál, tetszik, tud*, nagyfokú egyezést mutatnak a Kálmán C. et al. (1989) által azonosított csüggő-beférkőző igékkel: *akar, bír, fog, kell, kíván, lehet, mer, óhajt, szeretne, szokott, talál, tetszik, tud* és *szabad*. (Az eltéréseket félkövér szedéssel jelöltük; a *szándékozik* esetében nem volt nagyságrendi különbség az elvárt és a megfigyelt értékek között, a *szabad* lexémánál kicsit több mint tizenegyszerese volt az elvárt érték a megfigyelt értéknek.)

6. Összegzés és kitekintés

A vizsgálat célja az 'igekötő + ige + infinitívusz' beférkőző szerkezettel leg-erősebben asszociált igék megállapítása volt korpuszelemzésre épülő vizsgálat segítségével. Ehhez egy kollostrukciós elemzést, azon belül is az egyszerű kolexémaelemzés módszerét alkalmaztuk. A vizsgálat speciális követelményeket támasztott az elemzendő korpusz felé, mivel olyan annotációs eszköz felhasználására volt szükség, amely megbízhatóan azonosítja az egymástól távolabb elhelyezkedő igekötőt és igenévi alakokat; ezt az MNSZ2-UD változat biztosította.

Az egyszerű kolexémaelemzés a szakirodalomban formális-disztribúciós alapon leírtakhoz képest (vö. Kálmán C. et al. 1989) kisszámú lexémát azonosított szignifikánsan a beférkőző szerkezethez vonzott igeként; ezek a *kell*, *tud*, *van/lehet*, *akar*, *fog*, *mer* és *talál* igék. A beférkőző szerkezettel a vizsgálat szerint szorosan asszociált igék szinte teljes egyezést mutatnak a szemantikai és grammatikalizációs alapon definiált segédige-listával (Tolcsvai Nagy 2009), és a korábban korpuszalapon, százalékos érték alapján nyert igelista 10% feletti beférkőzést mutató nyolc igéjével (Modrián-Horváth 2009; a 10%-os határ megállapítása utólagos az adatsor alapján).

A szerkezettel kevésbé szorosan asszociálódó, további beférkőző igék feltárása a vizsgálatban alkalmazható szignifikanciaszűkítők mellett nem kivitelezhető egzakt módon, noha a magyar nyelv szórendi jellegzetességei szükségessé tennének ezzel kapcsolatos további vizsgálatokat.

A kollostrukciós elemzés egy másik típusa, a kovariációs kolexémaelemzés több információt nyújthatna arról, mely beférkőző igék mely infinitívuszokkal fordulnak elő szignifikáns gyakorisággal a korpuszbeli előfordulásukhoz mérten. További vizsgálati lehetőség a Gries (2019) által javasolt rendezett *n*-esek (tuplezation) használata. Ezekkel a módszerekkel a jövőben a mostaninál is nagyobb pontossággal mutatható majd be, mely lexémák asszociálódnak jellemzően a beférkőzés jelenségével, ami végső soron a beférkőzés mint konstrukció jelentésének árnyalt leírásához is elvezethet.

Irodalom

Bajzát Tímea Borbála 2020. „Hogy csakugyan nem tudják megkülönböztetni.” A *tud* + főnévi igenév szerkezet által nyelvileg aktivált modális tartományok konstrukciós vizsgálata. *Ösvények* 2020. 5–27.

- Bajzát Tímea Borbála 2022. A premodális tartományokkal összekapcsolódó segédige/melléknév + főnévi igenév konstrukciók mondatszintű szemantikai vizsgálata. In: Tátrai Szilárd – Tolcsvai Nagy Gábor (szerk.): *A magyar mondat és kontextuális környezete*. Budapest: ELTE BTK – ELTE Eötvös Kiadó. 141–188.
- Croft, William 2001. *Radical Construction Grammar*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198299554.001.0001>
- de Marneffe, Marie-Catherine – Manning, Christopher D. – Nivre, Joakim – Zeman, Daniel 2021. Universal Dependencies. In: *Computational Linguistics* 47 (2): 255–308. https://doi.org/10.1162/coli_a_00402
- Dér Csilla Ilona 2008. *Grammatikalizáció*. Budapest: Akadémiai Kiadó.
- Desagulier, Guillaume 2014. Visualizing distances in a set of near synonyms: rather, quite, fairly, and pretty. In: Glynn, Dylan – Robinson, Justyna A. (eds): *Corpus methods for semantics: quantitative studies in polysemy and synonymy*. Amsterdam, Philadelphia: John Benjamins. 145–178. <https://doi.org/10.1075/hcp.43.06des>
- Diewald, Gabriele 1997. *Grammatikalisierung. Eine Einführung in Sein und Werden grammatischer Formen*. Tübingen: Max Niemeyer Verlag. <https://doi.org/10.1515/9783110946673>
- Flach, Suzanne 2017. *Collostructions: An R implementation for the family of collostructional methods*. R package version 0.1.0. url: www.bit.ly/sflach
- Goldberg, Adele 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199268511.001.0001>
- Gries, Stefan Th. 2015. More (old and new) misunderstandings of collostructional analysis: On Schmid and Küchenhoff (2013). *Cognitive Linguistics* 26 [3]: 505–536. <https://doi.org/10.1515/cog-2014-0092>
- Gries, Stefan Th. 2019. 15 years of collostructions. Some long overdue additions/corrections (to/of actually all sorts of corpus-linguistics measures). *International Journal of Corpus Linguistics* 24 (3): 385–412. <https://doi.org/10.1515/cog-2014-0092>
- Gries, Stefan Th. – Durrant, Philip 2020. Analyzing co-occurrence data. In: Paquot, Magali – Gries, Stefan Th. (eds.): *A practical handbook of corpus linguistics*. Cham: Springer. 141–159. https://doi.org/10.1007/978-3-030-46216-1_7

- Gries, Stefan Th. – Hampe, Beate – Schönefeld, Doris 2015. Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics* 16 (4): 635–676. <https://doi.org/10.1515/cogl.2005.16.4.635>
- Heine, Bernd 1993. *Auxiliaries: Cognitive forces and grammaticalization*. New York, Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780195083873.001.0001>
- Imrényi András 2013. A beférkőző segédigék szerkezetek függőségi nyelvtani elemzéséhez. *Magyar Nyelv* 109: 291–308.
- Imrényi András 2017. Az elemi mondat viszonyhálózata. In: Tolcsvai Nagy Gábor (szerk.): *Nyelvtan*. Budapest: Osiris Kiadó. 664–760.
- Kalivoda Ágnes 2021. *Igeköötős szerkezetek a magyarban*. Doktori (PhD) értekezés. Budapest: PPKE BTK.
- Kalivoda, Ágnes – Malouf, Robert – Ackerman, Farrell – Sass, Bálint 2024. *Building a dependency treebank from the Hungarian Gigaword Corpus*. Kézirat, megjelenés előtt.
- Kalivoda Ágnes – Palágyi László 2024. Szimultán igeköötözés és igeképzés – konstrukcióegyesítés a magyarban. In: Kugler Nóra – Modrián-Horváth Bernadett – Simon Gábor (szerk.): *Korpusznyelvészeti módszerek a kognitív nyelvészeti kutatásban*. Budapest: ELTE BTK – ELTE Eötvös Kiadó. 29–57.
- Kalivoda Ágnes – Prószéky Gábor 2023. Infinitívuszi szerkezetek a magyarban: adatvezérelt megközelítés. *Nyelvtudományi Közlemények* 119: 223–245.
- Kalivoda, Ágnes – Prószéky, Gábor 2024. Hungarian Auxiliaries Revisited. *Acta Linguistica Academica* 71 [1]: Accepted Manuscript / Online first. <https://doi.org/10.1556/2062.2023.00701>
- Kálmán C. György – Kálmán László – Nádasdy Ádám – Prószéky Gábor 1989. A magyar segédigék rendszere. *Általános Nyelvészeti Tanulmányok* XVII: 49–103.
- Kálmán László – Rádai Gábor 1996. Megszakított összetevők a magyarban konstrukciós nyelvtani szempontból. *Acta Universitatis Szegediensis* 37: 83–110.
- Kenesei István 2000. Van-e segédige a magyarban? Esettanulmány a grammatikai kategória és a vonzat fogalmáról. In: Kenesei István (szerk.): *Igei vonzatszerkezet a magyarban*. Budapest: Osiris Kiadó. 157–196.
- Lengyel Klára 2000a. A segédigék és származékaik. In: Keszler Borbála (szerk.): *Magyar grammatika*. Budapest: Nemzeti Tankönyvkiadó. 252–258.

- Lengyel Klára 2000b. *Az igenevek helye a szófaji rendszerben*. Budapest: Akadémiai Kiadó.
- Levshina, Natalia 2015. *How to do linguistics with R. Data exploration and statistical analysis*. Amsterdam, Philadelphia: John Benjamins. <https://doi.org/10.1075/z.195>
- McEnery, Tony – Brezina, Vaclav 2022. *The fundamental principles of corpus linguistics*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781107110625>
- M. Korchmáros Valéria 1997. Ige vagy segédige? In: Büky László (szerk.): *Nyíri Antal kilencvenéves*. 109–123.
- Modrián-Horváth, Bernadett 2009. Gesichtspunkte zu einer funktionalen Typologie der ungarischen Infinitiv regierenden Hilfsverben. *Acta Linguistica Hungarica* 56 (4): 405–439. <https://doi.org/10.1556/ALing.56.2009.4.3>
- Modrián-Horváth Bernadett 2020. Beférkőzés és keretképzés a magyar nyelvben. In: Balázs Géza – Imrényi András – Simon Gábor (szerk.): *Hálózat kutatás: Hálózatok a nyelvben*. Budapest: Magyar Szemiotikai Társaság. 281–295.
- Modrián-Horváth Bernadett – Kalivoda Ágnes – Tanos Bálint 2024. *A magmondlat szerepű, passzív jelentésű sémák produktivitásának kvantitatív vizsgálata*. In: Kugler Nóra – Modrián-Horváth Bernadett – Simon Gábor (szerk.): *Korpusznyelvészeti módszerek a kognitív nyelvészeti kutatásban*. Budapest: ELTE BTK – ELTE Eötvös Kiadó. 173–223.
- Novák, Attila – Siklósi, Borbála – Oravecz, Csaba 2016. A new integrated open-source morphological analyzer for Hungarian. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Portorož, Slovenia: European Language Resources Association (ELRA). 1315–1322.
- Oravecz, Csaba – Váradi, Tamás – Sass, Bálint 2014. The Hungarian Gigaword Corpus. In: *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*. Reykjavík: European Language Resources Association (ELRA). 1719–1723.
- Péteri Attila 2015. *Satzmodusmarkierung im europäischen Sprachvergleich. Interrogativsätze im Deutschen und im Ungarischen mit einem typologischen Ausblick auf andere europäische Sprachen*. Berlin: Peter Lang. <https://doi.org/10.3726/978-3-653-05318-0>

- Pethő Gergely – Sass Bálint – Kalivoda Ágnes – Simon László – Lipp Veronika 2022. Igeköötő-kapcsolás. In: Berend Gábor – Gosztolya Gábor – Vincze Veronika (szerk.): *XVIII. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi Tudományegyetem, Informatikai Intézet. 77–91.
- Qi, Peng – Zhang, Yuhao – Zhang, Yuhui – Bolton, Jason – Manning, Christopher D. 2020. Stanza: A Python natural language processing toolkit for many human languages. In: *Proceedings of the 58th Annual Meeting of the ACL: System Demonstrations* (ACL 58). ACL, online. 101–108. <https://doi.org/10.18653/v1/2020.acl-demos.14>
- R Core Team 2021. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>
- Schmid, Hans-Jörg – Küchenhoff, Helmut 2013. Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings. *Cognitive Linguistics* 24 (3): 531–577. <https://doi.org/10.1515/cog-2013-0018>
- Simon, Gábor 2022a. Dogs know it, trees wait for it, and the wind snatches it: verbal collexemes and semantic domains of noun + verb personifications in Hungarian. *Studia Linguistica Hungarica* 34: 5–21. <https://doi.org/10.54888/slh.2022.34.5.21>
- Simon Gábor 2022b. *Ez elment vadászni* – a mutató névmási deixis funkcióinak és metaforizációjának korpuszvezérelt vizsgálata. In: Laczkó Krisztina – Tátrai Szilárd (szerk.): *Tanulmányok a deixisről*. Budapest: ELTE Eötvös József Collegium. 71–98.
- Simon Gábor 2024. Konstrukció és/vagy mintázat: az azonosítás kérdései egy használatalapú, korpuszvezérelt nyelvreírásban. In: Tátrai Szilárd – Laczkó Krisztina – Tolcsvai Nagy Gábor (szerk.): *Magyar igei konstrukciók. Használatalapú függőségi vizsgálatok*. Budapest: ELTE. 31–62.
- Stefanowitsch, Anatol 2006. Negative evidence and the raw frequency fallacy. *Corpus Linguistics and Linguistic Theory* 21: 61–77. <https://doi.org/10.1515/CLLT.2006.003>
- Stefanowitsch, Anatol 2013. Collostructional Analysis. In: Hoffman, Thomas – Trousdale, Graeme (eds.): *The Oxford handbook of Construction Grammar*. New York: Oxford University Press. 290–306. <https://doi.org/10.1093/oxfordhb/9780195396683.013.0016>
- Stefanowitsch, Anatol 2020. *Corpus linguistics. A guide to the methodology*. Textbooks in Language Sciences 7. Berlin: Language Science Press. DOI: 10.5281/zenodo.3735822

- Stefanowitsch, Anatol – Gries, Stefan Th. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8 (2): 209–243. <https://doi.org/10.1075/ijcl.8.2.03ste>
- Tolcsvai Nagy Gábor 2009. A magyar segédige + igenév szerkezet szemantikája. *Magyar Nyelvőr* 133: 373–93.



Collostructional analysis of Hungarian verbs in intervening position

The study presents the exploration of Hungarian finite verbs appearing between a preverb and its associated infinitive using the method of collostructional analysis. The simple collexeme analysis determines the degree of association between the verb and the constructional pattern in relation to the frequency of the particular verbs in the corpus, and thus it may contribute to the description of the basic meaning of the construction. In the literature, occurring in an intervening position is considered to be one of the key formal features of auxiliaries. With the analysis we aim at highlighting which verbs behave in this way in the HGC-UD corpus, since it can be one indicator of the functioning of these verbs as auxiliaries.

As a result, seven verbs proved to be statistically significant collexemes of the investigated construction. They have modal and temporal meaning. This group of verbs is nearly identical with the list of core auxiliary verbs in Hungarian proposed by Tolcsvai Nagy (2009) on the basis of semantic criteria and grammaticalization processes. Moreover, some verbs are not associated statistically with the construction, nevertheless they occur in intervening position frequently. Consequently, collostructional analysis not only supports previous findings with new data, but it also sheds new light on the auxiliary-like behaviour of Hungarian finite verbs, as well as the scalarity of this feature.

Keywords: intervening position, auxiliary, collostruction, simple collexeme analysis

Kalivoda Ágnes

HUN-REN Nyelvtudományi Kutatóközpont

kalivoda.agnes@nytud.hun-ren.hu

ORCID: 0000-0003-2520-5523

Modrián-Horváth Bernadett

Szegedi Tudományegyetem

bernadett.modrianhorvath@gmail.com

ORCID: 0000-0003-3834-9163

Simon Gábor

Eötvös Loránd Tudományegyetem

simon.gabor@btk.elte.hu

ORCID: 0000-0001-5233-6313

DÖMÖTÖR ANDREA

**Hozzászólás a *Beférkőző igék kollostrukciós vizsgálata.*
Az igekötő és a főnévi igenév közé ékelődő igék
kolexémaelemzése az MNSZ2-UD korpuszban című
tanulmányhoz**

A tanulmány az igekötő és a főnévi igenév közé való igei beférkőzés jelenségét vizsgálja korpuszalapú módszerrel: arra keres választ, hogy mely igék mutatnak statisztikailag szignifikáns hajlamot az igekötő és a főnévi igenév közé való beférkőzésre. A beférkőzés jelenségét a szakirodalom gyakran említi a segédigeiség egyik fő kritériumaként, így a kutatás hozzájárul a segédige mint kategória pontosabb definiálásához, azonban hangsúlyozottan nem célja, hogy egzakt listát adjon a magyar segédigékről. A korpuszalapú módszerek sajátja, hogy objektív, számszerű tényeket közölnek a nyelvről, és ez a tanulmány is ezt teszi: a korpuszatatok alapján egy spektrumot állít fel az igék beférkőzésre való hajlamosságáról, arra azonban nem vállalkozik, hogy ezen a spektrumon meghúzza a határt segédige és nem segédige között, ez ugyanis már szubjektív döntés lenne.

A tanulmány választott módszere a kollostrukciós elemzés, melyet a megfelelő részletességgel mutatnak be a szerzők, és a választás melletti érvelés is meggyőző.

Az adatgyűjtés módszerének leírásánál felmerül a kérdés, hogy miért nem használták a szerzők a kereséseikhez a függőségi elemzést, hiszen az általuk használt korpusz tartalmaz függőségi annotációt, és a vizsgált szerkezetek kinyeréséhez két ponton is kézenfekvő lenne a szintaktikai szerkezetre támaszkodni a morfológiai címkék mintázatai helyett. A függőségi elemzés alapján ugyanis mind a főnévi igenevek és a hozzájuk tartozó igekötők, mind pedig az igék és azok infinitívuszi vonzatainak összekapcsolása elvileg triviális feladat kellene hogy legyen. Az előbbi feladat esetén a tanulmány Pethő et al. (2022) mérésére hivatkozik, mely szerint a morfológiai címkékre alapozó emPreverb eszköz érte el a legjobb eredményt az igék és az elváló igekötők összekapcsolásában. Az említett mérésben

valóban szerepel a vizsgált módszerek között a függőségi elemzés alapján történő igekötő-ige kapcsolás, és ez valóban gyengébben teljesített, mint az emPreverb. Ez (és a saját tapasztalataim) alapján azt gondolnám, hogy az igék és a főnévi igenévi vonzatok összekapcsolásánál is a függőségi annotációk nem megbízható minősége lehet az ok, amiért nem ezt a módszert választották a szerzők. Ezt érdemes lenne reflektálni a tanulmányban, mert véleményem szerint hasznos tapasztalat lehet más korpuszhasználók számára is a keresési módszerük megválasztásához.

További, a korpuszt érintő észrevétel, hogy a beférkőző igék listájában szerepel a szabad „igeként annotálva”. Ez valószínűleg meghaladná a jelen tanulmány kereteit, de érdekes és fontos kérdés lenne, hogy pontosan mely esetekben szerepel a szabad igeként annotálva a korpuszban, és ez mennyire következetes. Csak abban az esetben tekintjük igének, ha igei inflexiót hordoz (például *szabadott*, *szabadna*), vagy ezek nélkül is előfordul az igei annotáció? Függ-e a szófajcímke például a kopula jelenlététől? (Vö. *Nem szabad kimenni*. vs. *Nem volt szabad kimenni*.)

Másik kopulához kapcsolódó probléma, hogy a beférkőző igék listázásánál a szerzők egy szótó alá sorolták a *lehet*, *van* és *volna* töveket, a *lesz* azonban külön tőként szerepel a listában. Nem világos, hogy a *lehet* miért a *van*, *volna* csoportba került, és miért nem a *lesz* változataként van kezelve, holott morfológiailag ez indokoltabb lenne – már amennyiben egyáltalán indokolt a szótövek bármilyen összevonása. Ezt a döntést maguk a szerzők is felülbírálják egy lábjegyzetben, elsősorban a lehet várható dominanciájára hivatkozva. Az én meglátásom szerint is érdemes a lehet alakot külön igeként, és nem a van vagy a lesz alakváltozataként kezelni.

A tanulmányban ismertetett eredmények nagyban átfednek Tolcsvai Nagy (2009) segédige-listájával, holott az nem formális-szintaktikai kritériumok alapján készült. Ez mutatja, hogy a korpuszalapú módszereknek helye van az elméleti nyelvészetben. Jelen kutatás fő hozzáadott értéke az introspekciós módszerű vizsgálatokhoz képest az, hogy képes megkülönböztetni a beférkőzésre statisztikailag szignifikáns hajlamot mutató igéket a beférkőzést megengedő igéktől, azaz nem csak azt mondja meg, hogy jellemző-e egy igére a beférkőzés, hanem azt is mennyire jellemző (vagy épp nem jellemző). Mindez, a szerzők megfogalmazása szerint, a segédigék „nyitott, prototípusulvú kategóriaként történő értelmezéséhez” járul hozzá.

Dömötör Andrea

Eötvös Loránd Tudományegyetem

domotor.andrea@btk.elte.hu

IMRÉNYI ANDRÁS

**Hozzászólás a *Beférkőző igék kollostrukciós vizsgálata.*
Az igekötő és a főnévi igenév közé ékelődő igék
kolexémaelemzése az MNSZ2-UD korpuszban című
tanulmányhoz**

1. Áttekintés. A hozzászólás témája

Kalivoda Ágnes, Modrián-Horváth Bernadett és Simon Gábor hiánypótló, úttörő munkát végzett el a beférkőző (segéd)igés szerkezetek kollostrukciós vizsgálatával. A DiAGram funkcionális nyelvészeti kutatócsoport futó NKFI-pályázatának célkitűzéseivel összhangban (NKFI K129040, „A magyar nyelv igei konstrukciói. Használatalapú konstrukciós nyelvtani kutatás”), a használatalapú nyelvtan módszertani implikációit is következetesen érvényesítve arra a kérdésre keresik a választ, hogy a Magyar Nemzeti Szövegtár igéi milyen mértékben, milyen valószínűséggel mutatnak beférkőző szórendet (például *meg találja kérdezni*). A munka elméleti háttérét a konstrukciós nyelvtan adja (vö. Goldberg 1995), amelyhez szervesen illeszkedik a Stefanowitsch és Gries (2003) kezdeményezte kollostrukciós elemzés. A konstrukciós nyelvtant a kezdetektől foglalkoztatja az igék és a köréjük szerveződő, illetve őket befogadó mondat szerkezeti mintázatok (konstrukciók) viszonya. A kollostrukciós elemzés ehhez szorosan kapcsolódva kvantitatív módszert ad annak vizsgálatára, hogy a különböző lexémákat (a hozzászólás tárgyát képező cikk esetében: igei lexémákat) milyen mértékben vonzza, illetőleg taszítja egy adott konstrukció (jelen esetben a beférkőző szórendi mintázat).

A módszer lehetséges változatai közül a szerzők az ún. egyszerű kolexémaelemzést választották. (A „kolexéma” terminus egy adott konstrukcióval társuló, azzal kompatibilis lexémára utal, például a beférkőző szórendű konstrukció kolexémája a *fog*, a *kell* stb.) A szerzők a kollostrukciós módszernek

megfelelően vizsgálták „az adott ige korpuszbeli gyakoriságát a vizsgált főnévi igenévi beférkőző mintázatban (a), az adott ige összes előfordulását más (nem beférkőző) konstrukciókban (b), más igék gyakoriságát az adott konstrukcióban (c), illetve más igék előfordulásait más konstrukciókban (d)” (Kalivoda et al. 2024: 115), és ezek arányaiból vontak le következtetést a beférkőző konstrukció, illetve különféle lexémák közötti vonzás/taszítás mértékével kapcsolatban. A szerzők kizárólag az igekötővel rendelkező főnévi igenevek mellett előforduló igék körében vizsgálták a beférkőzést. Mint írják, „a teljes alapsokaságot ezeknek a főnévi igeneves adatoknak a gyakorisága jelentette a vizsgálatunkban, ezt tekintettük tehát a korpuszbeli (vizsgált) konstrukciók számának, és ebben az alapsokaságban értelmeztük a beférkőzés kialakulását” (Kalivoda et al. 2024: 115).

A kutatás alapja az MNSZ2-UD korpusz (Kalivoda et al. megj. e.), amely több innovatív eszközzel, többek között a Universal Dependencies (UD, de Marneffe et al. 2021) formalizmusát követő függőségi elemzéssel gazdagította, valamint korpusztisztító eljárásokkal is továbbfejlesztette a szövegtárat (Oravecz–Várad–Sass 2014). Az, hogy ez az új adatbázis áll a vizsgálat hátterében, önmagában is rendkívül izgalmassá teszi a tanulmányt.

A szerzők számára is meglepő módon korpusznyelvészeti vizsgálatuk eredménye az eltérő módszertan ellenére „szinte teljes egyezést mutat a központi segédigéknek Tolcsvai Nagy (2009) által szemantikai és grammatikalizációs alapon kiemelt listájával” (Kalivoda et al. 2024: 105). Mint kiemelik, „az egyszerű kolecémaelemzés a szakirodalomban formális-disztribúciós alapon leírtakhoz képest (vö. Kálmán C. et al. 1989) kisszámú lexémát azonosított szignifikánsan a beférkőző szerkezethez vonzott igeként; ezek a *kell*, *tud*, *van/lehet*, *akar*, *fog*, *mer* és *talál* igék” (Kalivoda et al. 2024: 128). A szerzők tisztában vannak a módszer további finomításának szükségességével, erre utal például a következő megjegyzésük: „A szerkezettel kevésbé szorosan asszociálódó, további beférkőző igék feltárása a vizsgálatban alkalmazható szignifikanciaszintek mellett nem kivitelezhető egzakt módon, noha a magyar nyelv szórendi jellegzetességei szükségessé tennének ezzel kapcsolatos további vizsgálatokat” (Kalivoda et al. 2024: 129).

A jelen hozzászólás egyetlen fő problémafelvetést tartalmaz. E felvetés ahhoz kapcsolódik, hogy a korpuszelemzés nemcsak hogy aránylag kisszámú lexémát azonosított szignifikánsan beférkőző igeként, hanem meglehetősen központinak látszó elemeket is „taszított” (alacsony mértékben vonzott) lexémaként értékelt. Ide tartozik mindenekelőtt a *szokott* és a *tetszik*, de akár a *szeretne* is. Ezek kimaradása véleményem szerint a szerzők által

alkalmazott módszer finomításának, továbbfejlesztésének igényét alapozhatja meg.

2. A szokott és a tetszik hiányáról. Szempontok az elemzés finomításához

A kognitív nyelvészet alapvető tézise, hogy a kategóriák a prototípuselv szerint szerveződnek, azaz vannak központi és periférikusabb példányaik is. Általánosabban fogalmazva, a kognitív megközelítés a nyelvi rendszer leírásában nagy hangsúlyt helyez a kontinuumok, fokozatos átmenetek létezésére, elutasítva a vagy–vagy típusú gondolkodást (*exclusionary fallacy*, lásd Langacker 2008: 13). Ennélfogva önmagában sem nem meglepő, sem nem problematikus az, hogy a szerzők kevesebb elemet azonosítottak beférkőző elemként, mint a korábbi, más módszertannal dolgozó munkák. Az azonban nagyon feltűnő, ha nemcsak a kategória periferiáján elhelyezkedő elemek maradnak ki, hanem központinak látszóak is. Bár Tolcsvai Nagy (2009) – a szerző által is jelzett módon „nem teljes” – listája (*tud, bír, akar, mer, fog, kell, lehet*) nem tartalmazza a *szokott* és a *tetszik* elemeket, ennek oka leginkább a lista szemléltető jellege lehet. Még olyan más munkák is, amelyek igencsak szűkre szabják a segédigék csoportját, ide tartozónak minősítik a *szokott* és a *tetszik* elemeket. Lengyel (2000) például a *fog, szokott, talál* és *tetszik* elemeket tekinti a magyar nyelv ún. szóalakteremtő segédigéinek, e négy elemből pedig Kalivoda és munkatársai csak kettőt, a *fog*-ot és a *talál*-t azonosították szignifikánsan beférkőző használatuként.¹

Természetesen nagyon is lehetséges, hogy a korpusznyelvészettől megtudunk valami olyasmit, amit korábban nem tudtunk (nem tudhattunk), vagy nem jól tudtunk – rendszerint éppen ez a célja az ilyesfajta vizsgálatnak. Magam mégis úgy vélem, hogy ez esetben kissé más a helyzet: olyasmiről kérdezzük (kérdezik a szerzők) a korpuszt, amivel kapcsolatban viszonylag megbízhatónak tűnik, feltételezhetően nagyfokú interszubjektív megerősítéssel bír az introspektív evidencia. Egy kérdőíves vizsgálat minden bizonnyal azt mutatná ki, hogy a magyar beszélők túlnyomórészt a *Le tetszik ülni?*

¹ Az eltérés még akkor is feltűnő, elgondolkodtató, ha a szerzők megjegyzik: „azoknak a lexémáknak az esetében, amelyeknél a megfigyelt adatok megközelítik az elvárt adatokat, joggal állíthatjuk, hogy beférkőző, de a konstrukcióval nem szignifikánsan asszociált igékről van szó. Ilyenek különösen a nagyobb gyakoriságú *kelletik, merészel, bír, óhajt, szeretne, szokik/szokott, szándékozik, tetszik, kíván*” (Kaliyoda et al. 2024: 125).

(nem pedig a *Tetszik leülni?* vagy a *Leülni tetszik?*) pozitív formájú kérdés-sel kínálnák fel ülőhelyüket egy idős embernek a villamoson. A *tetszik* (és hasonló érvelés alapján a *szokott*) kimaradása a (szignifikánsan) beférkőző elemek köréből ezért legalábbis elgondolkodtató. Mi lehet az oka a hiánynak? Az alábbiakban megkísérlek néhány felvetéssel hozzájárulni ehhez a gondolkodáshoz.

Első és legfontosabb észrevételem a tanulmánynak ahhoz a megjegyzéséhez kapcsolódik, hogy míg a beférkőzés kutatásában „Kalivoda–Prószéky (2023) kizárja a vizsgálatból a vélhetően nem semleges szórendű (nem kijelentő modalitású és bal perifériával rendelkező) mondatokat”, addig a cikkben ismertetett kutatás „figyelembe vesz minden kontextustípusban való előfordulást, mivel nincs abszolút „semleges” mondat, minden megnyilatkozás információszerkeztét befolyásolja az adott setting (alap)” (Kalivoda et al. 2024: 113).

Véleményem szerint a tagadó, kontrasztív azonosító, felszólító stb. mondatok kiszűrése az alapsokaságból mindenképpen indokolt: a beférkőzési hajlamot azon kontextusok és funkciók esetében érdemes kutatni, amelyekben a beférkőzés lehetősége, esélye egyáltalán fennáll. Például a *nem fogok belenyugodni, tessék leülni* típusú mintázatokat kiszűrendőnek vélem, mivel a tagadás és a felszólítás szerkezeti sémája eleve gátolja a (vizsgált típusú) beférkőzést. Bár Kalivoda et al. (2024: 147) kiemeli, hogy a Kalivoda–Prószéky (2023)-ban alkalmazott módszerrel érdekes módon még kevesebb elem tűnt ki beférkőzőként (ezek az *akar*, *fog*, *kell* és *kíván*), megítélésem szerint a jövőben mégis ez az út lehet produktívabb. Ahogyan a *Jánosnak tetszik Mari* típusú példákban nincs helyük az alapsokaságban (mivel esetükben – igekötős infinitívusz hiányában – a beférkőzés feltételei nem állnak fenn), úgy a beférkőzést gátló mondat típusok példányait is érdemesebb volna mellőzni. A probléma felvetéséhez használt példához visszatérve, ha a *Le tetszik ülni?* mellett a *Nem tetszik leülni? Miért tetszett leülni? Tessék leülni! Tetszett volna leülni!* és más hasonló példákat is figyelembe vesszük, akkor az eredmény nem annyira (vagy nemcsak) a *tetszik* beférkőzési hajlamát fogja tükrözni, hanem inkább azt (is), hogy milyen valószínűséggel jelenik meg a semleges pozitív kijelentő mondatot (Imrényi 2017: 669–672) funkciójában és szórendjében felidéző eldöntendő kérdésben.

Ehhez kapcsolódóan vitatnám a szerzőknek azt az érvét – amellyel az összes kontextustípus vizsgálatát indokolják –, miszerint „nincs abszolút „semleges” mondat, minden megnyilatkozás információszerkeztét befolyásolja az adott setting (alap)” (KaliVoda et al. 2024: 113). A nyelvtudást

forma-funkció-kontextus triádok rendszerében tartom célszerűnek megközelíteni, azt vizsgálva, hogy milyen formák milyen funkciókkal társulnak a kontextus függvényében (vö. Ellis 2006: 8; Gries 2017: 593). Kétségtelen, hogy minden megnyilatkozás információszerkezetét befolyásolja az adott setting (alap), de ebből véleményem szerint nem következik, hogy nincsen semleges mondat. A kontextustípusok és funkciótipusok változatossága nem végtelen: létezik egy olyan kontextustípus és ehhez társuló funkciótipus, amely a megfelelő formai jegyekkel együtt jó alapot ad a „semleges mondat” (vagy szerintem inkább: semleges pozitív kijelentő mondat) kategóriájának a feltételezésére (Imrényi 2017: 669–672).

További észrevételem, hogy az eredmények reprezentativitását hátrányosan befolyásolhatták a korpusz műfaji és mediális kötöttségei, például a *tetszik* tipikus beférkőző használatát előhívó beszédhelyzetek és funkciók (ehhez vö. Domonkosi–Kuna 2015) alulreprezentáltsága, ezért célszerű lehet további korpuszokat is bevonni a kutatásba.

3. További észrevételek

Végezetül úgy vélem, hogy az eredmények pontosítását segíthetnék az alábbiak:

1. a beférkőző szerkezet első elemének tágabb értelmezése a kategorizációban; az igekötők mellett a más, igekötőszerűen viselkedő, pusztán főnévi igemódosítóknak is nevezett elemeknek a figyelembe vétele (vö. Komlósy 2015/1992: 462, Kiefer 2003)
2. a beférkőző szerkezet mellett annak vizsgálata, hogy igekötő/igemódosító nélküli infinitívuszt milyen gyakorisággal követ valószínűsíthetően hangsúlytalan ejtéssel egy adott végesealak. A segédigei használat szempontjából éppolyan fontos a *tudni szeretném*, mint a *meg szeretném tudni* (vö. Kálmán C. et al. 1989: 61); mindkettő egy semleges mondat magmondatát, protoállítást képviseli (Imrényi 2017).

Akár ezek, akár más szempontok hoznak továbblépést, személyes várakozásom az, hogy a jövőben – esetleg más korpuszok bevonásával vagy finomított módszertannal – a kutatás olyan eredményekhez vezethet el, amelyek még nagyobb mértékben harmonizálnak a korábbi, más módszerrel nyert eredményekkel. Mint jeleztem, ezt különösen a *szokott* és a *tetszik* beférkőző hajlammal kapcsolatban látom reálisnak. E személyes várakozás hátterében

pedig az a meggyőződés áll, hogy a jelenség központi vonatkozásaival kapcsolatban viszonylag megbízhatónak látszik (feltételezhetően nagyfokú interszubjektív megerősítéssel bír) az introspektív módszerrel nyert evidencia is.

4. Zárszó

Hozzászólásom végén ismét hangsúlyozni szeretném Kalivoda Ágnes, Modrián-Horváth Bernadett és Simon Gábor tanulmányának hiánypótló, úttörő jellegét. Kvantitatív, korpuszvezérelt kutatásuk fontos mérföldkő a beférkőző szerkezet vizsgálatában. A módszer továbbfejlesztésének szükségességét a szerzők több ponton hangsúlyozzák, és cikkük záró bekezdésében kijelölik ennek irányait is. Minden bizonnyal csak idő kérdése egy továbbfejlesztett elemzési eljárás kidolgozása, amelynek eredményei személyes várákozásom szerint még nagyobb összhangban lesznek a más módszerrel elért eredményekkel.

Irodalom

- de Marneffe, Marie-Catherine – Manning, Christopher D. – Nivre, Joakim – Zeman, Daniel 2021. Universal Dependencies. In: *Computational Linguistics* 47 (2): 255–308. https://doi.org/10.1162/coli_a_00402
- Domonkosi Ágnes – Kuna Ágnes 2015. A tetszikelés szociokulturális értéke. A tetszikelő kapcsolattartás szerepe az orvos-beteg kommunikációban. *Magyar Nyelvőr* 139: 39–63.
- Ellis, Nick C. 2006. Language acquisition as rational contingency learning. *Applied Linguistics* 27 (1): 1–24. <https://doi.org/10.1093/applin/ami038>
- Goldberg, Adele 1995. *Constructions: A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.
- Gries, Stefan Th. 2017. Corpus approaches. In: Dancygier, Barbara (ed.): *The Cambridge Handbook of Cognitive Linguistics*. Cambridge: Cambridge University Press. 590–606. <https://doi.org/10.1017/9781316339732.037>
- Imrényi András 2017. Az elemi mondat viszonyhálózata. In: Tolcsvai Nagy Gábor (szerk.): *Nyelvtan*. Budapest: Osiris Kiadó. 664–760.
- Kalivoda, Ágnes – Malouf, Robert – Ackerman, Farrell – Sass, Bálint 2024. *Building a dependency treebank from the Hungarian Gigaword Corpus*. Kézirat, megjelenés előtt.

- Kalivoda Ágnes – Prószyk Gábor 2023. Infinitívuszi szerkezetek a magyarban: adatvezérelt megközelítés. *Nyelvtudományi Közlemények* 119: 223–245.
- Kalivoda Ágnes – Modrián-Horváth Bernadett – Simon Gábor 2024. Beférköző igék kollostrukciós vizsgálata. Az igekötő és a főnévi igenév közé ékelődő igék kolexémaelemzése az MNSZ2-UD korpuszban. In: Kugler Nóra – Modrián-Horváth Bernadett – Simon Gábor (szerk.): *Korpusznyelvészeti módszerek a kognitív nyelvészeti kutatásban*. Budapest: ELTE BTK – ELTE Eötvös Kiadó. 105–134.
- Kiefer Ferenc 2003. A kétféle igemódosítóról. *Nyelvtudományi Közlemények* 100: 177–186.
- Kálmán C., György – Kálmán László – Nádasdy Ádám – Prószyk Gábor 1989. A magyar segéd-igék rendszere. *Általános Nyelvészeti Tanulmányok* XVII: 49–103.
- Komlós András 2015/1992. Régensek és vonzatok. In: Kiefer Ferenc (szerk.), *Strukturális magyar nyelvtan 1. Mondattan*. Budapest: Akadémiai Kiadó. Javított digitális kiadás. 274–486. <https://real.mtak.hu/35148/1/StrukturalisMagyarNyelvtan1.pdf>
- Korompay Klára 2017. Emlékeim Bárczi tanár úrról. *Helynévtörténeti Tanulmányok* 14: 115–124. <https://doi.org/10.35528/Helynevtort/14/06>
- Langacker, Ronald W. 2008. *Cognitive Grammar: A basic introduction*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195331967.001.0001>
- Lengyel Klára 2000. A segédigék és származékaik. In: Keszler Borbála (szerk.): *Magyar grammatika*. Budapest: Nemzeti Tankönyvkiadó. 252–258.
- Levshina, Natalia 2015. *How to do linguistics with R. Data exploration and statistical analysis*. Amsterdam, Philadelphia: John Benjamins. <https://doi.org/10.1075/z.195>
- Oravecz, Csaba – Váradi, Tamás – Sass, Bálint 2014. The Hungarian Gigaword Corpus. In: *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*. Reykjavík: European Language Resources Association (ELRA). 1719–1723.
- Stefanowitsch, Anatol – Gries, Stefan Th. 2003. Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8 (2): 209–243. <https://doi.org/10.1075/ijcl.8.2.03ste>
- Tolcsvai Nagy Gábor 2009. A magyar segédige + igenév szerkezet szemantikája. *Magyar Nyelvőr* 133: 373–393.

Imrényi András

Eötvös Loránd Tudományegyetem

imrenyi.andras@btk.elte.hu

ORCID: 0000-0001-6203-9304

SIMON GÁBOR – MODRIÁN-HORVÁTH BERNADETT –
KALIVODA ÁGNES

Válasz Imrényi András lektori hozzászólására

Imrényi András lektori hozzászólásában fontos, előremutató javaslatokat tett az általunk elemzett jelenség (a beférkőzés mint konstrukció társulása igei lexémákkal) további vizsgálatához, ezért mindenekelőtt szeretnénk megköszönni konstruktív szemléletmódját. A kutatás kiterjesztése a nem központi vagy prototipikus, de igeikötőszerűen viselkedő komponensekre újabb adatokkal egészítheti majd ki az általunk felvázolt képet. A nem igeikötővel, de főnévi igenévvel való együttes előfordulás adatvezérelt feltérképezése pedig még inkább stabilizálhatja majd a segédigei kategória leírását. Megjegyzendő, hogy az előbbi kiterjesztésnek együtt kell majd megvalósulnia az igeikötők automatikus azonosítását végző nyelvtechnológiai eszközök fejlődésével, a további, a segédigei funkcionálásra jellemző mintázatok elemzése pedig nem annyira felülírja, mint inkább kiegészíti a jelen vizsgálat eredményeit, ám ezek a lépések valóban nélkülözhetetlenek, ha a kutató a magyar segédigék kategóriáját a korpuszadatokra építve kívánja bemutatni.

Hasonlóan perspektivikus az a javaslat is a bemutatott elemzés léptékét tekintve, hogy más korpuszok adatain is célszerű elvégezni a vizsgálatot. Ez egyrészt összhangban van a korpusznyelvészet egyik módszertani alapelvével („principle of cooperation”, Sass 2022: 606–608), miszerint több különböző korpusz használata jobb eredményre vezet a kutatásban. Másrészt valóban érdemes tesztelni egyes igeik beférkőzésre való hajlandóságának regiszterspecifikusságát, e tekintetben a Magyar Nemzeti Szövegtárnak a kutatásunkban használt precízebb változata is csak kiindulópontja lehet további, specifikus korpuszokon végzett elemzéseknek.

Imrényi András hozzászólásában problémásnak tekinti, hogy a konstrukcióval szignifikánsan gyakori (és nem pusztán kompatibilis, ahogy a lektori véleményben áll) együttes előfordulást mutató igeik között nem szerepelnek a szakirodalom által (központi) segédigének minősített igeik,

így a Lengyel (2000a) által felsorolt *szokott* és *tetszik*, melyeket egyúttal beférkőzőnek is tekint a szakirodalom java (például Kálmán C. et al. 1989). Ezzel kapcsolatban két kérdés merül fel: valóban a szakirodalom fősodrával ellentétesen nem beférkőző igékről, illetve nem központi segédigékről lenne szó? Az első kérdésre a vizsgálat tükrében azt tudjuk válaszolni, hogy beférkőző, a konstrukcióval „kompatibilis”, de nem szignifikánsan vonzott igék a nevezett igék. Félrevezető lehet a vizsgálatban szokásosan használt „taszított” terminus, amely azt sugallhatja, hogy a szóban forgó igék nem szívesen társulnak a vizsgált konstrukcióval. Nem erről van szó, hanem arról, hogy az igék más konstrukciókkal is gyakran fordulnak elő.¹ Ez egy olyan empirikus tény, amely éppen a pusztán introspekción alapuló meglátásokhoz képest hoz újdonságot a „segédige-gyanús” igék kutatásába. A központi segédigeségre vonatkozó felvetésre a vizsgálat célkitűzése miatt csak részben tudunk reflektálni. A segédigeség – elméletileg is megalapozott!, vö. Kalivoda et al. (2024) 1.3. pontja – kritériumainak való megfelelés egy olyan prototípuselvű kategóriát definiálhat, amelynek tagjai a wittgensteini családi hasonlóság szerint rendeződnek el: a kategória tagjai más-más tulajdonságokat osztanak, nem tudunk felsorolni néhány szükséges és elégséges feltételt a segédigeség definíciójául. A *szokott* igével kapcsolatban például a deszemantizáció (vö. Tolcsvai Nagy 2009: 385) kritériuma kevésbé figyelhető meg és nem képes jelen idejű lehorgonyzásra – talán éppen e tulajdonságok miatt nem került bele a központi segédigéket szemléltető listába. Lengyel (2000a) és (2000b) egyaránt utal a segédigék behatárolásának nehézségeire, sőt lehetetlenségére, és az általa leírtakból nem bontakozik ki világosan, melyek lennének azok a fő szempontok, amelyek alapján éppen a *fog*, *szokott*, *talál* és *tetszik* minősülnek segédigének.

Imrényi András felveti, hogy a vizsgálat nemcsak a beférkőzési hajlandóságot vizsgálja, hanem azt is, hogy a vizsgált igék milyen gyakorisággal jelennek meg a „semleges pozitív kijelentő mondatot” tükröző szórendi mintázatokban, és ezzel összefüggésben utal arra is, hogy véleménye szerint szükséges lenne e kategória megkülönböztetése mind elméleti, mind metodológiai szempontból. Amint azt a tanulmány is kifejti, véleményünk szerint a „semleges” kategória nehezen értelmezhető a nyelvhasználat szempontjából, tekintettel a korpuszban megjelenő kontextusok sokféleségére.

¹ Vö. *le szoktam tölteni* (az MNSZ2-ben 1 találat, lásd Oravecz et al 2014) vs. *szoktam letölteni* (ugyanitt 11 találat, leginkább felülíró konstrukciókban); ellenben *le akartam tölteni* (6 találat) vs *akartam letölteni* (4 találat).

Az igekötős igékkel megjelenő beférkőző típussal kapcsolatban talán szerencsebb lenne „jelöletlen” szórendi mintázatról beszélni, ez utóbbi terminus ugyanis nem implicálná azt, hogy létezik olyan szórend, amelynek nincs információszerkezeti értéke. E terminológiai kérdéstől függetlenül az adott vizsgálati keretben nem tudunk különbséget tenni a kontextusban megjelenés mértéke és a beférkőzés mértéke között, de ezt nem érezzük hátránynak. Szeretnénk kiemelni, hogy minden (jelölt vagy jelöletlen) kontextustípus egyaránt megjelenik és potenciálisan befolyásolja az összes vizsgált ige elvárt és mért értékeit, a *kell* és *akar* igékét csakúgy, mint a *szokott* és *tetszik* igékét. Amennyiben a vizsgált (referencia-)korpusz megfelelően reprezentálja a mai magyar nyelvhasználatot, annyiban helyesek az eredmények; a korpusz összetételére nincs ráhatásunk, de reflektáltunk a tanulmányban a lehetséges anomáliákra, mint ahogy Imrényi András is jogosan említi bírálatában bizonyos diskurzustípusok alulreprezentáltságát.

Végezetül érdemes utalni Dylan Glynn (2014: 27) érvelésére is a „korpuszvezérelt kvantitatív kutatás” és az „introspekción alapuló sugaras hálózatkutatás” viszonyáról. Glynn szerint az előbbi „kizárólag gyakoriságra alapul, így viszont előnyben részesíti a tipikust” azzal szemben, ami lehetséges. Az intuíció jelentőségét abban látja, hogy az feltevésekhez vezet arról, mi fordulhat elő egyáltalán a nyelvben. Kiegészítik tehát egymást ezek a megközelítések, és éppen ebben rejlik a kognitív korpusznyelvészet kivételes lehetősége, mert nem a megbízhatóság mentén szembesíti az egyes adattípusokat, hanem aszerint, melyikből milyen következtetések vonhatók le. „A korpuszvezérelt vizsgálat – miként Glynn megjegyzi – megerősítheti a hálózatelemzés részleteit, de nem valószínű, hogy ugyanazt a képet festené”, miként az sem, hogy a megerősítés hiánya egyben a korábbi (vagy más adattípusokra épülő) elemzések elutasítását jelentené.

Bízunk abban, hogy a hozzászólásra készített válaszuk ehhez az együttműködéshez kínál alapot, és a szükséges egyensúly felé mozdítja el a magyar nyelvre irányuló kognitív korpusznyelvészeti kutatásokat.

Irodalom

Glynn, Dylan 2014. Polysemy and synonymy. Cognitive theory and corpus method. In: Glynn, Dylan – Robinson, Justina A. (eds.): *Corpus methods for semantics. quantitative studies in polysemy and synonymy*. Amsterdam, Philadelphia: John Benjamins. 7–38. <https://doi.org/10.1075/hcp.43.01gly>

- Heine, Bernd 1993. *Auxiliaries: Cognitive forces and grammaticalization*. New York, Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780195083873.001.0001>
- Kalivoda Ágnes – Modrián-Horváth Bernadett – Simon Gábor 2024. Beférköző igék kollostrukciós vizsgálata. Az igekötő és a főnévi igenév közé ékelődő igék kolexémaelemzése az MNSZ2-UD korpuszban. In: Kugler Nóra – Modrián-Horváth Bernadett – Simon Gábor (szerk.): *Korpusznyelvészeti módszerek a kognitív nyelvészeti kutatásban*. Budapest: ELTE BTK – ELTE Eötvös Kiadó. 105–134.
- Kálmán C. György – Kálmán László – Nádasdy Ádám – Prószéky Gábor 1989. A magyar segédigék rendszere. Általános Nyelvészeti Tanulmányok XVII: 49–103.
- Lengyel Klára 2000a. A segédigék és származékaik. In: Keszler Borbála (szerk.): *Magyar grammatika*. Budapest: Nemzeti Tankönyvkiadó. 252–258.
- Lengyel Klára 2000b. *Az igenevek helye a szófaji rendszerben*. Budapest: Akadémiai Kiadó
- Oravecz, Csaba – Váradi, Tamás – Sass, Bálint 2014. The Hungarian Gigaword Corpus. In: *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*. Reykjavík: European Language Resources Association (ELRA). 1719–1723.
- Sass Bálint 2022. Principles of corpus querying: A discussion note. *Acta Linguistica Academica* 69: 599–614. <https://doi.org/10.1556/2062.2022.00581>
- Tolcsvai Nagy Gábor 2009. A magyar segédige + igenév szerkezet szemantikája. *Magyar Nyelvőr* 133: 373–93

**A magmondat szerepű,
passzív jelentésű sémák
produktivitása**



MODRIÁN-HORVÁTH BERNADETT – KALIVODA ÁGNES – TANOS BÁLINT

A magmondat szerepű, passzív jelentésű sémák produktivitásának kvantitatív vizsgálata¹

Kivonat

A tanulmány célja négy passzív jelentésű konstrukció: a *V-ódik/-ődik*, a *V-(t)atik/-(t)etik*, a *V-va/-ve lesz* és a *V-va/-ve van* konstrukció produktivitásának korpuszalapú vizsgálata. A kutatás az MNSZ2, valamint az újraannotált MNSZ2-UD korpuszokon végzett gépi és kézi elemzéseken alapul, és a kvantitatív keretben meghatározható globális produktivitási indexet (vö. Baayen 1992) kvalitatív elemekkel ötvözi. A globális produktivitási index mérése egy, a használatalapú nyelvészetben gyakran alkalmazott módszer, mely a konstrukciós sémák típus- és példánygyakoriságának, valamint a sémákkal képzett hapax legomenonok számának segítségével számítható ki. A vizsgálat kvalitatív része az egyes sémák produktív jelentését leginkább tükröző hapaxokból nyert minta igenemek szerinti besorolására irányult.

A vizsgálat eredményei azt mutatják, hogy a korpusz egészében a legnagyobb globális produktivitással rendelkező séma a *V-va/-ve van* konstrukció, a legnagyobb típusgyakorisággal a *V-ódik/-ődik*, míg a legnagyobb

¹ A dolgozat elkészítését Modrián-Horváth Bernadett részéről az NKFIH 129040 számú pályázata (A magyar nyelv igei konstrukciói. Használatalapú konstrukciós nyelvtani kutatás) támogatta. Kalivoda Ágnes a 4. fejezetben ismertetett korpuszvizsgálatot és a 6.2. fejezethez a random mintavételezést végezte. A kézi korpuszvizsgálatban a szöveg tisztítási feladatokat és a hapaxok arányainak számításait Tanos Bálint végezte. Modrián-Horváth Bernadett az 5. és 6. fejezetben bemutatott kézi korpuszvizsgálatokat kivitelezte, valamint a dolgozat elméleti részeit írta. Ezúton fejezzük ki hálánkat lektorainknak, Kugler Nórának és Palágyi Lászlónak, akik számos segítő és inspiráló megjegyzéssel, korrektúrával, részletes kommentárral segítettek jobbra tenni a tanulmányt. A fennmaradó hibák természetesen a szerzőket terhelik.

potenciális produktivitással a *V-va/-ve lesz* rendelkezik; az egyes alkorpuszokra vonatkozó eredmények ettől részben eltérnek. A hapaxokból nyert véletlenszerű minta elemzése alapján a legstabilabban passzív jelentésű séma a rendkívül alacsony produktivitású *V-tatik/-tetik* konstrukció, ezt szorosan követi a *V-va/-ve lesz*, a túlnyomórészt passzív jelentésű a *V-va/-ve van*, és a hapaxok kevesebb mint felében találunk passzív jelentést a *V-odik/-ődik* esetében. A tanulmány további, elméleti és metodológiai megfontolásokat is tartalmaz.

Kulcsszavak: magyar nyelvtan, igenem, passzívum, korpusznyelvészet, konstrukciós nyelvtan

1. Bevezetés

1.1. Célkitűzés és elméleti háttér

A tanulmány négy, nem modális, részben vagy elsősorban passzív jelentésű,² magmondat (állítmányi)³ szerepű konstrukció kvantitatív produktivitási vizsgálatát tűzi ki célul. Konstrukció alatt általánosságban tanult, illetve itt különösen grammatikai típusú, azaz részben kitöltött vagy nem-kitöltött forma-jelentés párokat értünk (vö. Goldberg 2006, 2013). A vizsgált négy konstrukció: a *V-odik/-ődik*, a *V-(t)atik/-(t)etik*, a *V-va/-ve lesz* és a *V-va/-ve van* konstrukció. A vizsgálatok elméleti alapvetése a használatalapú konstrukciós nyelvészethez kötődik (vö. Bybee 2010, 2013; Diessel 2014; Langacker 1987). A konstrukciós nyelvészet számára a bipoláris (fonológiai és szemantikai pólussal rendelkező) nyelvi jel fogalma nemcsak a morféma és lexéma alkalmazható, hanem különböző komplexitású és absztrakciós fokú konstrukciókat foglal magába (vö. Goldberg 2006, 2013), így a grammatika és a lexikon szerveződése kontinuumként modellálható. A használatalapúság azt tételezi fel, hogy az egyes konstrukciók a nyelvhasználati események során keletkeznek, a reprezentációk stabilizálódnak vagy éppen meggyengülnek. A példányalapú reprezentációk között felismert fonológiai és szemantikai

² A továbbiakban az egyszerűség kedvéért „passzív jelentésű konstrukciók” néven említjük a négy vizsgált szerkezetet, tekintet nélkül arra, hogy az aktuális konstrukciós jelentések milyen arányban képesek passzív jelentést kifejezni.

³ A magmondat fogalmához lásd Imrényi (2017).

kapcsolatok alapján a konstrukciók viszonya hálózatos. Az asszociatív viszony létesülése nem tudatos szinten zajlik, a konstrukciókon gyakorisági hatások érvényesülnek.

Amikor azt állítjuk, hogy egy nyelvben „van passzív”, ez alatt rendszerint azt értjük, hogy létezik az adott nyelvben egy vagy több olyan produktív konstrukciós séma, melynek segítségével cselekvő igei bázissal ágensdefokuszált perspektívából konstruálhatunk meg eseményeket, ezért a passzívumot mindig jelölt igenemnek szokták tekinteni. A passzív fogalmának meghatározására a jelen tanulmány céljainak megfelelően egy inkább operacionális definíció szolgál: passzívnak tekintjük a morfoszintaktikailag jelölt fókuszváltást egy alapige két fokális résztvevője között, más szóval azt az esetet, amikor az alapige jelentése szerinti elsődleges és másodlagos figura (trajektor és landmark) egy konstrukció nézőpontszerkezetében „felcserélődik” (vö. Langacker 2006).⁴ A konstrukciós nyelvészeti megközelítés alapján a passzív kifejezőeszközök (is) alaki és szemantikai megfeleléseken alapuló, kapcsolódásokban gazdag hálózatot alkotnak (a hálózat modellálásához vö. Modrián-Horváth 2024). A jelen vizsgálatok pusztán metodológiai célból emelnek ki e hálózatból négy konstrukciót, miközben működésükre, elmebeli tárolódásukra egyaránt kihatással vannak a hálózat többi tagjai, így például a *V-ódik/-ődik* sémában is működő passzív jelentésű *-ó/-ő/-* (vö. Kugler-Simon 2018), a passzív jelentésű, kopula nélküli igenévi magmondatok (például *leves tálalva*, *küldetés teljesítve*, *a ház eladó*), a modális passzív konstrukciók (például *ehető*) vagy a nem magmondat szerepű passzív konstrukciók (például *Pontosz legyőzése*).

A passzív diatézis vizsgált kifejezőeszközeinek az inflexiós vagy szóképzési elemek közé való besorolása a konstrukciós nyelvészeti háttérnek megfelelően nem tartozik a jelen tanulmány fő kérdésfeltevései közé, legfeljebb annyiban, hogy az egyes konstrukciók produktivitási mértékét összehasonlítjuk egy tipikusan inflexiós jellegű analitikus konstrukció, a *V-ni fog* mutatóival.⁵ Az inflexiós elemekre a szakirodalom szerint magas, míg a derivációs

⁴ Langacker (2006) a passzív konstrukciót a „trajector-landmark reversal” egyik eseteként tartja számon.

⁵ Ladányi (2017: 567kk., különösen 102. lj.) a *V-ni fog* sémára illeszkedő, „jövő idejű, összetett, segédigés alakok”-at (lásd Ladányi 2017: 567-568) morfológiai szavaknak tekinti, leírásukat a morfológia feladatának tartja a szóalakokkal való funkcionális hasonlóság (lehorgonyzás), valamint az erős grammatikalizáltság miatt. Palágyi László lektorunk e besorolást többek között a korábbi szakirodalom alapján sem tartja problémamentesnek (Palágyi 2024, jelen kötetben).

elemekre alacsonyabb produktivitas jellemző, így ez az összehasonlítás közvetett módon segíthet árnyalni a vizsgált konstrukciók hovatartozásáról alkotott képet (az analitikus szerkezetek esetében is).

A produktivitas több dimenzióban meghatározható, összetett fogalom (vö. Ladányi 2008; Baayen 2009). A jelen tanulmány – számos produktivitas kutatáshoz hasonlóan (vö. Baayen–Lieber 1991; Baayen 1992; Dressler–Ladányi 2000) – Schultink (1961) van Marle (1985) által angolra fordított definícióját veszi alapul, mely az eredeti változatban csak a morfológiai produktivitasra terjed ki: produktivitas alatt a nyelvhasználók számára rendelkezésre álló, alak–jelentés párok összefüggésein alapuló, elméletileg korlátlan mennyiségű képzési produktum szándékolatlan létrehozásának lehetőségét értjük.⁶ A produktivitas méréséhez a kvantitatív szempontú globális produktivitas használjuk (lásd Baayen–Lieber 1991; Baayen 1992;⁷ Zeldes 2012; bővebb kifejtését lásd a 2. fejezetben). A produktivitas ebben a felfogásban graduális jellegű, viszonyfogalom. A tanulmány kérdésfeltevése elsősorban a négy vizsgált konstrukció produktivitasának egymáshoz képest való elhelyezése irányul.

1.2. Felépítés

Az 1–2. fejezet bemutatja a tanulmány kérdésfeltevést, az elméleti alapvetést és a használt korpusznyelvészeti módszert. A 3. fejezetben a passzív konstrukciók szemantikai tulajdonságait tárgyaljuk, ennek során kitérünk a konstrukciók poliszémiájára és a korpuszvizsgálat során jelentőséggel bíró elhatárolási nehézségekre is. A 4–7. fejezet az empirikus kutatásokról számol be. A 4. fejezetben elsőként bemutatjuk a vizsgálathoz használt korpuszt (MNSZ2-UD) és az adatgyűjtés folyamatát. Ezután ismertetjük az eredményeinket, előbb a teljes korpuszra, majd az egyes alkorpuszokra vonatkozóan. Az 5. fejezet az MNSZ2 egy alkorpuszából, a beszélt nyelvi alkorpuszból származó adatok kézi elemzésének folyamatát írja le és annak eredményeit mutatja be. A 6. fejezet arra a kérdésre keresi a választ, hogy az elemzett konstrukciók mely sémajelentéssel fordulnak elő produktívan,

⁶ „By productivity as a morphological phenomenon we understand the possibility for language users to coin, unintentionally, a number of formations which are in principle uncountable, by means of the morphological process which underlies the form–meaning correspondence of some words already known to them” (van Marle 1985: 100–101).

⁷ A globális produktivitas meghatározása Baayen (1992)-ben két mutatóra is vonatkozik, az itt használt formula a $P^* (n1/N; V)$.

ehhez a hapaxokból nyert minta kvalitatív vizsgálatát, valamint a magas gyakoriságú elemek áttekintését végeztük el.

Az empirikus eredményekre a 7. fejezetben reflektálunk: összevetjük a produktivitási mutatókat a szakirodalomban megfogalmazottakkal, összehasonlítjuk a gépi és kézi vizsgálat eredményeit, következtetéseket vonunk le a hapaxok vizsgálatának eredményeiből és állást foglalunk a kutatás metodológiai újdonságával, a szintetikus és analitikus konstrukciók produktivitásának együttes vizsgálatával kapcsolatban. A tanulmányt összefoglalással és a nyitott kérdések megfogalmazásával zárjuk (8. fejezet).

2. A vizsgálati módszerről

A vizsgálat kvantitatív jellege kapcsolódik egyrészt a kötet célkitűzéséhez, azaz a magyar igei konstrukciók nyelvtechnológiai módszerekkel történő kutatásához, másrészt illeszkedik ahhoz a tágabb, kognitív nyelvészeti keretben megfigyelhető trendhez, mely szerint a nyelvészek a használatalapúságot szem előtt tartva egyre erőteljesebben támaszkodnak empirikus, kvantitatív vizsgálatokra (vö. Janda 2013).

A főként Baayen által kidolgozott kvantitatív produktivitási modelleket, illetve mutatókat (vö. Baayen–Lieber 1991, Baayen 1992, 1993) számos használatalapú, illetve konstrukciós nyelvészeti tanulmány alkalmazza (például Stefanowitsch 2006 egy esettanulmányban, Zeldes 2012 és Hilpert 2013). Ennek egyik oka az, hogy a használatalapú nyelvtanok egyik centrális feltevése szerint a gyakoriság (típus- és példánygyakoriság) fontos szerepet játszik az egyes konstrukciós sémák mint mentális reprezentációk létrejöttében, stabilizálásában és működési mechanizmusában. Ezenkívül a produktivitást mindkét megközelítési mód fokozati fogalomnak ismeri el, és nem diszjunkt kategóriaként kezeli (produktív vs. nem produktív vagy szemiproduktív), felismerve, hogy a produktivitás hátterében nagyfokú individuális és szociokulturális nyelvi komplexitás áll (vö. Baayen 2009). További közös vonás, hogy a kvantitatív modellálás és a használatalapú nyelvtan kiindulópontja egyaránt nem egy „ideális” beszélő kompetenciája, hanem a nyelvi adatokon alapuló nyelvreírás, mely a kvantitatív modellálás esetében mérhető is.

Baayen (1992) a kvantitatív modellálási lehetőségek bemutatását leginkább pszicholingvisztikai összefüggésekbe helyezi és a lexikális reprezentáció erőssége, az elérés, aktiváció felől magyarázza, kitér azonban Bybee (1985) használatalapú elméletére, illetve az abban feltételezett gyakorisági hatásokra

is. Alapjában véve Baayen egyetért a gyakoriságnak a használatalapú morfológiai modellálásban tulajdonított szerepével, de túlságosan leegyszerűsítőnek tartja a gyakoriság és a szabályosság kapcsolatát Bybee modelljében. Míg Bybee abból indul ki, hogy az alacsony és közepes gyakoriságú típusok szabályos képzési módokat reprezentálnak, és a magas gyakoriságúak szabálytalanok, Baayen (1992: 140) megjegyzi, hogy a magas gyakoriságú típusok lehetnek szabályos képzésűek is, és ebben az esetben erősítik az adott képző produktivitását. (Ehhez a kérdéshez bővebben lásd a 6. pontot.)

A kvantitatív produktivitási mutatók célja a fentiek alapján nem diszjunkt besorolás a „produktív” vagy „nem produktív” kategóriába, értelmezéssel kapcsolatban mégis felmerül a *tertium comparationis* kérdése: mihez képest nevezhető egy konstrukciós séma inkább vagy kevésbé produktívnak? Baayen (1992) bevezet egy produktivitási „nullpontot”: a *per definitionem* terméketlen szimplexek (vagyis nem képzett szavak) produktivitási mutatói alkotják a viszonyítási alapot az általa alkalmazott kvantitatív módszerekben, nem produktívnak azokat a képzőket nevezi, amelyek mutatói nem haladják meg szignifikánsan a szimplexek értékeit (Baayen 1992: 116). Tekintve, hogy Baayen vizsgálatának tárgya a főnévképzés, a jelen tanulmány viszont igei konstrukciókra vonatkozik, a szimplexek vizsgálata itt kevésbé lett volna célravezető. A szimplex igei hapaxai gyakorlatilag mind pusztán elemzési hibákból adódnak, ellenben számos nagy gyakoriságú szimplex ige van a magyarban (például *van, jön, megy*), ezekből az adatokból a potenciális produktivitás gyakorlatilag nulla; a típusgyakoriság megállapítását a számos elemzési hiba jelentősen megnehezíti.

Az abszolút adatoknál sokkal érdekesebb számunkra az egyes passzív jelentésű konstrukciók produktivitási mutatóinak összevetése, szem előtt tartva, hogy a produktivitás graduális fogalom; az egyes sémák produktivitási mértékének arányát össze tudjuk hasonlítani. Ezenkívül összehasonlítási alapul szolgálnak számunkra egy analitikus konstrukció, a *V-ni fog* konstrukció produktivitási mutatói. Az analitikus „kontroll”-konstrukciót egyrészt azért vontuk be, hogy teszteljük, befolyással van-e a konstrukció analitikus jellege a produktivitási mutatókra pusztán a nagyobb szintaktikai kombinálhatóság révén. Másrészt a konstrukció a deriváció-inflexió kontinuumon inkább az inflexiók tartományban helyezkedik el, melyre többek között magasabb fokú produktivitás jellemző (ezt Gaeta 2007 korpuszvizsgálata szabályos szintetikus elemekkel kapcsolatban igazolta).

2.1. Kvantitatív produktivitási vizsgálatok

A produktivitás kvantitatív vizsgálatának számos módszere van, ezek közül a globális produktivitás csak egy – ez azonban az egyik leggyakrabban használt, elméleti szempontból jól alátámasztható módszer (lásd fent), mely statisztikailag is az egyik legjobb közelítésnek számít (vö. Zeldes 2012). Zeldes számos kvantitatív produktivitási vizsgálati módszer összehasonlítása során azt találta, hogy két, többdimenziós módszer tükrözi legjobban a nyelvi produktivitást: a szókészlet-növekedési görbe (vocabulary growth curve, VGC; ez a folyamatos adatfeldolgozással nyert görbe azt mutatja, hogyan növekszik a korpusz nagyságával az adott sémával képezett konstrukciók száma), valamint a globális produktivitás (P^*). Mivel Baayen (1992) alapján tudjuk, hogy ez a kettő szorosan korrelál egymással – a VGC görbe érintőjének meredekségét adja az $n1/N$ képlet, azaz a sémával képzett hapaxok száma, osztva a sémával képzett tokenek számával (lásd Baayen 1992: 115) –, számunkra elegendőnek bizonyult a kétdimenziós P^* meghatározása a vizsgált konstrukciókra.

A globális produktivitás (P^*) két dimenzióját a realizált/megvalósult produktivitás és a potenciális produktivitás adja. A realizált produktivitás egyenlő az adott sémával képzett típusok számával egy adott korpuszban – a továbbiakban egyszerűen típusgyakoriságként hivatkozunk rá –, míg a potenciális produktivitást úgy kapjuk meg, hogy a sémával létrehozott hapaxok számát elosztjuk a sémával létrehozott tokenek számával. Az eredmény egy koordináta-rendszerben ábrázolható, ahol az origótól mért távolság mérvadó a produktivitás szempontjából. További, a produktivitás kvantitatív mérésére használatos módszerek például:

- a TTR (type-token-ratio, típus-példány arány), mely klasszikusan a lexikai változatosságot fejezi ki, de nem ad számot például a hapaxok, illetve neologizmusok előfordulásáról, tehát nem ad teljes képet a produktivitásról, amennyiben a produktivitás fogalmát az új szavak létrehozásának lehetőségeként definiáljuk;
- a (kétfajta) HTR, melyet hapax/token ratioként, azaz hapax-token arányként (Stefanowitsch 2020: 318), illetve hapax/type ratioként, azaz hapax-típus arányként (vö. Tanos 2012, van Wettre 2022) szoktak számítani;
- a szintén a hapaxok számával operáló „terjeszkedő” produktivitás (expanding productivity, a magyar terminushoz lásd Kalivoda 2019), ez az adott sémával képzett hapaxok aránya a korpuszban található összes

hapaxhoz; a számítási képletnek megfelelően annak a valószínűségét méri, hogy egy új képzésű szó, azaz hapax, éppen a vizsgált sémával valósul meg;

- a potenciális produktivitás, mely a globális produktivitás egyik dimenzióját képezi, számítása: a hapaxok száma osztva a sémával létrehozott tokenek számával; annak a valószínűségét vizsgálja, hogy egy adott sémával létrehozott példány új lexéma;
- más globális produktivitási index például a Baayen által alkalmazott S/V, illetve az Aronoff (1976) által alkalmazott V/S, ahol V a sémával létrehozott típusok számát, S a teljes szókészletet jelöli.

A kvantitatív produktivitás-számítási módszerek átfogó ismertetését – az itt nem szereplő módszerekét is – lásd Baayen (2009), Zeldes (2012) és Hilpert (2013) műveiben.

2.2. A produktivitási mérések kiterjesztése szintaktikai konstrukciókra

A fent ismertetett kvantitatív produktivitási indexek elsősorban morfológiai produktivitás mérésére lettek kidolgozva, joggal merülhet fel tehát a kérdés, hogy mennyiben alkalmazhatóak ezek a mutatók analitikus igealakok – szintaktikai jellegű – konstrukcióira. Ehhez a következőket érdemes figyelembe venni. Az általunk választott elméleti keret (használatalapú kognitív nyelvészet) nem feltételez lényegi különbséget a különböző komplexitású és absztrakciós szintű nyelvi jelek működési elvében, ezért – ahogy Stefanowitsch (2006) is megállapítja – nemcsak morfológiai jellegű konstrukciókra alkalmazható Baayen indexe. Stefanowitsch esettanulmánya is egy grammatikai jellegű szószerkezetet vizsgál (a német *haben zu* + Infinitiv konstrukciót). Zeldes (2012) a lexikogrammatikai kontinuum elvére építve egy kifejezetten szintaktikai jellegű konstrukciótipusra, az argumentum-szerkezetre alkalmazza a produktivitási vizsgálatokat. Tanulmányunk metodológiai újdonsága abban áll, hogy itt két-két, hagyományosan morfológiai, illetve szintaktikai jellegű⁸ konstrukció produktivitásának összevetése a cél, melyek funkcionálisan hasonlóak egymáshoz, és egyaránt egy-egy igei komponens található bennük.

⁸ Bár a hagyományos grammatika (például Keszler szerk. 2000: 27–28) használja a „morfológiai típusú szószerkezet” terminust a segédigés szerkezetekre vonatkozólag, a V-*va/-ve* *van* típusú konstrukciót nem sorolja ezek közé. Imrényi (2017) ellenben e konstrukciókat olyan komplex magmondattípusnak tekinti, amelynek elemei együtt reprezentálják a lehorgonyzott folyamatot, valamint megemlíti passzív funkcionális jellegüket is.

3. A vizsgált konstrukciók szemantikai jellemzése

3.1. Konstrukciós poliszémia

A természetes nyelvek konstrukcióinak egyik alapvető tulajdonsága a poliszémia, nincs ez másként a passzív kifejezőeszközök esetében sem. A funkcionális és kognitív elméletek fokozott figyelmet fordítanak a poliszémia modellálására a sematikus („grammatikai”) jelentésű konstrukciók esetében is (például Janda 1990). A passzív jelentésű konstrukciók szemantikai leírásához az elméletet Modrián-Horváth (2024) tárgyalja részletesen, a jelen tanulmányban csak nagy vonalakban vázoljuk a konstrukciók által lefedett fogalmi tartományokat.

A vizsgált két szintetikus és két analitikus kifejezőeszköz is, eltérő mértékben, a passzívon kívül más, például mediális, rezultatív vagy akár progresszív jelentéstartalmak megkonstruálására is szolgál (Modrián-Horváth 2024). Ez a *V-ódik/-ődik* és a *V-va/-ve van/lesz* konstrukciók esetében nyilvánvalóbbnak tűnik, mint a par excellence passzívként számontartott *V-(t)atik/-(t)etik* sémával kapcsolatban. Azonban, ahogy H. Tóth (1991) Bécsi Kódexen végzett korpuszvizsgálata kimutatta, a *-(t)atik/-(t)etik* képző is számos jelentésben volt használatos (passzív, mediális, bennható, reflexív, kiható), melyek közül H. Tóth szerint a passzív jelentés produktívan (elavult, régies használatban), a mediális jelentés pedig nem produktívan jelen van a mai magyar nyelvben. A *V-ódik/-ődik* konstrukciós sémával kapcsolatban Szili (1999: 356) megállapítja, „hogy az *-ódik*, *-ődik* a kifejezendő szituációktól függően íveli át a passzívtól a mediálisig terjedő rokon jelentéstartományokat”.

A *V-va/-ve van* konstrukció esetében Honti–H. Varga (2006/2013) alapján megállapíthatjuk, hogy nagyfokú poliszémia jellemzi ezt a kifejezőeszközt. Honti és H. Varga (2013) Szepeszy (1972) alapján a konstrukció következő funkcióit sorolják fel: állapotjelölés, állapotváltozás; befejezettség, eredmény; személytelenség; a cselekvés intenzitásának a kiemelése; tömörítő árnyalás; a cselekvés befejezettségének kifejezése; a mozgás befejezettségének a jelölése; testhelyzet, biológiai állapot szemléltetése; állandósult kifejezések (Honti–H. Varga 2013: 260).

A jelen tanulmány a konstrukciós poliszémiát a konstrukciók egyik alapvető tulajdonságaként kezeli, melyre jellemző a radiális jelentésszerkezet (vö. Lakoff 1987, Lewandowska-Tomaszczyk 2007, grammatikai vonatkozásban Janda 1990), valamint a családi hasonlóság. Közös jelentésnyalábok figyelhetők meg például a *V-va/-ve van* konstrukció használataiban: passzív és

rezultatív például a *be van zárva* jelentése, reflexív/mediális-rezultatív a *meg van borotválkozva*, míg passzív-imperfektív/progresszív az *éreztetve van*. Bár az empirikus elemzésnél ez operacionalizálási nehézséget okoz, mégis természetes, hogy az egyes jelentések – különösen az olyan absztrakt szemantikai dimenziókban, mint a mediális és a passzív jelentés – nem különülnek el egymástól élesen (vö. 3.2.; a passzív és mediális igenem fokozati jellegéhez lásd Tolcsvai Nagy 2017: 331).

3.2. Passzív és mediális igenem: elhatárolási nehézségek

Egy hármas igenemfelosztás (aktív, mediális,⁹ passzív) esetén a passzívum határterülete a mediális igenemmel érintkezik, itt merülnek fel elhatárolási nehézségek az elemzésnél.

Tolcsvai Nagy Gábor (2017: 334) jellemzése szerint a prototipikus szenvedő (passzív) igenem trajektora páciens, „a folyamat rajta megy végbe, nyíltan jelen van, a mondatban kidolgozást kíván”, a landmark „erőforrás, rejtett (implicit) cselekvő, amely az eseményt végrehajtja”. A meghatározás szerint „[a] szenvedő igében az aktív cselekvő figurája háttérbe kerül, a folyamatot elszenvedő figura előtérbe helyeződik (a cselekvő ige jelentésszerkezetének fordított változataként)”. A Tolcsvai Nagy szerkesztette Nyelvtan (ONyt.) morfológiai fejezetében Ladányi (2017) említi az alacsony gyakoriságú passzívképzőt *-(t)atik/- (t)etik*, míg a mondatnyi fejezetben Imrényi (2017: 688k.) említést tesz az analitikus *V-va/-ve van/lesz* konstrukcióról a „régies” szintetikus passzív „ma is általános [...] típusú” alternatívájaként. Imrényi (2017: 689) szerint a „passzív [...] igék (például *becsukatik*) abban különböznek a szintén Páciens alanyú ún. mediális igéktől (például *becsukódik*), hogy a passzív igék esetében a konstruálás része az Ágens is (vö. *Az ajtó magától becsukódott. *Az ajtó magától becsukattott*). Ez az okozó szerepű, **aktív, tudatos szereplő** azonban 'nincs profilban' [...] [kiemelés a szerzőktől]. Az ONyt. tehát a **prototipikus**¹⁰ szenvedő ige leírására törekszik, amely egy **intencionális cselekvést** és cselekvőt involvál.

A mediális igék jelentésszerkezetében ezzel szemben a trajektor végpont szerepű, experiensként, témaként vagy elszenvedőként konstruált figura,

⁹ Az ONyt.-ban szereplő felosztásból (lásd Tolcsvai Nagy 2017) a visszaható és kölcsönös igéket jelentéstől függően, Kemmer (1993) alapján rokoníthatjuk a mediális igékkel, az elemzés során a vizsgált konstrukciókban adatolt visszaható igéket a mediális csoportba soroltuk (például *meg van borotválkozva*). A műveltető igék a cselekvőkkel rokoníthatók.

¹⁰ A prototípus elvű kategorizációhoz vö. Rosch (1978).

„amely vagy aki valamilyen történésnek, állapotváltozásnak, állapotnak az elsődleges szereplője. A másik figura kezdeményező szerepű, energiaforrásként konceptualizálva” (Tolcsvai Nagy 2017: 332). A landmark többnyire nem nyílt, de fokozati különbségek figyelhetők meg a mediális igék között arra nézve (is), hogy „a kezdeményező szerepű figura mennyire cselekvő és mennyire nyílt a jelenléte az ige jelentésszerkezetében” (lásd uott). A mediális igék jellemzése során Tolcsvai Nagy (2017: 333) megállapítja, hogy egy részük (ilyen például a *leng* ige) a szenvedőkhöz áll közel; a passzív és a mediális határán van „a *bevágódik* (az ajtó a huzattól), *borul* (az építmény)”. Egy másik kognitív szempontú csoportosításban Laczkó Krisztina (Laczkó 2023: 79) leginkább Kemmer (1993) nyomán, de E. Abaffy (1978) felosztásával is kompatibilis módon, azokat az igéket, melyeket „a **teljes szándékolatlanság** és az eseményszerkezet és/vagy szereplők kifejtetlenségének legerősebb foka” jellemez, passzív mediálisoknak nevezi;¹¹ ide tartozik a *letörik* és a *kiderül* ige (lásd uott, kiemelés a szerzőktől).

A passzív és mediális igék elhatárolását a legtöbb szerző a szándékosság/szándékolatlanság mentén próbálja elvégezni: a prototipikus passzív (szenvető) ige egy cselekvés fókuszváltással létrejött megkonstruálása, a cselekvés esszenciális összetevője egy intencionális ágens, mely landmarkként konstruálódik meg (a közvetlen hatókörben van, azaz a jelenet szereplőjeként akkor is feldolgozzuk, ha nincs explikálva), míg a mediális iginél nem található ilyen. Az intencionalitás kívánalma más megközelítésekben is meghatározó szerepet játszik, így például Siewierska (2013) szándékosságot kifejező határozókkal teszteli a passzív konstrukciókat a mediális (antikauzatív) konstrukcióktól való elhatárolás során. Ennek az alapján az (1) mondat passzív, míg a (2) mondat antikauzatív minősítést kap.

- (1) The door was opened deliberately.
- (2) The door opened (*deliberately).

A tesztelés a jelen tanulmány keretében metodológiailag nem megoldható, ellenben a tényleges nyelvi tevékenységben vizsgálható.¹² A „szándékosan

¹¹ Vö. Lengyel (2000a: 86 és 2000b: 69) mediopasszív kategóriáját és E. Abaffy (1978) mediális igéinek első két csoportját is.

¹² Hasonlóan a tesztelés elkerülése céljából Modrián-Horváth (2024) bemutatja, hogy az MNSZ2 adatai szerint mind a négy itt vizsgált konstrukció használatos az által névutós kifejezéssel, noha a szakirodalom szerint ez csak a V-(t)atik/-(t)etik esetében lehetséges; például „a hatóságok által volt beépítve” (doc#406), „Anne Frank keze által íródott” (doc#1057).

megrongálódott” szókapcsolatra a Google kereső segítségével hat különböző találatot kaptunk (lásd (3)–(4)).

- (3) Legkésőbb szeptember vége óta az Oroszországból Németországba irányuló gázszállítások tömegesen akadályoztatva vannak, miután az Északi Áramlat 1 és 2, az Oroszország és Németország közötti két fő gázvezeték feltehetően *szándékosan megrongálódott* szeptember 26-án.¹³
- (4) Ekkor a festmény *szándékosan megrongálódott*.¹⁴

A *szándékosan/direkt elfelejtődött* kifejezésekre is találunk példákat (lásd (5)–(6)).

- (5) Éppen ellenkezőleg: ebben az esetben azt gondolom, hogy ami elveszett, vagy *szándékosan elfelejtődött*, az éppen a cseh filmművészet identitásának és sajátosságának lényegéhez tartozhat.¹⁵
- (6) Ja, mintha *direkt elfelejtődött* volna..., mint az örmény nyelvtudásom...¹⁶

Azonban a *szándékosan*-teszt véleményünk szerint mégsem árul el sokat a cselekvés intencionalitásáról, már csak azért sem, mert különböző kontextusokban a szándékosság foka eltérő lehet, tehát egyazon igének létezhet mediális és aktív vagy mediális és passzív olvasata is. Ezenfelül a konstruálásban nagyfokú metapragmatikai tudatosság is szerephez juthat, ironikus olvasatot indukálva (vö. (5)–(6)) azáltal, hogy az aktív perspektívához kötődő *szándékosan* a mediopasszív perspektívájú ige által előhívott folyamat körülményét dolgozza ki.¹⁷

A passzív és mediális szemantikai tartomány elhatárolása könnyebb, ha az alapbeállítás szerinti trajektor (legtöbbször ágens, néha experiens)

¹³ https://forum.index.hu/Search/showArticleResult?topic_id=9136261&aq_ext=1&aq_text=megrong%E1l%F3dott

¹⁴ <https://hu.blogotirni.com/articles/elado-vaszil-iliev-utolso-portreja-bulgaria-today.html> [2023. 07. 27.]

¹⁵ <https://www.apertura.hu/2014/tel/hanakova-filmek-melyeket-szegyellunk/> [2023. 07. 25.]

¹⁶ <https://holdkatlan.hu/index.php/szepirodalom/proza/9392-piro-m-peter-mozgo-unnep> [2023. 07. 25.]

¹⁷ Köszönjük, hogy Palágyi László e körülményre felhívta a figyelmünket lektori véleményében.

a fókuszváltó konstrukcióban is a közvetlen hatókörben marad (vö. Imrényi 2017: 688–689).

Gyakran megfigyelhető azonban a passzív mintázatok metonimikus és metaforikus használata. A prototipikusnak tekinthető, intencionális (humán) ágens által végzett cselekvés sokszor valamilyen eszköz közvetítésével irányul az elszenvedőre, az oksági láncolat (causal chain; Croft 1991, 2001) alapján; az ágensi „behatás” itt az eszközhasználatról (például *átolajozódik*) akár egyetlen kattintásig terjed (egy tartalom *letöltődik*). A ráhatottság egyértelműen megvalósul, de a cselekvés kivitelezője ezekben az esetekben nem határolódik körül egyértelműen. Nem prototipikusan, de passzív jellegű a konstruálás például nem ágensi okozó esetében is (*elfelejtődik*); míg sok nem konkrét folyamat konstruálása (például *konstruálódik*, *elsáncolódik*) már szintén a mediális tartomány irányába mutat, mert az energiaforrás ezekben az esetekben nehezen azonosítható.

A passzív jelentéseknek a mediálistól való elhatárolása tekintetében érdemes egy pillantást vetni az „elismerten” grammatikalizálódott passzív konstrukciókkal rendelkező nyelvekre is. Az angol passzív használata mellett megjelenő, illetve kitehető *by*-konstrukciót szimptomatikus értékűnek szokták tekinteni, mely azt bizonyítja, hogy „valódi” passzívról és nem más predikatív konstrukcióról van szó (vö. például Siewierska 1984). Azonban rendszeresen találunk olyan megvalósulásokat, amelyeknél nem (intencionális) ágensi okozó mellett fordul elő a passzív forma, az angol *by*-frázis nem prototipikus ágens¹⁸ fejez ki. Az alábbi mondatpárok közül az első mondat mediopasszív (lásd (7), (9)), a második ágenszt tartalmazó mondat (lásd (8), (10)).

(7) Last weekend’s hiking trip was ruined *by the weather*.¹⁹

(8) I keep having dreams about the people who hurt me on my birthdays, and my birthday was ruined *by them*.²⁰

¹⁸ Az ágensi dimenziók differenciált megközelítéséhez lásd Dowty (1991) protoágens koncepcióját.

¹⁹ https://www.mmpublications.com/kazakhstan/site/Content/FlipBooks/FBKazGrade8_sb/files/basic-html/page140.html [2023. 08. 17.]

²⁰ <https://www.quora.com/I-keep-having-dreams-about-the-people-who-hurt-me-on-my-birthdays-and-my-birthday-was-ruined-by-them-I-always-wake-up-panicking-and-full-of-guilt-Is-that-a-sign> [2023. 08. 17.]

- (9) You have already been physically injured *by the accident*²¹
 (10) A second boy was shot and injured *by soldiers* in a nearby village²²

A pontosabb szemantikai besoroláshoz fontos, a használatalapúságból következő alapelv a konstrukcióknak támogató közegükben való vizsgálata, így többek között a diskurzus szerepének, a kontextusnak és a nyelvi variációnak a figyelembevétele (vö. Ladányi–Tolcsvai Nagy 2008: 33). A nagy korpuszok használata során ez a követelmény (például a produktivitas mutatóinak megállapításánál) nem vagy nehezen kivitelezhető. Kézi elemzéssel azonban kisebb adathalmazon elvégezhető, a jelen dolgozatban is az egyes példányok – hapaxok – elemzése támogató közegben történt, ezáltal az egyes konstrukciós példányoknak a passzív-mediális-aktív kontinuumon való besorolása pontosítható volt. A *fűrészselődik* szó mint típus nem árul el semmit arról, hogy milyen ágens/ráható erő által történik a fűrészelés, ezáltal nehézségekbe ütközik a szó besorolása a mediális vagy passzív kategóriába; míg az alábbi kontextus figyelembevételével mediopasszív kategóriába sorolható az előfordulás (11):

- (11) Aztán amikor mélyebbre fűrészselődött a völgy, a barlang teljesen szárazzá vált. (doc#2573)

3.3. A V-(t)atik/-(t)etik konstrukció

A -(t)atik/-(t)etik képzőt a hagyományos nyelvtan az egyetlen, produktív passzívképzőként tartja számon (vö. Keszler szerk. 2000). Kenesei et al. (1998) korlátozottan tartja a képző produktivitását, azonban az esetleges alternatív passzív konstrukció tekintetében nem foglal állást. Ladányi (2017: 610) szerint a képző elvileg korlátozás nélkül használható minden tárgyas igével, azonban az egyes ide tartozó igék példánygyakorisága alacsony, és használata a hivatali nyelvbe, valamint a szépirodalom nyelvbe szorult vissza.

²¹ <https://donaghuelabrum.com/2022/02/17/post-car-accident-checklist-for-pennsylvania-drivers/> [2023. 08. 11.]

²² <https://www.reuters.com/article/uk-israel-palestinians-violence-idUKBRE90E0F020130115> [2023. 08. 11.]

3.4. A V-ódik/-ődik konstrukció

A V-ódik/-ődik konstrukciót számos leírás a szintén szintetikus V-(t)atik/-(t)etik séma potenciális utódjaként említi meg (lásd Szili 1999, Grétsy 2014, Ladányi 2017), ez a séma vehetné át „a” passzívképző szerepét a magyar nyelvben. Emellett szólhatna az is, hogy a konstrukciót produktívan használják nyelvjárási beszélők passzívképzőként a Tisza, Szamos környékén – például *a fa kivágódik* (Grétsy 2014). Amennyiben ez bekövetkezne, a következő, egy gazdasági szakkönyvből származó, a fák oltásáról szóló idézet félkövérrel kiemelt igéi nem tűnnének idegenszerűnek az olvasó számára (Tsötönyi 1831: 131, kiemelés a szerzőktől).

- (12) „A’ fa, vagy ág, melybe oltanak, egyenesen kerekre *fűrészeliódik* ’s *faragódik*. Éppen a’ székén, velején, vagy közepén keresztül, ha lehet tsak annyira *hasitódik* be, a’ mennyire az oltó ág ék formára faragott vége le ér. Hogy a’ hasítás össze ne menjen, tömpe vastag ék *verődik bele*. Az oltó ág gyűrűjén alól, két kis uj szélességre *vágódik el*, úgy *faragódik* laposra és élesre, hogy a bele vagy közepe meg nem *sértődik*.”

3.5. A V-va/-ve lesz és van konstrukciók

A két konstrukciót a *van* és *lesz* tövek szuppletív használata (például *lesz* mint a létige jövő ideje, *lenne/lett volna* mint a létige feltételes módú alakjai, *legyen* mint felszólító-kötő mód) miatt nem lehet egyértelműen elhatárolni egymástól, egyetlen konstrukcióként való kezelésük mégsem elfogadható. Ennek oka leginkább abban rejlik, hogy a *van* statikus, a *lesz* dinamikus sematikus jelentéssel bír. Nehéz lenne kijelenteni, hogy az *át lesz irányítva* konstrukció kizárólag jövő időben horgonyzódik le, időjelentése inkább aktuális vagy általános jelen vonatkozású (míg más előfordulások természetesen a jövőre vonatkoznak). Továbbá a V-va/-ve *volt* mellett az elkülönült V-va/-ve *lett* múlt idő is az önálló konstrukcióként való vizsgálatot támogatja. Elővizsgálataink szerint (és ezt a jelen kutatás is igazolja) a V-va/-ve *van* konstrukció magasabb fokú poliszémiát mutat a V-va/-ve *lesz* konstrukciónál, valamint a passzív/mediális konstrukciós jelentés megoszlása tekintetében is erőteljes különbségek tételezhetők fel a két séma produktív használatában. Az ONYt. szerint (Imrényi 2017: 688–689) az *Ajtó be lett csukva* típusú konstrukció a „régies” szintetikus passzív „ma is általános [...] típusú” alternatívája

(Imrényi 2017: 688k.). A *V*-*va/-ve van/lesz* konstrukciókat Kádár–Németh (2009) szerint produktívan használják a moldvai magyar nyelvjárásokban.

3.6. További passzív jelentésű magmondat-konstrukciók

Meg kell említeni, hogy a nem modális (medio)passzív tartományban való konstruálásra további (magmondat-) konstrukciók is alkalmasak, így például a befejezett melléknévi igeneves magmondatok (például *a torony felállványozott, a gázellátás biztosított*), vagy más, alapvetően mediális jelentésű sémák használata (például a *V-ít* tranzitív sémával kontrasztot képező *V-ul/-ül* használata a (13–15) példákban).

- (13) Várjon, amíg a frissítés *települ!* (Adobe Reader rendszerüzenet, 2019. márc.)
- (14) A Microsoft Edge *bezárult*, amíg néhány lap meg volt nyitva. (Microsoft rendszerüzenet, 2023. 07. 21.)
- (15) Híreink pedig folyamatosan *frissülnek* az interneten (Kossuth rádió, Déli krónika, 2019. márc. 9., 12.39)

Ezeket a konstruálási sémákat jelen korpuszelemzés nem vizsgálja.

4. A konstrukciók globális produktivitási vizsgálata az MNSZ2-UD korpuszban

Az első eredményeink ismertetése előtt nagy vonalakban bemutatjuk a gépi módszerek alkalmazásának előnyeit és hátrányait azzal a céllal, hogy a kutatás lehetőségeit és korlátait kontextusba helyezzük. A gépi módszer előnyei között elsőként azt említenénk, hogy ezáltal rövid idő alatt hatalmas mennyiségű adat elemezhető, és olyan tendenciák rajzolódnak ki, amelyeket hagyományos módszerrel nem lehetne elérni. Másodszor, több műfajt képes magába foglalni, mint a kézzel végezhető elemzés. Stefanowitsch (2006) minél nagyobb, minél több műfajt tartalmazó korpusz használatát javasolja. Ennek kiemelt jelentősége van a produktivitás vizsgálatakor, hiszen Baayen (2009) figyelmeztet arra, hogy a produktivitás igen eltérő lehet az egyes műfajokban (sőt egyedi különbségeket is mutat az egyes szerzők repertoárjában). Harmadszor, az elemzéshez használt forrásanyagok, szoftverek, programkódok birtokában a teljes elemzési folyamat reprodukálható.

A kapott eredmények könnyebben ellenőrizhetők mások által is, és könnyebben felhasználhatók későbbi kutatásokban. Hátrányként pedig azt kell elkönyvelnünk, hogy a gépi elemzés minősége nem éri el egy kézzel összeállított, átnézett adathalmaz minőségét.

4.1. A felhasznált korpusz

A Magyar Nemzeti Szövegtár 2 (Oravecz et al. 2014) egy 1,04 milliárd szövegszóból álló, automatikusan elemzett, általános célú korpusz, amely a 20. század végének és a 21. század elejének magyar nyelvhasználatának reprezentálására törekszik. A kutatásunkban ennek a legújabb – megjelenés előtt álló – változatát vettük alapul, az MNSZ2-UD korpuszt (Kalivoda et al. 2024). Ez a verzió három szempontból tér el az eredeti MNSZ2-től. Egyrészt számos korpusztisztítási lépésen ment keresztül: ilyen például a bekezdésszintű duplumszűrés, a nem magyar nyelvű bekezdések automatikus azonosítása és törlése, valamint az extrém hosszú mondatok kiszűrése. Ezáltal a szöveg minősége jobb, de a korpusz mérete kisebb az MNSZ2-höz képest (776,9 millió szövegszó). Másrészt a stílusregiszter és a régió mellett elérhetővé vált a publikálás évére vonatkozó metaadat is a szövegek 93,8%-a esetében. Harmadrészt a korpusz új elemzési szintekkel gazdagodott: az emMorph (Novák et al. 2016) által létrehozott, részletes morfológiai elemzéssel, valamint függőségi elemzéssel, amely a Universal Dependencies formalizmusát követi (de Marneffe et al. 2021).

A jelenlegi kutatáshoz elsősorban arra volt szükség, hogy a lehető legpontosabban tudjuk azonosítani az igét, amelyhez egy-egy elvált igekötő tartozik. Ezért úgy döntöttünk, hogy a korpusz szövegeit végigelemezzük az emPreverb eszközzel is (Pethő et al. 2022). Ez a szabályalapú algoritmus kifejezetten a magyar igék és igekötők összekapcsolására készült, és ezzel érhető el a legpontosabb annotáció Pethő et al. (2022: 86–87) kiértékelése alapján.

4.2. Az adatgyűjtés folyamata

A teljes korpuszból lekérdeztünk minden olyan szóalakot, amely számunkra releváns képzőt tartalmaz, tehát a *-va/-ve*, *-atik/-etik/-tatik/-tetik* vagy *-ódik/-ődik* szuffixumok valamelyike azonosítható benne. Minden szóalak esetében megvizsgáltuk, hogy tartozik-e hozzá prefixált vagy elvált igekötő. Az utóbbit az emPreverb annotáció segítségével találtuk meg. Ennek a lépésnek az volt a célja, hogy az igekötős igelemmák egyforma reprezentációt

kapjanak, függetlenül az igekötő pozíciójától (például az *elkezdődött* és a *kezdődjön el* lemmája egységesen *el#kezdődik* lett). Egymás alakvariánsainak tekintettük és azonos alakra normalizáltuk a következő igekötőket: *belé* → *bele*, *bé* → *be*, *föl* → *fel*, *fölül* → *felül*, *fönn* → *fenn*, *odább* → *odébb*, *rá* → *rá*, *teli* → *tele*.

A *-va/-ve* képzős alakok esetében megvizsgáltuk azt is, hogy létigével állnak-e (amelyet maximum három tokenes bal és jobb kontextusban kerestünk), és ha igen, pontosan milyen alakban áll a létige. A 2.5. pontban megfogalmazottak alapján arra számítottunk, hogy a *V-va/-ve* + *lesz*, illetve a *V-va/-ve* + *van* szerkezetek használata, ezáltal a produktivitása is számottevő eltéréseket mutathat. A korpuszban különálló lemmaként jelenik meg a *lehet*, amelyet *lesz*-re, valamint a *nincs* és a *sincs*, amelyeket *van*-ra konvertáltunk át. Az igekötős *-va/-ve* képzős alak jobb kontextusában megjelenő létige kapcsán azzal szembesültünk, hogy szinte minden találat fals pozitív, azaz nem a keresett konstrukció egy példánya (például *a szavazatok közel 53%-át megszerelve lett a Francia Köztársaság XXIII. elnöke*, ahol a *megszerelve lett* hibás pozitív találat a keresett konstrukció szempontjából). Ezért a vizsgálatunkhoz csak azokat a találatokat őriztük meg, ahol a létige megelőzi a *-va/-ve* képzős alakot (például *lett megszerelve*, *meg lett szereve*), vagy ahol a létige előtt álló *-va/-ve* képzős alak igekötő nélküli határozói igenév (például *tájékoztatva lett*).

Minden találat esetében megőriztük a teljes mondatot, valamint a publikálás éveire és a dokumentum regiszterére vonatkozó metaadatokat. A kvantitatív produktivitási vizsgálatunkban csak azokat a találatokat vettük figyelembe, amelyek 2000-ből vagy azt követő évekből származnak.

A bemutatott adatgyűjtési folyamatot elvégeztük a *V-ni fog* konstrukció példányainak azonosítására is. Itt a (ragozott és ragozatlan) infinitívuszi szóalakot kerestük elsőként, majd azonosítottuk ennek a lehetséges igekötőjét, és maximum három tokenes bal és jobb kontextusban kerestük a *fog* lemmát. Ezután elvégeztük a passzív jelentésű konstrukciók, valamint kontrollcsoportként a *V-ni fog* szerkezet globális produktivitásának vizsgálatát, egyrészt a teljes korpuszra vetítve, másrészt az egyes regiszterek szerinti bontásban. A következő két fejezetben ismertetjük a méréseink eredményét.

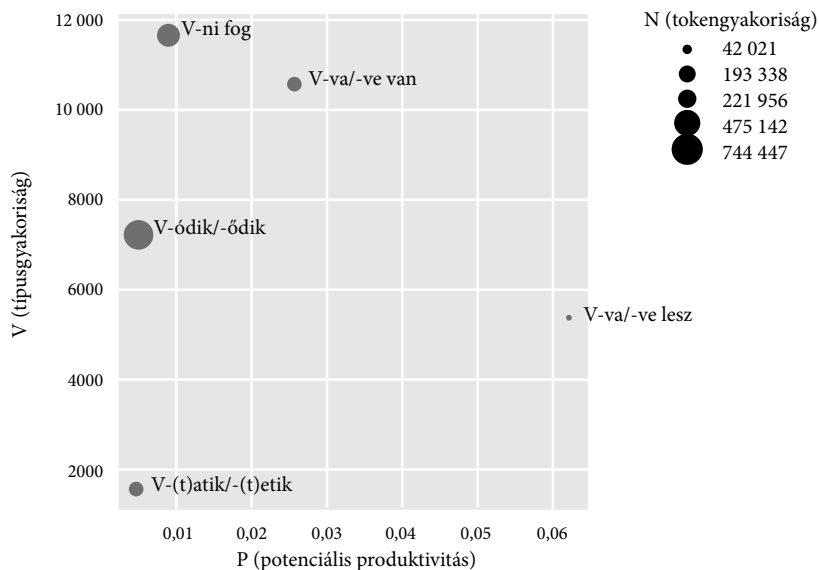
4.3. Az egész korpuszra vetített eredmények

A globális produktivitás megállapításához négy értékre van szükség minden vizsgált konstrukció esetében. Ezek a tokengyakoriság (hány példányban adatható a szerkezet a korpuszban), a típusgyakoriság (hány igelemméval jelenik meg), a hapaxok száma (hány egyszer előforduló igelemméval szerepel), valamint a potenciális produktivitás, amelyet úgy kapunk, hogy a hapaxok számát elosztjuk a tokengyakorisággal.

1. táblázat. A vizsgált konstrukciók tokengyakorisága (N), típusgyakorisága (V), hapaxainak száma (n₁) és potenciális produktivitása (P) a teljes korpuszban

szerkezet	N	V	n₁	P
V-ódik/-ődik	744 447	7203	3647	0,0049
V-va/-ve van	193 338	10 539	4978	0,0257
V-(t)atik/-(t)etik	221 956	1528	1033	0,0047
V-va/-ve lesz	42 021	5341	2616	0,0623
V-ni fog	475 142	11 625	4202	0,0088

A globális produktivitást úgy kapjuk meg, hogy a potenciális produktivitást és a típusgyakoriságot egy-egy dimenzióknak tekintjük, ezáltal kétdimenziós térben ábrázolni tudjuk (az előbbi egy koordinátarendszer X-tengelyén, az utóbbit az Y-tengelyén szokás megjeleníteni). A vizsgált konstrukciók egy-egy adatpontként elhelyezhetők ebben a térben az értékeiknek megfelelően. Még összetettebb képet kaphatunk az adatról akkor, ha az egyes konstrukciók tokengyakoriságát az adatpontok nagyságával szemléltetjük.



1. ábra. A vizsgált konstrukciók globális produktivitása a teljes korpuszban

Az 1. ábráról a következők olvashatók le:

- A *V-(t)atik/-(t)etik* képző a mai magyar nyelvben igen korlátozottan produktív: mind a potenciális produktivását, mind a típusgyakoriságát nézve alacsony értékeket látunk. Egyedül a tokengyakorisága számottevő (a 3. helyen végzett a vizsgált öt konstrukció között). Ennek a konstrukciónak egyúttal a legalacsonyabb a TTR-aránya (típus-példány aránya) is, mely nagyságrendi különbséget mutat a *V-va/-ve van* és *V-va/-ve lesz* sémákhoz képest (lásd 2. táblázat).
- A *V-ódik/-ődik* globális produktivitása már jelentősebb: típusgyakoriságát tekintve a középmezőnyben szerepel, de a potenciális produktivitása alacsony (a magas tokengyakoriságához képest kevés hapaxban jelenik meg).
- A vizsgált passzív jelentésű konstrukciók közül leginkább a *V-va/-ve van* és a *V-va/-ve lesz* produktív, de más-más dimenziók mentén: az előbbinek a típusgyakorisága, az utóbbinak a potenciális produktivitása a kiemelkedő.

- Potenciális produktivitás szempontjából 0,005 alatti értéket mutat a *V-ódik/-ődik* és a *V-(t)atik/-(t)etik*; 0,01 alatti a *V-ni fog*, melynél csaknem háromszor produktívabb a *V-va/-ve van*, és utóbbinál is csaknem két és félszer magasabb értéket mutat a *V-va/-ve lesz*.
- Az összehasonlítási pontként használt *V-ni fog* szerkezet a legnagyobb típusgyakoriságú, de az alacsony potenciális produktivitását tekintve inkább a szintetikus szerkezetekhez közelít.

2. táblázat. A vizsgált konstrukciók típus-példány aránya

szerkezet	N	V	V/N
<i>V-ódik/-ődik</i>	744 447	7203	0,0097
<i>V-va/-ve van</i>	193 338	10 539	0,0545
<i>V-(t)atik/-(t)etik</i>	221 956	1528	0,0069
<i>V-va/-ve lesz</i>	42 021	5341	0,1271
<i>V-ni fog</i>	475 142	11 625	0,0245

4.4. Az alkorpuszok eredményei

A konstrukciók globális produktivitását megvizsgáltuk alkorpuszok, azaz regiszterek szerinti bontásban is. A következőkben csak azokat a megfigyeléseket emeljük ki, amelyek a teljes korpuszon végzett vizsgálathoz képest nagy eltérést mutatnak. Az egyes regiszterekre vonatkozó részletes eredményeket mellékletben közöljük.

- Szépirodalom (I. sz. melléklet): a teljes korpuszhoz nagyon hasonló eredményt kaptunk, csak annyi a lényegesebb eltérés, hogy a *V-va/-ve lesz* konstrukció típusgyakorisága alacsonyabb.
- Ezzel szemben a személyes alkorpusz (II. sz. melléklet) esetében a *V-va/-ve lesz* típusgyakorisága magas, felülmúlja a *V-ódik/-ődik* sémáét.
- A hivatalos alkorpuszban (III. sz. melléklet) a passzív jelentésű konstrukciók egymáshoz viszonyított távolsága hasonlóan alakul a teljes korpuszhoz; a *V-ni fog* szerkezet távolodik el minden másától a nagyobb típusgyakoriság irányába.
- A sajtónyelvi alkorpuszban (IV. sz. melléklet) megjelenő tendenciák egyeznek a hivatalos alkorpuszban látottakkal.

- Az eddigi mintáktól jelentősebben eltér a tudományos ismeretterjesztő alkorpusz (V. sz. melléklet): az összehasonlítási alapul használt *V-ni fog* szerkezet kisebb típusgyakoriságú, míg a *V-va/ve van* konstrukció feltűnően produktív mindkét dimenzió mentén.
- A beszélt nyelvi alkorpusz (VI. sz. melléklet) olyan szempontból kilóg a sorból, hogy valójában az „írott nyelvvel” állna kontrasztban, tehát a korábbi öt regiszterrel együttesen. A beszélt nyelvi alkorpusz a regisztert tekintve heterogén, informális és formális szövegeket egyaránt tartalmaz. A passzív jelentésű konstrukciók egymáshoz viszonyított globális produktivitása nem tér el jelentősen a teljes korpuszon mérttől. Talán azt lehetne kiemelni, hogy a *V-ni fog* szerkezet értékei alakultak máshogy (kimagasló típusgyakoriság és token gyakoriság – a korábbi mérésekben rendszerint a *V-odik/-ődik* token gyakorisága volt a legmagasabb).

5. Kézi elemzéses vizsgálat a beszélt nyelvi alkorpuszon

Ahogy a 4. fejezet bevezetőjében említésre került, a gépi elemzésnek nemcsak hatalmas előnyei, de hátrányai is vannak. Gyakran fogalmazódik meg a gépi elemzéssel szembeni kritikaként, hogy az azt végző nyelvésznek nincs elég alapos betekintése, rálátása azokra az adatokra, amelyekkel dolgozik. Másrészt az automatikusan kinyert adatok között sok téves vagy irreleváns találat is lehet. Baayen (2009: 907) figyelmeztet a pusztán gépi elemzés „zajosságára”, illetve arra, hogy torzulhatnak az eredmények a gépi elemzésnél. A jelen tanulmány ezért egy korábbi, kézzel végzett elemzést is tartalmaz 3 sémával kapcsolatban (*V-va/-ve van*, *V-va/-ve lesz* és *V-(t)atik/-(t)etik*, amelyet az MNSZ2 (Oravecz et al. 2014) egyik alkorpuszából, az ún. beszélt nyelvi alkorpuszból származó adatokon végeztünk (szemben a 4. fejezetben bemutatott, gépileg duplumszűrt változattal). A korpusz mérete mintegy 76,2 millió szövegszó, ez az MR1 Kossuth rádió bizonyos anyagait öleli fel az 2004–2012 évekből, felolvasott szöveget (hírek) és spontán beszélgetést (riportok) vegyesen tartalmaz. Az alkorpusz választását az indokolta, hogy várhatóan itt volt a legnagyobb a (fél)spontán megnyilatkozások aránya.²³ A negyedik, *V-odik/-ődik* sémára nézve az adatokat kiegészítettük

²³ Ezenfelül véletlenszerű minták vizsgálata alapján ez volt a duplumokkal legkevésbé terhelt alkorpusz, szemben más alkorpuszokkal: a személyes alkorpuszban nagy mennyiségben fordul elő többszörösen beágyazott idézet (a forrásként szolgáló internetes fórumok egyik jellegzetessége), így ugyanaz a szövegdarab újra és újra megismétlődhet. Itt például egy

az MNSZ2-UD korpusz megfelelő alkorpuszán mért anyaggal, mivel ez kézilag feldolgozhatatlan mennyiségű előfordulást tartalmazott; a gépi adatokat csak összehasonlításképpen tüntettük fel és csillaggal jelöltük.

5.1. Az adattisztítás és -feldolgozás folyamata

Ahhoz, hogy a hapaxok pontos száma megállapítható legyen, szükséges volt a rendelkezésre álló szövegeket megtisztítani és elemezni.

Ennek első lépésében a kiválasztott grammatikai tulajdonságokkal rendelkező lemmákat 300 karakteres kontextusukkal együtt lekértük az MNSZ2-ből. Az analitikus igealakok vizsgálatához a határozói igenévként annotált szavakat kerestük, az eredményt a *van/lesz* lemmával való együttes előfordulásra pozitívan szűrve egy mondaton belül.²⁴ Az igenévvél záruló elváló ige-kötős konstrukciók esetében (például *be van zárva*) 2, a nem elváló ige-kötős vagy ige-kötő nélküli esetben 3 szó beférkőzését engedték meg (azaz az ige-kötő legfeljebb 4 szóra távolodott el az igenévtől, ami az ige-kötős konstrukciók esetében Kalivoda (2021) eredményei alapján a korpusztalálatok 99,88%-át lefedi – azzal a kitételrel, hogy a beszélt nyelvi korpusz esetében a találati arány némileg alacsonyabb lehet). A *van*, illetve *lesz* igével záruló konstrukciók esetén 1 szó beférkőzését engedték meg (például *zárva lesz*). A *V-(t)atik/-(t)etik* séma vizsgálatához *-atik*, *-etik*, *-tatik* vagy *-tetik* végződésű lemmákra kerestünk, majd negatívan szűrtük az eredményeket *-hatik*, illetve *-hetik* végződésre.²⁵ A *V-odik/-ődik* séma kézi vizsgálata (kézi ige-kötőkapcsolás) a találatok magas száma miatt ebben a körben nem volt megoldható, ezért itt a korábbiakban bemutatott, automatikus ige-kötőkapcsolást tartalmazó részkorpuszt használtuk; itt is az *-odik* vagy *-ődik* végződésű lemmákra kerestünk.

Az így kapott szövegeket Google Sheets táblázatokban és Python (v.3.10) programnyelvben dolgoztuk fel a következő lépésekben.

esetben (az *előadat* hibásan *előadatik*-ként lemmatizált szó esetében) egyetlen valós példány másolatai alkották csaknem az összes találatot, pontosan 260-at a 265-ből (a maradék öt találat között is van duplum, és mindegyik hibás lemmatizálásnak köszönhető).

²⁴ Keresőkifejezések: *V-va/-ve van* és *lesz* esetében inverz szórenddel: szóalak: határozói igenév, pozitív szűrés [+1, +2] ablakban *lesz*, illetve *van* szótőre ige szófajjal, tetszőleges ragozással, negatív szűrés [0, +1] ablakban tetszőleges írásjelre. Semleges szórenddel a pozitív és a negatív szűrési ablak egyaránt [-3, -1] volt, mert a felhasznált korpuszban az írásjelek nem számítanak külön tokennek.

²⁵ Keresőkifejezés: szótó: **(atik|etik|tatik|tetik)* és szófaj: ige, negatív szűrés 0 szó távolságban szótó: **(hatik|hetik)*

Elsőként a szövegeket kisbetűsítettük, eltávolítottuk az összes írásjelet (kivéve azokat, amelyekre technikai okokból szükség volt egy-egy esetben), és az így keletkezett változatokból töröltük a felesleges szóközlőket.

Az így egységesített szövegekben a következő szabállyal kerestünk duplikátumokat: ha két találati mintában legalább egy ötven karakter hosszúságú szakasz megegyezett, akkor az adott elemeket azonosnak tekintettük, és a második példányt töröltük. A viszonylag szigorú kritériumot az indokolta, hogy a rádiós szövegekből álló korpusz sokszor ugyanannak a diskurzusrészletnek több, részben eltérő lejegyzését tartalmazza, így bizonyos mértékű eltéréssel itt is ismétlődéseket találhatunk. Az ötven karakteres határ bevallottan önkényes, azonban kézi ellenőrzéssel validáltuk több kisebb (≈3000 kontextus/minta) találatlistán, és nem produkált fals pozitív eredményeket. A (16a-b) szöveg példa az olyan minimális eltérésű ismétlődésre, ahol az egyetlen elhangzott diskurzusrészlet két különböző transzkripció miatt kétféleképpen került a korpuszba (például egyszer *szemész*, egyszer *zenész* alakban szerepelt ugyanaz a hangsor; az elütéseket a korpusz tartalmazza).

- (16) a. Tehát ez teljesen elképesztő, ugyanúgy működött mint ma, mondjuk például ismerek egy szemészt, aki ilyen reklámcsészeket készítettett és *a saját szemészeti pecsétje van beleütve*, az ő általa forgalmazott szemkenőcsöket akarta ezzel reklámozni. (doc#2608)
 b. Ez teljesen elképesztő. Ugyanúgy működött, mint ma. Például, ismerünk egy zenészt, aki reklám csészeket készítettett, és *a saját pecsétje van beleütve*. Az ő által forgalmazott szemkenőcsöket akarta ezzel reklámozni. (doc#2647)

A duplikátumok eltávolítása jelentős változásokat hozott a fennmaradó szövegek mennyiségében, gyakran 25% körüli szövegről derült ki, hogy duplum. Ez a megfigyelés fontos kérdéseket vet fel a tisztítatlan korpuszokon végzett statisztikai elemzések megbízhatóságával kapcsolatban, amennyiben az eredeti korpusz nincsen valamilyen módon hatékony zajszűréssel ellátva.

Az így kapott találati listából kinyert hapaxok és alacsony tokengyakoriságú elemek ezután egyenkénti ellenőrzésre kerültek, kiszűrendő a fals találatokat, például bibliai vagy irodalmi idézeteket. A bizonyított, számunkra fellelhető forrásból származó idézeteket eltávolítottuk az adatsorból, mert sokszor többszáz éves vendégzsövegek kerültek volna be az adatok közé,

a vizsgálat viszont a mai magyar nyelvhasználatban tetten érhető produktivitást hivatott vizsgálni. A szállóigévé vált idézetek (például *a kocka el van vetve*) a listában maradtak. Ellenőrzésre szorult az is, hogy nem történt-e „véletlen” veszteség, amikor is egy szövegben két keresett kifejezés is szerepel, és az egyik alapján a szöveg duplumnak bizonyult, így törlődik, ami által a másik – releváns – találat elvész. Ezekben az esetekben a szöveget visszaállítottuk és a példányokat felhasználtuk a számításokhoz.

A gépi és kézi szűrés után fennmaradt példányok és hapaxok számát ezután határoztuk meg, és számoltuk ki a produktivitásra utaló értékeket.

A *V-ódik/-ődik* konstrukcióba tartozó megvalósulások nagy száma miatt (hozzávetőleg 106 000 token duplikátumok kiszűrése előtt és 76 000 utána) a kézi ellenőrzés nem volt lehetséges, így a duplikátumok eltávolítása után ezeknél csak az elveszett hapaxok visszaállítására került sor a produktivitási értékek kiszámítása előtt. Ebben a csoportban az igekötőkapcsolás is gépileg történt a beszélt nyelvi korpuszrészben (a módszerhez lásd 4.), ezért a *V-ódik/-ődik* sémára vonatkozó adatok csak korlátozottan hasonlíthatók össze a többi eredménnyel.

Az analitikus alakok és a *V-tatik/-tetik* séma vizsgálatánál kézi igekötőkapcsolás történt (utóbbinál ettől csak a legnagyobb, 100 fölötti gyakoriságú lemmák esetében tekintettünk el, mivel ezek az adatokba való betekintés alapján nem befolyásolták érdemben a produktivitási mutatót, például *születik, megszületik, adatik, megadatik*). A folyamat során az igekötős igéken kívül (például *be van zárva*) egy lemmához tartozónak tekintettük a következőket:

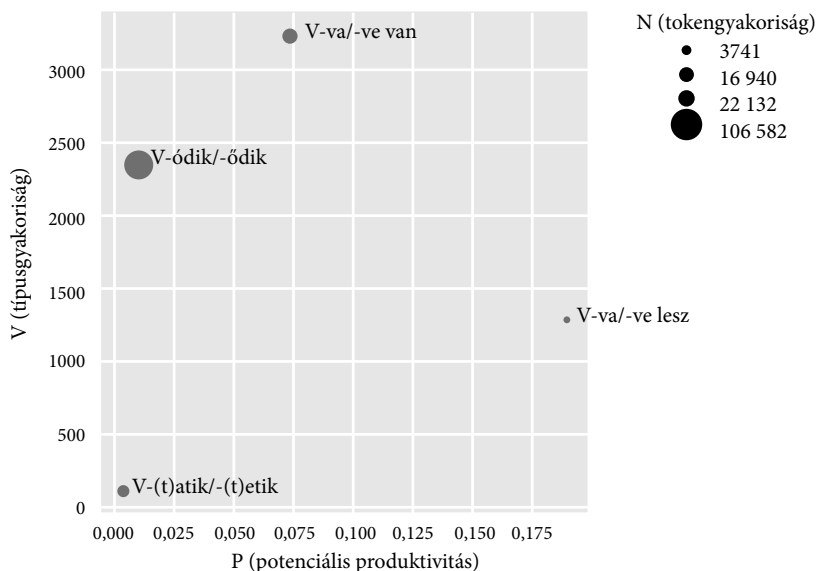
- Személyes névmások és igék kapcsolata (például *rám van bízva*), mely analóg a *rá van bízva* előfordulással, így egy lemmához soroltuk; egyekben az egységesség kedvéért a helyesírási előírásokat követtük (például a *számontart* és a *tart* külön lemmát képvisel, de a *férjhez megy* a *megy* lemmához tartozik).
- A regionális, stilisztikai alakváltozatokat, például *fel/föl, bocsát/bocsájt*, egy lemmához soroltuk, hogy ne gyarapítsuk indokolatlanul a hapaxok számát; hasonlóképpen az MNSZ2 az *alkuszik/alkud-* tőlváltozatokat automatikusan egy lemmához sorolja.
- Néhány konstrukciónál még a kontextus pontos ismeretében sem lehetett megállapítani a bemeneti igetövet, ilyen például a *tör/török* igék *V-va/-ve van* alakja, itt az MNSZ2 lemmatizálását követve egy tőhöz soroltuk az adott igét.

5.2. Eredmények

A globális produktivitás mutatóját itt is a 4.2. pontban bemutatott számítással mértük. A 3. táblázat az egyes sémák adatait, a 2. ábra ezek grafikus ábrázolását mutatja be.

3. táblázat. Globális produktivitás a beszélt nyelvi korpuszban, kézi elemzés
(A *-gal jelölt sor adatai automatikus elemzésből származnak)

szerkezet	N	V	n_1	P
V-ódik/-ődik*	106 582	2358	1094	0,0103
V-va/-ve van	22 132	3255	1618	0,0731
V-(t)atik/-(t)etik	16 940	99	61	0,0036
V-va/-ve lesz	3741	1286	707	0,1889



2. ábra. Globális produktivitás a beszélt nyelvi korpuszban

- A legmagasabb globális produktivitás a kézi vizsgálatban is a *V-va/-ve van* sémánál található.
- A legmagasabb potenciális produktivitása a *V-va/-ve lesz* konstrukciónak van, csakúgy, mint az automatikus elemzés eredményeinél.
- *V-(t)atik/-(t)etik* globális produktivitása a legalacsonyabb, ebben a vizsgálatban is mindkét tengelyen rendkívül alacsony (a többinél a típusgyakoriság tekintetében nagyságrendekkel alacsonyabb, a potenciális produktivitás tekintetében a *V-ódik/-ődik* sémát kivéve szintén nagyságrendekkel alacsonyabb) értékek kapcsolódnak ehhez a sémához.
- *V-ódik/-ődik* típusgyakorisága magasabb, mint a *V-va/-ve lesz* sémáé, de potenciális produktivitása jóval alacsonyabb; a hapaxok között gyakori a passzív jelentésű konstrukció (vö. 5.). A 4.1. részben említetteknek megfelelően a *V-ódik/-ődik* sémára vonatkozó eredményeket azonban fenntartásokkal kell kezelni az automatikus feldolgozás miatt.

További, elsősorban nem kvantitatív jellegű eredmények:

- a *V-va/-ve van* konstrukció nem csak tranzitív és a generatív szakirodalomban „unakkuzatívnek” nevezett (például *el van sápadva, ki van fáradva*) igei bemenettel fordul elő (vö. Alberti 1996, Kenesei 1998), hanem ún. „unergatív”²⁶ (például *salakoz/va van*) igékkel is (vö. Németh 2012 is), lásd a (17) példát.

(17) Amikor volt a nagy havazás, akkor nem tudtam jönni. Egyetlen alkalommal, mert nagy volt a hó. *És nem volt salakozva*, felolvadt, és nem fogtak a láncok se, úgyhogy nem tudtam jönni. (doc#2608)

- sőt nem csak telikus igékkel, hanem **atelikus** igékkel is, például

(18) éreztetve/sulykolva/motiválva/lebegtetve/népszerűsítve/foglalkoztatva van,

- valamint nem csak rezultatív/perfektív, hanem **imperfektív** jelentésű konstrukciókban is előfordul, lásd a (18)–(19) példákat.

(19) folyamatosan szorongatva vagyunk

²⁶ Az „unergatív” és „unakkuzatív” igék elhatárolási kérdéseire vö. Halm (2012).

Az imperfektív jelentés részben, de nem kizárólag, a **V-va/-ve van szórendű** variációhoz köthető. A V-va/-ve van szórendű konstrukciók a konstrukciós séma megvalósulásainak nem elhanyagolható részét: a tokengyakoriság mintegy 10%-át, a típusgyakoriság mintegy 15%-át teszik ki.

6. A hapaxok és a nagy gyakoriságú típusok kvalitatív vizsgálata

6.1. A hapaxok szerepe a (potenciális) produktivitásban

Térjünk vissza egy kicsit a választott elméleti keretnek és a vizsgálati módszernek a bevezetőben már taglalt kapcsolatához. Pusztán matematikailag nézve a potenciális produktivitási mutatóban a hapaxok száma a számlálóban található, azaz egyenes arányosságban áll a potenciális produktivitás fokával, míg a nevezőben álló tokengyakoriság fordított arányosságban áll a produktivitási mutatóval, tehát matematikailag csökkenti azt. Ez Baayen (1992) szerint összefüggésbe hozható a pszicholingvisztikai kutatások szófeldolgozáshoz kapcsolódó eredményeivel: a ritka komplex szavak feldolgozása lassabb, mivel feltehetőleg a szabály aktiválásával férünk hozzá a szó jelentéséhez, míg a gyakoribb komplex szavak feldolgozása gyorsabb, mivel ezek erős önálló lexikai reprezentációval rendelkeznek. Bybee (1985; 1988) modellje is hasonlóképpen erős lexikális autonómiát tulajdonít a gyakori, míg gyenge autonómiát a ritka komplex szavaknak, utóbbiak esetében azonban nem egy szabály működtetését, hanem az elmében tárolt hasonló hangzású és jelentésű példányok által alkotott séma aktivációját feltételezi.

A nagyobb fokú lexikális autonómiával rendelkező elemek esetében lexicizáció figyelhető meg, ami önálló reprezentáció létrejöttével és gyakran a séma/szabály alapján levezethető („kompozicionális”) jelentés megváltozásával jár. Baayen (1992: 140) szerint azonban vannak szabályos jelentésű nagy tokengyakoriságú elemek, és ezek erősítik a morfológiai szabály produktivitását. A kérdés a mi esetünkben az, hogy mi ez a „szabály”, mi a sémához kapcsolható jelentés, illetve mely jelentés(ek)ben produktív egy-egy vizsgált séma. Mind Baayen, mind Bybee érvelése alapján (vö. Kalivoda 2021 is) feltételezhető, hogy az alacsony gyakoriságú elemek, jelen esetben a hapaxok vizsgálatával információkat szerezhetünk a szerkezet produktivitását biztosító elemekről, valamint a sémához kapcsolódó, szabályosságként felfogható műveletről. Ettől a feltételezéstől eltér Kay (2013) nézőpontja, aki szerint a hapax jellegű adatok eltérnek a normától, ezért nem játszanak szerepet

a konstrukcionálódásban (Simon 2024). A jelen tanulmányban – Schultink (1961) Baayen–Lieber (1991) által alkalmazott definíciójának megfelelően – a nagy számban, nem szándékosan létrehozott új lexémák jelentőségét emeljük ki. Kognitív keretben a „szándékolatlanság” kritériumát leginkább alacsony metapragmatikai tudatossággal tudnánk jellemezni, melyet azonban a korpuszelemzés keretében nem tudunk elemezni.

A fenti kettősség rámutat a „hapax legomenon” terminus polisziemiájára: a filológiai és korpusznyelvészeti definíció jelentősen különbözik egymástól. Míg a filológiai értelemben vett hapax valóban csak egy-egy előfordulással bír akár egy adott nyelvben, nyelvváltozatban (sokszor holt nyelvekről, így lezárt korpuszokról lévén szó), a korpusznyelvészeti értelemben vett hapax egy korpuszban fordul elő egyszer, ahol a korpusz a(z élő) nyelv alapsokaságát hivatott reprezentálni.²⁷ Természetesen lehetséges, hogy a korpuszban előforduló hapax valóban egy egyszeri előfordulás még a nyelv alapsokaságát tekintve is (azaz csak egyvalaki csak egyszer, az adott mintában használta az adott kifejezést), de ennek – a korpusz nagyságától függően – elég csekély a valószínűsége.

A korpusz tehát csak modellál, az, hogy benne egy konstrukció hapax legomenon vagy más igen alacsony gyakoriságú előfordulás, sokszor csak véletlen kérdése. Baayen (2009: 912) mindenesetre megállapítja, hogy a hapaxok esetén ritka szavakról van szó, amelyek lexikális reprezentációja gyenge, ezért a legvalószínűbb, hogy e szavak megértése és produkciója során jobban támaszkodunk a szabályalapú feldolgozásra, míg a magas gyakoriságú szavak esetében erős a lexikális reprezentáció, ezeknél csökken a szabályalapú feldolgozás jelentősége. A használatalapú elméleti keretben természetesen inkább konstrukciós sémáról, mintsem szabályról beszélhetünk, Bybee elméletében „a morfológiai szabályok [...] nem szimbolikus szabályként, hanem lexikai kapcsolatokból kirajzolódó sémákként értendők” (Tanos 2008). Amikor „szabályalapú” feldolgozásról beszélünk, elsősorban a szabályszerű (sémát aktiváló) feldolgozásra helyezzük a hangsúlyt az autonóm reprezentáció aktiválásával szemben.

A szabályosság tekintetében szerepet játszik még a bázisnak és a derivátumnak az egymáshoz viszonyított gyakorisága (lásd Baayen 2009): minél nagyobb a bázis gyakorisága a derivátuméhoz képest, annál fontosabb

²⁷ Zeldes (2012: 60, 39. l.) szerint a hapax legomenon kifejezést a klasszikus görög művekkel kapcsolatban a szokatlan, homályos jelentésű, sokszor archaikus szavakra vonatkoztatva használták, míg a modern korpuszokban a legtöbb neologizmus hapax lehet.

szerepet játszanak a szabályok, és fordítva: minél gyakoribb a derivátum a bázishoz képest, annál nagyobb az esélye az önálló reprezentáció létrejöttének. Ezt a hatást kollostrukciós vizsgálattal lehetne jól megfigyelni (vö. 8. pont).

6.2. A hapaxok kézi elemzése igenem szerint

Az alábbiakban egy, a gépi elemzés (lásd 4.) során nyert hapaxokból véletlenszerűen kiválasztott, 4x200 elemszámú (minden sémához 200 elemet tartalmazó) minta elemzését mutatjuk be. A minta elemzését kézzel végeztük Microsoft Excel táblázatokban, a kontextus figyelembevételével. Az elemzés során kizártuk a helyesírási vagy annotációs anomália miatt keletkezett hibás találatokat (ez a V-(t)atik/-(t)etik séma esetében kiugróan magas volt), például elgépelés miatt lett hapaxként besorolva a (20) mondatban a *kisodródik* („kisodpródott”) szó, és hibás annotálás miatt tekintette a program a V-va/-ve lesz konstrukcióhoz tartozónak a *pihenve* – *lesz* kifejezést (20):

- (20) A mexikói *kidodpródott* a Sauberrel, de vissza tudott térni a pályára és a boksza. (doc#2333)
- (21) Mígnem a szülei házában *pihenve* figyelmes *lett* egy érdekes hírre. (doc#2338)

A találatok kategorizálásánál a következő csoportosítást alkalmaztuk: főként a V-ódik/-ődik sémánál jellemzően mediopasszívnak tekintettük a nem szándékos ágensi, de ráhatott jegyekkel rendelkező használatokat (például *kirotálódhatott*, *elfuserálódik*), passzívnak az ágens által végrehajtott, átperspektivált folyamatokat (például *idevarázsolódik*, [*kilincs*] *lenyitódik*), mediálisnak az alany szándékától független állapotot, folyamatot (például *stabilizálódik*, *elburjánzódik*) és aktívnek a saját erőforrásból végrehajtott cselekvéseket (például *hátraérdeklődik*, *kászálódik*). Az alábbi táblázatban a szortírozás utáni abszolút elemszámok, valamint százalékos értékek láthatók (4. táblázat).

4. táblázat. Hapaxok konstrukciós igeneme az egyes képzési sémákban, abszolút elemszám és százalékos megoszlás

	V-(t)atik/-(t)etik		V-va/-ve lesz		V-va/-ve van		V-odik/-ődik	
passzív	55	95%	128	88%	96	75%	43	35%
medio-passzív	0	0%	0	0%	1	1%	17	14%
mediális	3	5%	18	12%	31	24%	58	47%
aktív	0	0%	0	0%	0	0%	6	5%
összes elemszám	58	100%	146	100%	128	100%	124	100%

A legnagyobb százalékban passzív jelentésben a V-(t)atik/-(t)etik séma fordult elő (az összes előfordulás 95%-a passzív jelentésű); ezzel a sémával mediális jelentésű hapaxot csak a *születik* és *méltóztatik* ige összetételeiben, valamint egy nyilvánvalóan archaikus előfordulásban (*megszabadultatik*) találunk. A passzív jelentésű hapaxok százalékos arányát tekintve második helyen a V-va/-ve lesz konstrukció áll (88% passzív jelentésű használattal). A V-va/-ve van séma esetében már a hapaxok mintegy negyedénél mediális vagy mediopasszív használatot regisztráltunk, háromnegyede passzív jelentésű. A V-odik/-ődik séma eredményeiben a passzív és mediopasszív jelentésű elemek számának összege meghaladja a mediális elemek számát. A kisszámú aktív megvalósulás szándékos, önerőből történő cselekvésre utal (például *bekászálódik*).

6.3. A magas gyakoriságú elemek

A 2. részben említésre került, hogy feltehetőleg nemcsak az alacsony, de a magas gyakoriságú elemek sémajelentésének is kitüntetett szerepe van a sémák produktivitásában. Bár ezen elemeket inkább egységként tároljuk és dolgozzuk fel, Baayen (2009) megállapítása szerint a magas gyakoriságú, rendszeres képzésű (transzparens) elemek jelenléte növeli az adott konstrukciós séma produktivitását.

Jelen tanulmányban a magas gyakoriságú elemek beható vizsgálatára nem került sor, ennek egyik oka az, hogy a konstrukciós jelentések igen erősen kontextusfüggőek. A gépi vizsgálat során nyert, konstrukciós sémánként tíztíz leggyakoribb lemma alapján a következőket lehet megállapítani:

A V-(t)atik/-(t)etik sémában használt 10 leggyakoribb lemmából 7 nem passzív, hanem mediális jelentésű (például *születik*, *viseltetik* 'valamilyen

érzülettel van', *méltóztatik*), ez éles kontrasztot mutat a hapaxok által ki-rajzolt rendszeres sémajelentéssel, mely 95%-ban passzív. Ez a jelenség arra utal, hogy a Baayen által említett tendencia – miszerint a magas gyakoriságú, szabályos képzésű elemek erősítik a séma produktivitását (vö. Baayen 1992: 140, lásd 2.) – „megfordítása” is igaz lehet, tehát látható, hogy a leggyakoribb elemek nem transzparenssek, és ezzel párhuzamosan megállapítható a konstrukció alacsony globális produktivitása.

A V-*odik/-odik* séma esetében többnyire cselekvő igéből képzett, de már jelentésváltozáson átment konstrukciókat figyelhetünk meg, például *folytatódik*, *megkezdődik*, *befejeződik*; már nem transzparens a 9. és 10. leggyakoribb lemma, a *fejlődik* és az *aggódik*.

A V-*va/-ve van* sémával használt 10 leggyakoribb lemma között vannak mediális jelentésűek is, például *meg van győződve*, *fel van készülve*. Ahogy ezt Modrián-Horváth (2024) is kifejti, például a *meg van győződve* konstrukció esetében is megfigyelhető a *meggyőződik* (*valamiről*) ige jelentésétől való nagyfokú eltérés. Kontextusfüggő igenemet tételezhetünk fel az *el van foglalva* konstrukciónál (például a *vár el van foglalva* passzív jelentésű, míg a *valaki el van foglalva* külső hatásra létrejövő belső erődinamikájú, ezért inkább mediális jellegű „elfoglaltságra” utal, vö. *elfoglalja magát*), és 5 passzív jelentésű konstrukció is található a leggyakoribb lexémacsoportban. Mivel e csoportban a hapaxok vizsgálata is jelentős arányban (25%) mutatott mediális konstrukciós jelentést, feltételezhető, hogy a V-*va/-ve van* séma kevésbé specifikus a passzív jelentésre nézve, mint a V-*(t)atik/-(t)etik*, illetve a V-*va/-ve lesz* konstrukciók.

A V-*va/-ve lesz* sémával létrehozott leggyakoribb konstrukciók között kizárólag tranzitív kiinduló igéből képzett konstrukciókat találunk, melyek többnyire passzív jelentésűek (például *biztosítva lesz*, *ki lesz alakítva*); a *meg lesz lepve* kontextusfüggően mediális jelentésű is lehet. Látható, hogy a konstrukció passzív jelentése, amely a hapaxok vizsgálatánál is a második legerősebbnek bizonyult (88%) itt a nagy gyakoriságú elemek passzív jelentésével társul, amely összességében a legerősebb passzív jelentésű produktív sémát adja ki.

7. Az eredmények diszkussziója

7.1. Az egyes sémák globális produktivitása

Az eredmények értelmezése során némileg problémássá válik az, hogy a globális produktivitás egy kétdimenziós, „vizuális természetű” mérési módszer, amely a kapott számok természetén alapján (a potenciális produktivitás 0 és 1 közötti szám, míg a realizált gyakoriság sokszor többes nagyságrendű) nem redukálható egyetlen értékre. Mi alapján lehet eldönteni, hogy két konstrukció közül melyik a produktívabb? Baayen (1993: 190) arra a megállapításra jut, hogy a magasabb típusgyakoriság előnyt élvez, mert ez egy elsődleges, közvetlenül a korpuszadatokból származó érték, míg a potenciális produktivitás egy származtatott érték, amelyet a tokengyakoriságok továbbelemzése útján kapunk. Tehát ha mindenáron egy globálisan „legproduktívabb konstrukciót” akarunk megnevezni, a *V-va/-ve van* szerkezetet van okunk annak tekinteni. Azonban a potenciális produktivitási érték az, amelynek inkább van prediktív értéke, mivel a magas típus-token arány önmagában akár egy korábban produktív séma „lenyomata” is lehet a korpuszban (vö. Stefanowitsch 2020), és ugyanezt feltételezhetjük a magas típusgyakorisággal kapcsolatban is. A potenciális produktivitás szempontjából megfigyelt értékek és arányok azt mutatják, hogy a *V-ódi/-ődi* és a *V-(t)atik/-(t)etik* konstrukciók igen alacsony potenciális produktivitással rendelkeznek; ezen értékek nem egészen kétszerese, de így is 0,01 alatti a *V-ni fog* potenciális produktivitási mutatója. A kontrollkonstrukcióhoz képest majdnem háromszor produktívabb a *V-va/-ve van*, és ennél is majdnem két és félszer produktívabb a *V-va/-ve lesz* a potenciális produktivitás szempontjából az egész korpuszra nézve. A *V-ódi/-ődi* és a *V-ni fog* értékei talán meglepőek lehetnek, tekintettel arra, hogy az előbbi a *V-(t)atik/-(t)etik* egyre inkább terjedő alternatívájának tekinti a szakirodalom egy része, illetve a *V-ni fog* szerkezettel kapcsolatban inflexiós jellegénél fogva a passzív konstrukciónál magasabb globális produktivitást várhatnánk. A potenciális produktivitás az alkalmazott képlet alapján fordítottan arányos a tokengyakorisággal, és minél nagyobb a vizsgált korpusz, annál alacsonyabb értéket mutat (az összefüggést problematizálja Gaeta 2007 is). Nem zárhatjuk ki a lehetőségét annak, hogy ezt a tokenszenzitivitást figyelhetjük meg az egyes konstrukciók értékét illetően is, ezt cáfolni látszik azonban az a tény, hogy a *V-(t)atik/-(t)etik* és a *V-va/-ve van* tokengyakorisága között nincs nagyságrendi, sem többszörös különbség (mintegy 15%-os az eltérés),

a potenciális produktivitásban mégis drámai eltérés mutatkozik. A *V-ódik/-ődik* séma mediális, mediopasszív és passzív jelentéseket fog át, melyek nem különülnek el egymástól élesen. Magas típusgyakorisága összefüggésbe hozható e széles jelentésbeli spektrummal, valamint egy korábbi magasabb fokú produktivitással. A potenciális produktivitást illetően a hapaxok vizsgálata alapján valószínűsíthető, hogy a konstrukciós jelentésnek csak bizonyos (passzív és mediopasszív) szegmensei mutatnak nagyobb fokú produktivitást, míg a kevésbé produktív, de igen magas példánygyakoriságú mediális konstrukciók csökkentik a potenciális produktivitási mutatót.

Az egész korpuszra vetítve tehát a legmagasabb globális produktivitás a *V-va/-ve van* sémánál található, a konstrukcióval kapcsolatos közismert nyelvi babonák ellenére. A *V-va/-ve lesz* típusgyakorisága ezzel szemben az egész korpuszban alacsonyabb, melynek számos oka lehet, például a személytelen T/3 szerkezet (lásd F. Gulyás 2016) használata a dinamikus passzív helyett, a rugalmas szórend kiaknázása a páciens előtérbe helyezésére dinamikus események konstruálásánál. A *V-va/-ve lesz* séma megvalósulásainak egy része szemantikai kapcsolódások révén részben a *V-va/-ve van* konstrukcióhoz tartozhat (a *V-va/-ve lesz* konstrukcióból vett hapaxmin-tában mintegy 53% volt a biztosan a *V-va/-ve lesz*-hez tartozó, múlt idejű, kijelentő módú konstrukciók aránya). Ha a két konstrukciót egyesítenénk, a *V-va/-ve van* magas példánygyakorisága miatt a *V-va/-ve van* értékeihez közeli, kissé magasabb potenciális produktivitású és alacsonyabb típusgyakoriságú értékeket kapnánk. Az informális regisztert képviselő személyes alkorpuszban azonban a *V-va/-ve lesz* konstrukciónak nemcsak a potenciális, de a típusgyakorisága is kiugróan magas (a típusgyakoriság meghaladja a *V-ódik/-ődik* sémáét). Bár az MNSZ2 alkorpuszainak elnevezése azt engedné sejtetni, hogy a beszélt nyelvi alkorpusz szövegei állnak a spontán mindennapi társalgás nyelvhasználatához a legközelebb, ez az alkorpusz (ahogy az 5. fejezetben említettük) rádiós beszélgetések leiratát tartalmazza, melyek csak részben spontán megnyilatkozások. Ezért joggal feltételezhetjük, hogy a személyes alkorpusz adatai az ún. beszélt nyelvihez hasonlóan, sőt esetleg jobban tükrözik a hétköznapi nyelvhasználatot, tehát a *V-va/-ve lesz* sémának kiemelt szerepe van a spontán, informális nyelvhasználatban a (dinamikus) passzív eseménykonstruálásban.

A *V-(t)atik/-(t)etik* séma produktivitásának megítélése a szakirodalomban nem volt egységes, ennek oka abban keresendő, hogy az említett grammatikák más-más produktivitási fogalommal dolgoznak, így a potenciális szavak vagy a filológiai értelemben vett hapaxok létrehozásának lehetősége

a kompetencianyelvtanok számára bizonyítékként szolgálhat egy adott képző vagy séma produktívítására. A *V-(t)atik/- (t)etik* produktívításának kvantitatív jellegű mérése során megerősítést nyert, hogy a képzési séma produktívítása igen korlátozott, ez egybevág a grammatikák többségében megfogalmazott feltételezésekkel (vö. Kenesei et al. 1998, Ladányi 2017) és a tanulmány szerzőinek intuíciójával is. E ponton kiemelnénk, hogy a nyelvi intuíciónak való megfelelés Baayen–Lieber (1991) megközelítésében is fontos tényező. Ugyanakkor meg kell említeni azt a tényt is, hogy a *V-(t)atik/- (t)etik* séma alacsony típusgyakorisága és potenciális produktívítása ellenére közepes tokengyakorisággal bír. A tokengyakoriság hatása akár erősíthetné is a produktivitást, amennyiben a séma elsődleges, passzív jelentésmintázatára illeszkedne (melyet a hapaxok vizsgálata is kimutatott), azonban a legnagyobb példányszámú típusok (például *születik*) jellemzően nem passzív jelentésűek.

A produktívítás kutatásában jelentős tényezőként merül fel a nyelv szociokulturális komplexitásából adódó, regiszterek közti különbség vizsgálata, melyet az MNSZ2, illetve az MNSZ2-UD egyes alkorpuszai reprezentáltak. Megfigyelhettük, hogy a *V-va/-ve lesz* konstrukció produktívítása a teljes korpuszon mért értékeknél jóval alacsonyabb a szépirodalmi, és sokkal magasabb a személyes alkorpuszban. A tudományos alkorpuszban a *V-va/-ve van* konstrukció bizonyult feltűnően produktívnak mindkét dimenzió mentén. Az összehasonlítási alapként használt *V-ni fog* konstrukció a hivatalos, a sajtónyelvi, valamint a (rádiós szövegeket tartalmazó) ún. beszélt nyelvi alkorpuszokban kimagasló típusgyakorisággal és a beszélt nyelvi alkorpuszban kimagasló tokengyakorisággal fordult elő, míg a teljes korpusznál alacsonyabb volt a típusgyakoriság a tudományos alkorpuszban.

7.2. Szintetikus és analitikus konstrukciók globális produktívítása

A vizsgálatok megkezdése előtt felmerült az a kérdés, hogy összehasonlítható-e a szintetikus és analitikus konstrukciók produktívítási indexe. Bár a szakirodalomban nem ismeretlen az eredetileg szintetikus kifejezőeszközök produktívításának mérésére kifejlesztett globális produktívítási index szintaktikai jellegű konstrukciókra való használata, a két konstrukciótípus mérőszámainak egymással való összevetése mégis metodológiai kihívást, újítást jelentett. Vizsgálatunkba ezért bevontunk összehasonlítási alapként egy nem passzív jelentésű, grammatikalizálódott, igeidőjelölő analitikus konstrukciót, a *V-ni fog* konstrukciót. Az eredmények azt mutatták, hogy a *V-ni fog* típusgyakorisága kissé magasabb a *V-va/-ve van* konstrukciónál

(a *V-ódik/-ődik* és a *V-va/-ve van* között van), potenciális produktivitása viszont jóval alacsonyabb mindkét vizsgált passzív analitikus sémáénál. Eből arra lehet következtetni, hogy a konstrukció szintetikus vagy analitikus jellege nem áll közvetlen összefüggésben sem a realizált, sem a potenciális produktivitás mértékével, ami megnyugtató a jelen vizsgálat metodológiájával kapcsolatban. A *V-ni fog* séma értékeivel való összehasonlítás eredményei újabb kérdéseket vetnek fel a *V-va/-ve van* konstrukció státuszára nézve: ez a konstrukció is az igei paradigma részét képezi, a morfológiai leírásokban van a helye?

7.3. Kézi és gépi elemzések eredményei

A 4. fejezetben gépi elemzéssel nyert, a beszélt nyelvi alkorpuszra vonatkozó adatokat összevetettük az 5. fejezetben található kézi elemzéssel nyert adatokkal. Ennek során azt találtuk, hogy a sémák egymáshoz viszonyított produktivitása nem változott, ellenben az abszolút értékekben jelentős különbségek is mutatkoztak. A kézi elemzésnél problémát jelentett, hogy a kiemelkedően magas tokengyakoriságú *V-ódik/-ődik* sémánál a jelen projekt keretében kivitelezhetetlen volt az adatok kézi feldolgozása, így az e sémára vonatkozó adatok kizárólag gépi elemzésen alapulnak. – Az összehasonlítás eredménye a teljes vizsgálat validitására nézve biztató az egyes konstrukciók egymáshoz viszonyított produktivitásával kapcsolatban.

A nagy adatmennyiség miatt a teljes korpuszra, illetve a hat alkorpuszból ötre nézve kizárólag a gépi elemzéssel lehetett adatokat nyerni. Mindamellett néhány jelenség vizsgálata kézi elemzést igényelt, ilyen (az eredmények validálásán túl) a hapaxok jelentéseinek vizsgálata. Ezenkívül a kézi korpuszvizsgálat során derült fény arra, hogy a *V-va/-ve van* szórendű konstrukciók (például *lebegtetve van*), melyek egy részét az automatikus elemzésnél (4. fejezet) a magas hamis pozitív találati arány miatt ki kellett zárni a vizsgálatból, a konstrukciós séma jelentős részét: a tokengyakoriság mintegy 10%-át, a típusgyakoriság mintegy 15%-át teszik ki (lásd 5.2). A *V-va/-ve van* szórendű konstrukciók feltehetőleg külön típust képviselnek aspektuális tekintetben a *V-va/-ve van* konstrukciókon belül, erre szintén a kézi elemzés során lettünk figyelmesek. A manuális korrekció során ismertük fel azt is, hogy az igekötős *V-va/-ve van/lesz* szórendű találatok szinte kivétel nélkül hamis pozitívak. Mindez alátámasztja annak fontosságát, hogy ha a vizsgálat nem is folytatható le egészében kézi vizsgálattal, mégis fontos az adatok beható ismerete.

7.4. Hapaxok és nagy gyakoriságú elemek vizsgálata

A hapaxok és nagy gyakoriságú elemek sémajelentésének vizsgálata hasznosnak bizonyult annak megállapításánál, hogy mely jelentésekben a legproduktívabbak a vizsgált sémák. A hapaxokból nyert minta, valamint a leggyakoribb igelemmák vizsgálata alapján megállapítható, hogy a *V-(t)atik/- (t)etik* produktivitása (mely igen alacsony) szinte kizárólag a passzív jelentésű elemekhez köthető, míg a legnagyobb példányszámú lexémák mediális jelentésűek. A másik legstabilabban passzív jelentésű séma a *V-va/-ve lesz* (a mintában 88%-ban passzív jelentésű hapaxokkal és többnyire passzív jelentéssel a leggyakoribb lemmák között), ezt követi a *V-va/-ve van* (itt 75%-ban volt passzív a hapaxok jelentése, és a leggyakoribb lemmák között is több mediális bemeneti igét találtunk). A *V-va/-ve van* és *V-va/-ve lesz* jelentős különbségeket mutat passzív/mediális megoszlás szempontjából, miközben tudvalevő, hogy a *lesz* tövű szavak gyakran a *van* szuppletív alakjaiként szerepelnek (*lenne, lesz, lett volna, legyen* stb.).

A 2. fejezetben írottak alapján a *V-va/-ve van*, illetve *V-va/-ve lesz* konstrukciók egyértelmű elkülönítése nehézségekbe ütközik mind kézi, mind gépi vizsgálat során: az utóbbi esetben a nagy adatmennyiség is közrejátszik abban, hogy csak nagyon nehezen, vagy egyáltalán nem kivitelezhető a tövek elkülönítése, a kézi elemzésnél pedig a kontextus alapján számos esetben nem egyértelmű, hogy a *van* vagy a *lesz* ige paradigmájához van közelebb az aktuális használat. Az egyetlen egyértelmű esetet a *V-va/-ve lesz* kijelentő módú, múlt idejű alakjai képviselik. A hapaxok kézi vizsgálata alapján ezek aránya mintegy 53% a mintában, azaz a *lesz* lemmához sorolt előfordulások több mint feléről van szó (76 eset a 146-ból; a múlt idejű használatok aránya hasonló a *V-va/-ve volt* vonatkozásában). Figyelemre méltó módon ezek közül mindössze egy tartozott a mediális jelentésű konstrukciók közé (*fel lettünk világosodva*), az összes többi *V-va/-ve lett* múlt idejű, kijelentő módú paradigmatis alak passzív jelentéssel bírt. Ez az eredmény abba az irányba mutat, hogy valóban rendszeres különbség lehet a *V-va/-ve van* és *V-va/-ve lesz* konstrukciók sémajelentései között.

A *V-ódik/-ődik* séma, melyben sokan az új szintetikus passzívképző lehetőségét látják, az esetek mintegy 35%-ában mutatott passzív jelentést a hapaxok között, ehhez járul még mintegy 14%-nyi mediopasszív (ráhatott) használat. A produktivitási tendenciák megállapításánál érdemes lehet figyelembe venni, hogy a mediopasszívnak tekintett (ráhatott) előfordulásokat más nyelvekben tipikusan passzív alakokkal fejezik ki.

8. Összegzés és kitekintés

A tanulmány célja négy passzív jelentést mutató konstrukció korpuszalapú vizsgálata volt. A globális produktivitás kvantitatív vizsgálatának eredménye leginkább a hapaxok és a magas gyakoriságú elemek vizsgálatával összevetve értékelhető a vizsgálni kívánt jelenségre nézve. Az elemzés során azt lehetett megállapítani, hogy a legmagasabb globális produktivitás a *V-va/-ve van* sémánál található, mely a hapaxok vizsgálata alapján túlnyomórészt (a minta 75%-ában) passzív jelentéstartalmat fejez ki. A személyes alkorpuszban, amely feltehetőleg jól tükrözi a spontán hétköznapi nyelvhasználatot, a *V-va/-ve lesz* globális produktivitása is kiugróan magas, e séma jelentése a hapaxok mintegy 88%-ában passzív, ezért arra következtethetünk, hogy a passzív jelentéstartalom nem modális, dinamikus predikatív megkonstruálására elsősorban a *V-va/-ve lesz* séma alkalmas produktívan. Az egyértelműen ehhez a sémához tartozó *V-va/-ve lett* múlt idejű, kijelentő módú paradigmatis alakoknál csaknem kivétel nélkül passzív sémajelentést találunk, ami erősíti ezt a feltételezést. A vizsgálatokban legalacsonyabb globális produktivitása – a nyelvi intuíciónak megfelelően – a *V-(t)atik/-(t)etik* szinte kizárólag passzív jelentésű sémának volt. A *V-odik/-odik* séma nagy típusgyakorisággal, de igen alacsony potenciális produktivitással rendelkezik, és a hapaxok kvalitatív vizsgálata alapján a séma jelentése megosztott: csaknem 50%-ban mediális, 35%-ban passzív, és mintegy 15%-ban mediopasszív. A magas gyakoriságú elemek vizsgálata a hapaxokon a fent említettekkel egybevágó eredményeket mutatott a sémák jelentésére, illetve a produktív sémajelentésekre vonatkozóan.

A *V-va/-ve van* séma vizsgálata során az eddigi leírásokban feltételezettnél gazdagabb spektrumú igei bázisokra lettünk figyelmesek: kiinduló igeként atelikus igeik realizációi is megjelennek, gyakran egy speciális szórendi mintázattal (*motiválva/lebegtetve van*). A *V-va/-ve van* és *V-va/-ve lesz* sémáknál egyaránt megfigyelhettük a személytelen passzív két tulajdonságát (vö. F. Gulyás 2016: 119kk.): az alanytalan passzívumot és az intranszitiv igekből képzett konstrukciókat.

A globális produktivitásra vonatkozó gépi és kézi elemzések összevetése hasonló eredményeket mutatott az egyes sémák egymáshoz képesti produktivitását tekintve, azonban a számszerű adatokra nézve eltérések mutatkoztak. A kutatás kvalitatív elemzésből nyert eredményei is aláhúzzák az adatok beható ismeretének fontosságát. A hapaxok manuális vizsgálatával értékes eredményeket nyertünk az egyes sémák megvalósulásainak igenemével

kapcsolatban, így fény derült például arra, hogy a *lesz* ige múlt idejű tövével alkotott *V-va/-ve lesz* konstrukciók gyakorlatilag mindegyike passzív jelentésű, illetve összefüggéseket találtunk a *V-va/-ve van* konstrukció szérendi megvalósulása és aspektuális tartalma között. Egy másik metodológiai kérdés volt a szintetikus és analitikus konstrukciók globális produktivitási indexének összehasonlíthatósága. Itt az összehasonlítási alapként vizsgált *V-ni fog* elemzése alapján (magas típuszám, alacsony potenciális produktivitás) úgy tűnik, nincs összefüggés a szintetikus/analitikus jelleggel.

A vizsgálat további kérdéseket, lehetőségeket nyit a jövőbeni kutatások számára. Ilyen kérdés például a *van* és *lesz*, *V-va/-ve van* és *V-va/-ve lesz* paradigmák szemantikai kapcsolatainak vizsgálata, az utóbbi analitikus konstrukciókat a jelen tanulmányban metodológiai célból különválasztottuk. Elméleti jelentőségű az a megállapítás, hogy a *V-va/-ve van* konstrukció potenciális produktivitása jóval magasabb, a típusgyakorisága közel van a kontrollként vizsgált *V-ni fog* összetett igeidő-konstrukció értékeihez, ez ugyanis felveti azt a kérdést, vajon a *V-va/-ve van* konstrukció is analitikus (összetett, segédigés) igealakként kezelendő, az igei paradigma részét képezi-e. Ennek megítélésénél azt is érdemes tekintetbe venni, hogy egy jelölt igenem- és aspektuális komponensekkel bíró konstrukcióról van szó, amelyvel kapcsolatban nem várnánk, hogy produktivitása felveszi a versenyt egy igeidő-jelölő konstrukcióéval, vagy akár felül is múlja azt.

Nyitott kérdésként kezelhető a passzív és mediális elhatárolásának problémája, amelynél a jelenleg használt kritériumok finomítására lehet szükség, tekintettel a ráhatottságot kifejező mediopasszív használatokra. Itt elsősorban az ágentív passzív metaforikus és metonimikus kiterjesztése által létrejövő poliszémia behatóbb vizsgálatára lenne szükség az olyan, nem nyíltan hozzáférhető folyamatok konceptualizációjában, amelyekben a folyamat kivitelezője nem állapítható meg (például *letöltődik*, *át lesz irányítva*, *elsáncolódik*). Végül, de nem utolsó sorban, az analitikus konstrukciókkal végzendő kollostrukciós vizsgálat alkalmas lehetne a konstrukció és a benne megjelenő lemmák közötti affinitások kimutatására.

Irodalom

- Alberti Gábor 1992. Passzivizálási művelet a magyarban. *Néprajz és Nyelvtudomány* 37: 7–45.
- Aronoff, Mark 1976. *Word formation in Generative Grammar*. Cambridge, MA: MIT Press.

- Baayen, Harald 1992. Quantitative aspects of morphological productivity. In: Booij, Geert Evert et al. (eds.): *Yearbook of Morphology 1991*. 109–149. https://doi.org/10.1007/978-94-011-2516-1_8
- Baayen, Harald R. 1993. On frequency, transparency and productivity. In: Booij, Geert – van Marle, Jaap (eds.): *Yearbook of Morphology 1992*. Dordrecht: Kluwer Academic. 181–208. https://doi.org/10.1007/978-94-017-3710-4_7
- Baayen, Harald 2009. Corpus linguistics in morphology: Morphological productivity. In: Lüdeling, Anke – Kytö, Merja (eds.): *Corpus Linguistics. An international handbook. Vol. 2.* (HSK 29.2). Berlin, New York: Mouton de Gruyter. 899–919. <https://doi.org/10.1515/9783110213881.2.899>
- Baayen, Harald – Lieber, Rochelle 1991. Productivity and English derivation: A corpus based study. *Linguistics* 29: 801–843. <https://doi.org/10.1515/ling.1991.29.5.801>
- Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: Benjamins. <https://doi.org/10.1075/tsl.9>
- Bybee, Joan L. 1988. Morphology as lexical organization. In: Hammond, Michael – Noonan, Michael (eds.): *Theoretical morphology: Approaches in modern linguistics*. London: Academic Press. 119–141. https://doi.org/10.1163/9789004454101_010
- Bybee, Joan 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511750526>
- Croft, William 1991. *Syntactic categories and grammatical relations: The cognitive organization of information*. Chicago: University of Chicago Press.
- Croft, William 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198299554.001.0001>
- de Marneffe, Marie-Catherine – Manning, Christopher D. – Nivre, Joakim – Zeman, Daniel 2021. Universal Dependencies. *Computational Linguistics* 47 (2): 255–308. https://doi.org/10.1162/coli_a_00402
- Diessel, Holger 2014. Usage-based linguistics. In: Mark Aronoff (ed.): *Oxford Research Encyclopedia of Linguistics*. New York: Oxford University Press. <https://doi.org/10.1093/obo/9780199772810-0068>
- Dressler, Wolfgang – Ladányi, Mária 2000. Productivity in word formation (WF): A morphological approach. *Acta Linguistica Hungarica* 47 (1): 103–44. <https://doi.org/10.1023/A:1014010530824>
- E. Abaffy Erzsébet 1978. A mediális igéről. *Magyar Nyelv* 74: 280–293.

- F. Gulyás Nikolett 2016. *Személytelen szerkezetek finnugor nyelvekben funkcionális és tipológiai megközelítés*. Budapest: ELTE.
- Gaeta, Livio 2007. On the double nature of productivity in inflectional morphology. *Morphology* 17 (2): 181–205. <https://doi.org/10.1007/s11525-007-9117-7>
- Goldberg, Adele 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199268511.001.0001>
- Grétsy László 2014. Megivódik. *Szabad Föld* [2014. 11. 14.] <https://szabadfold.hu/orszag-vilag/2014/11/gretsy-laszlo-megivodik>
- Halm Tamás 2012. Unergatív és/vagy unakkuzatív: a szemelfaktív igék argumentum-szerkezetéről, szemantikájáról és szintaxisáról a magyarban. In: Gécseg, Zsuzsanna (szerk.): *LingDok 11.: Nyelvészdoktoranduszok dolgozatai*. Szeged: JATEPress. 73–88.
- Hilpert, Martin 2013. *Constructional change in English. Developments in allomorphy, word formation, and syntax*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139004206>
- Honti László – H. Varga Márta 2013. *Meg van írva! A határozói igenév és a létige alkotta szerkezet funkciójáról és háttéréről*. In: Honti László (szerk.): *Magyar nyelvtörténeti tanulmányok*. Budapest: Károli Gáspár Református Egyetem, L'Harmattan. 255–266.
- Imrényi András 2017. Az elemi mondat viszonyhálózata. In: Tolcsvai Nagy Gábor (szerk.): *Nyelvtan*. Budapest: Osiris Kiadó. 664–760.
- Janda, Laura 1990. The radial network of a grammatical category – its genesis and dynamic structure. In *Cognitive Linguistics* 1 (3): 269–288. <https://doi.org/10.1515/cogl.1990.1.3.269>
- Janda, Laura A. 2013. Quantitative methods in Cognitive Linguistics: An introduction. In: Janda, Laura A. (ed.): *Cognitive Linguistics: The quantitative turn – The essential reader*. Berlin, Boston: Walter de Gruyter. 1–32. <https://doi.org/10.1515/9783110335255.1>
- Kalivoda Ágnes 2019. Véges erőforrás végtelen sok igekötős ígére. In: Berend Gábor – Gosztolya Gábor – Vincze Veronika (szerk.): *XV. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2019)*. Szeged: Szegedi Tudományegyetem, TTIK, Informatikai Intézet. 331–344.
- Kalivoda Ágnes 2021. Igekötös szerkezetek a magyarban. Doktori (PhD) értekezés. Budapest: PPKÉ BTK.

- Kalivoda, Ágnes – Malouf, Robert – Ackerman, Farrell – Sass, Bálint 2024. *Building a dependency treebank from the Hungarian Gigaword Corpus*. Kézirat, megjelenés előtt.
- Kay, Paul 2013. The limits of (Construction) Grammar. In: Hoffmann, Thomas – Trousdale, Graeme (eds.): *The Oxford handbook of Construction Grammar*. New York: Oxford University Press. 32–48. <https://doi.org/10.1093/oxfordhb/9780195396683.013.0003>
- Kádár Edit – Németh Boglárka 2009. Predikatív határozói igeneves szerkezetek csángó beszélt nyelvi szövegekben. In: É. Kiss Katalin – Hegedűs Attila (szerk.): *Nyelvelmélet és dialektológia*. Piliscsaba: Pázmány Péter Katolikus Egyetem. 189–212.
- Kemmer, Suzanne 1993. *The middle voice*. Amsterdam, Philadelphia: John Benjamins. <https://doi.org/10.1075/tsl.23>
- Keszler Borbála (szerk.) 2000. *Magyar grammatika*. Budapest: Akadémiai kiadó.
- Komlósi András 1992. Régensek és vonzatok. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 1. Mondattan*. Budapest: Akadémiai Kiadó. 299–528.
- Kugler, Nóra – Simon, Gábor 2018. The schematization of Hungarian participle-noun compounds. *SKY Journal of Linguistics* 31: 35–69.
- Laczko Krisztina 2022. Adalékok a mediális igés kompozitumszerkezetekhez. In: Tátrai Szilárd – Tolcsvai Nagy Gábor (szerk.) *A magyar mondat és kontextuális környezete*. Budapest: ELTE BTK. 76–106.
- Ladányi Mária 2008. Produktivitás és analógia a szóképzésben: a magyar főnévképzés néhány esete. Általános Nyelvészeti Tanulmányok XXII. Tanulmányok a funkcionális nyelvészet köréből. 349–430.
- Ladányi Mária – Tolcsvai Nagy Gábor 2008. Funkcionális nyelvészet. *Általános Nyelvészeti Tanulmányok XXII. Tanulmányok a funkcionális nyelvészet köréből*. 17–58.
- Lakoff, George 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago, London: The University of Chicago Press. <https://doi.org/10.7208/chicago/9780226471013.001.0001>
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. I. Theoretical prerequisites*. Stanford California: Stanford University Press.
- Langacker, Ronald W. 2006: Dimensions of defocusing. In: Shibatani, Masayoshi et al. (eds.): *Voice and grammatical relations: In honor of Masayoshi Shibatani*. Amsterdam, Philadelphia: John Benjamins. 115–137. <https://doi.org/10.1075/tsl.65.08lan>

- Lengyel Klára 2000a. A segédigék és származékaik. In: Keszler Borbála (szerk.): *Magyar grammatika*. Budapest: Nemzeti Tankönyvkiadó. 252–258.
- Lengyel Klára 2000b. *Az igenevek helye a szófaji rendszerben*. Nyelvtudományi Értekezések 146. Budapest: Akadémiai Kiadó.
- Lewandowska-Tomaszczyk, Barbara 2007. Polysemy, prototypes, and radial categories. In: Geeraerts, Dirk – Cuyckens, Hubert (eds.): *The Oxford handbook of Cognitive Linguistics*. Oxford: Oxford University Press. 139–169.
- Modrián-Horváth Bernadett 2024. Vázlat a magyar passzív jelentésű konstrukciók vizsgálatához a használatalapú nyelvtan szemszögéből. In: Tátrai Szilárd – Laczkó Krisztina – Tolcsvai Nagy Gábor (szerk.): *Magyar igei konstrukciók. Használatalapú függőségi vizsgálatok*. (Megjelenés alatt.)
- Németh Boglárka 2012. *Az aspektus a magyar nyelvben. Különös tekintettel a statikusságra*. Budapest: Tinta Könyvkiadó.
- Novák, Attila – Siklósi, Borbála – Oravecz, Csaba 2016. A new integrated open-source morphological analyzer for Hungarian. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Portorož, Slovenia: European Language Resources Association (ELRA). 1315–1322.
- Oravecz, Csaba – Várad, Tamás – Sass, Bálint 2014. The Hungarian Gigaword Corpus. In: *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*. Reykjavík: European Language Resources Association (ELRA). 1719–1723.
- Pethő Gergely – Sass Bálint – Kalivoda Ágnes – Simon László – Lipp Veronika 2022. Igekötő-kapcsolás. In: Berend Gábor – Gosztolya Gábor – Vincze Veronika (szerk.): *XVIII. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi Tudományegyetem, Informatikai Intézet. 77–91.
- Rosch, Eleanor 1978. Principles of categorization. In: Rosch, Eleanor – Lloyd, Barbara B. (eds.): *Cognition and Categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates. 27–48. <https://doi.org/10.4324/9781032633275-4>
- Schultink, Henk 1961. Produktiviteit als morfologisch Fenomeen. *Forum der Letteren* 2: 110–125.
- Siewierska, Anna 2013. Passive constructions. In: Dryer, Matthew S. – Haspelmath, Martin (eds.) *WALS Online* (v2020.3) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.7385533> (<http://wals.info/chapter/107> [2023. 10. 05.])
- Simon Gábor 2024. Konstrukció és/vagy mintázat: Az azonosítás kérdései egy használatalapú, korpuszvezérelt nyelvleírásban. In: Tátrai Szilárd – Laczkó Krisztina – Tolcsvai Nagy Gábor (szerk.): *Magyar igei konstrukciók. Használatalapú függőségi vizsgálatok*. (Megjelenés alatt.)

- Stefanowitsch, Anatol 2006. Konstruktionsgrammatik und Korpuslinguistik. In: Fischer, Kerstin – Stefanowitsch, Anatol (Hg.): *Konstruktionsgrammatik I: Von der Anwendung zur Theorie*. Tübingen: Stauffenburg.
- Stefanowitsch, Anatol 2020. *Corpus linguistics. A guide to the methodology*. Textbooks in Language Sciences 7. Berlin: Language Science Press. DOI: 10.5281/zenodo.3735822
- Szepeszy Gyula 1972. A *lenni* + *-va*, *ve* igeneves szerkezet funkciói. *Magyar Nyelvőr* 96: 404–414.
- Szili Katalin 1999. Valahol a passzívum és a mediálisok között ... (A szenvedő jelentéstartalom változó megjelenésmódja kapcsán). *Magyar Nyelvőr* 123 (3): 350–357.
- Tanos Bálint 2008. Joan Bybee nyelvelmélete. *Általános Nyelvészeti Tanulmányok XXII*: 537–566.
- Tanos Bálint 2012. Határozószó-képzési produktivitás kvantitatív keretben. In: Tolcsvai Nagy Gábor, Tátrai Szilárd (szerk.): *Konstrukció és jelentés. Tanulmányok a magyar nyelv funkcionális kognitív leírására*. Budapest: ELTE. 59–69.
- Tolcsvai Nagy Gábor 2017. Jelentéstan. In: Tolcsvai Nagy Gábor (szerk.): *Nyelvtan*. Budapest: Osiris Kiadó. 207–499.
- Tsötönyi Márton 1831. *Leg hasznosabb és új fel fedezésekkel irott rövid gazdasági munka*. Pest: Wigand Ottó.
- van Marle, Jaap 1985. *On the paradigmatic dimension of morphological creativity*. Dordrecht: Foris. <https://doi.org/10.1515/9783111558387>
- van Wette, Niek 2022. The hapax / type ratio. An indicator of minimally required sample size in productivity studies? *International Journal of Corpus Linguistics* 27 (2): 166–190. <https://doi.org/10.1075/ijcl.19114.van>
- Zeldes, Amir 2012. *Productivity in Argument Selection: From Morphology to Syntax*. Berlin, Boston: Walter de Gruyter. <https://doi.org/10.1515/9783110303919>



Quantitative productivity analysis of Hungarian passive predicative constructions

The aim of the study is to investigate the quantitative productivity of four passive constructions: *V-ódik/-ődik*, *V-(t)atik/-(t)etik*, *V-va/-ve lesz*, and

V-vá/-ve van. The research is based on automatic and manual analyses using the HGC and the reannotated HGC-UD corpora, and combines the global productivity index (cf. Baayen 1992), which can be calculated in a quantitative framework, with qualitative elements. The Global Productivity Index is a measure often used in usage-based linguistics, which can be calculated using the type and token frequency of constructional schemas and the number of hapaxes legomena fitting into the schemas. The qualitative part of the study was aimed at classifying the sample of hapaxes that best reflect the productive meaning of each schema according to their voice.

The results of the study show that the schema with the highest global productivity in the totality of the corpus is the *V-vá/-ve van* construction, the one with the highest type frequency is *V-ódik/-ődik*, while the one with the highest potential productivity is *V-vá/-ve lesz*; the results for individual sub-corpora differ partially. The analysis of a random sample of hapaxes shows that the schema with the most persistent passive meaning is the extremely low productivity construction *V-tatik/-tetik*, closely followed by *V-vá/-ve lesz*, the *V-vá/-ve van* predominantly passive meaning, and finally less than half of the hapaxes have passive meaning in *V-ódik/-ődik*. Further theoretical and methodological considerations are also discussed in the paper.

Keywords: Hungarian grammar, voice, passive, corpus linguistics, construction grammar

Modrián-Horváth Bernadett

Szegedi Tudományegyetem

bernadett.modrianhorvath@gmail.com

ORCID: 0000-0003-3834-9163

Kalivoda Ágnes

HUN-REN Nyelvtudományi Kutatóközpont

kalivoda.agnes@nytud.hun-ren.hu

ORCID: 0000-0003-2520-5523

Tanos Bálint

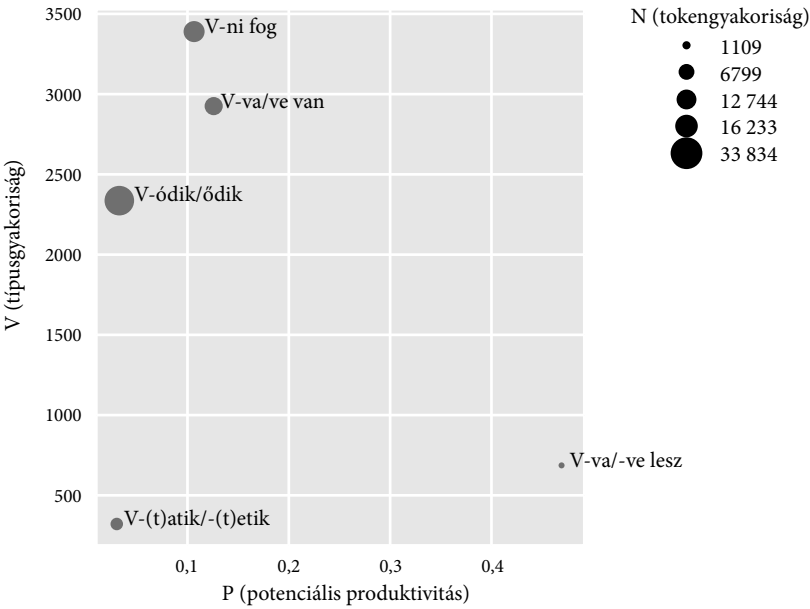
batanos@gmail.com

ORCID: 0009-0005-6070-3490

I. számú melléklet: Szépirodalmi alkorpusz

5. táblázat. A vizsgált konstrukciók tokengyakorisága (N), típusgyakorisága (V), hapaxainak száma (n₁) és potenciális produktivitása (P) a szépirodalmi alkorpuszban

szerkezet	N	V	n ₁	P
V-ódik/-ődik	33 834	2325	1177	0,0348
V-va/-ve van	12 744	2915	1606	0,1260
V-(t)atik/-(t)etik	6799	330	213	0,0313
V-va/-ve lesz	1109	688	518	0,4671
V-ni fog	16 233	3379	1761	0,1085

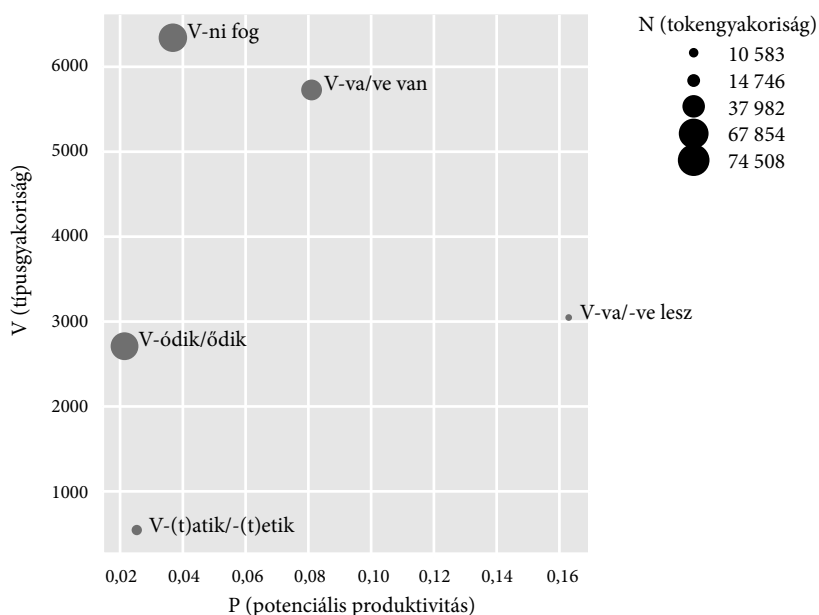


3. ábra. A vizsgált konstrukciók globális produktivitása a szépirodalmi alkorpuszban

II. számú melléklet: Személyes alkorpusz

6. táblázat. A vizsgált konstrukciók tokengyakorisága (N), típusgyakorisága (V), hapaxainak száma (n₁) és potenciális produktivitása (P) a személyes alkorpuszban

szerkezet	N	V	n ₁	P
V-ódik/-ődik	67 854	2687	1439	0,0212
V-va/-ve van	37 982	5682	3080	0,0811
V-(t)atik/-(t)etik	14 746	535	368	0,0250
V-va/-ve lesz	10 583	3022	1723	0,1628
V-ni fog	74 508	6293	2735	0,0367

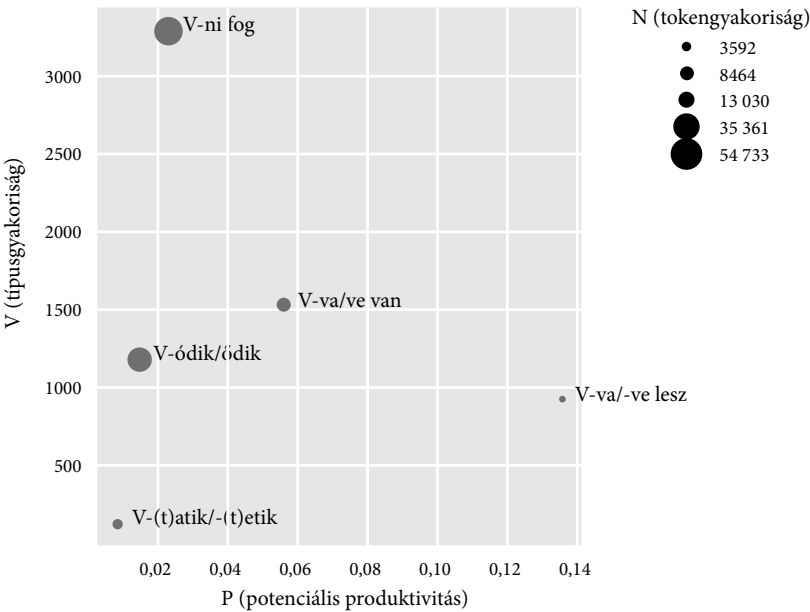


4. ábra. A vizsgált konstrukciók globális produktivitása a személyes alkorpuszban

III. számú melléklet: Hivatalos alkorpusz

7. táblázat. A vizsgált konstrukciók tokengyakorisága (N), típusgyakorisága (V), hapaxainak száma (n₁) és potenciális produktivitása (P) a hivatalos alkorpuszban

szerkezet	N	V	n ₁	P
V-ódik/-ődik	35 361	1179	533	0,0151
V-va/-ve van	13 030	1528	739	0,0567
V-(t)atik/-(t)etik	8464	131	75	0,0089
V-va/-ve lesz	3592	928	489	0,1361
V-ni fog	54 733	3280	1293	0,0236

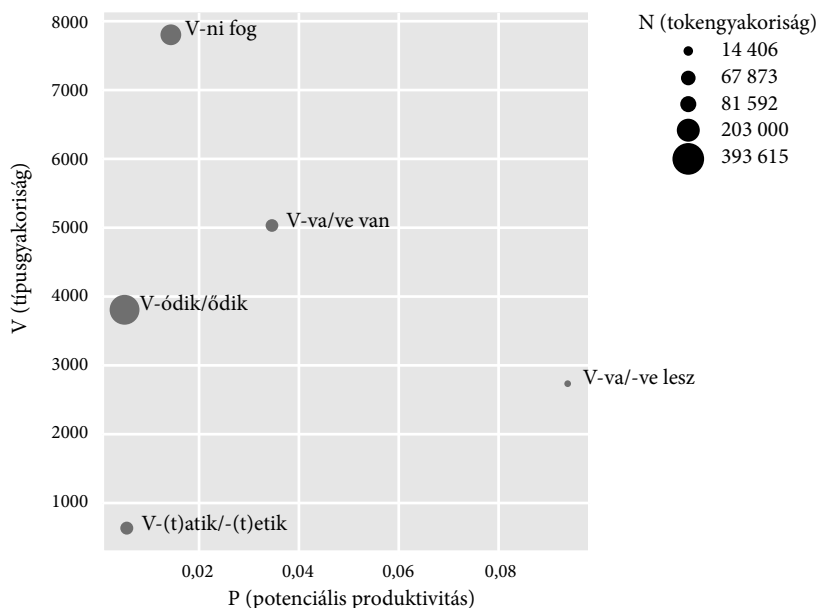


5. ábra. A vizsgált konstrukciók globális produktivitása a hivatalos alkorpuszban

IV. számú melléklet: Sajtónyelvi alkorpusz

8. táblázat. A vizsgált konstrukciók tokengyakorisága (N), típusgyakorisága (V), hapaxainak száma (n₁) és potenciális produktivitása (P) a sajtónyelvi alkorpuszban

szerkezet	N	V	n ₁	P
V-ódik/-ődik	393 615	3768	1800	0,0046
V-va/-ve van	67 873	5008	2336	0,0344
V-(t)atik/-(t)etik	81 592	598	415	0,0051
V-va/-ve lesz	14 406	2696	1354	0,0940
V-ni fog	203 000	7772	2823	0,0139

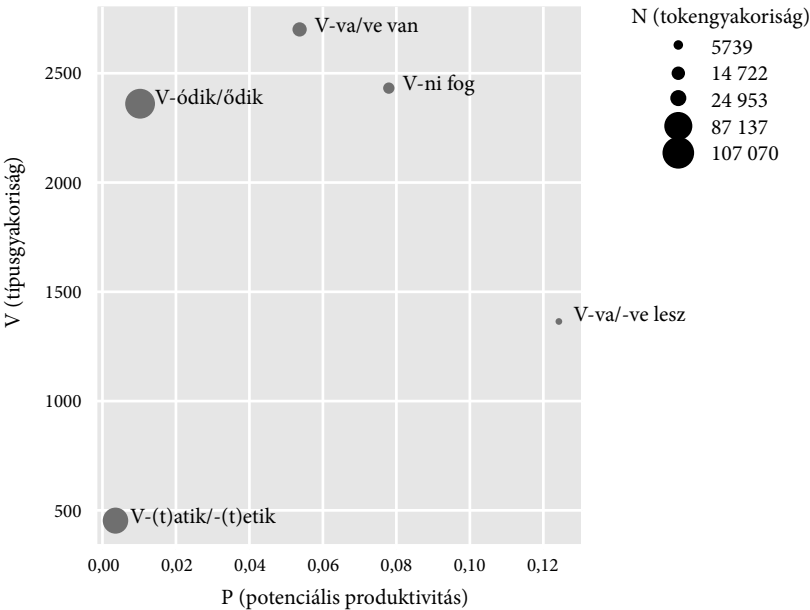


6. ábra. A vizsgált konstrukciók globális produktivitása a sajtónyelvi alkorpuszban

V. számú melléklet: Tudományos ismeretterjesztő alkorpusz

9. táblázat. A vizsgált konstrukciók tokengyakorisága (N), típusgyakorisága (V), hapaxainak száma (n₁) és potenciális produktivitása (P) a tudományos ismeretterjesztő alkorpuszban

szerkezet	N	V	n ₁	P
V-ódik/-ődik	107 070	2355	1059	0,0099
V-va/-ve van	24 953	2698	1339	0,0537
V-(t)atik/-(t)etik	87 137	444	285	0,0033
V-va/-ve lesz	5739	1359	716	0,1248
V-ni fog	14 722	2427	1148	0,0780

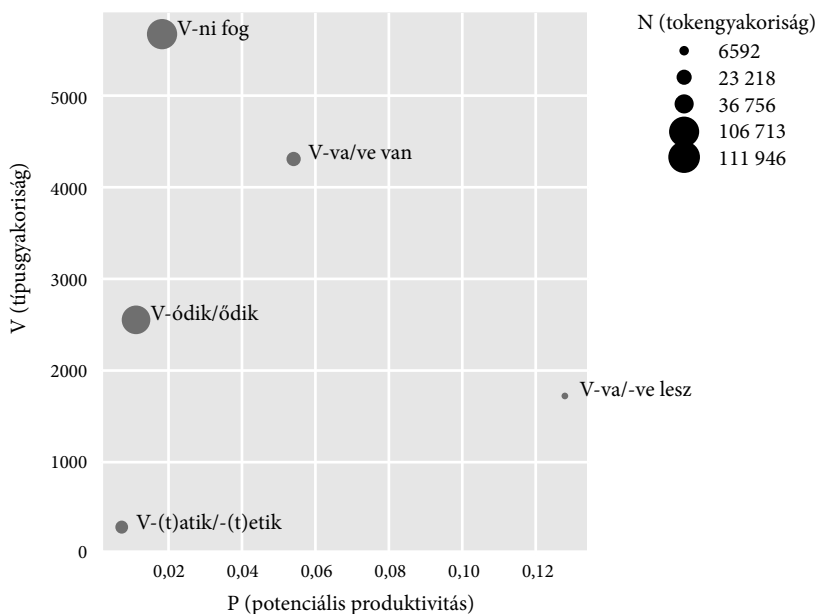


7. ábra. A vizsgált konstrukciók globális produktivitása a tudományos ismeretterjesztő alkorpuszban

VI. számú melléklet: Beszélt nyelvi alkorpusz

10. táblázat. A vizsgált konstrukciók tokengyakorisága (N), típusgyakorisága (V), hapaxainak száma (n1) és potenciális produktivitása (P) a beszélt nyelvi alkorpuszban

szerkezet	N	V	n ₁	P
V-odik/-ődik	106 713	2537	1215	0,0114
V-va/-ve van	36 756	4321	1983	0,0540
V-(t)atik/-(t)etik	23 218	270	172	0,0074
V-va/-ve lesz	6592	1718	844	0,1280
V-ni fog	111 946	5675	2048	0,0183



8. ábra. A vizsgált konstrukciók globális produktivitása a beszélt nyelvi alkorpuszban

KUGLER NÓRA

**Hozzászólás *A magmondat szerepű, passzív jelentésű
sémák produktivitásának kvantitatív vizsgálata* című
tanulmányhoz**

1. A tanulmány főbb újdonságai, eredményei

A tanulmány egyik fő módszertani újdonsága annak kidolgozása, hogyan lehet mérni és összevetni szintetikus igealakok (morfológiai konstrukciók, például *betömetik*, *betömődik*), illetve analitikus igealakok (szintaktikai jellegű konstrukciók, például *be lett tömve*) produktivitását.

A dolgozat összetettségét mutatja a gépi (lásd 4. fejezetpont)¹ és a finom, kvalitatív, kézi elemzés (lásd 5. és 6.) együttes alkalmazása. Az adatok humán elemzése ráirányította a figyelmet arra, hogy a konstrukciók szórendi változatosságát is szükséges figyelembe venni, mivel a szórendi konstruálás különbsége bizonyos szerkezetek esetében – adatolhatóan – funkcionális különbséggel asszociálódik (7.3.).

Újdonság, hogy az MNSZ2-UD-ben az emPreverb eszközzel a Szerzők megoldották, hogy az elemző az igekötős igék szórendi megvalósulásait egy lemmához tartozó alakokként kezelje (lásd 4.1.).

A keletkezés éve metaadat alapján a Szerzők el tudták végezni nagyszámú adat lehető legbiztosabb időbeli elhelyezését. Ez akkor is nagy jelentőségű, ha tudjuk, hogy a dokumentum metaadatainál szereplő év sokszor megtévesztő az adat keletkezési ideje szempontjából (például forrásidézés esetében).

Új eredmény a vizsgált konstrukciók esetében a produktivás adataalapú mérése. „A vizsgálat eredményei azt mutatják, hogy a korpusz egészében

¹ A fejezetpontok Modrián-Horváth-Kalivoda-Tanos címben szereplő tanulmányának megfelelő részeire mutatnak. A lektorált tanulmány szerzőire Szerzők formával hivatkozom a továbbiakban.

a legnagyobb globális produktivitással rendelkező séma a *V-va/-ve van* konstrukció, a legnagyobb típusgyakorisággal a *V-odik/-ődik*, míg a legnagyobb potenciális produktivitással a *V-va/-ve lesz* rendelkezik; az egyes alkorpuszokra vonatkozó eredmények ettől részben eltérnek” (Modrián-Horváth-Kalivoda-Tanos, Kivonat). A kézi elemzés révén a Szerzők megfigyeléseket tettek az egyes konstrukciók megvalósulásainak igenemek szerinti osztályozásával és a hapaxok, valamint a leggyakoribb kifejezések alapján az igenemek szerinti belső változatossággal, arányokkal kapcsolatban.

A Szerzők rendszeres különbséget mutattak ki a *V-va/-ve van* és a *V-va/-ve lesz* konstrukciók között, az utóbbit dinamikusság, általánosan és egyértelműen erős passzív jelentés (passzív igenembe tartozás) jellemzi.

A hozzászólás a Szerzők további új eredményeit is részletezi, az első pont csak kiemelte a tanulmány főbb hozadékát, és érzékeltetni kívánta a munka jelentőségét. A továbbiakban témák köré szervezve fogalmazok meg észrevételeket, továbbgondolásra szánt javaslatokat. Néhány esetben reflektálok a Szerzők és a Lektor között zajló közös gondolkodásra is.

2. A gépi és a kézi elemzés, az adatvételezés és az elemzés módszertana

Az adatkezelés és az elemzés bemutatása szempontjából a tanulmány igényes, gondos.

A Szerzők a 4.2. pontban azt írják, hogy mivel felfigyeltek arra, hogy a *-va/-ve* képzős alak jobb kontextusában megjelenő létige nem minden esetben a keresett konstrukció része, ezért „csak azokat a találatokat őriztük meg, ahol a létige megelőzi a *-va/-ve* képzős alakot (például *lett megszerezve*, *meg lett szerezve*), vagy ahol a létige előtt álló *-va/-ve* képzős alak igekötő nélküli határozói igenév” (4.2.). Kézi ellenőrzés nélkül azonban így sem zárható ki a téves találatok. Például az *olvasva* igekötő nélküli határozói igenév és a *lesz* vele szomszédos, múlt idejű alakjai (3 találat az MNSZ2-ben, alkorpuszra szűkítés nélkül) mind ilyenek (#527778009, #986735265, #1082416488).

- (1) [...] József Attila Róheim korábban idézett cikkét *olvasva lett* kíváncsi arra, [...] (#527778009)

A duplumszűrővel kapcsolatos tapasztalatok alapján a Szerzők megalapozottan mutatnak rá arra, hogy az MNSZ2, illetve bármely „tisztítatlan” korpusz korrekció nélküli felhasználása önmagában a duplumok nagy száma miatt megbízhatatlanná teheti a statisztikai elemzést (4.1., 5.).

A Szerzők is reflektálnak arra a problémára, hogy „a V-ódik/-ődik sémára vonatkozó adatok csak korlátozottan hasonlíthatók össze a többi eredmény-nyel” (5.1.).

A 6.3. pontban az olvasható, hogy „az igekötős V-va *van/lesz* szórendű találatok szinte kivétel nélkül hamis pozitívak”. Igekötős határozói igenévre, és azt követő *van* szótőre keresve azonban sok jó találatot kaptam a teljes korpuszban, de a beszélt nyelvi alkorpuszban is. Ezekből néhány példát közlök illusztrációként.

- (2) És mindazok a nehézségek egyik pillanatról a másikra szemünk elé kerültek amelyek eddig *elfedve voltak*. (#853351106)
- (3) ne csak a tornaterembe, vagy ahol az emberek éppen *összegyűjtve vannak* (#855014252)
- (4) Tehát Magyarországon ez a galopp sport teljesen *elnyomva van* (#869686612)

Elfogadtam azonban a Szerzőknek azt az érvelését, hogy a rossz találatok nagy aránya rontott volna az automatikus elemzésen.

A példák közlésének gyenge pontja az, hogy a korpuszadatok mellett dokumentumazonosító szerepel. Ez ugyan nem akadálya annak, hogy az adat visszakereshető legyen (a probléma természetesen fel sem vetődik a könnyen kinyerhető hapaxok esetében), de nem kérdezhető le közvetlenül, ahogy szó-azonosító megadásával ez lehetséges volna.

3. A produktivitás mérése, az értékek értelmezése, avagy mérhető is a globális produktvitas, meg nem is

A produktivitás méréséhez a Szerzők olyan módszertant választottak, amely nagymennyiségű adat gépi elemzésével számszerűsíteni tudja annak a lehetőségességét, hogy az adott konstrukciós séma (amely a már ismert példányokból szerkezetként és műveletként is sematizálódik) a nyelvhasználatban spontán megvalósul (lásd globális produktivitas). A vizsgált konstrukciók a típusgyakoriság és a potenciális produktivitas

által meghatározott koordináta-rendszerben elhelyezhetők, a kapott értékek mindkét dimenzióban (és a globális produktivitás értéke szerint is) összevethetők. A Szerzők külön is közlik a hapaxok számát és a példánygyakoriságot (táblázatosan mindkettőt, az ábrák esetében kiemelten a példánygyakoriságot), amelyek alapján a potenciális produktivitás értéke kiszámolható. Annak érdekében teszik ezt, hogy a konstrukciókról az olvasó összetett, pontos képet kapjon. Ezzel az eljárással elvileg skalárisan értelmezhető a vizsgált konstrukciók produktivitása. A konstrukciók közötti viszonyok ezen tényezők, illetve a kétféle számítás (típusgyakoriság v. realizálódott produktivitás, valamint potenciális produktivitás) alapján jól modellálható a globális produktivitás. A kérdés már csak az, hogy mit jelentenek ezek az eredmények, mit jelentenek a kapott értékek.

Az értelmezéssel kapcsolatban több kérdés is felvethető. Tudjuk, hogy mit jeleznek a realizált/megvalósult produktivitás és a potenciális produktivitás mérőszámai, valamint hogy ezek együttesen alakítják a globális produktivitást, de az nem magától értetődő, hogy a produktitásnak ezek a „dimenziói” hogyan viszonyulnak egymáshoz, hogyan járulnak hozzá a skalaritás, gradualitás értelmezéséhez.

Vajon mit jelent az (hogyan lehet értelmezni azt), hogy a sémák a produktivitás szempontjából eltérő aspektusokból emelkednek ki? Mit jelent az, hogy a globális produktivitás szempontjából a *V-va/-ve van*, a potenciális produktivitás szempontjából a *V-va/-ve lesz*, a típusgyakoriság szempontjából pedig a *V-odik/-ődik* emelkedik ki? Mit jelent az, ha egy konstrukció esetében a potenciális és a realizált produktivitás épp ellentétes irányban tér el a többi konstrukció értékeitől? Ezeket a kérdéseket a Szerzők is felteszik a 7.1. fejezetpontban, például amikor az egyik konstrukció (*V-va/-ve van*) a típusgyakoriság, a másik (*V-va/-ve lesz*) a potenciális produktivitás mutatója alapján emelkedik ki. Baayen plauzibilis érvelésére támaszkodva az előbbi, a típusgyakoriságot tekintik a produktivitás szempontjából jelentősebbnek. Ez azt mutatja, hogy bár két dimenzióban tájékozódunk, ezek nem egyforma súllyal esnek latba a produktivitással kapcsolatos értelmezés, következtetés során. A Szerzők a végleges szövegben (7.1.) rámutatnak arra is, hogy a kérdés nyitottságának a magyarázata a jelenség komplexitásában rejlik (például a tokenérzékenység játszhat szerepet az értékek különbségében, de nem egy tényező önmagában idézi elő vagy magyarázza az adatolt eltéréseket).

A 7.2. fejezetpont úgy értelmezi az analitikus szerkezetek produktivitási értékeit a *V-ni fog* kontrollszerkezet értékeinek a viszonyában, hogy „a konstrukció szintetikus vagy analitikus jellege nem áll közvetlen összefüggésben

sem a realizált, sem a potenciális produktivitás mértékével” (7.2.). A szóképzési mintázatokkal szemben a *V-ni fog* kontrollszerkezetet épp annak inflexióis jellege, ugyanakkor analitikus felépítése, az inflexióis jelleggel összefüggésben várhatóan kiemelkedően nagy produktivitása motiválta. Ennek a várakozásnak azonban részlegesen ellentmondanak az adatok (erre a Szerzők is reflektálnak). Az analitikus passzív jelentésű konstrukciók potenciális produktivitása rendre nagyobb, mint a szintetikus konstrukcióké, a *V-va/-ve van* esetében pedig a realizált produktivitás (típusgyakoriság) minden viszonylatban kiemelkedő, és alig marad el a *V-ni fog* realizált produktivitásától. Hiányérzetem van annak értelmezésével kapcsolatban, hogy a *V-ni fog* potenciális produktivitása lényegesen alacsonyabb, mint az analitikus passzív jelentésű konstrukcióké. A Szerzők szerint ez „megnyugtató a jelen vizsgálat metodológiájával kapcsolatban” (7.2.), én inkább nyugtalanítónak gondolom. Ha a *V-ni fog* megfelelő viszonyítást tesz lehetővé az analitikus konstrukciók esetében a produktivitás skalaritása szempontjából, és arra számítunk, hogy inflexióis jellege miatt erősen produktív, ezért várható a spontán megvalósulása a nyelvhasználatban, akkor ezt a várakozást a típusgyakoriság erősíti meg, ugyanis ez az a dimenzió, amelyben a *V-ni fog* (kissé) kiemelkedik. Érthető, hogy a nagy példánygyakoriság csökkenti a potenciális produktivitás értékét, és nem kell fennállnia annak az összefüggésnek, hogy egy produktív inflexióis konstrukciónak mindkét produktivitási értéke nagy legyen, mindenesetre engem elbizonytalanított az eredmény a potenciális produktivitás súlyával kapcsolatban. A Szerzők csak az összefoglalásban jelzik, hogy az eredmény számukra is váratlan (a *V-va/-ve van* esetében „egy jelölt igenem- és aspektuális komponensekkel bíró konstrukcióról van szó, amellyel kapcsolatban nem várnánk, hogy produktivitása felveszi a versenyt egy igeidő-jelölő konstrukcióéval, vagy akár felül is múlja azt” lásd 8. fejezetpont), de ebből nem módszertani, hanem kategorizációs következtetésre jutnak: a *V-va/-ve van* is besorolható lehet az analitikus igealakok közé. Ez egyrészt megerősíti a korábbi hasonló felvetéseket (lásd például Lengyel (2000)-ben is grammatikai érvekkel alátámasztva), de (megítélésem szerint) nem magyarázza a *V-va/-ve van* mutatóit.

A dolgozatban szó van arról, hogy melyik konstrukciónak a legnagyobb a globális produktivitása (4.2.), de nehéz értelmezni a két számolt érték alapján egy olyan P^* minőséget/mértéket, amely vizuálisan „reprezentálódik” („a globális produktivitás egy »vizuális természetű« mérési módszer” 7.1.). Tehát a két dimenzió együttes, de nem nyilvánvalóan egyenrangú figyelembevételével kell érzékelnie az olvasónak a globális produktivitás mértékét.

Számomra zavarba ejtő az, hogy topológiai elem táruhat ez az összefüggés, de ez nem létesülő új érték vagy tartomány a koordináta-rendszerben, hanem olyan viszony, amely az indexek alapján van, ugyanakkor számítással nem adható meg (vagy nem érdemes megadni), következésképp az alkalmazott ábrázolási keretben nem is jeleníthető meg, vagyis ez a „vizuális reprezentáció” még sincs adva, nem ábrázolódik (de tudhatjuk, hogy nagy-e vagy kicsi).

További kérdés az is, hogyan jelölhetők ki a grádusok a globális és/vagy a realizált és a potenciális produktivitás mutatóinak az alapján. Milyen megalapozottsága van annak, hogy az értékek alapján fokozatokat lehessen megalapítani a produktivásban azon túl, hogy a vizsgált konstrukciók az értékek alapján egymáshoz viszonyíthatók? Vajon a hapaxok milyen aránya jelezne egyértelműen nagy produktivitást? Köszönöm a Szerzőknek, hogy a végleges szövegváltozatban lépéseket tettek ennek kidolgozására (7.1.).

A lehetőségesség kalkulációja a potenciális produktivitás számításán alapul (viszonyban az adatolt tényleges megvalósulások típusgyakoriságával). A potenciális produktivitást a hapaxok és a tokengyakoriság hányadosa adja. A dolgozat pontosan közli a mérési módszertan részleteit, az tehát világos, hogyan történt a számolás, a konstrukciók adatainak a kinyerésével foglalkozó részből pedig az is transzparenssé válik, hogy mi az a nyelvi anyag, amelynek a mérését a Szerzők elvégzik. A bizonytalanság akkor lép be, ha a statisztikai elemzés eredményével modellálni akarjuk a nyelvhasználatot, mert akkor értelmeznünk kell, hogy mit jelent az adott morfológiai konstrukció spontán megvalósításának a lehetőségessége.

Mit jelent az, hogy a *V-ni fog* kontrollkonstrukció „a legnagyobb típusgyakoriságú, de az alacsony potenciális produktivitását tekintve inkább a szintetikus szerkezetekhez [értsd a *V-ódik/-ődik*, *V-tatik/-tetik* értékeihez, magyarázat a hozzászólótól] közelít” (4.3.)? Látjuk az értékeket, de úgy tűnik, hogy az nem nyilvánvaló, hogyan kell értelmezni ezt az adott konstrukció műveletére vonatkozóan.

Tudhatjuk-e azt valójában, hogy a *V-tatik/-tetik* konstrukció P vagy P* „értéke” elég vagy nem elég ahhoz, hogy a szóképzési mintázat műveleti oldala is részét képezze a nyelvhasználók tudásának, és megnyilvánuljon az innovatív alkalmazásban?

A baayeni módszertannak a magyar igeképzési mintázatokra történő kiterjesztő alkalmazásával kapcsolatban is értékes hozzájárulást tesznek a Szerzők, rámutatnak ugyanis ennek korlátaira. Az adatok alapján úgy értékelik, hogy az igeképzés produktivitásának meghatározásában a szimplex

igék hapaxai és tokengyakorisága alapján számolt potenciális produktivitás értéke nem alkalmas arra, hogy egyfajta viszonyítási pontként szolgáljon a szóképzési produktivitás értelmezése számára. Ez a megfigyelés a főnévi és az igei képzési sémák közötti különbség megragadása révén hozzájárulhat ahhoz, hogy pontosabb, kidolgozottabb legyen a leírása a lexikális kategóriák és a morfológiai konstrukciók összefüggésének. Ebben a kidolgozási folyamatban Palágyi László munkáira érdemes lehet hivatkozni (főként Palágyi 2022: 132).

A dolgozat 4.4. fejezetpontja lényeges megfigyeléseket közöl az alkorpuszok mentén kirajzolódó hasonlóságokról és különbségekről az öt konstrukció (négy cél-, egy kontrollkonstrukció) vonatkozásában. Bizonyára óvatosságból, a Szerzők visszafogottan értelmezték a megfigyeléseket. Köszönöm, hogy a diszkussziós részt ebben a vonatkozásban is kifejtettebbé tették a Szerzők a végleges változatban.

4. A passzív jelentés

A passzív és a mediális igék közötti átmenet kezeléséhez, az erőforrás (tágabban az erődinamikai interakció szereplőinek) pontosabb leírásához jól alkalmazhatónak látom a tranzitív paramétereket (például az erőforrás ágenciája, akaratlagossága, a szereplők egyeditettsége; a nem erőforrásként konceptualizált másik szereplő ráhatottsága; az energetikai interakció pontszerűsége, mozgásos-cselekvéses jellege [Kinesis]; Hopper–Thompson 1980). Ez segíthetne az azonosításra való hivatkozások pontosításában. Lásd például (a 3.2. fejezetpontban): „A ráhatottság egyértelműen megvalósul, de a cselekvés kivitelezője ezekben az esetekben **nem világos**. Nem prototipikusan, de passzív jellegű a konstruálás például nem ágensi okozó esetében is (*elfelejtődik*); míg sok nem konkrét folyamat konstruálása (például *konstruálódik*, *elsáncolódik*) már szintén **a mediális tartomány irányába mutat**, mert az energiaforrás ezekben az esetekben **nehezen azonosítható**” (kiemelés tőlem, KN).

A passzív jelentés körülhatárolásánál, a mediális és a passzív igenem viszonyának tárgyalásánál hiányoltam Lengyel Klára értekezését (2000: 67–69), mivel az érvel mind az analitikus igei elemzés, mind a *V-va/-ve van* konstrukció statikussága mellett, továbbá azt állítja, hogy „A tárgyas igékből létrehozott analitikus igealakok passzív és mediális jelentésűek egyszerre; igenem szempontjából ezeket a mediopasszív kategóriába sorolhatjuk” (Lengyel

2000: 69). Ez eléggé egybecseng a Szerzők kategorizációs konklúziójával. Ugyanakkor az is lényeges, hogy a Szerzők az adatok finom elemzésével, kontextuális egyértelműsítéssel ennek az átmeneti tartománynak a pontosítására törekszenek.

Mivel az analitikus szerkezetekben határozói igenévi alak jelöli az állapotváltoz(tat)ás folyamatát, ez a konstruálási mód felszámol olyan megkülönböztetéseket, amelyeket a szintetikus igealakok, az igei lemma alakjai megjeleníthetnek (ez főképp a jelen idejű, kijelentő módú, egyes számú alakokra igaz). Ilyen például az aktív : passzív különbség ices ragozással való kifejezése (lásd *tör : törík*, 5.2.), illetve a műveltető és a passzív képzés különbsége (lásd *felolvastat : felolvastatik*).² Különösen az előbbinek szerepe lehet abban a mediopasszív meghatározatlanságban vagy mediális : passzív többértelműségben, amely a keretnek ('frame') megfelelően kidolgozott szereplői viszonyok mellett is fennállhat (*le van törve az ág / ez az ember*). Ezért egyfelől indokolt a Szerzőknek az az eljárása, hogy a *tör/törík* igei komponensek összevonásával egy igei lemmához sorolódnak a (*le*)*törve van* szerkezetek, másfelől viszont az is érdekes volna, hogy maga a konstruálási mód hogyan kezdeményezi a passzív jelentés terében a poliszémian alapuló többértelműséget vagy meghatározatlanságot, illetve hogyan járul hozzá a kategória határok életlenségéhez. Az ige és a folyamatjelentéssel összefüggő szereplői viszonyok (vö. más megközelítésben az argumentumszerkezet) kérdése nem független attól sem, hogy unergatív vagy unakkuzatív a szerkezet megítélése (lásd 5.2., például a *salakozva van* esetében *vki salakozik* vagy *vki salakoz vmit* folyamat eredményezi az adott állapotot).

Érdemes lenne a mediálissal való érintkezés miatt bevonni a *V-ÖbAn van* konstrukciót is (például *vmi elázóban van*), amelyet Lengyel (2000: 66) termékeny analitikus mediális képzésmódnak³ tart, bár megfontolandó érveket hoz fel az analitikus igealakként való értelmezéssel szemben is (Lengyel 2000: 67). A *V-ÖbAn van* konstrukcióról a 3.6. (*További passzív jelentésű magmondat-konstrukciók* című) fejezetpont nem tesz említést, igaz, nem is törekszik teljességre. A konstrukció Lengyel szerint produktívan „mediális, folyamatos jelentésű, állapotváltozást kifejező igékből hozható létre” (Lengyel 2000: 66); a léteigi komponens nem a statikusság kifejezését végzi el (az igei grammatikai kategóriák felvétele mellett), hanem a folyamatos

² Természetesen az eltérő jelentések a konstruálás egyéb jellemzőivel is összefüggnek (a figurák/szereplők jellege, viszonyuk, erődinamika, eseményszerkezet, vö. Tolcsvai Nagy 2017: 317), nem az igealak önmagában kapcsolódik össze ezekkel az értelmekkel.

³ Szándékosan nem alkalmazom a mozgószabályt.

értelmezést erősíti. A konstrukció egyértelműen komplex magmondatként funkcionál (vö. Imrényi 2017). Lengyel egy kapcsolathálózatot is felvázol (a nyomtatott formában inkább sorozatszerűen), amelyet az állapot(változás)-jelentés tart össze, és a dinamikus : statikus aspektuális különbségek is alakítanak: „*felborulóban van – felborul – fel van borulva; kibomlóban van – kibomlik – ki van bomolva*” (Lengyel 2000a: 67).

A *V-ÓbAn van* konstrukciók dominánsan mediálisak, nem nyilvánvalóan passzívak (a Szerzők értelmezésében azonban a passzív jelentés tartományában működnek). Az MNSZ2-ben a legnagyobb, 100 fölötti példánygyakorisággal adatolt megvalósulások jellemzően olyan szerkezetek, amelyek alapszava mediális ige, amellyel nem feltétlenül asszociálódik egy kauzatív, előidéző esemény.

1. táblázat. A *V-ÓbAn van* konstrukció leggyakoribb típusai az MNSZ2-ben

kialakulóban	504
visszaszorulóban	387
átalakulóban	359
csökkenőben	357
fogyóban	290
eltűnőben	286
terjedőben	273
alakulóban	231
leáldozóban	225
megszűnőben	199
kimerülőben	186
fellendülőben	176
függőben	175
feljövőben	164
változóban	148
hanyatlóban	135
készülőben	132
kihalóban	130
lecsengőben	123
növekvőben	112

Például ha valami változik (lásd *változóban van*), és az elsődleges figura a belső folyamatai révén (a környezettel való nem száliens kapcsolata mellett) változásra képes (például élő szervezet), akkor kevésbé várható, hogy a teljes hatókör részeként aktiválódjon egy körülhatárolt környezeti tényező mint változtató erő. Azonban még az ilyen típusok esetében is mediopasszív a szerkezet, ha a változás csak külső hatásra következhet be, például a vélemény és a családi költségvetés nem önmagától változik meg.

- (5) Talán a véleménye is *változóban van*, és azt hiszem hajlik rá, hogy jelen legyen. (#124376184)
- (6) a családi költségvetés összege és felhasználásának módja is *változóban van*. (#160216600)

Ugyanakkor a leggyakoribb kifejezések között is ott van a *készülőben*, amely tipikusan (de nem minden esetben, például (8)) feltételez külső hatóerőt, méghozzá tipikusan intencionális ágenst, készítőt (7).

- (7) Valami fehér könyv is *készülőben van*. (#86174684)
- (8) maga a költő is *készülőben van* (#25538785)

Nem elhanyagolható a kevésbé gyakori kifejezések között sem a külső hatóerőt, stimulust (például *felszáradóban* (39)) vagy szándékos cselekvő ráhatását feltételező konstrukció (például *épülőben* (38)).

- (9) A megye útjain jelentős vízátfolyások voltak, de a burkolatok már *felszáradóban vannak*. (#158219760)
- (10) meglehetősen sok társasház *épülőben van* a fővárosban. (#176703591)

A mediopasszív, vagy legalábbis nem egyértelműen mediális szerkezetek esetében gyakori az *-ít* képzős kauzatív „pár” (vö. *épít, felszárít*), de ez nem kizárólagos (lásd *formálódóban*).

- (11) Mindenesetre működésük nyomán *formálódóban van* egy újfajta „nemzetközi büntetőjog”, amit jogászok szerint a jövő törvény-széknek kellene igazán kiteljesítenie. (#151537172)

5. Gondolatok a vizsgált konstrukciók közötti viszonyok értelmezéséhez

A vizsgálati célnak kiválóan megfelelő döntés az, hogy a Szerzők az analitikus szerkezetek produktivitásának megítélésében viszonyítási pontnak a *V-ni fog* konstrukciót választották. Az analitikus igealakok segédigéinek a körében a *fog* a legerősebben grammatikalizálódott, ezzel összefüggésben a leginkább absztrakt kifejezések közé tartozik, a folyamatnak az időviszonyok kifejezése szempontjából alakilag jelöltebb megoldását képviseli (a szintetikus igealakok alaki jelelőtlenségével összevetve, például *megyek* mind a diskurzus idejével érintkező, mind jövőbeli cselekvés kifejezésében). Mivel a *fog* segédige az igei deiktikusidő-jelölésben, az episztemikus lehorgonyzásban funkcionál, a *V-ni fog* konstrukció nem válogat a *V* komponens igenemének a vonatkozásában (aktív: *építeni fog*; mediális: *el fog keseredni*, *le fog esni* stb.).

Valószínűsíthetően a *V-tatik/-tetik* ritkaságával és a passzív ige infinitivusi alakjában a műveltetővel való egybeesésével (vö. *ki fog takarítottatni az üzem/igazgató*) függ össze az, hogy passzív igével nem (vagy kevésbé) várható az analitikus jövő idő kifejezése. A leggyakoribb, nem teljesen lexikalizálódott *V-tatik/-tetik* típus („*[ae]tik” kereséssel; a (meg)születik kizárásával, igekötő(kapcsolás) nélkül) az *adatik* (8834) és az *íratik* (5947). A passzív *adatni fog* (az infinitivusra irányuló lekérdezés után -2 +2 ablakban *fog* igelemmára szűrve) összesen 10 példánnyal csak funkcióigés szerkezetekben (*tudtára fog adatni* #188955485) és régi (például tudományos, szépirodalmi) szövegek közlésében (25 *p.ftnyi jutalom fog adatni* #243318980) fordult elő, egy példány az erősen egységként viselkedő *meg fog adatni* (#863546868) volt. Volt egy további találat homonim funkcióigés műveltető szerkezetre is (*valakinek a népi demokráciából díjat fog adatni* #295210971). Az *íratni fog* esetében csak műveltető szerkezetre volt találat (2 előfordulás duplum nélkül, például *mással fog íratni* [sajtóközleményt] #280913782). Ebből már lehet óvatosan következtetni arra, hogy a még kevésbé gyakori *V-tatik/-tetik* típusigék analitikus jövő idejű alakjának a megvalósulására kicsi az esély. Ez pedig jelezhet egy olyan összefüggést is, hogy mivel nem várható a *dolgozat fog íratni*, szemben az adatható *dolgozatot fog íratni* szerkezettel, mert az alaki egybeesés nem kedvez a szintetikus passzív ige analitikus jövő idejű, az ige-névi homonímia miatt nem egyértelműen passzív olvasatú realizációjának, ezt a funkciót is a *V-va/ve lesz* tudja átvenni.

Korpusz

MNSZ2 = Magyar Nemzeti Szövegtár v2.0.5 (<http://corpus.nytud.hu/mnsz/>)
(Lásd Oravecz–Váradi–Sass 2014.)

Irodalom

- Hopper, Paul J. – Thompson, Sandra A. 1980. Transitivity in grammar and discourse. *Language* 2: 251–299. <https://doi.org/10.1353/lan.1980.0017>
- Imrényi András 2017. Az elemi mondat viszonyhálózata. In: Tolcsvai Nagy Gábor (szerk.): *Nyelvtan*. Budapest: Osiris Kiadó. 664–760.
- Lengyel Klára 2000. *Az igenevek helye a szófaji rendszerben*. Nyelvtudományi Értekezések 146. Budapest: Akadémiai Kiadó.
- Oravecz, Csaba – Váradi, Tamás – Sass, Bálint 2014. The Hungarian Gigaword Corpus. In: *Proceedings of LREC 2014*.
- Palágyi László 2022. *Morfológia az egésztől a részekig. A magyar főnév esete*. Budapest: Szépirodalmi Figyelő Alapítvány.
- Tolcsvai Nagy Gábor 2017. Jelentéstan. In: Tolcsvai Nagy Gábor (szerk.): *Nyelvtan*. Budapest: Osiris Kiadó. 207–499.

Kugler Nóra

Eötvös Loránd Tudományegyetem

kugler.nora@btk.elte.hu

ORCID: 0000-0002-8223-2926

PALÁGYI LÁSZLÓ

**Hozzászólás A magmondat szerepű, passzív jelentésű
sémák produktivitásának kvantitatív vizsgálata című
tanulmányhoz**

A tanulmány négy passzív jelentésű konstrukciós séma adatorientált, kvalitatív szempontokat is érvényesítő vizsgálatára vállalkozik. A vizsgált jelenség önmagában is alapvető nyelvészeti kategóriákat relativizál, illetve tesz folyamatos újragondolás tárgyává. Egyfelől az aktív és passzív konstruálást hagyományosan az inflexiós dimenziók közé sorolják, ugyanakkor az ún. *voice* mégsem sorolható feltétlenül a leginkább általános inflexiós jelentések közé a világ nyelveiben (Haspelmath 2023). Másfelől a magyarban nincs olyan „uralkodó” vagy „alapbeállítású” szerkezet (legalábbis nyelvtanaink szerint), amely passzív jelentés kifejezésére szolgálna. A helyzetet tovább bonyolítja, hogy a passzív jelentést a magyarban szó szintű (morfológiai jellegű) és szónál nagyobb (szintaktikai jellegű) sémák is képviselik. A tanulmány a használatalapú nyelvelméletek kontinuum-elvének értelmében nem tesz alapvető különbséget a különböző nagyságú konstrukciók működésében, a négy passzív jelentésű sémát egységesen kezeli. A tanulmány egyik módszertani tétje, hogy az analitikusabb konstrukciók termékenysége vizsgálható-e adekvát módon – az eredetileg a morfológiai produktivitás megragadására kidolgozott – globális produktivitás indexével. Ennek eldöntésére a szerzők bevontak egy segédigés szerkezetet, a *fog V-ni* konstrukciót is a vizsgálatba. A tanulmány egyik fontos eredménye, hogy az analitikus és szintetikus sémák globális produktivitási indexe összehasonlítható, vagyis a grammatikai kontinuumon különböző helyet elfoglaló sémák termékenysége egységesen kezelhető. A másik kiemelendő eredmény, hogy a *V-va/-ve van* séma alkalmazási hatóköre a korábbi, elsősorban introspekcióra építő elemzések megállapításaihoz képest jóval szélesebbnek bizonyult, produktivitása összevethető a kontrollként vizsgált *fog V-ni* sémáéval. Noha a különböző passzív konstrukciók közötti szemantikai különbségek (elsősorban a passzív és mediális jelentés kérdése mentén) kirajzolódni látszanak,

egyetérttek a tanulmány összegzésében megfogalmazott felvetésével, miszerint azok feltérképezése már a kollostrukciós elemzés irányába mutat. További rész kérdésekkel kapcsolatos megjegyzéseimet és javaslataimat a szerzők beépítették a dolgozat végső változatába.

Megjegyzések, javaslatok

- Ugyan az idő tipikusan inflexiók jelentésként határozható meg a világ nyelveiben (Haspelmath 2023), a *fog V-ni* sémát képviselő analitikus szerkezetek inflexiók szóalakként való értelmezése nem magától értetődő. Egyfelől ez a megközelítés feltételezi a dolgozatban nem szereplő *morfológiai* szó fogalmát, amely általában a szó szintű és szónál nagyobb szerkezetek közötti átmenet problémájának kiküszöbölésére szolgál a szabályelvű leírásokban, másfelől a magyar szabályelvű leírások szerint a *fog* segédigés szerkezetek még csak nem is számítanak morfológiai szónak (Kiefer 2000) (szemben a múlt időben és feltételes módban megjelenő *volna* segédigés szerkezetekkel). Ezzel összefüggésben a 155. oldalon olvasható a következő: „A passzív diatézis vizsgált kifejezőeszközeinek az inflexiók vagy szóképzési elemek közé való besorolása a konstrukciós nyelvészeti háttérnek megfelelően nem tartozik a jelen tanulmány releváns kérdésfelvetései közé.” Noha a tanulmány az inflexió és a deriváció bináris értelmezését elutasítja, a kérdés az eredmények értelmezésekor áttételesen mégis visszaszüremkedik a gondolatmenetbe, amikor fölvetődik, hogy a *V-va/-ve van* séma magas produktivitása folytán kezelhető volna az igei paradigma részeként. Mivel az inflexiók megközelítés adekvátsága már a *fog V-ni* séma kapcsán sem egyértelmű, javaslom ennek a kérdéskörnek az elhagyását vagy további árnyalását.
- Nem egészen világos a tanulmány álláspontja a magas gyakoriság és a produktivitás összefüggésével kapcsolatban. A tanulmány leginkább Baayen azon megközelítését látszik elfogadni, hogy a magas tokengyakoriság nem zárja ki az elem hozzájárulását a séma produktivitásához. Véleményem szerint nem Bybee elképzelése leegyszerűsítő, hanem annak idézett interpretációja. Bybee nem egy törvényre, hanem tendenciára mutat rá, amely diakronia és szinkronia egységében, nyelvi működés és változás viszonylatában egyaránt értelmezhető, és az természetesen nem zárható ki, hogy gyakori szavak is transzparenssek legyenek (vagy éppen mintaként szolgáljanak új sémák kialakulásához). Baayen

pszicholingvisztikai párhuzama, miszerint „a ritka komplex szavak feldolgozása lassabb, mivel feltehetőleg a szabály aktiválásával férünk hozzá a szó jelentéséhez, míg a gyakoribb komplex szavak feldolgozása gyorsabb, mivel ezek erős önálló lexikai reprezentációval rendelkeznek” (180. oldal), ugyancsak Bybee megközelítését támasztja alá, hacsak nem feltételezzük, hogy az erős önálló(!) reprezentáció egyúttal a szabály/séma gyors aktivációjával is együtt jár.

- A dolgozatban az olvasható, hogy a „V-(t)atik/-(t)etik sémában használt 10 leggyakoribb lemmából 7 nem passzív, hanem mediális jelentésű (például *születik*, *viseltetik* ’valamilyen érzülettel van’, *méltóztatik*), ez éles kontrasztot mutat a hapaxok által kirajzolt rendszeres sémajelentéssel, mely 95%-ban passzív. Ez a jelenség arra utal, hogy a Baayen által említett tendencia – miszerint a magas gyakoriságú, szabályos képzésű elemek erősítik a séma produktivitását (vö. Baayen 1992: 140, lásd 2.) – „megfordítása” is igaz lehet, tehát látható, hogy a leggyakoribb elemek nem transzparenssek, és ezzel párhuzamosan megállapítható a konstrukció alacsony globális produktivitása.” Bár feltételezhető összefüggés a sémát képviselő elemek lexikalizálódása és a séma produktivitásának visszaszorulása között, nem világos, hogy az idézett rész is erre utal-e. Ahogy a tanulmányban is többször említésre kerül, a leggyakoribb elemek a produktív sémák esetében sem feltétlenül transzparenssek, vagyis a fenti „párhuzam” mibenléte homályos.
- A szerzők álláspontja szerint a passzív és mediális igék elhatárolásában a szándékosság/szándékolatlanság tesztelése „metodológiaiilag nem megoldható”. Ezzel egyet is értek, amennyiben a szerzők úgy vélik, hogy a teszt nem vihető végbe korpusz alapon. Azonban a korpuszadatok alapján mintha magát a módszert is megpróbálnák dekonstruálni, miközben azok interpretálhatóak a teszt eredményessége szempontjából is. Véleményem szerint éppen azért (nem pedig annak ellenére) nem ad a *szándékosan reggelizett* szókapcsolatra egyetlen találatot sem a google keresője, mert a reggelizés tipikusan intencionális cselekvés: nem szükséges az intencionalitás alszerkezetét kidolgozni annak megértéséhez. A *szándékosan megrongálódott* típusú kifejezésekben pedig irónia, nagy mértékű metapragmatikai tudatosság fedezhető föl, a konceptualizálók ezekben a kifejezésekben kétféle (passzív és aktív) perspektívát is hozzáférhetővé tesznek, és nagyon is tisztában vannak a passzív jelentésű konstrukció és a *szándékosan* határozó inkongruenciájával.

Irodalom

- Haspelmath, Martin 2023. Inflection and derivation as traditional comparative concepts. *Linguistics* 62: 43–77. <https://doi.org/10.1515/ling-2022-0086>
- Kiefer Ferenc 2000. *A ragozás*. In: Kiefer Ferenc szerk. *Strukturális magyar nyelvtan 3. Morfológia*. Akadémiai Kiadó, Budapest. 569–618.

Palágyi László

Eötvös Loránd Tudományegyetem

palagyi.laszlo@btk.elte.hu

ORCID: 0000-0002-1437-4746

A tagmondatkapcsolattól a komplex birtokos szerkezetig



BAJZÁT TÍMEA BORBÁLA – SZEMES BOTOND –
SZLÁVICH ESZTER

**A tagmondatkapcsolattól a komplex birtokos
szerkezetig: hasonlattípusok azonosítása és vizsgálata**

Kivonat

A kutatás három célt tűz ki maga elé: egyrészt a különböző típusú hasonlatok számítógépes azonosítását, másrészt ezeknek a hasonlattípusoknak a kognitív jelentéstani elemzését, harmadrészt kvantitatív vizsgálatukat a kanonikus magyar regényirodalomban. A bemutatott konstrukciók eltérő mértékben okoznak nehézséget a formai elkülönítésben, ezért egy több lépcsőből álló, reguláris kifejezéseken alapuló, valamint a szerkezetek komponenseinek morfológiai és szemantikai tulajdonságait is figyelembe vevő módszer kidolgozásával kívántunk eljutni a célkitűzésünk megvalósításához. A lekérdezéseket 100 kanonikus magyar regényen végeztük el, ez összesen 173 évet ölelt fel, ami átlagosan 1,7 évenként egy regényt jelent. Mindemellett a sűrített hasonlatok azonosítása lehetővé teszi kvantitatív vizsgálatukat is a korpuszban, valamint a hasonlító összetett mondatokra vonatkozó gyakoriségelemzés eredményeinek pontosítását.

Kulcsszavak: korpusznyelvészet, stilometria, hasonlatok, konstrukciók, irodalomtörténet

1. Bevezetés

A jelen tanulmány célkitűzése összetett. Egyfelől arra vállalkozik, hogy számot adjon a különböző hasonlattípusok számítógépes azonosításáról. Ahogyan a cím sejteti, nem csupán a két tagmondatból álló, formailag konvencionálisabb, „hagyományos” hasonlatok azonosítása a célunk, amelyek esetében a hasonlítást kidolgozó mellékmondat a főmondat elemi jelenetével (elemi jelenet vö. Imrényi 2017: 666) párhuzamba állítható (Kugler 2017: 850), hanem olyan „sűrített” szerkezeteknek is, amelyek névszókból álló szókapcsolatokként vagy csupán egyetlen szóalakként dolgozódnak ki a formai oldalon. Ez utóbbiakra az alábbi példákat említhetjük, egyelőre még konkrét elemzés és kategorizáció nélkül: „...*az először szerető sokbeszédűségével mondanám el egész boldogságomat*” (Eötvös József, A karthauzi), „*s ajkai e szót dörgék, mely mennyei zene gyanánt visszhangzott fülembe*” (Nagy Ignác, Magyar titkok).

Másfelől az azonosító metódus részletes ismertetése mellett szeretnénk felkínálni az általunk vizsgált konstrukciók (a konstrukció fogalmához vö. Goldberg 1995; Langacker 2005) jelentéshálózatának funkcionális szemléletű (Langacker 1987; Tolcsvai Nagy szerk. 2017) leírását a forma-jelentés pár felől a kontinuum-elv érvényesítésével. Emellett, harmadrészt, ismertetjük a kutatásnak az irodalom- és a stílustörténetre reflektáló eredményeit is, valamint kezdeményezzük a vizsgált szerkezetek kognitív stilisztikai leírását (Tátrai 2012), de nem ez utóbbi áll a tanulmány fókuszában. Az alább bemutatott módszerrel lehetőség nyílik arra, hogy a hasonlatkutatás és a kvantitatív alapokon nyugvó stílustörténeti leírások ne csupán a prototipikusnak tekintett, tagmondat által kidolgozott hasonlatokat vegyék figyelembe, hanem azokat más nyelvi adatokkal ütköztessék, illetve egészítsék ki. Másrészt a különböző típusú sűrített hasonlatok sajátos stíluspotenciállal rendelkeznek, ezért vizsgálatuk nagyban hozzájárulhat egy szöveg stílusának jellemzéséhez. Hiszen egyszerre jellemző ezekre a szerkezetekre a figurák részletezése és egyfajta tömörség, valamint e tömörségből adódó jelentéssűrítés. Vegyük például Konrád György *A városalapító* című regényéből az alábbi mondatot: „*Elmetszett falaival, üvegrepedésként szétrajzó utcáival kitárul az egész város.*” E szakasz prototipikus hasonlattá való átírásakor („*Elmetszett falai- val, szétrajzó utcáival, melyek olyanok voltak, mint az üvegrepedések, kitárul az egész város.*”) éppen az eredeti megfogalmazásmód sűrítettsége veszne el, amely szorosabb szemantikai kapcsolatba rendezi a konstrukció elemeit (az üvegrepedés, a szétrajzás és az utcák egysége), miközben mintegy

észrevétlenül simul bele a hasonlat a képileg meglehetősen gazdag, mégis egyszerű és gördülékeny mondatba. Ez a kettősség a sűrített hasonlatok stíluspotenciáljának az alapja: a jelentések egymáshoz való közelítése (ezt a mondatban való közelség ikonikusan is jelezi), ami a közöttük lévő viszony kidolgozása helyett a megfogalmazás képszerűségét állítja az előtérbe. A prototipikus hasonlatok és más alárendelő mellékmondatok ezzel szemben nem a jelentések „összedszolgáltatásért”, hanem a köztük lévő kapcsolat formailag jelöltebb kidolgozásáért felelősek. Még inkább érvényes ez a birtokos szerkezetekbe sűrített hasonlatokra:

- (1) Az először szerető sokbeszédűségével mondanám el egész boldogságomat.
- (2) Olyan sokbeszédűen mondtam el a boldogságomat, mint az először szerető.

Az (1) mondat átalakítható tagmondatokkal kidolgozott hasonlattá (2), ám igencsak körülményesen, a szerkezet tömörségének fellazításával és így a stíluspotenciáljának megváltozásával. Ezek a hasonlattípusok azon túl, hogy a befogadó részéről nehezebben feldolgozható és stílusosan erősen jelölt konstrukciót jelentenek (a jelöltséget a konstrukció ritkasága és szemantikai sűrűsége eredményezi), megengedik továbbá a jelzőhalmazt és az abból eredő stíluspotenciál megváltozását, amennyiben mind a birtok, mind a birtokos kaphat jelzői bővítményt – például a következő mondatban:

- (3) Papírköteget nyomott az orrom elé és a döllyfös szerények prófétikus büszkeségével mondta... (Határ Győző, Heliáne)

Az egyes típusokról és a rájuk jellemző formai és stílusos sajátosságokról az alábbiakban részletesebben is értekezünk. Kutatásunk további mozgatórugója, hogy olyan nyelvi jelenségeket tegyünk automatikusan felismerhetővé és leírhatóvá, illetve olyan jelenségekre hozzunk nyelvi adatokat a leírás számára, amelyekkel korábban elvétve foglalkozott a nyelvészet és a hazai stílus kutatás. Ám mivel a stílus kérdése „az elméletalkotás évtizedeinek nagy túlélőjeként” (Herrmann et al. 2015: 39) a digitális bölcsészet ösztönzésére ismét egyre inkább az irodalom- és nyelvtudományok érdeklődésének homlokterébe került (vö. a stilometria fogalmával Szemes–Byszuk 2021 alapján), úgy véljük, hogy módszereink a szűken vett nyelvészetén túl is fontos kvalitatív és kvantitatív vizsgálatokhoz járulhatnak hozzá. Kvalitatív szempontból

hasonló vállalkozásokként említhetők Simon Gábor közelmúltban publikált munkája a hasonlatkutatás terén (Simon 2022), valamint Herczeg Gyula stílustörténeti írásai a 20. század közepéről, amelyek alapvető inspirációforrásként szolgáltak tanulmányunk számára is (Herczeg 1958, 1981).

A vizsgálat során törekedtünk a korpuszalapúságra, azért, hogy nagyszámú nyelvi példányon azonosíthassunk mintázatok közötti összefüggéseket és tendenciákat (a korpuszunk felépítéséről lásd 4.2. alfejezet), valamint a korpuszalapúságot korpuszvezérelt eljárásokkal is kiegészítettük. A korpuszalapúság és a korpuszvezéreltség nem csupán abban valósult meg, hogy a kialakított protokollokat nagy mennyiségű, reprezentativitásra törekvő nyelvi anyagon alkalmaztuk, hanem abban is, hogy az eljárások kialakításánál minden egyes elemzői döntést az adatok elemzésekor összegyűjtött tapasztalataink alapján hoztunk meg a minták manuális ellenőrzése után.

2. A kiindulópont

2.1. A hasonlatok és a hasonlatkutatás

A hasonlatok nyelvészeti kiindulópontból történő megragadásához Kocsány Piroska (2008) által adott definícióból indultunk ki. Azért ezt a leírást választottuk, mert egyfelől kezdeményezést mutat a jelentéskonstruálás dinamikus természetének érvényesítése felé, másfelől tágan értelmezi a konstrukcionális jelentést, illetve részletesen szemléli a prototipikusnak tekintett több tagmondatból álló, *mint* kötőszóval jelöltté tett hasonlító szerkezettől eltérő példányokat a forma és jelentés oldaláról egyaránt: „Egy tárgy vagy egy tényállás megragadása (leírása, magyarázata, értékelése, hangulatának felidézése) egy másik tárgy vagy tényállás révén, a két tárgy/tényállás között felfedezett részleges azonosság (= hasonlóság), illetve analógia alapján. A hasonlat szintaktikai szerkezet, amelyben a hasonlított (az a tárgy vagy tényállás, amelyet megragadni kívánunk) és a hasonló (az a tárgy vagy tényállás, amely szemléletes, érzékletes magyarázatként kínálkozik) összefüggését grammatikai és lexikális elemek jelzik” (Kocsány 2008: 268, vö. még Nate 2006: 741). Azonban érdemes tovább árnyalnunk ezt az analógiából és a részleges azonosságból kiinduló elképzelést, hiszen feltehetőleg nem egyszerűen arról van szó, hogy egy entitást egy másik entitáson keresztül, a közöttük felfedezett valamilyen fokú részleges hasonlóságon alapulva teszünk hozzáférhetővé a befogadó számára, hanem a feldolgozott relációkon keresztül

lehetővé válik a közös figyelemirányítási aktus terében új tudást létrehozása is a két entitás reflexív viszonyában (Mpouli 2016: 45–46). Erre a jelenségre a tanulmány későbbi fejezeteiben is kitérünk, leghangsúlyosabban akkor, amikor a birtokos szerkezetbe sűrített hasonlatok konstrukciós jelentéseinek elmozdulását figyelhetjük meg a jelentéssűrítő funkció felé.

Fishelov (1993), illetve Simon (2022) külön kategóriaként kezeli az ún. poétikus hasonlatokat, amelyek eltérő működést mutatnak az alapnak tekintett nem-poétikus hasonlatoktól. Fishelov (1993: 4) hipotézise szerint a poétikus hasonlatok modellje levezethető a nem-poétikus (alacsonyabb poétikai potenciálítású) hasonlatok tipikus jegyeiből. Az ettől való elmozdulást a poétikus hasonlatok modelljének a részeként kezeljük.

A hasonlatkutatás hagyományában mind a hazai és mind a nemzetközi szinten a tárgyalt konstrukcionális sémákat négy komponensre tagolják, amelyek eltérő mértékben kaphatnak nyelvi kidolgozást, a kidolgozási szintek feltételezeten összefüggésben állnak a nem-poétikus és poétikus tengelyen való elmozdulással. Másfelől a konstruálás nem csupán a kidolgozottság fokából kiindulva válhat tematizálhatóvá, hanem például a sorrendiség felől megragadva is. Az általánosságban tárgyalt részei a tipikus hasonlatoknak: a hasonlított, a hasonló (vehikulum), a hasonlítás alapja (ground, tehát az a konceptuális összetevő, amely a hasonló és a hasonlított közötti hasonlító mentális műveletet aktiválja a feldolgozásban) és a hasonlító funkció nyelvi kidolgozója (indikátor/modalizátor vö. Simon 2022: 146). Fishelov nyomán a poétikus és nem-poétikus hasonlatok közötti eltérés alapja, hogy a tipikus, kevésbé poétikus hasonlatok esetében a hasonlított fogalmi entitás az, amely mentálisan nehezen hozzáférhető, s ezt a fajta kognitív elérést teszi lehetővé a hasonlóval való együttes előfordulás (Fishelov 1993: 5; vö. irányulási elv Shen 2008: 296–297). A nem-poétikus hasonlatokban továbbá a hosszabban kidolgozott hasonlított megelőzi a nyelvileg rövidebben kidolgozott hasonlót, valamint a hasonlat konstrukciós megvalósulásában a tagok standardizáltak, konvencionális együttállásokat alkotnak, amely relációk a feldolgozás mentális oldaláról tekintve egyszerűbbek, jól hozzáférhetők. A kötőszó kidolgozottsága szintén jellemzője a nem-poétikus hasonlatoknak.

A modalizátor nyelvi kidolgozottságának változatosságát a szakirodalom (Fishelov 1993: 5, Mpouli 2017, Simon 2022: 147) is tematizálja mint potenciális kutatási irányt a poétikussági potenciál vizsgálatában. A jelen tanulmány is amellet érvel, hogy a konstruálás, a modalizátor nyelvi kidolgozása közvetlen kapcsolatban áll a megvalósuló példány poétikai és stilisztikai potenciálításával és funkcionalitásával (lásd (4a)–(4b)).

- (4) a. Úgy libbent be a terembe, mint egy kecses balettáncos.
b. A balettáncosok kecsességével libbent be a terembe.

A poétikusság tengelyén való elmozdulás nem feltétlen következménye a tipikus konstruálási módoktól való eltérésnek. A nem-poétikus hasonlatok sorrendiségére jellemző a hasonlított, hasonlítási alap, modalizátor és hasonló mintázat (Fishelov 1993: 5), amely egyértelműen asszociálható a FORRÁS-ÖSVÉNY-CÉL sémaként (Tolcsvai Nagy 2011: 50–51) modellálható mentális elérési útvonallal (Simon 2022: 148). Shen (2008) szerint a kognitíven szaliensebb, konkrétabb, jobban elérhető fogalom képezi a hasonlítás forrását, alapját. A hasonlítási alap (ground) implicitsege továbbá szintén egyfajta lehetőséget teremt a poétikus konstruálási mód érvényesítésére, ezekben az esetekben „nyílt” hasonlatokról beszélhetünk (Fishelov 1993: 6, Mpouli 2017). A kifejtetlenség azonban nem feltétlen jelenti azt, hogy a feldolgozás akadályoztatva lenne az észlelés felől, hiszen a konvencionizálódott sémákat megvalósító példányokban a közös enciklopédikus tudás aktiválódik/aktiválódhat (Tolcsvai Nagy szerk. 2017: 30–31). A hasonlítás alapja a tipizálható mintákban egy szaliens tulajdonsága a vehikulumnak/hasonlónak (amely egyúttal könnyebben azonosítható, mint a hasonlított). A két entitás (hasonló és hasonlított) közötti reláció nem bizonyul ellentmondónak a nem-poétikus hasonlatok esetében. Fishelov a megfigyelései alapján arra a következtetésre jut, hogy a hasonlatok stilisztikai potencialitásának növekedése nem abból fakad, hogy a vizsgált példány minél több, a nem-poétikus hasonlatokra jellemző tulajdonságot (nála: szabályt) sért meg, hanem a konvencionizálódott konstruálási módok és az ettől való elmozdulások együttes, speciális mintázatokban való megvalósulásai válnak azonosíthatóvá (Fishelov 1993:14).

A hazai korpuszalapú, illetve korpuszvezérelt, (fél)automatizált NLP-eszközöket alkalmazó hasonlatkutatás területe még egyelőre kiaknázatlan. Nemzetközi viszonylatban azonban nem precedens nélküli a törekvés a hasonlatok korpuszalapú és korpuszvezérelt vizsgálatára. Kiemelhető Suzanne Mpouli munkássága, aki angol és francia nyelvre dolgozott ki félautomatizált annotációs eljárást a hasonlatok azonosítására (Mpouli 2016), amely protokoll képes megalapozottá tenni szintaktikai és szemantikai oldalról egyaránt a gépi feldolgozás felé való elmozdulást grammatikai és szemantikai jegyeket egyaránt érvényesítve (vö. Mpouli 2016: 82–90). Nem csupán a vizsgálatának célja egyezik az általunk kialakított protokoll célkitűzéseivel, hanem az elméleti kiinduló probléma is, miszerint önmagában a kötőszó – annak

kiterjedt poliszém hálózatából és multifunkcionalitásából adódóan – nem alkalmas arra, hogy a példányt további szűrések nélkül hasonlatként dektálhassuk. Ugyan számos nyelvspecifikus vonatkozást azonosíthatunk a francia *comme*, az angol *as* és *like* (Mpouli 2016: 54–55) és a magyar *mint* lexémák jelentéshálózatának összevetésében, azonban az kijelenthető, hogy a nyelvtechnológiai, illetve a korpuszalapú azonosítási eljárásban egyaránt számolnunk kell a fent körvonalazott problémával. Az általunk kidolgozott módszerek célja az, hogy viszonylag kevés technológiai és humán-erőforrás bevonásával tudjuk maximalizálni a korpuszokból kinyerhető hasonlatok számát oly módon, hogy redukáljuk a fals pozitív találatok begyűjtésének lehetőségét.

2.2. A kutatás elméleti háttere és fókusza

A kutatás nyelvelméleti háttere a használatalapú funkcionális kognitív konstrukciós nyelvtan, amelynek alapvetése, hogy a nyelvi sémák mögött nem a grammatikai kompetencia működtetése áll, hanem olyan általános képességek, melyek más kognitív folyamatokban is részt vesznek. Ebből a perspektívából szemlélve a nyelvtan egy olyan emergens jelenség, amely be van ágyazódva a percepció és a produkció műveleteinek kölcsönös dinamikus egymásra hatásába, s olyan műveletek alakítják, mint például a sematizáció, absztrakció, kategorizáció és analógia (Diessel 2015: 295–300; Hilpert 2014: 1–22).

A konstrukció, megfelelően a kutatás által érvényesített elméleti keretnek, olyan szimbolikus forma-jelentés pár, amelynek emergens jelentése nem vezethető le a lexikai és morfológiai komponensek jelentéséből. A konstrukcióra tehát egy olyan szerkezeti mintázatként tekintünk, amely egészen áll szoros kapcsolatban annak funkciójával, jelentésével (Goldberg 1995; Langacker 2005). A konstrukciós mintázatok sematikussága skalárisan értelmezhető (Brdar 2020; Traugott 2008). A sémák emergens nyelvi viselkedésére hatással van a konstruálás dinamikus természete, tehát újabb és újabb konstrukciós mintázatok jöhetnek létre a nyelvi változás során, illetve a sematikus mintázatok egymással vegyülhetnek, valamint újabb funkciókkal bővíthetnek. Egyes konstrukciós forma-jelentés párok grammatikálizálódhatnak, funkciójuk sematizálódhat ('bleaching' vö. Bybee 2010: 144, 167). A konstrukcióba integrálódó komponensek az adott konstrukció-típustól függően eltérő mértékben reprezentálódnak a tudásban (Langacker 2005; Diessel 2015: 300). Természetesen ez az integrálódás nem szabályalapú,

hanem sokkal inkább a prototipikalitási skálával és a flexibilis konvencionálódással áll összefüggésben, amelyek szorosan összekapcsolódnak a produkciós gyakorisággal (Bybee 2010: 9; 24–25; 32).

A konstrukciós jelentés a hagyományos értelemben vett lexikális jelentéssel analóg módon poliszém (a poliszémia fogalmához vö. Tolcsvai Nagy 2017: 268–288) hálózatokba rendeződik. A kutatásban kiindulópontként érvényesítettük azt az elvet, hogy az egymással szinonimnak tekinthető szerkezetek között funkcióbeli eltérések válnak észlelhetővé (Brdar 2020: 144), s ezek a funkcióeloszlási különbségek elsősorban a szándékosan vagy intuitívan kiaknázott figyelemirányításbeli és feldolgozási mintázatok felől, másrészt pedig a stíluspotenciál felől szemlélve vázolhatóak fel.

A jelen tanulmányban vizsgált konstrukciókban a közös az, hogy olyan figuratív nyelvhasználati sémákat valósítanak meg a példányaik, amelyek két entitás relációjában a hasonlítás műveletét dolgozzák ki. Nem vállalkozunk arra, hogy a görcső alá vett konstrukciók komplett jelentéshálózatának poliszémiáját felvázoljuk, csupán a hasonlítás műveletével asszociálódó hálózati csomópontok részletesebb elemzésére koncentráltunk a forma-jelentés pár oldaláról. A tanulmány a két tagmondatból álló, prototipikus hasonlító mintázatból indul ki.

- (5) a. Úgy szaladt el, mint ahogy egy róka menekül.
- b. Úgy szaladt el, akár a róka menekülés közben.

A vizsgált szerkezetben ((5a)–(5b)) megfigyelhető, hogy a hasonlítás művelete két tagmondat által nyelvileg kidolgozott elemi jelenet (elemi jelenet vö. Imrényi 2017: 391) között dolgozódik fel. A formai oldalon megfigyelhetjük, hogy a szerkesztésmód analitikusabb, a hasonló és a hasonlított közötti jelentésviszony nem a szóalakokon belül válik nyelvileg kidolgozottá, hanem modalizátorként egy kötőszó funkcionál. A nyelvi konstruálás ikonikus összefüggésben áll a mentális elérés ösvényével és a jelöltséggel. A fenti példamondatokban az analógia hosszabb szekvenciában, összetett egységben dolgozódik ki. A konstruáló szubjektum a figyelem előterébe helyezi magát a hasonlítást. Ez nem csupán az analitikus kidolgozási megoldásban figyelhető meg, hanem abban is, hogy a példányban a modalizátorok (*mint* és *akár* kötőszavak, úgy kontextualizáló tagmondatot bevezető utalószó) poliszém hálózatában elsődlegesként kategorizálhatjuk a hasonlító funkciót. Ezért a szerkezet morfoszintaktikai elaborációjában szemantikailag kevésbé tekinthető jelöltnek az általunk vizsgált konstrukciós sémák között (vö.

jelöltségi elmélet Ladányi 2017: 550–524). Az ebbe a típusba sorolható hasonlatfajták között elsősorban a kötőszó jelenti a különbséget. Ezek közül is kiemelkedik a *mintha* kötőszó, amely a feltételes jelentésáryalattal szubjektivizálja is a hasonlítás művetét, egyúttal diskurzusjelölőként is működhet.

A *-ként* esszívuszi esetragos szerkezet példányai képezték a második vizsgálati csoportot a kutatásunkban. A szerkezet konstrukciós mintázata a következő: N + *-ként*. Ezzel el is érkezünk a „sűritett”, névszókból álló szókapcsolatokként vagy csupán egyetlen szóalakként kidolgozódó hasonlító szerkezetekhez, amelyek a kutatásunk fókuszában állnak. Ezekre a formai oldalról általános jellemző a szintetikus szerkesztésmód felé való elmozdulás, tehát ilyenkor a modalizátor formailag és szemantikailag erősebb integrált viszonyt tükröz a konstrukció bizonyos alszerkezeteivel.

(6) Rókaként szaladt el a tett helyszínéről.

Ugyanakkor a *-ként* esszívuszi esetraggal létrehozott névszói konstruktumok feldolgozásánál kevésbé beszélhetünk arról, hogy a szerkezet szemantikailag feltűnően jelöltebb, hiszen a vizsgált morféma koncenvionális jelentése az azonosítás mellett magába foglalja a hasonlítást is. A hasonlítás-tól elkülönböző jelentés konstrukcionális elsáncolódása megfigyelhető a két tagmondatból álló hasonlat esetében is, azonban ott a szórenden és a prozódian keresztül jelölt a szemantikai elkülönbözés (lásd (7)).

(7) Mint tanárnő dolgozott az iskolában.

Ezzel szemben az N + *-ként* konstrukció megvalósulásainak funkcióeloszlásai kevésbé kapnak jelöltséget a formai oldalon (8a, b), amely a funkcionális oldalon a kategóriába való besorolást valósítja meg.

- (8) a. Tanárnőként dolgozott.
b. Rókaként szaladt el a mezőn.
c. Gyerekként örült a meglepetésnek.

A hasonlító és a besoroló/kategorizáló konstrukcionális jelentés közötti választás nem minden esetben egyértelműen eldönthető a példányok elemzésénél, hipotézisként felvázolható, hogy bizonyos példák feldolgozásánál számolnunk kell a szemantikai kétértelműséggel, vagy az előtér-háttér viszonyok profilálásokor létrejövő elmosódással (8c). A (8c) példamondat

viisszatükrözi azt, hogy az elemi mondat dekontextualizált feldolgozása nem támogatja egyértelműen sem a hasonlító, sem pedig az azonosító értelmezést, azaz a példamondat egyaránt jelentheti azt, hogy 'egy felnőtt gyermeki örömet mutatott', illetve azt is, hogy 'a gyermekből az életkorának megfelelő, tipikus örömet váltott ki a meglepetés'. A kutatás során a konstrukcióra kapott minták elemzésére alapozva arra a következtetésre jutottunk, hogy a kategorizáló/besoroló jelentés (8a) általában olyan esetekben valósul meg a hasonlítás helyett, amikor a ragozott főnév az alany egy szerepkörét, jellemzőjét ragadja meg; prototipikusan ilyenek a foglalkozás, életkor/életszakasz, nemiség kategóriái.

A harmadik csoportot az N + *-szerű* + *(-en)* képzőszerű utótagos konstrukció példányai képezték a kutatásban.

(9) Rókaszerű alak szaladt át a mezőn.

A *-szerű* utótag egyfajta „hasonlatzáslóként” jelöli azt (vö. *metaphor flag* Steen et al. 2010: 26; *metaforazászló* Simon et al. 2019: 240), hogy az összetétel előtagjához hasonlítunk egy másik entitást valamilyen felfedezett analógia alapján. A konstrukció példányai hajlamosak a lexikalizációra, tehát a hasonlító funkciójukban elhomályosult példányok szűrése volt kihívás a vizsgálatban (például észszerű, újszerű). A forma felől felfedezhetők kidolgozásbeli tendenciák a lexikalizációra erősebb potenciállal bíró szerkezettagokra vonatkozóan (lásd 3.3. szakasz).

A negyedik vizsgált szerkezet az N + *gyanánt* névutós konstrukciós mintázat volt. Ugyan itt maga a modalizátor írásban elkülönül a névszói szerkezettagtól, azonban funkcióját tekintve nem elválasztható az esetragos szerkesztésmódtól (Tolcsvai Nagy 2017: 378). „A névutó egy dolgot egy másik dologhoz viszonyít szemantikailag, azaz két dolog (jellegzetesen két fizikai tárgy) egymáshoz való viszonyát fejezi ki, az egyiket előtérbe állítva.” (Tolcsvai Nagy 2017: 378).

(10) a. Róka gyanánt szaladt át a mezőn.

b. Kalapács gyanánt alkalmazta a csavarhúzó.

Az N + *gyanánt* konstrukció szintén poliszém jelentéshálózattal rendelkezik. Az azonosítás problémáiról lásd részletesebben a 3.4. szakaszt. Hasonlóan az N + *-ként* szerkezeti mintázathoz a példányok által aktivált jelentések a feldolgozásnál nem egyértelműen kategorizálhatók a funkciómegoszlást

tekintve, mivel a hasonlításon túl észlelhetünk olyan jelentésárnyalatot, amely már nem két entitás egymással analógia alapján történő párhuzamba állításán alapul, hanem 'valami helyett, valamilyen funkcióban alkalmazott' értelmezéssel, azaz azonosító jelentéssel társul a (8a) mondat mintájára. Ilyen például az alábbi mondat:

- (11) Elköszönés gyanánt kezet nyújtott.

A vizsgált ötödik konstrukciót birtokos szerkezetbe sűrített hasonlatnak neveztük el (Herczeg 1958: 313 alapján), ezzel a terminussal fogunk hivatkozni rá; a konstrukció felépítése pedig: DET – NGEN – NINS – V, azaz egy névelő; egy (morfológiailag jelöletlen) birtokos esetű főnév (genitívus); 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*.

- (12) Az előkelők magabiztosságával állt meg.
 (13) A színpadra belibbenő balett-táncosok üde kecsességével kínálta helyet az idős hölgyet.

A szerkezetet azért tartjuk sűrített hasonlatnak, mert a legtöbb esetben parafrázálhatók hasonlító összetett mondatokkal: *olyan magabiztosan állt meg, mintha előkelő lenne / olyan magabiztosan állt meg, mint egy előkelő; olyan finom eleganciával lépett be, mintha táncosnő lenne / olyan finom eleganciával lépett be, mint egy táncosnő*. Feltételezésünk szerint a hasonlító funkcióval asszociálódó szerkezetek mentális hálózatában ezen konstrukciós séma az, amely a szemantikai jelöltség¹ kontinuumán a jelölt mintázattípusok felé a leg-radikálisabban elmozdul. A szerkezetben működtetett referenciapont-szerkezet (Tolcsvai Nagy 2017: 250) az entitások viszonyában az elérési útvonalat birtokos szerkezeten keresztül nyitja meg. A birtokos szerkezetből jelentéskiterjesztéssel absztrahálódott konstrukciós jelentés feldolgozása vélhetően nagyobb mentális erőfeszítéssel jár, hiszen itt a szerkezeten belül korlátlanul közbeékelődhetnek részletező funkciót betöltő jelzők, emellett az analógia komplex morfológiai mintázaton keresztül aktiválódik. Az ebben a típusban

¹ A tanulmány az általa alkalmazott szemantikai jelöltség fogalom alatt azt érti, hogy a szemantikailag kevésbé jelölt kifejezések könnyebben, gyorsabban feldolgozhatók a kommunikációs szituációban, míg a szemantikailag jelöltebb kifejezések konceptuálisan nehezebben érhetők el. A szemantikai jelöltség skaláris jelenség (Ladányi 2017: 520–521).

megjelenő *előkelő és magabiztos*), vagy a köztük lévő hasonlóság felmutatását végzi el, és nem prototipikus birtokviszonyt dolgoz ki. Esetenként nem is magától értetődő a birtokviszonyban rejlő egyéb relációk felfejtése: a (13) mondat kétféleképpen is parafrázálható, aszerint, hogy a *kecsesség* főnevet jelzőnek, vagy a hasonlítás alapjának tesszük-e meg: „úgy kínálta helytel az idős hölgyet, mint egy kecses balett-táncos”; vagy „*olyan kecsesen kínálta helytel az idős hölgyet, mint egy balett-táncos.*” Ugyanakkor nem állítható, hogy ezek az átirat változatok teljesen megfeleltethetők volnának a birtokos konstrukcióknak. Herczeg Gyula meglátását idézve: „Egy *lepedő hideg fehérsége* más és több, mint: egy *hideg, fehér lepedő*, a főnevesítés a szokványos jelzős szerkezet egyhangúságát megbontja. *Fehérség*, ill. a *lepedő fehérsége* kifejezőbb, szemléletesebb, mint a közbeszédben bármikor előforduló melléknévi jelző + főnév. Ugyanakkor a *fehérség* elé tett *hideg* melléknévi jelző jelentése is módosulhatott. *Hideg fehér lepedő* esetében a két melléknévi jelző a főnév két konkrét tulajdonságát jelöli, a *hideg fehérség* azonban olyan kapcsolatot, amelyben a *hideg* átvitt értelemben, nagyobb szemléletességgel áll, s ’érzéketlen, komor, lelketlen’ stb. jelentésárnyalattal egészül ki.” (Herczeg 1974: 173)

Végezetül a konstrukció stíluspotenciáljához tartozik, hogy a hasonlító elem határozott főnévként szerepel szemben a prototipikus hasonlatokkal (*a balett-táncos* vs. *egy balett-táncos*), ezáltal egy, a befogadók számára közös, aktívan jelenlévő tudásként lépnek fel a kommunikáció során – mind tudjuk, milyenek a balett-táncosok. Ennek köszönhető, hogy a hasonlat során kiemelt tulajdonságot (*kecsesség*) a hasonlított tipikus tulajdonságaként ismerjük fel.

3. A módszertan és a kutatói korpusz

3.1. A kutatás technikai apparátusa

Ahogy fent megfogalmaztuk, olyan keresési metódus létrehozása volt a célunk, amely lehetővé teszi a hasonlatként funkcionáló szerkezetek automatikus, illetve félautomatikus azonosítását. Mivel a vizsgált formai megvalósulások gazdag poliszém hálózattal rendelkeznek, ezért az automatizált szűrés során elkerülhetetlen, hogy olyan példányok is bekerüljenek a mintáinkba, amelyek nem az általunk keresett funkciót valósítják meg. Minden egyes módosítás után manuálisan elemeztük újra a mintáinkat, azért, hogy az eljárások hatékonyságát tesztelhesük.

A szűrések elvégzésénél első lépésben nagy bizalmat fektettünk a reguláris kifejezésekkel történő lekérdezésekbe. A reguláris kifejezés (*regular expression, regex*) működésének bemutatására nem vállalkozunk. Az elmondható, hogy a regex szintaxis megértése nem kívánja meg a magas szintű programozási ismereteket, viszonylag könnyen elsajátítható eszközként áll rendelkezésünkre a nyelvi mintázatazonosításban. Természetesen az adott programkörnyezet meghatározhat specifikus szintaktikai szabályokat, de a reguláris kifejezés lényegében egy olyan karakterekből és logikai kifejezésekből álló szekvencia, amely illeszkedik az általunk keresni kívánt nyelvi mintázatra, viszont azokat a további mintázatokat nem fedi le, amelyeket nem kívánunk a szűrés során kinyerni. Előnye az egyszerűségben rejlik. A logikai feltételek megadásával tudjuk a szabályainkat ráilleszteni az adott karakterszekvenciára, ezáltal képesek vagyunk a magasabb fokú absztrakció elérésére, tehát arra, hogy a konkrét példányok formai jegyeiből következtetve hozzunk létre olyan generáló szabályokat, amelyek a kidolgozott séma alapján a kategória többi példányát is megtalálják a szövegtörzsünkben. Az egyszerű szintaxison túl nagy előnye, hogy bármely programkörnyezetben működtethető (némi változtatásokkal), ezzel összefüggésben például a szövegszerkesztő programok is képesek a regexeken alapuló szűrések lefuttatására, és azok alapján történő szövegmanipulációra. A reguláris kifejezések alkalmazásával kíséreltük meg azt, hogy viszonylag nagy elemszámú mintákat hozzunk a vizsgált konstrukcióra vonatkozóan, s az ezekből nyert konzekvenciák integrálhatóak legyenek a módszertani döntéseinkbe.

A kereséseinket a rendszerfüggetlen R-studio programkörnyezetben készítettük el (RStudio Team 2020), de a későbbiekben opcionálisan felmerülhet a kódok Pythonba történő implementációja is. A kódjaink nyílt hozzáférések, a lábjegyzetbe beillesztett GitHub repozitóriumban érhetőek el.² A reguláris kifejezésekkel történő keresések az e-magyar (Indig et al. 2019; Simon et al. 2020) automatizált nyelvi elemzőlánccal feldolgozott szövegeken futnak. A grammatikai tulajdonságok figyelembevételével tudjuk csökkenteni a fals pozitív találatok számát. Az emtsv-vel való feldolgozás kimenete alapján a szövegeket TSV formátumban kezeljük, amelynek az oszlopai a következők: 1. a szövegben előforduló szóalakok, 2. az előforduló szóalakok lemmái, 3. szófajcímkek és 4. morfológiai annotáció (Novák et al. 2016).

Voltak szerkezetek, amelyek szűrésénél bebizonyosodott számunkra, hogy az MNSZ2-ből nyert gyakorisági listák alapján összeállított stop-szó listák

² <https://github.com/SzemesBotond/s-r-tett-hasonlatok>

alkalmazása a félautomatizációs eljárás megbízhatóságát nagymértékben növelheti. A stop-szó lista olyan sztringekből összeállított halmaz, amelyeket kizártunk a keresésből. A stop-szó listák összeállítása azért volt kulcsfontosságú, mert ezzel a viszonylag egyszerű eljárással képesek vagyunk a fals pozitív találatok mintába kerülésének esélyét csökkenteni, hiszen negatív-szűrőként fel tudjuk venni azon kifejezéseket, amelyek olyan lexikalizálódott komponensek, melyek az éppen vizsgált modalizátorral együttesen előfordulva nem a hasonlatok által kidolgozott szemantikai műveletet valósítják meg (például *hívőként*, *adjunktusként*, *öszöndíjasként*). A stop-szó listák szerkezetspecifikus összeállításáról a 3. szakaszban írunk a típusok szemléltetésénél – a konkrét listák elérhetők az említett GitHub repozitóriumban. Összefoglalóan elmondhatjuk, hogy a stop-szó listáink összeállításánál a Magyar Nemzeti Szövegtár 2. (v2.0.5, Oravecz–Váradi–Sass 2014) gyakorisági listáit vettük alapul. Az MNSZ2-ből letöltött gyakorisági listákat manuálisan elemeztük, a konkrét problémához összeállított szempontrendszer szerint. Gyakoriság szempontjából 50 előfordulás volt az a küszöbérték, amely alá nem mentünk a lista elemeinek elemzésében. A gyakoriságon alapuló stop-szó lista alkalmazása egy olyan eljárás, amely viszonylag kevés erőforrást von be, de a hozadéka elvitathatatlan a hatékonyság növelésében, mindamellet, hogy belátható a limitációja. Mivel a konstrukciós jelentésekkel asszociálódó konkrét lexikai kifejezések az ilyen listákban jól detektálhatóak, hozzáférhetővé válnak azok az elemek, amelyek az adott konstrukciós helyen megjelenve önmagukban is radikálisan csökkentik a valószínűségét a hasonlítás funkció megvalósulásának.

Minden módszertani változtatás után kézi annotációt végeztünk a kvalitatív, exploratív elemzéssel, amely során egyrészt ellenőriztük, hogy a ki nyert példányok valóban funkcionálisan hasonlatként működnek-e, illetve milyen formai és szemantikai jegyeik azok, amelyek támogatják a hasonlatként való értelmezést, s hogyan lehet elkülöníteni őket a fals pozitív találatoktól a formai oldalon. Ennek köszönhetően az automatizációra törekvés kéz a kézben járt a finomabb, nyelvészeti elemzés által inspirált módszertani igényekkel. Minden mintát hárman elemeztünk, egymástól függetlenül, majd az eredményeinket egymáséval összevetve alakítottuk tovább a módszerünket konszenzusosan. Az így lefolytatott elemzői feldolgozás lehetővé tette azt is, hogy a forma-funkció oldaláról összefüggéseket tárhassunk fel az egyes hasonlító, vagy a hasonlító funkcióval érintkező konstrukciós funkciók között.

3.2. A kutatói korpusz

A kutatás korpuszaként szolgáló nyelvi anyag 100 kanonikus magyar regény szövegét tartalmazza, 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 tagmondatokkal megvalósuló hasonlatok gyakorisági vizsgálatát is elvégezte a korpuszban található regényeken. Az, hogy a mondatstilisztikai vizsgálatok során a korpusz összetételét nem változtatjuk, azzal indokolható, hogy így a kutatások eredményei egymással jól összevethetők, s ez további adalékokat szolgáltathat a magyar regény stílustörténetének számítógépes stilisztikai feltárásához.

A korpuszba beválogatott legkorábbi szövegek a magyar regényirodalom kezdetével kapcsolódnak össze, azaz az 1830-as évekkel (Szilágyi–Vaderna 2010: 370). A végpontot Nádas 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 napjaink magyar irodalmából kiválasztható. Azért esett a választásunk a kanonikus művek vizsgálatára,³ mert a 19. század első feléből ezek érhetők el a legnagyobb számban digitalizált formában. Ugyanígy az ebből a korszakból elérhető alkotók száma szabta meg azt is, hogy a 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 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

³ A kanonicitás kritériumát az ELTeC előírása szerint határoztuk meg (Distant Reading for European Literary History COST Action CA16204 vö. Schöch et al. 2020), amely így megfeleltethető az ELTE Regénykorpusz által használt kanonikusság-értelmezéssel. Eszerint azok a művek számítanak kanonikusnak, amelyek 1979 után minimum két újrakiadásban jelentek meg (Bajzát et al. 2021). A szövegek összetételét az irodalomtörténeti konszenzuson túl a Magyar Elektronikus Könyvtár (Magyar Elektronikus Könyvtár - Magyar Elektronikus Könyvtár (oszk.hu)) és a Digitális Irodalmi Akadémia (DIA | Digitális Irodalmi Akadémia) adatbázisai határozták meg, valamint az Akadémia Kiadó Magyar irodalom című kézikönyvéhez (Gintli szerk. 2010) és a magyar prózastílus történetét tárgyaló korábbi monográfiákhoz (elsősorban: Herczeg 1981) fordultunk.

ezáltal vizsgálható legyen az is, hogy az egyes tendenciákat követi-e az adott alkotó életműve, vagy inkább időszaktól független szerzői „ujjlenyomatokról” (Baayen 2002) beszélhetünk.

4. A hasonlattípusok elkülönítése

A hasonlattípusokat többféle szempontból osztályozhatjuk: a hasonlító és a hasonlított nyelvi kidolgozóinak egymáshoz képesti távolsága (egy vagy több tagmondat), a hasonlítás nyelvi eszköze (morfémával, lexémával, vagy konstrukciós sémával történő megvalósulás) is kiindulópont lehet az elkülönítés során. A két tagmondatból álló, hasonlító összetett mondatot tartjuk a leginkább begyakorlott típusnak, az egy tagmondatból álló hasonlatokat sűrített hasonlatoknak nevezzük. A hasonlatok elkülönítése formai szempontok alapján történik, de a különböző típusokról megállapíthatók stilisztikai jellemzők is.

4.1. A két tagmondatból álló, prototipikus hasonlatok szűrési módja

Kutatásunk során Szemes (2021) eljárását vettük alapul, amely a nyelvileg kidolgozott tagmondatkapcsolatok gyakoriságvizsgálatából jutott irodalomtörténeti következtetésekre. Az eljárás a különböző típusú kapcsolatokat kötőszavak/vonatkozó névmások és mondatközi írásjelek pozíciója alapján azonosította reguláris kifejezések segítségével. A reguláris kifejezéseknél az alábbi jellemzők meghatározása vált kiemelten fontossá: (1) a kötőszavak/névmások pozíciója, azaz az írásjeltől való lehetséges távolsága; (2) a poliszém kötőszók funkcióeloszlása és prototipikus jelentése, valamint (3) az adott kötőszó prototipikusan tagmondatok vagy mondatrészek között létesít-e inkább kapcsolatot. Ezeknek a kérdéseknek a megválaszolásához jelen tanulmány szerzői 100 mondatos mintákat elemeztek a korpuszból mindegyik kötőszó (*mint*, *mintha*, *akár*, *akárha*, *akárcsak*, *minthogyha*) esetében manuálisan. Az elemzés a hasonlító alárendelés (azaz a prototipikusnak tekintett hasonlatok) esetében az alábbi szabályok megalkotásához vezetett.

Hasonlító alárendelést a mondatközi írásjelet közvetlenül, vagy 1 szó után követő, valamint a mondat elején vagy az első szó után álló *mint*, *mintha*, *akár*, *akárha*, *akárcsak*, *minthogyha* kötőszavak valamelyike dolgoz ki. Kivételt képeznek azok az esetek, amikor a *nem* vagy az *a* előzi meg a *mint*

kötőszót, mert az előbbi esetben a *nem mint* + *főnév* szerkezetek, utóbbiban pedig a régiesen használt *amint* (*a mint*) megvalósulása a jellemző.

Az *akár* hasonlító jelentése csak akkor van túlsúlyban, ha nem a választó *akár...*, *akár...* sémájú konstrukcióban szerepel, így ennél a kötőszónál előzetesen biztosítani kell, hogy az *akár* egyszer szerepeljen a mondatban – ezáltal 82-szer hasonlító kötőszóként és 18-szor partikulaként volt azonosítható a 100-as példányszámú mintában. Ezt a kritériumot a reguláris kifejezésben is meghatározhatjuk, ám a teljesítmény optimalizálása miatt célszerűbb a reguláris kifejezést megelőzően a korpusz mondatait egy új változóhoz rendelni, és a *nem mint* és az *a mint* szószerkezetekkel együtt azokat a mondatokat is eltávolítani ebből a korpuszból, amelyekben több *akár* szerepel – majd ezen a szűkített korpuszon végezni a keresést.

Amennyiben egy szó áll a mondatközi írásjel és a kötőszó között, 70-szer hasonlító kapcsolat a 100 mondatos mintában. A hasonlító összetételnél figyelembe kell venni továbbá, hogy a főmondatrész gyakran csak odaértett módon jelenik meg a mondatban:

- (14) [Olyan vagy,] Mintha megváltoztál volna.

A mondat elején álló *mintha* kötőszó ebben az esetben 100-ból 31-szer jelentett tipikus hasonlatot (15), 16-szor nem tagmondatok közötti viszonyt írt le, és 53-szor, leginkább a *mintha* kötőszó esetében odaértett főmondatrészeiről volt szó (16). (*a mint* kötőszó a mondat elején 50-es mintában 37-szer vonatkozott kidolgozott hasonlító összetételre, 7-szer nem tagmondatok közötti viszonyra, 7-szer pedig odaértett főmondatra).

- (15) Mintha kihúzták volna alólam a talajt, olyan lettem. (Füst Milán: Feleségem története)

- (16) Mintha soványodott volna egy kicsit. (Szabó Magda: Freskó)

Az eljárás nem tesz különbséget a valóban kidolgozott és az odaértett hasonlatok között, ám a későbbiekben érdemes lehet a hasonlító összetett mondatok differenciáltabb tipológiáját megadni. A *mintha* szerkezet bemutatott példányai is illusztrálják, hogy mindenféleképpen érdemes a későbbiekben további vizsgálatokat végezni a jelentésalkotási műveleteik konstrukcionális feltáráshoz, mivel ezen mintázat megvalósulásai érintkezhetnek az episztemikus inferenciális értelmezéssel is (az inferencialitás fogalmához vö. Kugler 2015).

A két tagmondatból álló hasonlatok szűrésére alkalmazott reguláris kifejezés továbbá tévesen ismer fel hasonlító összetett mondatként mondatrészek közötti kapcsolatot is, például: „*Találkozott Jánossal, kivel mint szakemberrel beszélhetett.*” Az ilyen hibák azonban statisztikailag elenyészők a hasonlító összetételek számához képest. Nem jelent hibát továbbá, hogy a minősítő minőségjelzői kapcsolatok közül csak a kötőszóval kidolgozott tagmondatkapcsolatokat sorolja be a keresés a hasonlító összetett mondatok alá („Egy olyan nyakkendőre lenne szükségem, mint [*amilyen*] az öné”), szemben azokkal, amelyeknél a viszonyt nem dolgozza ki kötőszó („Olyan vagyok, [*mint*] *amilyen te*”), hiszen a kutatás éppen a formai oldalon is kidolgozott és a figyelem előterében álló kapcsolatokra irányul.

4.2. A -ként esetragos szerkezetek korpuszvezérelt szűrési módszere

A -ként esetragos szerkezetek szűrése látszólagosan nem jelent a formai oldalról kihívást a korpuszban való kereséskor, hiszen a mintázat azonosítása a keresett kifejezések végződése alapján könnyedén elvégezhető. Azonban, ha az így kapott mintát megvizsgáljuk, akkor szembesülünk azzal a problémával, hogy a -ként bizonyos alkonstrukciós sémái a jelentéshálózatban eltávolodnak a hasonlítás funkciójától⁴, és erőteljes a hajlandóságuk arra, hogy kategorizáló/besoroló szerkezetként működjenek (17b).

- (17) a. A Kis-Szamos óriási ezüstkígyóként hajlongott réteken, (Jósika Miklós: Abafi)
 b. [...], háziasszonyként nincsen tere az ő ízlésének és finomságának, (Esterházy Péter: Harmonia Caelestis)

Az (17a) példában a *Kis-Szamos* hasonlítottként vesz részt a hasonlítás kidolgozásában, míg az óriási ezüstkígyó lesz a névszói szerkezetben a hasonló, a -ként esetrag pedig a hasonlatot jelölő nyelvi modalizátor. Feltehetőleg a példányban a folyó kanyargó térbeli elhelyezkedése (*hajlongott*), és a víztükrön megfigyelhető fénysugarak csillámló játéka kerül hasonlító jelentésvizonyba a kígyók tekergő, csúszó mozgásával és pikkelyes felületével. A (17b) példában nem beszélhetünk valódi hasonlatról, hiszen a *háziasszonyként*

⁴ Alkonstrukció alatt a jelen tanulmány azt érti, hogy magasabb absztrakciós szinten ezek a mintázatok megegyező formai kidolgozásokat valósítanak meg, viszont ha a kevésbé absztrahált sémáikat vizsgáljuk, akkor jól/jobban azonosítható a forma-funkció oldaláról történő elkülönböztetés a használatukban.

kompozitumban a megnyilatkozói perspektívából a nyelvi szerkezet elsődleges figurája valamilyen szereppel azonosítódik a háttérben. Látható, hogy a (17a) és (17b) példamondatokban megfigyelhető hasonlat (17a) és a kategorizáló/besoroló szerkezet (17b) között nincs olyan jól megragadható formai különbség, amely alapján a reguláris kifejezésekkel kialakított szűrési eljárást fejleszteni tudtuk volna. Azonban hatékonyan kiküszöbölhetőek voltak az olyan találatok azonosítása, amelyek valamilyen időperiódusokhoz rendelt gyakoriságot fejeznek ki (például *óránként*, *hetenként*), hiszen a formai oldalról közelítve elégségesnek bizonyult az az esetükben, hogy az *-nként* szekvenciára végződő szóalakokat kizártuk a formalizált keresésből. Ezen kívül a szűrés első lépéseként az e-magyar nyelvi elemzővel azonosítottuk a szövegben előforduló főneveket, hiszen tipikusan ebbe a szófaji kategóriába tartoznak azok a lexémák, amelyek a *-ként* esetraggal kompozitumot alkotva potenciálisan hasonlat kidolgozását kezdeményezhetik.

Az előzőekben felvázolt (lásd (17a)–(17b) példamondat) problémához visszatérve fontos hiányosságként tűnik fel az, hogy a magyar nyelvre vonatkozóan még nem készült olyan automatikus szemantikai elemző eszköz, amely korpuszon futtatva képes lenne az előforduló lexémák, illetve lexikai egységek jelentéstani osztályozására. Feltűnő tendenciaként megfigyelhető, hogy a tipikus azonosító szerkezetekben előforduló *-ként* esetraggal kompozitumot alkotó főnevek általában valamilyen foglalkozást (*tanárnként*), rangot (*királyként*), nemet (*férfiként*), életkort (*gyermekként*), családi szerepet (*anyaként*) jelölnek. Azaz ebben az esetben a konstrukció a hasonlítottat nem valami máshoz viszonyítja, hanem a figurát **valamilyen szerepben, nézőpontból** láttatja, mint amely a megszólalás kezdetén a közös figyelmi jelenetnek az előtérben állt. Ez azt is jelenti, hogy a hasonlítás két oldala azonosítódik egymással – azonban ebben az esetben is felfedezhető a hasonlatok működése a hasonlított két szerepe közötti viszony kidolgozásában. Még pontosabb az a megfogalmazás, miszerint a prototipikus hasonlítás és a kategorizáló/besorolószerkezetek között fokozati különbség áll fent: nem minden esetben lehetséges éles határt húzni a két kategória közé. Ez különösen a *gyanánt* névutóval kidolgozott esetekben válik nyilvánvalóvá (3.4.), az N + *-ként* konstrukció esetében az említett jelentéstani kategóriák segíthetnek az elkülönítésben.

Ugyan a saját nyelvi intuícióinkat nem teljesen redukálva, de viszonylag megbízható, a nyelvi adatokat előtérbe helyező módszerként a Magyar Nemzeti Szövegtár 2-höz fordultunk, mégpedig úgy, hogy gyakorisági listát kérünk le a *-ként*-re végződő szóalakokra (az *-nként* végződéseket kihagyva),

50-es gyakorisági küszöbre állítva a megjelenítést. Így egy összesen 2754 elemből álló listát kaptunk, amelyből manuálisan válogattunk ki minden olyan lexémát, mely potenciálisan a fentebb leírt jelentéstani kategóriákba sorolhatóak. Ezzel a módszerrel egy összesen 1662 elemből álló stop-szó listát állítottunk össze.

A stop-szó lista alapján szűrtük tovább a formai kritériumoknak megfelelő szerkezeteket a kutatói korpuszunk teljes anyagát, és az így kapott eredmények nagymértékben hasonlító szerkezeteket tartalmaztak, az azonosítás hatékonysága a vizsgált mintában 78,28% volt (a vizsgált minta mérete 300 mondat volt). A szerkezet gyakorisága bizonyos esetekben korrelál a prototipikus hasonlatok számával, azaz a két tagmondatlal kidolgozott hasonlatokban gazdag művek gyakran előszeretettel alkalmazzák az N + *-ként* típusú szerkezeteket is. Ilyenek például Mészöly Miklós szövegei a korpuszból – a *Film* című regényben például éppen a párhuzamba állítást hasonlító, összekapcsolást szolgáló a filmes technikák nyelvi kifejezőiként szolgálnak. Ezáltal pedig a regény stílusának kulcseszközeinek tételezhetők például a következő idézetek (18a)–(18b).

- (18) a. Panoráma-háttérként hagyjuk kiélesedni a mélyebben fekvő Moszkva teret (Mészöly Miklós: Film)
b. valamilyen homályos elképzelés egy másik sétáról, amit pontosan ugyanígy teszünk meg, csak fordítva, izgalmas párhuzamként a mostanival (Mészöly Miklós: Film)

Van azonban példa olyan szerzőre, aki a szerkezetet annak ellenére alkalmazza gyakran, hogy a korpusz egészéhez viszonyítva ritkán dolgoz ki prototipikus hasonlatot. Ilyen Konrád György erősen nominális prózája, amelyben a mondatösszehasonlítás nem a tagmondatok, hanem a névszói szerkezetek halmozásából adódik, és amelyben sűrű névszói szerkezetekbe „rejtve” jelennek meg hasonlatok (lásd (19)).

- (19) Elmesztett falaival, üvegpedésként szétrajzó utcáival kitárul az egész város. (Konrád György: A városlapító)

De említhető Gárdonyi Géza kifejezetten rövidmondatos stílusa is, amelyben az N + *-ként* alakú, konceptuálisan gyorsabban hozzáférhető hasonlatok az egyszerű mondatok rövidségét megtartva képesek más kognitív tartományok aktiválására és a szöveg stilizációjára (ha visszafogott módon is):

- (20) A fák fekete óriásokként álltak az út mellett. (Gárdonyi Géza: Az egri csillagok)

4.3. A -szerű utótagos névszók hasonlatként

A -szerű esetén a korpuszkeresés viszonylag egyszerű feladat volt, hiszen olyan egyalakú utótagról van szó, amely gyakorlatilag hasonlatzászlóként működik – azaz kifejezetten azt jelöli, hogy az előtagjához hasonlítunk valamit (21).

- (21) ez a kettő volt olyan lényyszerű mozgás a közelükben, ami úgy hatott időnként, (Mészöly Miklós: Az atléta halála)

Az első ellenőrzés után a stop-szavak közé felkerültek olyan homonim szóalakok, amelyek valójában nem a -szerű utótagos névszók megvalósulásai voltak (például *idegrendszerű*, *módszerű*), valamint olyan kifejezések, amelyek lexikalizálódtak és egy sokkal általánosabb jelentést vettek fel, amely esetében már nincsen valódi hasonlítás (*korszerű*, *sportszerű*, *ésszerű/észesszerű*, *időszerű*, *értelemszerű* stb.). Ebben az esetben nem volt szükséges a szűrés során a szófaji sajátosságokat figyelembe venni, mert a -szerű általánosan a főnevekhez kapcsolódik – a mintában két kivétel volt, amelyekben melléknévi előtaghoz kapcsolódva fordult elő az utótag (*könnyűszerű* és *szépszerű*). Ezek azonban annyira hapax jellegűek voltak, hogy az MNSZ2-keresés sem hozott rájuk eredményt, így nem volt szükséges tovább pontosítani a módszert. Az azonosítás hatékonysága 100 elemből álló mintán 97% volt.

Ebben a vizsgálatban nem foglalkoztunk a határozóragos (N-szerűen) konstrukciókkal, azok ugyanis a hasonlatok egy olyan speciális típusát alkotják, amelyek a cselekvések módját, jellemzőit hasonlítják az adott névszóhoz asszociálódó tulajdonságokkal, szokásokkal.

Mészöly Miklós említett *Film* című regényében e konstrukcióból is találunk példát a filmes vágások és a kameramozgás nyelvi érzékeltetésére: „Mindezt még világosabbá teszi egy *cinéma direct-szerű véletlen*.” Még érdekesebb azonban az alábbi szöveghely, amelyben különböző típusú hasonlatok halmozását láthatjuk: a *mintha* kötőszóval bevezetett hasonlítás egy újabb hasonlatot bont ki (*feladatszerű*):

- (22) De kis ideig még utána is egymáson marad a szájuk, *mintha a feladatszerű együttműködés lazulásával hirtelen fellépne az ernyedtség, s nem volna erejük ellenállni neki.* (Mészöly Miklós: Film)

A (22) példa jól szemlélteti a névszóba sűrített hasonlatok egyik funkcióját: mintegy észrevétlenül, kevésbé feltüntetett pozícióban képesek belesimulni a mondat szerkezetébe, a tagmondatok által kidolgozott logikai kapcsolatokat (jelen esetben éppen egy hasonlítást) nem megtörve tudják beágyazni a háttérbe a hasonlítás műveletét azzal, hogy az elsődleges figurát (*együttműködés*) jelzős szerkezetben részletezi a *-szerű* utótagos kifejezés (*feladatszerű*), ezzel létrehozva egy több hasonlítási műveletből felépülő láncot. Tehát nem csupán az első és a második tagmondat által kidolgozott elemi jelenetek konstruálódnak meg párhuzamként, hanem a hasonlóként funkcionáló második tagmondaton belül is feldolgozódik egy hasonlítás. A *-szerű* poétikai funkciója és jelöltsége a szerkezeti beágyazódásának szintaktikai tulajdonságaiból fakad, hiszen nominatívuszi alakként képes arra, hogy jelzőként forduljon elő, s detektálhatunk olyan eseteket is, amikor a jelzők halmozásában (akár több *-szerű* alak mellérendelői láncolatában), valamely mondatbeli figura nagyfokú részletezettségének kidolgozásában vesz részt. Ez a jelzőhalmozás együtt járhat nagyobb fokú stilisztikai jelöltséggel is.⁵ A hosszú szerkesztettségű jelzői láncokon túl természetesen a kutatói korpuszunkban nagy arányban jelenik meg legalább egy, nem *-szerű* utótagos jelző környezetében (23).

- (23) guta még meg volt fagyva elbizakodott arcukon, amely fehér szemfedőszerű csipkébe volt szervírozva. (Bródy Sándor: Rembrandt)

Összefoglalóan kijelenthetjük, hogy a *-szerű* utótagos alakok azonosítása az általunk vizsgált többi szerkezethez képest nem bizonyult összetett kihívásnak. A konstrukció példányai közül a nagyobb mértékben lexikalizálódott alakok szűrése volt a feladatunk.

⁵ A stilisztikai jelöltség fogalma alatt azt értjük, hogy a stilisztikailag jelöltebb nyelvi kifejezés feldolgozásánál a befogadóban aktiválódik egy olyan stílusmintázat, amely az adott kontextusban feltűnő, kevésbé közömbös, s tipikusan más diszkurzussal asszociálódik, ezáltal nagyobb figyelem irányul rá a feldolgozás során (Tátrai 2012: 56).

4.4. A gyanánt névutós hasonlat

A *gyanánt* névutós konstrukció a kevésbé gyakori hasonlattípusok közé sorolható, amelynek oka az enyhe stilisztikai jelöltsége – ami fakadhat abból, hogy a *gyanú* főnév származéka (lásd TESz.), így a grammatikalizáltsága mellett valamelyest megmaradt a tartalmas jelentése. Az Értelmező kézisztár három jelentést különít el a *gyanánt* esetében (ÉKSz²): 1. Olyan minőségben, mintha a névszóval jelölt személy, tárgy v. dolog volna; 2. olyan formában, úgy, mintha a névszóval jelölt személlyel v. tárggyal azonos volna, bár tudja, hogy nem az; 3. A névszóval jelölt személyhez v. tárggyhoz v. dologhoz hasonlóan; úgy, mint. Esetünkben a mérvadó különbség a *gyanánt* poliszemiájában a hasonlító (3. jelentés) és az azonosító (1.) mozzanat megkülönböztetése volt. Az azonosító és a hasonlító szerkezetek viszonyát a *-ként* rag esetében már fejtegettük, a *gyanánt* esetében azonban nem lehetett olyan egyértelmű szerepkategóriákat (vö. foglalkozás: *tanárként*, családi szerep: *anyaként*) elkülöníteni, mint a *-ként*-nél. A *gyanánt* esetében nem olyan egyértelmű az azonosító mozzanat, inkább átmenet van az egyértelmű azonosítás és a hasonlítás között.

(24) Vigasz gyanánt áltató hazugságokat találtam ki

(25) Villám gyanánt hatolt elmémbe a gondolat

A (24) mondatnál a *vigasz* és a *hazugságok* azonosítódik: a hazugságok vigasz szerepbe lépnek, nem pedig hasonlóak egymáshoz – ugyanakkor a hasonlító jelleg is némiképp megmarad. A (25) mondatban a gondolat a villámhoz hasonló, a *villám* konvencionált jellemzői (például hirtelenség, váratlanság, félelmetesség) párhuzamba állíthatóak az adott *gondolat* tulajdonságaival (amelynek igéből képzett elvont főnév lévén nincsenek ilyen asszociatív jellemzői).

Vannak olyan mondatok a mintában, amelyeknél nem lehet csak a névszói szerkezet alapján eldönteni, hogy az azonosító vagy a hasonlító funkció a domináns:

(26) A jó szándékot egészen tett gyanánt veszem, átköltözteté leányomnak lelkét az én keblembé (Jósika Miklós: Abafi)

A *jó szándék* és a *tett* antonimikus viszonyát nehéz volna azonosításként értelmezni, intuitívan valószínűbbnek tűnik a *jó szándék olyan, mint a tett*

parafrázis szerinti értelmezés. A paradox viszonyt a mondatban a *vesz* ige oldja fel, amely 'valamiként értelmez, valaminek tekint' jelentése alapján a két névszó közti kapcsolat azonosító viszonyként írja le.

- (27) Egészen természetes immár, hogy ösztöneink ösztöneink ellen működnek, hogy úgyszólván *ellenösztöneink működnek ösztöneink helyett, sőt gyanánt* – szellemeskedem, ha ugyan szellemességnek tekinthető ez. (Kertész Imre: Kaddis a meg nem született gyermekért)

A (27) példában a szerző egy fokozó szerkezetbe helyezi az amúgy tartalmas jelentéssel nem bíró névutót (ezzel erősítve a szövegrészre is erősen jellemző iróniát), így a névutók jelentését nemcsak párhuzamosítja, hanem mintegy rangsorolja is, a fokozó szerkezet alapján a *helyett* a mondatban kevésbé intenzív jelentésű, míg a *gyanánt* a fokozott intenzitást jelöli. Esetünkben ez azért különösen izgalmas, mert éppen a hasonlítás esetén meghatározó helyettesítő funkcióval szemben a *gyanánt* azonosító funkcióját emeli ki – miközben az idézet szellemességéhez hozzájárul a *gyanánt* által kidolgozott hasonlítás is.

Az egyértelműen azonosító jellegű konstrukciókat nem tekintjük sűrített hasonlatnak, az azonosító konstrukciók kiszűréséhez szükség volt bizonyos megkötésekre, ehhez kollokációvizsgálat után stop-szó listát állítottunk össze. A kollokációmérés az MNSZ2 nyelvi anyagán történt, az 5,0 feletti LogDice értékkel rendelkező kifejezések voltak azok, amelyek esetében mérlegeltük az azonosító szerkezet komponenseként történő működést. A kollokáció mérés során lemmákra kértük le a *gyanánt* kollokáltjait -1 0 ablakban. Kétféleképpen tudtuk így szűkíteni a találati listát, egyrészt a névszókat szűrtük, ezáltal kizártuk azokat a zömmel elvont főneveket, amelyek generikusabb jelentésük miatt valószínűbben fordulnak elő azonosító szerkezetben (például a *segítség, érdekesség*). Szűrtük azokat az igéket is, amelyek részvétele a mondatban szintén azonosító szerkezetet feltételez, (például *fizet, fogyaszt, szolgál*). Az azonosító jellegű konstrukciókat a módszerrel ebben az esetben sem tudtuk teljesen kizárni, azonban a helyes találatok megtalálásának arányán javítottunk. A stop-szavakkal való szűréssel 130 példányszámú mintát tudtunk kinyerni az egész korpuszból, ennek 62,3%-a (81 db) volt sűrítő hasonlat. Mivel a *gyanánt* névutós szerkezet a legkevésbé gyakori hasonlattípus a vizsgált konstrukciós sémák közül, valamint mivel az azonosítást és a hasonlítást kevésbé lehet a formai oldalról megragadni, mint a *-ként*

konstrukciók esetében, a módszer kevésbé hatékonyabb – ugyanakkor ez mutatja is a *gyanánt* poliszemiájának komplexitását is.

4.5. A birtokos szerkezetekbe sűrített hasonlatok

Az alábbiakban a birtokos szerkezetbe sűrített hasonlatok számítógépes azonosítására kifejlesztett módszerünk bemutatására (vö. Bajzát–Szemes–Szlá-vich 2022), valamint a konkrét nyelvi példányok vizsgálatával a konstrukció jelentéshálózatának leírására vállalkozunk.

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

A fent idézett mondatpár (28a, b) rávilágíthat az egyes konstrukciók automatikus azonosításának nehézségeire is – ami pedig elengedhetetlen a későbbi kvantitatív vizsgálatok számára. A birtokos szerkezetekbe sűrített hasonlatok ugyanis – amelyek sajátos stíluspotenciáljára (a konstrukció tömörségére és a jelzőhalmazos 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 kendő-jével intett az idős hölgy felé.*” 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 poliszemiájának szemléltetésére is hozunk a korpuszunkból példányokat, hogy felvázolhassuk a konstrukció jelentéshálózatát.

Azon kívül, 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 hasonlítást végrehajtó példányoktól, azonban kevésbé parafrázálhatók a hasonlító összetett mondatokkal, valamint a kompozitum birtokos esetű főneve rendszerint valamilyen absztrakt entitás (lásd (29)).

- (29) [...], akkor sokkal kíméletesebben és főleg a megfontoltság ravasz-ságával kell eljárnia. (Nádas Péter: Emlékiratok könyve)

A forma-jelentés pár szimbolikus viszonyából kiindulva motivált az, hogy az előzőekben tárgyalt két jelentést/jelentésárnyalatot egymástól élesen nem elválasztva tárgyaljuk. Az efféle integrált, jelentéssűrítő funkcióit aktiváló szerkezetekre Herczeg Gyula (1958: 313) is felfigyelt, elkülönbötetve 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 ravaszán kell eljárnia*.” A két típus nem egyértelmű elhatárolhatóságát mutatja azonban, hogy bár kissé erőltetetten hat, nem tűnik teljesen hibásnak e példamondat hasonlattá alakítása sem: „*olyan ravaszán, 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.

A keresések az e-magyar nyelvi elemzővel feldolgozott szövegeken futottak le. Első lépésként az elvont jelentéstartalmat kifejező és a -vAl instrumentális 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 (például „*A színpadra belibbenő balett-táncosok kendőjével intett az idős hölgy felé*.”) Az elvont jelentéstartalmat a formai oldalról tipikusan 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 az e szavakat tartalmazó mondatok képezték a további elemzési lépések 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*.” (Mikszáth Kálmán: Új-Zrínyiász) 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 egyébként lehetővé teszi a képzőkre vonatkozó információk kinyerését, mert jelen esetben könnyen listázható a keresett főnevek végződése, így a legpontosabb találatokat ezen keresztül kaphatjuk.

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. Fontos, 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.

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 főfelülreprezentált a keresett konstrukciókban (például *méltó-ságával*). Az e-magyar által lemmaként azonosított formák a birtokos és a -val ragot nélkülöző eseteket, azaz tulajdonképpen magukat az elvont főneveket mutatják (például *méltóság*). Ezért ebben a fázisban a képzők eltávolításával hoztuk létre a kívánt tövet, amelyen ismételten lefutattuk az e-magyar elemzését. Így tudtuk csak azokra a szavakra korlátozni a keresést, amelyekben melléknévi alaptól jönnek létre a 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 mégis nagy számban fordulnak elő a konstrukcióban. Ezeket elsősorban Herczeg Gyula szintén erre irányuló szólistája (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, lánghozzás, észjárás, arckifejezés, illetve: nyugalom, izgalom, aggodalom, fájdalom, szánalom, léleknyugalom, unalom, vasfegyelem, buzgalom, értelem, félelem, szorgalom, 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, és így a pontosabb keresést segítik. 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 *-vall/-vel* instrumentális ragos eset, és így nem a sűrített hasonlat meglétét jelzi: *szembe(n), visszaél, helyettesít, összevet, versenyez, küzd, vetekszik, ellenkezik, ellentétben, megleged, áthidal, rendelkezik, kapcsolatos, tisztába(n), együtt, rendelkezik, összefügg, összetéveszt, összeegyeztet, arány, felruházt.*

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 egyértelműen eldönthető, hogy egy példány melyik kategóriába tartozik, mindhárman egymástól függetlenül végeztük a feldolgozást, majd összevetettük egymással az eredményeinket, és a kérdéses eseteket megvitattuk. Törekedtünk arra, hogy a funkcionális kiindulópontot folyamatosan érvényesítsük, ezért csak korlátozottan tudtuk igen-nem oppozícióra leszűkíteni a döntéseinket, ezért úgy döntöttünk, hogy ahelyett, hogy csak a sűrített hasonlatokat címkézzük találatként, a vele szoros hasonlóságot (a prototípuselvű kategorizációhoz vö. Langacker 1987: 16–17; Taylor 1995; Tolcsvai Nagy 2013; Tolcsvai Nagy szerk. 2017: 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%).

5. Diszkusszió

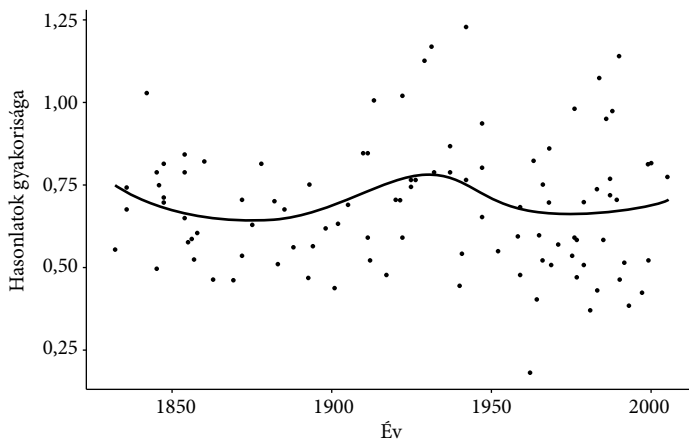
A dolgozatban bemutatott módszer a különböző, „sűrített” hasonlattípusok gépi azonosítására több szempontból is hozzájárulhat a hazai hasonlatkutatás eredményeihez. Egyfelől ezáltal nagy mennyiségű példa hozható korábban kevésbé tárgyalt szerkezetekre, ami fontos kiindulópont a szerkezetek grammatikai, stilisztikai leírása és a megismerésben betöltött szerepük megértése szempontjából. A dolgozatban kísérletet tettünk az első lépések megtételére ez irányba. A példák nagy száma továbbá fontos lehet későbbi módszerek, akár nem szabály alapon, hanem gépi tanuláson nyugvó eljárások számára is, ugyanis az általunk kialakított félautomatikus mintavételezési eljárás képes lehet tanítókorpuszok létrehozására is.

Másfelől az automatikus azonosítás segítségével a prototipikus, tagmondatok által kidolgozott hasonlatok mellett más típusok gépi elemzésére is lehetőség nyílik, ebből következően pedig a kidolgozott módszer további alkalmazási lehetőségeket kíván. Ilyen például a hasonlítás gyakoriságának történeti változásaira rákérdező poétikatörténeti kutatás is – a korábbi eredményeinkkel szemben (Szemes 2021) már egy átfogóbb képet tudunk adni a történeti folyamatokról. Az 1. ábra ezt szemlélteti: az összes hasonlattípus relatív gyakoriságát⁶ a kutatói korpusz regényeiben azok megjelenési idejével összekapcsolva mutatja be (a típusok eltérő gyakorisága miatt eltérő súllyal is szerepelnek az ábrán látható értékekben). Az adatpontokra illesztett regressziós görbe egyértelműen kirajszolja, hogy a 20. század első felének irodalmi nyelve a korábbi és későbbi szakaszokhoz képest inkább előnyben részesítette a hasonlítás műveletét a regényvilág elemeinek bemutatásakor. Ezáltal egy stílustörténeti korszak is elkülöníthetőnek tűnik, amely csak részben egyezik a korábban impresszionista prózának nevezett korszakkal. A hasonlatok számának növekedése összefügghet ugyanakkor a nominális, megnevező stílus elterjedésével (a világ elemeit „képszerűen” bemutatva), illetve a szereplői tudatra fókuszáló narráció előtérbe kerülésével (a hasonlatok gyakran egy szereplő benyomását viszik színre, anélkül, hogy a narrátor szólal megváltozna, lásd (30))

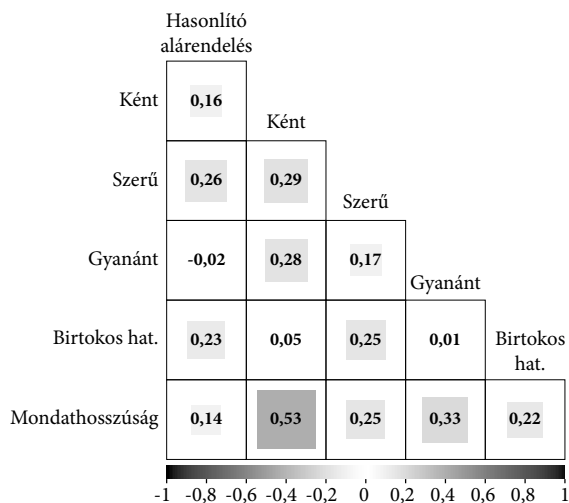
- (30) Rusz nem tudja, merre lépjen. Mintha levegő lenne alatta vagy szakadék, úgy forgolódik. Rámered a mosógépre s feléje indul. (...)

⁶ A relatív gyakoriság a típusok arányát a szöveg szavainak számához képest mutatja, 1000-rel megszorozva.

Rusz oldalt lógatja a fejét, mintha akasztófán lógna. Mintha csak ugyan hurok szorítaná a nyakát. Nem szól. (Gelléri Andor Endre: A nagymosoda)



1. ábra. Hasonlatok gyakoriságának alakulása a vizsgált 100 regény alapján



2. ábra. Az egyes hasonlattípusok relatív gyakorisága és az átlagos mondathosszúság korrelációs mátrixa

Érdekes tanulságokkal szolgálhat továbbá az egyes hasonlattípusok gyakoriságának korrelációját vizsgálni, azaz arra a kérdésre keresni a választ, hogy milyen összefüggés van az egyes szerkezetek használatában. A 2. ábra, amely a típusok relatív gyakorisága között fennálló kapcsolatot mutatja a Pearson-féle korreláció alapján, azt jelzi, hogy azok a szövegek, amelyek gyakran dolgoznak ki hasonlító kapcsolatot tagmondatok között, szintén előszeretettel alkalmazzák a *-ként*, a *-szerű* ragozott alakokat és a birtokos szerkezetbe sűrített hasonlatokat – bár a korreláció nem tekinthető magas fokúnak –, és nincs összefüggés a *gyanánt* névutó hasonlító szerkezetekben való megjelenésének gyakoriságával. Még beszédesebb az összefüggés, amely a mondatok átlagos hosszúsága és az egyes konstrukciók között fennáll, már csak azért is, mert ez esetben nem várnánk pozitív korrelációt, hiszen e konstrukciókra éppen a sűrítés és nem a kibontás jellemző. Ugyanakkor ez hívja fel a figyelmet az ilyen hasonlatok tágabb kontextusára és stilisztikai funkciójára: ekkor ugyanis a hasonlítás művelete a prototipikus esethez képest kevésbé kerül az előtérbe, mintegy belesimul a mondat logikai szerkezetébe és inkább egy mellékmondat vagy egy felsorolás egy figurájának kiegészítő jellegű kidolgozását végzi el. Azaz olyan mondatrészek és tagmondatok további jellemzéséhez járulnak hozzá, amelyek már amúgy is kidolgozott grammatikai viszonyokban szerepelnek – tehát nem valami (tagmondattal kidolgozott hasonlatok) helyett szerepelnek, hanem valaminek (más kapcsolattípusok, akár éppen hasonlítások, ahogy fentebb láttuk Mészöly Miklós egy mondatánál) a további, részletesebb bemutatásához járulnak hozzá. Vagyis azt is mondhatjuk, hogy ezekben az esetekben a sűrítéssel tagmondatok további bővülése valósul meg. A minták elemzése azt mutatja, hogy elsősorban a *-szerű* utótaggal ellátott névszói szerkezetek teszik lehetővé a hasonlítás beékelését egy felsorolásba, például: „*a vendéglő bolthajtásos, barlangszerű és ugyancsak homályos bejárója elé értünk*” (Gárdonyi Géza: A láthatatlan ember). Mindez a mondat nagyobb fokú stilisztikai jelöltségét is jelenti, amely annál hangsúlyosabb, minél ritkább, ezáltal stílusosan feltűnőbb szerkezet dolgozza ki a hasonlítást – legerősebben a birtokos szerkezetek esetében, amelyek váratlanabb, absztraktabb kapcsolatok létrehozását teszik lehetővé.

1. táblázat. Az azonosító eljárások összefoglalása

Összegző ábra a hasonlattípusok korpuszbeli azonosíthatóságának módszertanairól			
Típus	Eljárások		
A két tagmondatból álló, prototipikus hasonlatok	Reguláris kifejezések írása kötőszavak és az írásjelek lehetséges pozíciói alapján	Kvalitatív elemzés 100 elemből álló mintákon	A reguláris kifejezések finomítása
<i>-ként</i> esetragos szerkezet	<i>-ként</i> esetragos szerkezetet tartalmazó mondatokra való keresés reguláris kifejezésekkel	MNSZ2-ből gyakorisági listák lekérdezése (50-es küszöbértékel), tipikusan azonosító komponensként előforduló névszók-ból stop-szó lista összeállítása (1662 elem), <i>-nként</i> végződés szűrése	A stop-szó lista beemelése a keresési eljárásba
<i>-szerű</i> utótagos névszók	<i>-szerű</i> végződésű főnevekre való keresés, homonim alakok kiszűrése (például <i>módszerű</i> , <i>idegrendszerű</i>)	Stop-szó lista összeállítása MNSZ2-ből származó gyakorisági lista alapján, amellyel a lexikalizálódott, nem hasonlító funkciójú szerkezetek szűrhetővé válnak	A stop-szó lista beemelése a keresési eljárásba
<i>gyanánt</i> névutós hasonlat	<i>gyanánt</i> tokenre való keresés	Kollokációvizsgálat (LogDice 5,00 küszöbérték) alapján névszók meghatározása, amelyek azonosító szerkezetben vesznek részt (például <i>érdekesség</i>)	Igék szűrése, amelyek környezetében azonosító szerkezetként működik a névszó + <i>gyanánt</i> szerkezet (például <i>fizet</i> , <i>fogyaszt</i>)

A jelen tanulmányban közölt módszertani lépéseket a 3. táblázat összegzi. A bemutatott kutatás célja az volt, hogy a magyar nyelv különböző hasonlattípusainak számítógépes azonosítására dolgozzon ki félautomatikus eljárásokat. Másfelől célunk volt az elemzett típusok kognitív jelentéstani elemzése konstrukcionális kiindulópontból. Kutatói korpuszként a magyar kanonikus regényirodalomból összeállított gyűjteményt választottuk, azon kíséreltünk meg kvantitatív vizsgálatot a kialakított mintavételezési protokollok alkalmazásával. Módszertani konklúzióként kiemelhető, hogy a vizsgált szerkezetítípusok deduktív módszerekkel csak korlátozottan közzelíthetők meg, az azonosítási metódus kialakítása többlépcsős folyamatot igényelt, amely során becsatornázhatóvá váltak a kvalitatív adatelemzésből nyert ismereteink. A vizsgált szerkezetekre jellemző a nagyfokú formai variabilitás és ezzel összekapcsolva a funkcionális változatosság. A különböző jelentésárnyalatok megragadása ugyan a formalizált módszerekkel nem minden esetben egyértelmű, de az esettanulmányaink igazolják, hogy nagymértékben szűkíthetők a fals pozitív találatok egyszerű, magas szintű technikai tudást mellőző módszerekkel. A kutatás folytatásaként körvonalazódik egy gold standard korpusz építése, amelyen lehetővé válna a módszerünk hatékonyságának pontosabb mérése. Az azonosított szerkezetítípusok szisztematikus, részletesebb kognitív szemantikai vizsgálata szintén gyümölcsöző eredményekkel járhat a későbbiekben.

Irodalom

- Bajzát Tímea Borbála – Szemes Botond Bálint – Szlávích Eszter 2021. Az ELTE DH Regénykorpusz és lehetőségei. In: Tick József – Kokas Károly – Holl András (szerk.): *Online térben az online térért: Networkshop 30: országos online konferencia*. 2021. április 6–9. Eötvös Loránd Tudományegyetem. Budapest: HUNGARNET Egyesület. 63–72. <https://doi.org/10.31915/NWS.2021.7>
- Bajzát Tímea Borbála – Szemes Botond Bálint – Szlávích Eszter 2022. Adalékok a birtokos szerkezetekbe sűrített hasonlatok számítógépes azonosításához. In: Berend Gábor – Gosztolya Gábor – Vincze Veronika (szerk.): *XVIII. Magyar Számítógépes Nyelvészeti Konferencia: MSZNY 2022*. Szeged: SZTE, Informatikai Intézet. 93–104.
- Benkő Loránd 1967 (szerk.). *A magyar nyelv történeti-etimológiai szótára*. Budapest: Akadémiai Kiadó.

- Brdar-Szabó Rita 2020. Rezultatív mikrokonstrukciók és morfológiai vetélytársaik elmélet és empiria egységében. In: Simon Gábor – Tolcsvai Nagy Gábor (szerk.): *Nyelvtan, diskurzus, megismerés*. Budapest: ELTE Eötvös Kiadó. 141–166.
- Diessel, Holger 2015. Usage-based construction grammar. In: In: Dąbrowska, Ewa – Divjak, Dagmar (eds.): *Handbook of Cognitive Linguistics*. Berlin, Boston: De Gruyter Mouton. 295–321. <https://doi.org/10.1515/9783110292022-015>
- Fishelov, David 1993. Poetic and non-poetic simile: Structure, semantics, rhetoric. *Poetics Today* 14 (1): 1–23. <https://doi.org/10.2307/1773138>
- Gintli Tibor szerk. 2010. *Magyar irodalom*. Budapest: Akadémiai Kiadó.
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.
- Herczeg Gyula 1958. Határozóként álló, -val, -vel ragos utótagú birtokos szerkezetek. *Magyar Nyelvőr* 84: 311–573.
- Herczeg Gyula 1974. Elvont főnevek sajátos mondatstilisztikai szerepkörben egy századvégi írónál. *Magyar Nyelvőr* 98: 173.
- Herczeg Gyula 1981. *A régi magyar próza stílusformái*. Budapest: Tankönyvkiadó Vállalat.
- Herrmann, J. Berenike – van Dalen-Oskam, Karina – Christof, Schöch 2015. Revisiting style, a key concept in literary studies. *Journal of Literary Theory* 9 (1): 25–52. <https://doi.org/10.1515/jlt-2015-0003>
- Hilpert, Martin 2014. *Construction Grammar and its application to English*. Edinburgh: Edinburgh University Press.
- Ide, Nancy M. – Sperberg-McQueen, Michael. 1995. The TEI: History, goals and future. *Computers and the Humanities* 29: 5–15. <https://doi.org/10.1007/BF01830313>
- Imrényi András 2017. Az elemi mondat viszonyhálózata. In: Tolcsvai Nagy Gábor (szerk.): *Nyelvtan*. Budapest: Osiris Kiadó. 663–761.
- Indig Balázs – Sass Bálint – Simon Eszter – Mittelholcz Iván – Kundráth Péter – Vadász Noémi 2019. emtsv – Egy formátum mind felett. In: Berend Gábor – Gosztolya Gábor – Vincze Veronika (szerk.): *XV. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi Tudományegyetem Informatikai Intézet. 235–247.
- Kilgarrieff, Adam – Baisa, Vít – Buřt, Jan – Jakubíček, Miloř – Kovář, Vojtěch – Michelfeit, Jan – Rychlý, Pavel – Suchomel, Vít 2014. The Sketch Engine: Ten years on. *Lexicography* 1: 7–36. <https://doi.org/10.1007/s40607-014-0009-9>

- Kocsány Piroska 2008. Hasonlat. In: Szathmári István (szerk.): *Alakzatlexikon*. Budapest: Tinta Könykiadó. 267–280.
- Kugler Nóra 2015. *Megfigyelés és következtetés a nyelvi tevékenységben*. Budapest: Tinta Könykiadó.
- Kugler Nóra 2017. Az összetett mondat. In: Tolcsvai Nagy Gábor (szerk.): *Nyelvtan*. Budapest: Osiris Kiadó. 806–895.
- Ladányi Mária 2017. *Alaktan*. In: Tolcsvai Nagy Gábor (szerk.): *Nyelvtan*. 4. Budapest: Osiris Kiadó. 503–662.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Volume I. Theoretical prerequisites*. Stanford–California: Stanford University Press.
- Langacker, Ronald W. 2005. Construction Grammars: Cognitive, radical, and less so. In: Ruiz de Mendoza Ibáñez, Francisco J. – Peña-Cervel, M. Sandra (eds.): *Cognitive Linguistics Research* 32. Berlin, New York: Mouton de Gruyter. 101–159. <https://doi.org/10.1515/9783110197716.1.101>
- Mpouli, Suzanne 2016. Automatic annotation of similes in literary texts. Doktori disszertáció.
- Mpouli, Suzanne 2017. Annotating similes in literary texts. In: Bunt, Harry (ed.): *Proceedings of the 13th Joint ISO-ACL Workshop on Interoperable Semantic Annotation (ISA-13)*. Tilburg: Tilburg University. 26–36.
- Nate, Richard 2006. Simile. In: Sloane, Thomas O. (ed.): *Encyclopedia of rhetoric*. Oxford: Oxford University Press. 741.
- Novák, Attila – Siklósi, Borbála – Oravecz, Csaba 2016. A new integrated open-source morphological analyzer for Hungarian. In: Calzolar Nicoletta – Choukri Khalid – Mazo, Helene – Moreno, Asuncion – Declerck, Thierry – Goggi, Sara – Grobelnik, Marko – Odijk, Jan – Piperidis, Stelios – Maegaard, Bente – Mariani, Joseph (eds.): *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Paris: European Language Resources Association (ELRA). 1315–1322.
- Oravecz, Csaba – Várad, Tamás – Sass, Bálint 2014. The Hungarian Gigaword Corpus. In: Calzolari, Nicoletta – Choukri, Khalid – Declerck, Thierry – Hrafn–Maegaard, Bence – Mariani, Joseph – Moreno, Asuncion – Odijk, Jan – Piperidis, Stelios (eds.): *Proceedings of the ninth international conference on language resources and evaluation (LREC-2014)*. Reykjavik: ELRA. 1719–1723.
- Paradis, Carita 2005. Ontologies and construals in lexical semantics. *Axiomathes* 15: 541–573. <https://doi.org/10.1007/s10516-004-7680-7>
- Pusztai Ferenc szerk. 2003. *Magyar értelmező kéziszótár*. 2., átdolgozott kiadás. Budapest: Akadémiai Kiadó.

- R Studio Team 2020. *R Studio: Integrated Development for R. Rstudio*. Boston: PBC.
- Sass Bálint 2017. Keresés a korpuszban. In: Forgács Tamás (szerk.): *A nyelvtörténeti kutatások újabb eredményei IX*. Szeged: Szegedi Tudományegyetem Magyar Nyelvészeti Tanszék. 267–277.
- Schöch, Christof – Eder, Maciej – Rosario, Arias – Pieter, Francois – Primorac, Antonija 2020. *Foundations of distant reading. Historical roots, conceptual development and theoretical assumptions around computational approaches to literary texts*. <https://dh2020.hcommons.org/>
- Shen, Yeshayahu 2008. Metaphor and poetic figures. In: Gibbs, Raymond W. (ed.): *The Cambridge handbook of metaphor and thought*. Cambridge: Cambridge University Press. 295–307. <https://doi.org/10.1017/CBO9780511816802.019>
- Simon Gábor – Bajzát Tímea – Ballagó Júlia – Havasi Zsuzsanna – Roskó Míra – Szlávich Eszter 2019. Metaforaazonosítás magyar nyelvű szövegekben: egy módszer adaptálásáról. *Magyar Nyelvőr* 143: 223–247.
- Simon Gábor 2022. A megismerés buktatókövei: Hasonlatok Török Sophie narrációjában. In: Tátrai Szilárd – Melhardt Gergő – Szemes Botond – Beszkid Judit (szerk.): *Élveboncolás: Közelítések Péterfy Gergely Kitömött barbár című regényéhez*. Kalligram: Budapest. 142–173.
- Simon Eszter – Indig Balázs – Kalivoda Ágnes – Mittelholcz Iván – Sass, Bálint – Vadász Noémi 2020. Újabb fejlemények az e-magyar háza táján. In: Berend Gábor – Gosztolya Gábor – Vincze Veronika (szerk.): *XVI. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi Tudományegyetem. 29–42.
- Steen, Gerard – Dorst, Aletta G. – Hermann, Berenike J. – Kaal, Anna A. – Krennmayr, Tina – Prasma, Trijntje 2010. *A method for linguistic metaphor identification*. Amsterdam – Philadelphia: John Benjamins.
- Szemes Botond 2021. Kidolgozott viszonyok. A tagmondatkapcsolatok automatikus azonosításának hasznosíthatósága a stilisztikában és az irodalomtörténet-írásban. *Digitális Bölcsészet* 4 (1): 31–77. <https://doi.org/10.31400/dh-hun.2021.4.2415>
- Szemes Botond – Byszuk, Joanna 2021. A krakkói Computational Stylistics Group bemutatkozása. *Digitális Bölcsészet* 4 (2): 1–5.
- Szilágyi Márton – Vaderna Gábor 2010. A klasszikus magyar irodalom (kb. 1750-től kb. 1900-ig). In: Gintli Tibor (szerk.): *Magyar irodalom*. Budapest: Akadémiai Kiadó. 316–637.

- Tátrai Szilárd 2012. Viszonyulás és viszonyítás. Megjegyzések a stílus szociokulturális tényezőinek vizsgálatához. In: Tátrai Szilárd – Tolcsvai Nagy Gábor (szerk.): *A stílus szociokulturális tényezői*. Budapest: Eötvös Loránd Tudományegyetem. 51–72.
- Tolcsvai Nagy Gábor 2017. Jelentéstan. In: Tolcsvai Nagy Gábor (szerk.): *Nyelvtan*. Budapest: Osiris Kiadó. 207–466.
- Tolcsvai Nagy Gábor 2021. *Kognitív szemantika*. Budapest: ELTE Eötvös Kiadó.
- Tolcsvai Nagy Gábor (szerk.) 2017. *Nyelvtan*. A magyar nyelv kézikönyvtára 4. Budapest: Osiris.
- Traugott, Elisabeth Closs 2008. ‘All that he endeavoured to prove was ...’: on the emergence of grammatical constructions in dialogual contexts. In: Kempson, Ruth – Cooper, Robin (eds.): *Language change and evolution*. London: Kings College Publications. 143–177.



Semi-automatic identification and analysis of different types of similes

The research has three purposes: firstly, the computer-based identification of different types of similes, secondly, the cognitive stylistic and semantic analysis of these similes, and thirdly, their quantitative analysis in canonical Hungarian novels. The constructions cause varying difficulties in formal isolation, so our aim was to achieve our objective by developing a multi-step method based on regular expressions and taking into account the morphological and semantic properties of the components of the constructions. The queries were run on 100 canonical Hungarian novels over 173 years, an average of 1.7 novels per year. Meanwhile, the identification of the compressed similes allows their quantitative analysis in the corpus, as well as the refinement of the results of the frequency analysis for the similar compound phrases.

Keywords: corpus linguistic, stylometry, similes, constructions, history of literature

Bajzat Tímea Borbála

Eötvös Loránd Tudományegyetem

bajzat.timi9696@gmail.com

ORCID: 0009-0006-9196-6442

Szláwich Eszter

Eötvös Loránd Tudományegyetem

szlavich.eszter@btk.elte.hu

ORCID: 0000-0003-2481-3228

Szemes Botond

Bölcsészettudományi Kutatóközpont; Digitális Bölcsészet Nemzeti Labora-
tórium

szemes.botond@abtk.hu

ORCID: 0000-0002-0637-6776

HORVÁTH PÉTER

***Hozzászólás A tagmondatkapcsolattól a komplex
birtokos szerkezetig: hasonlattípusok azonosítása és
vizsgálata című tanulmányhoz***

1. Általános gondolatok

A tanulmány elméletileg és módszertanilag is jól megalapozott kutatást mutat be. A kutatás eljárásai és eredményei a hasonlatok konstrukciós leírásához és az irodalmi szövegekben betöltött funkcióinak pontosabb megértéséhez is hozzájárulnak. A tanulmány egyik általánosabb eredménye, hogy a hasonlatok példáján keresztül rámutat arra, hogy funkcionálisan poliszém és formailag heterogén, komplex nyelvi jelenségeket is jó hatásfokkal fel lehet ismertetni szabályalapúan. A vizsgált nyelvi jelenségek automatizált felismertetése mellett a szabályalapú módszerek alkalmazásának egyik további hozadéka, hogy azok a gépi tanulási módszertanokhoz képest a kérdéses nyelvi jelenségek pontosabb, a nyelvészeti leírás szempontjából is nagy fokú relevanciával bíró jellemzését kívánják meg. A tanulmány a hasonlatok vizsgálatán keresztül szemléletesen bemutatja, hogy egy nyelvi jelenség funkcionális és formai jellemzőinek nyelvészeti leírása hogyan járulhat hozzá a nyelvi jelenség reguláris kifejezésekkel, illetve stopword listákkal való minél hatékonyabb szabályalapú azonosításához, illetve az utóbbi miképpen követeli meg a minél szabatosabb nyelvészeti leírást. Bár manapság nyelvi jelenségek felismertetésére nagyobb mértékben használnak gépi tanulási módszereket, a szóban forgó tanulmány jó példája annak, hogy még mielőtt belekezdenénk egy tanítókörpusz építésébe, érdemes feltenni a kérdést, hogy nem lehet-e az adott nyelvi jelenség felismertetését a kutatás számára megfelelő minőségben szabályalapúan is megvalósítani.

2. Két felvetés a kutatás továbbgondolásához

- A tanulmány többször is utal arra, hogy a különböző típusú hasonlatok a feldolgozás nehézsége és szemantikai, illetve stilisztikai jelöltségük szempontjából is különböznek egymástól. Ebben a korpusznyelvészeti módszereket alkalmazó tanulmányban elfogadható az, hogy ezek az állítások pusztán szerzői introspekción alapulnak, ugyanakkor felvethető a kérdés, hogy a feldolgozás nehézségének, illetve a szemantikai és stilisztikai jelöltségnek a vizsgálatára milyen, az introspekción túllépő módszereket lehetne alkalmazni a hasonlatok vizsgálata során.
- Vajon lehetséges lenne a hasonlatok azonosítása során a hasonlított és a hasonló szerepében megjelenő entitások automatikus vagy félautomatikus kategorizálása valamilyen ontológia alapján és ezáltal a hasonlatok szemantikai mintázatainak a részletesebb leírása? Egy ilyen eljárás segítségével nagy adattömegben lehetne vizsgálni, hogy különböző entitástípusokat jellemzően milyen entitástípusokhoz hasonlítunk, ami nemcsak a szépirodalmi szövegek hasonlattípusainak, hanem a fogalmi rendszer szerveződésének a mélyebb megértéséhez is hozzájárulhatna.

Horváth Péter

Eötvös Loránd Tudományegyetem

horvath.peter@btk.elte.hu

ORCID: 0000-0002-3517-5623

SIMON GÁBOR

***Hozzászólás A tagmondatkapcsolattól a komplex
birtokos szerkezetig: hasonlattípusok azonosítása és
vizsgálata című tanulmányhoz***

A hasonlat „ismeretlen ismerős” a nyelvészeti kutatásban. Egyrészt a figuratív nyelvhasználat hagyományos alakzata, az iskolai nyelvtankönyvek egyik legegységértelműbb kategóriája, amely gyakran kapja meg a „kifejtett metafora” billogot. Másrészt viszont olyan formai változatosságot és jelentésbeli gazdagságot mutató jelenség, amelynek feltérképezése az elmúlt másfél-két évtizedben indult meg igazán, elsősorban a kognitív nyelvészet elméleti keretében, ám egyre inkább korpuszvizsgálatokra építve (lásd Fishelov 2007; Shen 2008; Tartakovsky–Shen 2018, 2019). Egy alkalommal összegyűjtöttem néhány kurrens elemzésre építve, milyen kategóriákat különítenek el napjainkban a hasonlaton belül. A hasonlítás alapjának kifejtettsége alapján zárt és nyílt hasonlatról beszélhetünk (az utóbbi Beardsley 1958-ra vezethető vissza); a hasonlított (target) és a hasonló (source) konvencionalitása alapján konvencionális vagy opák hasonlatokról; a konceptuális szerkezet fogalmi irányulása alapján kompatibilis (a hasonló konkrétabb és ismerősebb, mint a hasonlított) és ellentmondó (clashing, a hasonló kevésbé konkrét vagy ismerős, mint a hasonlított) hasonlatokról; végül az alapján, hogy a hasonlóság alapja mennyire szaliens jegye a hasonlónak, sztenderd, nem sztenderd humoros, nem-sztenderd ironikus és nem-sztenderd poétikus hasonlatról. És ezek még csak a szemantikai elemzés szempontjai, a formai megvalósulás mintázatairól (milyen az elemek sorrendje, mi a modalizátor stb.) nem is beszélve.

A hasonlat tehát távolról sem olyan egyértelmű, mint ahogyan azt a retorikai-stilisztikai hagyomány sejteti. Bajzát Tímea Borbála, Szemes Botond és Szlávich Eszter tanulmánya ennek az összetettségnek ered a nyomába, és kiterjedt módszertani apparátussal, nagy mennyiségű adat elemzésével, több szempont bevonásával mutat rá arra, hogy a hasonlat sokféle szerkezettel fejezhető ki a magyar nyelvben, ráadásul ezeknek a szerkezeteknek nem homogén a jelentésbeli funkciójuk. Ez már önmagában kivételes eredmény,

a szerzők ugyanis (i) elhelyezik a magyar nyelvet a szisztematikus hasonlat-elemzés térképén, (ii) módszertani fogódzót nyújtanak nyelvészeknek és digitális bölcsészeknek a további kutatásokhoz, és (iii) olyan részletességgel térképezik fel a hasonlat magyar nyelvű eszköztárát, amely egyúttal a hasonlítás mentális műveletének jobb megértését is szolgálja.

A vizsgálat kognitív korpusznyelvészeti elemzésre épül, amely a konstrukciós nyelvleírás eszközeit és szemléletmódját is alkalmazza. A jelen kötet alapgondolatához igazodva ebben a hozzászólásban csak általánosabb, a vizsgálat egészét érintő koncepcionális javaslatokat szeretnék megfogalmazni, melyek egy része a kognitív nyelvészethez, más részük a korpusznyelvészethez, megint más részük a konstrukciós nyelvleíráshoz kapcsolódik.

A szerzők arra vállalkoznak, hogy a magyar nyelvű hasonlatok kifejezésbeli sokféleségét és funkcionális változatait is felmutassák (törekedve a hasonlítás mint jelentéskonstruálás skaláris értelmezésére). Mint ahogyan az azonban a figuratív nyelvhasználat kutatásában gyakran történik, a formai változatosság megfigyelése jobban sikerül, mint a funkciók elkülönítése. Mindjárt a központi fogalom, a hasonlat körülhatárolása problematikus azoknál a szerkezeteknél, amelyeknek nem hasonlító olvasata is lehetséges. Elsősorban a birtokos szerkezeteknél találtam több olyan példát is a tanulmányban, melyeknél a konstruálás nem csupán analógián, de referenciális azonosságon is alapulhat. A boldogságáról valló bőbeszédű beszélő nem csak hasonlíthat az először szeretőhöz, hanem lehet ő maga először szerelmes. A (12) példabeli megállást végző figura lehet éppen előkelő is magabiztosságát tekintve. A Mikszáth-regényből idézett szereplők pedig, akik éppen árakról egyezkednek, lehetnek élelmes piaci árusok is, és nem csak ezen árusokra hasonlító. Persze a szerzők maguk is aprólékosan bemutatják a hasonlítás és az azonosítás különbségét, mégis az a benyomásom támadt, hogy túl szélesre nyitják az analógia értelmezését, amikor a referenciapont-szerkezeteket is beemelik az analógiaképzés esetei közé. Amennyiben a referenciapont és a célpont fogalmi kapcsolata (például az előkelők magabiztossága, vagy a piaci árusok leleményessége) a tulajdonsága a diskurzuszvilág résztvevőjének, úgy nem hasonlításról, hanem egy szerep, egy személyiségjegy előtérbe helyezéséről. Az analógia leginkább azoknál a birtokos szerkezeteknél érhető tetten, amelyekben a birtokos határozatlan névelővel (tehát típusként, például *egy balett-táncos*) jelenik meg: ezekben az esetekben a referenciapont lehet a hasonló, a célpont pedig a hasonlított, amely a referenciapont felől hozzáférhető, tipikus jegyet, szcenáriókat tartalmazó tartomány egy elemeként konstruálható meg.

A birtokos szerkezetek alkalmasak a hasonlítás kifejezésére, de a tanulmány nem kínálja fel ennek a funkciónak a körütekintő megkülönböztetését a referenciális azonosság és a fogalmi hozzáférhetőség más eseteitől.

Ugyancsak a hasonlítás fogalmának tágasságát mutatja, hogy a mintha kötszóval megvalósított feltételes hasonlatok is bekerültek a vizsgált anyagba. Itt az analógiával az episztemikus inferencia olvasata verseng, és a példák egy részénél szintén nem dönthető el, hogy a konstruáló két jelenet hasonlóságát kívánja jelezni, vagy egy lehetséges magyarázatot szeretne adni a látottakra. (16) például értelmezhető úgy is, hogy a beszélő a megfigyelt karakter soványodásáról szóló vélekedését osztja meg, (22)-ben pedig a hasonlító olvasat (úgy maradt a szájuk egymáson, hogy az feltételesen hasonlít valamilyen fiziológiai és/vagy mentális állapotra) és a következtető olvasat (egymáson maradt a szájuk, amely magyarázható (nak tűnik) valamilyen fiziológiai és/vagy mentális állapottal) között nem is igazán lehet választani. A *mintha*-hasonlatok külön vizsgálatot érdemelnének, mindenesetre ezeknél a példáknál a hasonlat kategóriája más jellegű jelentésalkotási műveletekkel érintkezik.

Az elemzés több vizsgált szerkezetnél is meggyőzően érvel a hasonlítás helyett az azonosítás mellett, és nagyon fontos az is, hogy az elemzők kísérletet tettek e két funkció megkülönböztetésére az adatoknál. Ezen a ponton inkább csak azt javasolnám, hogy az azonosítás helyett besorolásról, kategorizálásról beszéljünk a további elemzésekben, ugyanis a *tanárként*, *királyként*, *anyaként* típusú kifejezések nem azonosítják a diskurzusvilág egy résztvevőjét a tanár, király, anya stb. résztvevővel, hanem – ahogyan a szerzők is írják – a résztvevőt valamilyen szerepkategória példányaként mutatják be. Másként: besorolják a résztvevőt a kategória példányai közé, és a hasonlítás talán a kategória prototípusával való összevetésben kap csak szerepet. A szerepek működésére a szerzők is utalnak, de megítélésem szerint sem önazonosítás („önmagaként láttatja, csupán egy másik szerepben”), sem a szerepek közötti összehasonlítás nem tételezhető. Végül, a *gyanánt* kapcsán felvethető a helyettesítés funkciója: a jó szándék helyettesíti a valódi tettet (26), az áltató hazugságok a valódi vigaszt pótló gesztusok (24). Vélemezhető a cselekvő szándéka, hogy valamilyen cselekvést végrehajtson, amellyel pótolni tudja az elmaradt (ugyanakkor elvárt) viselkedésmódot. Itt tehát ismét nem annyira azonosításról lehet beszélni, miként erre utalnak a szerzők is („átmenet van az egyértelmű azonosítás és a hasonlítás között”).

Mivel az elemzés egyszerre próbál egy funkcióhoz (hasonlítás) több formai megoldást társítani, és ugyanakkor az egyes szerkezeteknél többféle (esetenként érintkező) funkciót is bemutatni, a forma és a funkció vizsgálata

nem mindig egészíti ki egymást, néhol inkább olyan láncszemeknek tűnnek, amelyek egymásba kapcsolódnak, de más irányba tartanak. Az is látszik, hogy a korpusz adatainál az elmosódó kategória határok sok esetben nem segítik az egyértelműsítést. Ha a formai gazdagság felmutatása a cél, akkor célszerű a funkciót jól körülhatárolni, míg ha a mintázatok használatára összpontosít az elemzés, az egyes funkciók árnyalt megkülönböztetésére érdemes törekedni. Azt a skalaritást, amely a hasonlatok szemantikájában a tanulmány elején tematizálódott, az elemzés inkább csak foltszerűen tudta bemutatni.

Ez át is vezet a vizsgálat korpusznyelvészeti jellegének legfőbb dilemmájához: korpuszalapúnak vagy korpuszvezéreltnek tekintendő-e az elemzés? A két terminust hol együttesen, hol váltogatva alkalmazzák a szerzők, ami azonban problematikus, hiszen ezek eltérő szemléletmódok (és ebből következő módszertani döntéssorozatok). A hasonlat korpuszalapú elemzésénél egy meghatározásból kell kiindulni (mit tekintünk általában véve hasonlatnak), amely egy teoretikus konstruktum, és amelyet aztán operacionálizálni kell a korpuszadatok elemzéséhez (el döntve, mit nem tekintünk hasonlatnak, illetve miként kezeljük a határeseteket). Minden olyan esetben, amikor a funkció felől közelített a tanulmány a formai megoldásokhoz, ezt a korpuszalapúságot véltem felfedezni. Ehhez képest a korpuszvezéreltség nyelvi mintázatokból indul ki, minden, a korpuszban elérhető adatot tekintetbe vesz, és azok módszeres elemzése révén jut el a vizsgálni kívánt jelenség meghatározásához. A tanulmányban ismertetett kutatási forgatókönyv (a lekérdezéssel kapott adatok áttekintése, a lekérdezés finomítása, majd az adatok újraelmzése) ez utóbbi szemléletmódot mutatja. A hasonlítás szemantikájának korpuszadatokra épített feltárása, valamint a hasonlat nyelvi lehetőségeinek átfogó feltérképezése együttesen monografikus vállalkozás lenne, a tanulmány vitathatatlan erénye azonban, hogy mindkét irányba tesz lépéseket. A kutatás fókuszában megítélésem szerint a nagyléptékű mintázatkinyerésre irányuló módszertani innováció áll, amely a hasonlatok korpuszvezérelt leírását alapozza meg. A tanulmánynak jót tett volna a fókusz következetes megtartása, ám a hasonlatkutatás elméleti javaslataival való előzetes számvetés, valamint a kapott adatok kiértékelése mindinkább a „mi a (tipikus/sztenderd) hasonlat?” kérdést helyezte előtérbe, amely korpuszalapú gondolkodásmódot vegyített a gondolatmenetbe. A jövőben érdemes lesz ezt a két irányt reflektáltabban megkülönböztetni: a korpuszalapú hasonlatalemzésnek a hasonlat elméleti modelljeiből kiindulva kell megpróbálnia minél több példányhoz eljutni a korpuszban, míg a korpuszvezérelt

megközelítésnek a mintázatkeresésre kell összpontosítania, hogy a centrum felől a periféria felé tágitva mutassa be a kategória szerveződését.

Pontosság (milyen mértékben kapunk valóban hasonlatokat az elemzésekkel) és fedés (megkapjuk-e a korpusz összes hasonlatát az elemzéssel) szempontjai feszülnek egymásnak tehát ebben a kutatásban is, a szerzők pedig egyértelművé teszik a tanulmány végéhez közeledve, hogy vizsgálatukban a magas fedést favorizálták, feláldozva némiképp az azonosítás precizitását. A korpuszvezéreltség általában a fedésnek kedvez, hiszen nincs előzetes meghatározásunk, amelyhez a találatok pontosságát mérhetnénk, és amely mentén az operacionalizálást finomíthatnánk. Innen nézve teljesen érthető, hogy a szerzők is a fedés terén ismerték fel innovációjuk eredményeit. Ugyanakkor a tanulmány a pontos azonosítás kérdéseivel is szembesíti az olvasót, és részben válaszokat is kínál ezekre a kérdésekre (a stopszók beemelésével, vagy a lexikális szemantikai kategóriák meghatározásával), részben pedig a funkciók elkülönítésének problematikus esetein keresztül nyitva is hagyja ezeket. Jól mutatja mindez, hogy az adatelemzés rendre elméleti belátásokhoz is vezeti a kutatót, elmélet és empiria kölcsönviszonyát látjuk a vizsgálatban.

Nem világos ugyanakkor, milyen értelemben tekinthető „referenciakorpusznak” a regényekből összeállított minta. Mivel ez a korpusz a lekérdezések forrása, így a hasonlatelemzés adatbázisa, sokkal inkább kutatói korpusznak tekinthető. A referenciakorpuszok egy szűkebb, kisebb terjedelmű és specifikusabb kutatói korpuszhoz (más terminussal fókuszkorpuszhoz) kínálnak összehasonlítási alapot. Ebben az esetben viszont nem láttam nyomát kétféle korpusz elkülönítésének vagy kialakításának: a módszertani fejlesztést a szerzők egyetlen korpuszon tesztelték, abban mutatták be az elemzés relevanciáját, amely bár valóban nagyléptékű, de nem szolgál összehasonlítási alapként, és így nem is referenciakorpusz.

A vizsgálat magát konstrukciós nyelvtani elemzésnek tekinti, forma és funkció egymásra vonatkoztatott feltárása erre a kiindulóponttra vezethető vissza. A szerzők a tanulmány elején leszögezik, hogy a konstrukciót holisztikus egységnek tekintik, amely „nem szegmentálható egyértelműen lexikai és morfológiai komponensekre”. Ám később a hasonlítást „morfémával, lexémával vagy konstrukciós sémával” tartják megvalósíthatónak, ez pedig azt implicálja, hogy (i) a morféma és a lexéma nem konstrukció, ezeknek nincsen konstrukciós sémájuk, illetve (ii) a konstrukció mindig több elem-ből álló egység. Vagyis, ha a hasonlat konstrukció, akkor hogyan valósulhat meg morfémával és lexémával? Továbbá, ha a konstrukció holisztikus egység,

milyen elméleti alapja van a vizsgálat azon részének, amely egyes szerkezetek tipikus főnévi összetevőit kollokációs adatokra vagy más korpuszbeli gyakorisági adatsorokra építve elemzi? Nyilvánvaló, hogy a konstrukció nem abszolút (vagy totális) holisztikus egység, hanem szegmentálható, komponensekre bontható (például *N+-ként*), vannak külön vizsgálható lexikális összetevői. Ugyanakkor természetesen ezek az egységek, komponensek egészlegesen működtethetők a konstruálás során. Mindenképpen jót tenne a tanulmánynak, ha a konstrukció fogalmát részletesebben bemutatnák a szerzők, már csak azért is, mert a goldbergi alapok mellett egy langackeri-ánus konstrukciófelfogás is több esetben megjelenik az elemzésekben, amely viszont szemantikai integrációra, kompozitumszerkezetek kölcsönös kidolgozási viszonyaira épül, és amely nem azonos a kognitív konstrukciós nyelvtan alapfogalmával. Ráadásul több olyan fogalom, amely a konstrukcióhoz kapcsolódik, például mezokonstrukció, alkonstrukció, meghatározatlan marad. A hasonlítás művelete köré vélhetően kiterjedt és rétegzett konstrukciós hálózat szerveződik a magyarban is, ennek felfejtésére a tanulmány nem vállalkozhat természetesen. Azt azonban tisztázni kellene, milyen értelemben konstrukció a hasonlat (lásd a goldbergi és a langackeri konstrukciófogalom különbségeit), a vizsgált mintázatok milyen viszonyban állnak ezzel a tételezett konstrukcióval, és a bemutatott formai variabilitás konstrukciókat vagy konstrukcióváltozatokat (herbsti terminussal allostrukciókat) különít-e el.

Végezetül, a vizsgálat alkalmaz stíluselméleti fogalmakat, kategóriákat is, amely magától értetődőnek tűnhet a vizsgálat tárgya, egy a priori stílus-eszköz esetében. Sajnos azonban a tanulmánynak nincs stíluselméleti kiindulópontja, legalábbis ilyen nem mutatnak be a szerzők, annak ellenére, hogy stílusról és poétikai funkcióról is tesznek megállapításokat. Különösen zavaróan hatott a stilisztikai jelöltség és jelöletlenség megállapítása az egyes szerkezeteknél, ugyanis nem világos, milyen értelemben válnak egyes megoldások jelöltebbé: a nyelvi potenciál előtérbe kerülésével, a konvencionális, elvárt megoldásoktól való eltávolodás révén, vagy a szociokulturális tényezők terén (például a gyanántnak tulajdonítható enyhe régiesség kapcsán). Mivel a vizsgálat igen nagy vállalást tesz a korpuszvezérelt mintázatkinyerés és -elemzés megvalósításával, kiértékelésével és adaptálásával a prózanyelv alakulástörténetének új szempontú bemutatásában, a stilisztikai, stíluselméleti aspektus végképp megterheli a kifejtést, anélkül, hogy támpontokat adna az olvasónak a hasonlattípusok rendszerszerű leírásában. Ennek az aspektusnak az átmeneti elhagyása javasolható, hiszen a vizsgálat a stílusnál általánosabb nyelvelméleti és kutatómódszertani kérdésekkel szembesíti az olvasót.

Ráadásul joggal remélhető, hogy a javasolt módszertani innováció elvezet majd a hasonlatok nagyléptékű elemzéséhez a stílustulajdonítás terén is.

Összességében véve azzal a közhelynek tűnő, de most nagyon is szó szerint értett formulával zárom a hozzászólást, hogy hiánypótló munkát végeztek a tanulmány szerzői. Egy, a kurrens kognitív nyelvészetben egyre nagyobb figyelmet kapó jelenség átfogó, korpusznyelvészeti eszközökkel elvégzett elemzését mutatták be, amely magának a jelenségnek az árnyalt megközelítését is felmutatta, eszközt adott a jelenség iránt érdeklődő kutatók kezébe, megvonta az eszköz eredményességének mérlegét (utat nyitva továbbfejlesztése előtt), és rámutatott annak alkalmazhatóságára tágabb értelemben vett digitális bölcsészeti kutatásokban. Ezek után csak bízni lehet abban, hogy az első lépéseket továbbiak követik, és hamarosan jól kitaposott úton indulhatnak el a kognitív korpusznyelvészet iránt érdeklődő kutatók.

Irodalom

- Beardsley, Monroe 1958. *Aesthetics: Problems in the philosophy of criticism*. New York: Harcourt, Brace & World.
- Fishelov, David 2007. Shall I compare thee? Simile understanding and semantic categories. *Journal of Literary Semantics* 36: 71–87. <https://doi.org/10.1515/JLS.2007.004>
- Shen, Yeshayahu 2008. Metaphor and poetic figures. In: Gibbs, Raymond W. (ed.): *The Cambridge handbook of metaphor and thought*. Cambridge: Cambridge University Press. 295–307. <https://doi.org/10.1017/CBO9780511816802.019>
- Tartakovsky, Roi – Shen, Yeshayahu 2018. 'Simple as a fire': Making sense of the non-standard poetic simile. *Journal of Literary Semantics* 47 (2): 103–119. <https://doi.org/10.1515/jls-2018-2002>
- Tartakovsky, Roi – Shen, Yeshayahu 2019. Meek as milk and large as logic: A corpus study of the non-standard poetic simile. *Language and Literature* 28 (3): 203–220. <https://doi.org/10.1177/0963947019865445>

Simon Gábor

Eötvös Loránd Tudományegyetem

simon.gabor@btk.elte.hu

ORCID: 0000-0001-5233-63

**A lexikai választékosság
és szemantikai
specifikusság mintázatai**



HORVÁTH PÉTER

A lexikai választékosság és szemantikai specifikusság mintázatai a magyar kanonikus költészetben¹

Kivonat

A tanulmány az idegen nyelv tanulását vizsgáló kutatásokban használt, a szövegben szereplő szavak referenciakorpuszbeli gyakoriságán alapuló lexikai választékosságnak (lexical sophistication) a mérése alapján vizsgálja 24 magyar kanonikus költő verseiben a szókincs és a terjedelem, a szókincs és a versforma, valamint a szókincs és a vers szerkezeti egységeinek az összefüggését. Egyrészt bemutatja, hogy a magyar kanonikus költészetben a közepes terjedelmű versek lexikai választékossága jellemzően valamivel alacsonyabb a rövid és hosszabb versek lexikai választékosságánál. Másrészt rámutat arra, hogy a formailag egyszerű versek lexikai választékossága jellemzően alacsonyabb, ami a versforma és a tartalom ikonikus összefüggésével magyarázható. Harmadrészt rámutat arra is, hogy a versek előrehaladásával a vers lexikai választékosságának a mértéke általában csökken, ami összefüggésbe hozható a vers tartalmi általánosulásával. Végezetül megállapítja azt az általános mintázatot is, hogy a versszakon belül a páratlan sorokra alacsonyabb, míg a páros sorokra magasabb lexikai választékosság jellemző, ami egy olyan, a sorpárok által alkotott struktúrát feltételez, amelyben a sorpár első sora által kifejezett általánosabb és a második sora által kifejezett specifikusabb jelentésegységek kerülnek egymással viszonyba. A kvantitatív eredmények alapján leírhatóvá válik a magyar kanonikus költészetre jellemző prototipikus vers lexikai választékosságának és a lexikai választékosságból levezetett szemantikai specifikusságának a versszakok és verssorok szintjén érvényesülő sémája.

¹ A tanulmány az NKFIH K 137659 sz. pályázat (A személyjelölés konstrukcióinak korpuszalapú, kognitív poétikai vizsgálata) és a Digitális Örökség Nemzeti Laboratórium támogatásával készült.

Kulcsszavak: típus-token arány, lexikai választékosság, szemantikai specifikusság, ELTE Verskorpusz, versforma, ikonikusság, versszak, verssor

1. Bevezetés

A tanulmány azt vizsgálja, hogy a magyar kanonikus költészetben milyen mintázatok tárhatók fel a versek, illetve a versek szerkezeti egységeinek a lexikai választékossága (lexical sophistication) és a lexikai választékosságból levezetett szemantikai specifikussága alapján, és hogy a megállapított mintázatokat milyen tényezők motiválhatják. A lexikai választékosság az idegen nyelv elsajátítását vizsgáló kutatások egyik gyakori mérőszáma, amely arra utal, hogy egy adott szöveg a szóhasználatát tekintve mennyire tekinthető választékosnak abban az értelemben, hogy a szövegben szereplő szavak mennyire gyakoriak egy referenciakorpuszból kinyert gyakorisági lista alapján.

A tanulmány egyrészt azt vizsgálja, hogy a versek bizonyos formai tulajdonságai mutatnak-e valamilyen összefüggést a versek lexikai választékosságával, illetve amennyiben mutatnak, mivel magyarázható ez az összefüggés. A kutatás másik fő kérdése az, hogy a versek különböző részei megvalósítanak-e valamilyen tipikus mintázatot a lexikai választékosság szempontjából. Vagyis az első kérdéssel szemben a második kérdés nem a versek egészét jellemző formai tulajdonságokra, hanem a versek belső szerkezetére koncentrál: a vizsgálat arra keresi a választ, hogy a vers előrehaladása során vannak-e tipikus mintázatai a lexikai választékosság változásának, illetve amennyiben vannak, mivel magyarázható ez.

A 2. rész először bemutatja a szókincsvizsgálatoknál tipikusan használt típus-token arány mérésének a legelterjedtebb módszereit, rámutatva arra, hogy ezek nem igazán alkalmazhatók hatékonyan versek vizsgálatára a versek hosszúságbeli különbségei, illetve a nagyszámú rövid terjedelmű vers miatt. Ezt követően ismerteti a lexikai választékosság vizsgálatára kidolgozott módszert, amely versek szókincsvizsgálatára is jól használható. A 3. rész mutatja be a kutatáshoz használt korpuszt és a lexikai választékosság mérésének az itt alkalmazott módszerét. A 4. rész ismerteti a vizsgálatok eredményeit. A 4.1. rész bemutatja, hogy a versek hosszúsága összefüggést mutat a versek lexikai választékosságával: a közepes terjedelmű versek lexikai választékossága jellemzően valamivel alacsonyabb a rövid és a hosszú versek lexikai választékosságánál. A 4.2. rész rámutat arra, hogy a versforma sem független a lexikai választékóságtól: a formailag egyszerű versek

lexikai választékossága jellemzően alacsonyabb a többi vers lexikai választékosságánál, ami forma és tartalom ikonikus összefüggésével magyarázható. A 4.3. rész rámutat arra, hogy a versek első, középső és utolsó versszakainak a lexikai választékossága egy meghatározott mintázatot követ: a vers előrehaladásával a lexikai választékosság mértéke jellemzően csökken, ami a versek tartalmi általánosulásával magyarázható. A 4.4. rész bemutatja, hogy a versszakok páratlan és páros sorainak a lexikai választékossága esetében is megjelenik egy jól meghatározható tendencia: a páros sorok lexikai választékossága jellemzően magasabb, mint a páratlan sorok lexikai választékossága, ami azzal magyarázható, hogy a páratlan és páros sorok által alkotott sorpárok olyan jelentésszerkezetek, amelyben a páratlan sor által kifejezett általánosabb jelentés kerül viszonyba a páros sor által kifejezett specifikusabb jelentéssel. Végezetül a 4.5. rész a lexikai választékosságnak a vers három fő része esetében és a sorok esetében megállapított, makro- és mikroszinten érvényesülő mintázata alapján leírja a magyar kanonikus költészetre jellemző prototipikus vers lexikai választékosságának és a lexikai választékosságból levezetett szemantikai specifikusságának a kétszintű sémáját.

2. A szókincsvizsgálat módszereinek alkalmazhatósága versek vizsgálatára

2.1. A típus-token és lemma-token arány mérésének problémái

A szókincs vizsgálatának egyik legelterjedtebb módszere a típus-token arány kiszámítása, aminek során a szövegben szereplő különböző szóalakok számát elosztják a szövegben szereplő tokenek, azaz a szöveg összes szavának a számával. Az osztás eredménye egy 0 és 1 közötti érték. Minél közelebb van ez az érték az 1-hez, annál több különböző szóalak jut az összes szóra, azaz a szöveg szókincse annál diverzebb. Ennek a módszernek egy továbbgondolt változata a lemma-token arány számítása. A lemma egy szó szótári alakja. Például az *ablak*, *ablakot* *ablakon*, *ablakkal*, *ablakunk* szavak mind külön típusok, ugyanakkor a lemmájuk mindegyiknek az *ablak* szó. Már ebből a példából is látható, hogy a magyar és a magyarhoz hasonló ragozó nyelvek esetében a típus-token arány helyett a lemma-token arányt érdemes használni, ami persze megköveteli a szövegek lemmatizálását. A típus-token és a lemma-token arány mérésének módszertani kérdései megegyeznek egymással. Az alábbiakban szereplő, típus-token aránnyal kapcsolatos

módszerek egy az egyben alkalmazhatók a lemma-token arány számítására is, és ugyanígy, a típus-token arány mérésével kapcsolatos problémák érvényesek a lemma-token arány mérésére is.

A típus-token, illetve a lemma-token arány számításának a fő problémája, hogy a mérőszám érzékeny a szövegek hosszúságára. A szövegek hosszúsága negatívan korrelál a típus-token arány értékével, azaz a hosszabb szövegeknek jellemzően alacsonyabb a típus-token aránya, mint a rövidebb szövegeknek. Ezt nem nehéz belátni, hiszen ha egy szöveg egymást követő szavait nézzük, akkor egyre kevesebb az esélye annak, hogy olyan szóval találkozunk, amely még nem fordult elő korábban. Mindez tehát azt jelenti, hogy különböző hosszúságú szövegek, például versek típus-token arányainak az összevetése valójában semmilyen értelmes információval nem tud szolgálni egy szókincs-vizsgálatra irányuló kutatás számára.

A típus-token arány szöveghosszúságra való érzékenységet két módon próbálták megoldani. Az egyik irányvonal arra törekedett, hogy a típus-token arány kiszámításához használt egyszerű osztást olyan további matematikai műveletekkel egészítsék ki, amelyek ellensúlyozzák a szöveghosszúság hatását. Emögött az az elképzelés húzódott meg, hogy van valamilyen matematikai állandó, amellyel a különböző hosszúságú szövegek típus-token arányai normalizálhatók, és így értelmesen összevethetők egymással. Ezeknek a megoldásoknak a nagy részéről azonban kiderült, hogy továbbra is érzékenyek maradnak a szöveghosszúságra (Tweedie–Baayen 1998).

A másik irányvonal nem a különböző hosszúságú szövegekből kapott típus-token arányok normalizálására törekedett, hanem magának a szöveghosszúságnak a normalizálására, megtartva a típus-token arány számításának az egyszerűségét. Ennek az irányvonalnak a legegyszerűbb megoldása az, hogy minden szövegnek csak az első X szavát veszik számításba. Nem nehéz ugyanakkor belátni, hogy az egyes szövegek felvezető részei nem reprezentatívak az egész szövegre vonatkozóan. Egy másik megközelítés felosztja egymást követő, egyenlő szószámú szegmensekre a szöveget, a szegmenseknek külön kiszámolja a típus-token arányait, majd az így kapott értékeket átlagolja (Johnson 1944; Scott 2007: 157–158.). Ez a megközelítés mindenképpen jobb, mint az előző, hiszen nem csak a szövegek elejét veszi figyelembe. Problematiszus ugyanakkor, hogy mivel a szövegek szószáma többnyire nem egész számú szorzata a szegmensek előre definiált szószámának, a szöveg vége kimarad a vizsgálatból. Egy másik problémája a módszernek, hogy a szöveg lexikailag eltérő szerkezeti részei és a szegmensek közötti véletlen megfelelés jelentősen torzíthatja az eredményt. E

problémákat jól kezeli a típus-token arány mozgó átlagának (moving-average type-token ratio) a kiszámítása (Covington–McFall 2010). Ebben az esetben is ugyanolyan hosszú, azonban egymást átfedő szegmenseknek számolják ki a típus-token arányát, majd az így kapott értékeket átlagolják. Például 100 szavas szegmensek használata esetén kiszámolják a szöveg 1.-től a 100. szóig, majd a 2.-től a 101. szóig, majd a 3.-től a 102. 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 szegmensenst számításba veszi az eljárás, a szöveg szerkezeti egységei és a szegmensek esetleges egybeesésének a torzító hatása is kiküszöbölhető.

Bár az utóbbi két módszer alkalmas különböző hosszúságú szövegek típus-token arányainak az értelmes összevetésére, bizonyos hosszúságnál rövidebb szövegek elemzésére ezek a módszerek sem megfelelőek. A szegmenseknek ugyanis minimum 100 szó hosszúságúnak kell lenniük ahhoz, hogy megbízható eredményekhez jussunk. Túl rövid szegmensek esetén ugyanis a módszer a szóismétlések nagy részét már nem érzékeli (Malvern–Richards 2002). A versek jelentős része azonban rövidebb 100 szónál. Mindez tehát azt jelenti, hogy a versek hosszúságának a jelentős különbségei, illetve a rövid versek nagy száma miatt a típus-token arányt mérő egyik módszer sem használható hatékonyan versek szókincsének vizsgálatára. A típus-token arány mérése mellett ugyanakkor egyéb lehetőségek is vannak egy szöveg szókincsének a kvantitatív vizsgálatára (lásd Jarvis 2013). Ilyen például az idegen nyelv elsajátítását vizsgáló kutatásokban kidolgozott lexikai választékosság mérése, amely jól használható versek szókincsvizsgálatára is.

2.2. Lexikai választékosság

Az idegen nyelv tanulásával, illetve tanításával foglalkozó kutatásoknak egyik fontos kérdése, hogy milyen módon lehet mérni a nyelvtanulók lexikai tudásának a szintjét. A nyelvtanulók idegen nyelvi szókincsének a felmérésére használatos egyik legtipikusabb módszer a tanulók által létrehozott szövegek lexikai választékosságának (lexical sophistication) a mérése. A lexikai választékosság mérésének alapja többnyire a szövegekben szereplő szavak gyakoriságának a megállapítása valamilyen referenciakorpuszból kinyert gyakorisági lista alapján (Laufer–Nation 1995). Az egyes tanulók lexikai profilját az alapján adja meg a módszer, hogy a tanuló szövegében szereplő szavak milyen arányban tartoznak különböző gyakorisági tartományokba,

például a leggyakoribb első ezer szóba, a leggyakoribb második ezer szóba stb. Minél több kevésbé gyakori szó szerepel a tanuló szövegében, annál választékosabb a szóhasználata, azaz annál magasabb szintű a lexikai tudása.²

A módszert számos módon továbbfejlesztették az utóbbi két évtizedben. Például módosított listákat használtak, figyelembe véve, hogy azokat a ritka szavakat könnyebb megtanulni, amelyek fonológiailag és szemantikailag hasonlítanak a tanuló anyanyelvének valamely szavára, illetve amelyek a nyelvórák, nyelvkönyvek tipikus témáinak a szavaiként jelennek meg (Bardel et al. 2012). Emellett anyanyelvi beszélők szövegeit tartalmazó korpuszok helyett nyelvtanulók szövegeit tartalmazó korpuszokat használtak a gyakorisági listák előállítására (Naismith et al. 2018), vagy éppen bigrammok és trigrammok gyakoriságának a mérését integrálták a módszerbe (Kyle–Crossley 2015). Mindazonáltal a módszer továbbgondolt változataiban is a lexikai választékosság mérésének alapja maradt a tanuló szövegében szereplő szavak kategorizálása vagy pontozása azoknak egy referenciakorpuszban mért gyakorisága alapján.

Szemben a típus-token arány mérésével, a lexikai választékosság mérése nem érzékeny a szövegek hosszúságára. Emellett a lexikai választékosság mérése – szemben a típus-token arányt szövegszegmensekben mérő megközelítésekkel – száz szónál rövidebb szövegeken is hatékonyan elvégezhető, a módszer alkalmazása szempontjából tulajdonképpen nincsen alsó határa a szövegek hosszúságának. Mindez tehát azt jelenti, hogy a módszer az eltérő hosszúságú, sok esetben kifejezetten rövid versek szókincsvizsgálatára is jól alkalmazható. Természetesen versek esetében a magasabb vagy alacsonyabb lexikai választékosság önmagában nem jelent semmiféle minőségbeli különbséget.

A későbbiekben szereplő magyarázatok egyik fontos kiindulópontja, hogy a lexikai választékosság mértéke összefüggésbe hozható a szemantikai specifikussággal. Jellemzően ugyanis a kevésbé gyakori, és ennek folytán nagyobb lexikai választékosságú szöveget, szövegrészt eredményező szavak specifikusabb jelentésűek, mint a gyakoribb szavak. Vagyis míg a gyakori nem funkciószavak általában a hétköznapi gondolkodás általánosabb kategóriáira vonatkoznak, addig a ritkább szavak többnyire ezeknél specifikusabb kategóriákra (entitásokra, minőségekre, eseményekre) utalnak, illetve olyan fogalmakra, amelyeket csak speciálisabb kontextusokban használunk (a lexémák eltérő szemantikai specifikusságáról lásd Cruse 1977; Langacker 2008: 55–57).

² A lexikai választékosságnak ez az értelmezése tehát különbözik a funkcionális stilisztikai értelmezéstől, amely a választékosságot a megnyilatkozónak, illetve a befogadónak a szöveg megformáltságához való, szociokulturálisan meghatározott viszonyulása mentén értelmezi (Tolcsvai Nagy 1996; Tátrai–Ballagó 2020).

3. Korpusz és módszer

3.1. Korpusz

A vizsgálat korpuszát 24 magyar költő összes ezer szónál rövidebb verse alkotja, amelyeket automatizáltan nyertem ki az ELTE Verskorpuszból (Horváth et al. 2022a). A kutatási korpuszba az ELTE Verskorpusz azon szerzői kerültek bele, akik nem születtek Csokonainál régebben, illetve akiknek legalább 100 versük van. Az ezer szavas limit oka az volt, hogy a vizsgálatot igyekeztem a lírai műnemre szűkíteni. A hosszú versek kiszűrésével az elbeszélő költemények egy jelentős részét el lehetett távolítani a vizsgálati anyagból. A kutatási korpuszban szereplő szerzőket az 1. táblázat mutatja be időrendben, az ezer szónál rövidebb versek számával együtt. A kutatási korpusz összesen 11 420 verset tartalmaz.

1. táblázat. A kutatási korpusz

Költő	Versek száma	Költő	Versek száma
Csokonai Vitéz Mihály	376	Komjáthy Jenő	246
Berzsenyi Dániel	136	Ady Endre	1116
Kisfaludy Károly	119	Kaffka Margit	99
Kölcsey Ferenc	150	Somlyó Zoltán	376
Vörösmarty Mihály	645	Juhász Gyula	1278
Arany János	388	Babits Mihály	510
Tompa Mihály	472	Kosztolányi Dezső	626
Petőfi Sándor	829	Tóth Árpád	451
Madách Imre	313	Reményik Sándor	667
Gyulai Pál	154	József Attila	597
Vajda János	186	Dsida Jenő	1055
Reviczky Gyula	334	Radnóti Miklós	297

Az XML-formátumú ELTE Verskorpuszban különböző automatikusan létrehozott annotációs rétegek szerepelnek. Egyrészt annotálva vannak a versek szerkezeti jellemzői: a versszakok és a sorok. Másrészt annotálva van a versek szavainak lemmája, szófaja és morfoszintaktikai jellemzői. Ezek az annotációk az e-magyar programmal lettek létrehozva (Váradi et al. 2017; Novák et

al. 2017; Indig et al. 2019). Harmadrészt a korpusz számos vershangzáshoz kapcsolódó poétikai jellemző annotációját is tartalmazza. Például szerepelnek a korpuszban a metrum (Horváth 2021), a rímképlet (Horváth 2023) és a verssorok szótagszámainak az annotációi. Ez a kutatás a versek szerkezeti jellemzőire, a lemmára, valamint a rímképletre és a sorok szótagszámaira vonatkozó annotációkat használja fel.

Mivel a kutatási korpusz elsősorban nagymértékben kanonikus magyar szerzők verseit tartalmazza, az nem alkalmas a magyar költészet egészére vonatkozó egyértelmű következtetések levonására. A kutatási korpusz tehát nem tekinthető egy nála nagyobb szövegcsoporthoz, az összes magyar vers reprezentatív mintájának, hiszen a korpuszban szereplő szerzők pont azért kanonizálódtak, mert a verseik a náluk korábbi és a velük kortárs szerzők verseihez képest eltérő nyelvi jellemzőket mutatnak.³ A kvantitatív vizsgálatok alább bemutatandó eredményei így a teljes magyar költészetre vagy éppen más nyelvek költészetére vonatkozóan pusztán erős hipotéziseknek tekinthetők, amelyeket további, kevésbé vagy nem kanonikus, illetve más nyelven alkotó költők verseinek a vizsgálatával lehetne igazolni.

3.2. Módszer

A vizsgálat során a versek szavainak a gyakoriságát a szavak két korpuszban való gyakorisága alapján határoztam meg. A szógyakoriságok megállapításához használt egyik referenciakorpusz az ELTE Regénykorpusz⁴ volt, amely 400 magyar regényt tartalmaz többek között a szavak lemmájának úgyszintén az e-magyar programmal létrehozott automatikus annotációival. Az ELTE Regénykorpusz az ELTE Verskorpuszhoz hasonló időszakot ölel fel, ugyanúgy a 20. század első feléig tartalmaz regényeket. A versekben szereplő szavak gyakoriságának a megállapításához használt másik korpuszt maga a kutatási korpusz adta, azaz az ELTE Verskorpusznak a fentebb ismertetett kritériumok alapján kiválasztott versei. A két korpuszt nem összevontan használtam: minden vizsgálat esetében megadom az ELTE Regénykorpusz és az ELTE Verskorpusz alapján meghatározott mértékét a lexikai választékoságnak.

³ Ugyanakkor persze bizonyos költők bizonyos újításai mintaként szolgálhatnak más szerzők számára.

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

A versek szavainak gyakoriságát a szavak Regénykorpuszbeli, illetve Verskorpuszbeli relatív gyakoriságai alapján határoztam meg. Ehhez tehát előzetesen ki kellett számítani, hogy a két referenciakorpuszban az egyes szavak lemmái milyen arányban fordulnak elő a referenciakorpusz összes szavához képest. A relatív gyakoriságokat egymillió szavankénti előfordulásra normalizáltam. A kapott relatív gyakoriságok alapján a lemmákat négy csoportba soroltam. Az első csoportba kerültek a leggyakoribb lemmák, amelyeknek az egymillió szóra normalizált relatív gyakorisága több, mint 1000. A második csoportba a 101 és 1000 közötti, a harmadik csoportba a 11 és 100 közötti, a negyedik csoportba pedig a 11-nél kisebb relatív gyakoriságú lemmák kerültek.⁵ A 2. táblázat mutatja be, hogy a Verskorpusz és a Regénykorpusz esetében hány lemma került bele a négy gyakorisági kategóriába.

2. táblázat. Az egyes gyakorisági tartományokba eső lemmák száma a két referenciakorpuszban

	1001– (1 pont)	101–1000 (2 pont)	11–100 (3 pont)	0–10 (4 pont)
ELTE Verskorpusz	122	1043	5354	77 826
ELTE Regénykorpusz	110	939	6135	415 718

A versek, illetve a versek részeinek a lexikális választékosságát egy pontosos módszerrel számítottam ki. Első lépésben a versekben szereplő szavakat pontosztam az alapján, hogy az egyes szavak a referenciakorpuszban való relatív gyakoriságuk alapján melyik gyakorisági csoportba kerültek. Minél ritkább egy versbeli szó a referenciakorpuszban, annál több pontot kapott. A leggyakoribb, ezernél nagyobb relatív gyakoriságú szavak egy pontot, a száznál nagyobb relatív gyakoriságú szavak két pontot, a tíznél nagyobb relatív gyakoriságú szavak három pontot, míg a legritkább, tízenegynél alacsonyabb relatív gyakoriságú szavak négy pontot kaptak. Ügyszintén négy pontot kaptak azok a szavak, amelyek egyáltalán nem szerepeltek a referenciakorpuszban. Ez az utóbbi eset természetesen csak akkor

⁵ A gyakorisági tartományok megállapításához azért nem azt az elsőre triviálisabbnak tűnő módszert használtam, hogy egyszerűen a gyakoriság alapján sorba rendezett lemmákból jelöltem ki tartományokat (például első 100 lemma, 101-től 1000-dik lemmáig stb.), mert számos lemma azonos előfordulással szerepel a két korpuszban, például a rangsorban az ezredik lemma előtti és utáni lemmák is ugyanannyi előfordulással szerepelnek, mint az ezredik lemma.

fordulhatott elő, amikor a Regénykorpusz volt a referenciakorpusz. Az (1) példa Ady Endre *Góg és Magóg fia vagyok én* című versének első versszakát mutatja be a szavakhoz tartozó gyakorisági pontszámokkal. A pontszámok kiszámításának referenciakorpuszát a Verskorpusz adta.

- (1) Góg [4] és [1] Magóg [4] fia [2] vagyok [1] én [1],
 Hiába [2] döngetek [3] kaput [2], falat [2]
 S [1] mégis [1] megkérdem [3] tőletek [1]:
 Szabad [2] -e [1] sírni [1] a [1] Kárpátok [3] alatt [1] ?

A pontozás következő lépése a versek, illetve bizonyos vizsgálatokban a versen belüli szövegrészek lexikai választékossági pontszámainak a kiszámítása. Ezt a versekben, illetve versen belüli szövegrészekben szereplő szavak pontjainak az átlagolásával végeztem el. Például az (1)-ben szereplő Ady-versszak szógyakorisági pontszámainak az átlaga 1,85. A versekre, illetve szövegrészekre kapott szógyakorisági pontok átlagai már az adott szöveget vagy szövegrészt jellemzik a lexikai választékosság szempontjából. Minél több kevésbé gyakori vagy ritka szó szerepel egy adott versben vagy szövegrészben, annál nagyobb lesz a vers vagy szövegrész lexikai választékossága.

A fent leírt összes műveletet, a szavak referenciakorpuszbeli relatív gyakoriságainak a kiszámítását, a vers szavainak a pontozását, illetve a versek és a versekben szereplő szövegrészek lexikai választékosságának a kiszámítását a szavak pontszámainak az átlagolásával automatizáltan hajtotta végre egy a kutatáshoz írt szkript, csakúgy, mint az egyes szerzők verseire, illetve a versek szerkezeti egységeire kapott lexikai választékossági értékek középértékeinek a kiszámítását.

4. Eredmények

4.1. Vershosszúság és lexikai választékosság

A kutatás első megválaszolandó kérdése az, hogy van-e valamilyen összefüggés a versek hosszúsága és lexikai választékossága között. A kérdés megválaszolásához a 24 költő verseit három terjedelmi kategória alapján vizsgáltam. Az első terjedelmi kategóriába az 1–49 szószámú versek kerültek, a második terjedelmi kategóriába az 50–199 szószámú, a harmadik terjedelmi kategóriába pedig a 200–999 szószámú versek. Első lépésben minden terjedelmi

kategória esetében szerzőnként külön kiszámítottam az egyes versek lexikai választékossági pontszámainak az átlagát. (Ezeket az értékeket, illetve az egyes terjedelmi csoportokba tartozó versek számát lásd a Függelékben szereplő 13. táblázatban.) Ezt követően a szerzőkre kapott átlagértékeket átlagoltam mind a három terjedelmi kategória esetében. Ezt a számítást mind a két referenciakorpusz, a Verskorpusz és a Regénykorpusz alapján is elvégeztem. A kapott értékeket a 3. táblázat mutatja be.

3. táblázat. A három terjedelmi kategória átlagos lexikai választékossága

	1–49	50–199	200–999
Verskorpusz	1,945	1,923	1,945
Regénykorpusz	2,108	2,102	2,101

A kapott átlagértékekből látható, hogy amennyiben a Verskorpuszt használjuk referenciakorpuszként az 50–199 szavas, közepes terjedelmű versek átlagos lexikai választékossága valamivel alacsonyabb, mint az 50 szó alatti rövid versek, illetve a 199 szó feletti hosszú versek átlagos lexikai választékossága. Ha a Regénykorpusz adja a referenciakorpuszt, akkor a három terjedelmi csoport átlagos lexikai választékossága közötti különbség minimális. Érdekes azt is megnézni, hogy a 24 szerzőből hány teljesíti teljesen vagy részlegesen ezt a mintázatot. A 4. táblázat ezt mutatja be a szerzőkre kapott átlagok és mediánértékek alapján. (A mediánértékeket lásd a Függelék 14. táblázatában.)

4. táblázat. A lexikai választékosság és a vershosszúság összefüggésének mintázatát részben vagy teljes egészében megvalósító szerzők száma (R: rövid versek;

K: közepes terjedelmű versek; H: hosszú versek)

	ÁTLAG			MEDIÁN		
	K < R	K < H	K < R & K < H	K < R	K < H	K < R & K < H
Verskorpusz	15	17	11	15	20	11
Regénykorpusz	14	12	6	12	13	8

A 4. táblázatból látható, hogy ha a Verskorpuszt használjuk referenciakorpuszként, akkor az átlagok és a mediánok esetében is a 24 szerzőből 15 szerzőnél érvényesül az, hogy a közepes terjedelmű versek lexikai választékosságának a középértéke alacsonyabb a rövid terjedelmű versek lexikai

választékosságának a középértékénél, és 17, illetve 20 szerző esetében jelenik meg az a tendencia, hogy a közepes terjedelmű versek lexikai választékosságának a középértéke alacsonyabb a hosszú versek lexikai választékosságának a középértékénél. A teljes minta, azaz az az eset, amikor a közepes terjedelmű versek lexikai választékosságának a középértéke alacsonyabb a rövid és a hosszú versek lexikai választékosságának a középértékeinél is, az átlag- és a mediánértékek esetében is 11 szerzőnél érvényesül. Ha a Regénykorpuszt használjuk referenciakorpuszként, akkor a mintázatnak csak azt a részét valósítja meg nagyobb számú szerző, amely szerint a közepes terjedelmű versek átlagos lexikai választékossága alacsonyabb a rövid terjedelmű versek átlagos lexikai választékosságánál: ezt 14 szerzőnél lehet megfigyelni.

Arra a kérdésre, hogy mivel magyarázható a Verskorpuszt referenciakorpuszként használó vizsgálattal megállapított azon tendencia, miszerint a kanonikus magyar költészetben a rövid és hosszú verseknek tipikusan valamivel nagyobb a lexikai választékossága, mint a közepes hosszúságú verseknek, ennek a vizsgálatnak a keretében csupán hipotézisértékű válaszok adhatók, amelyeket további vizsgálatok erősíthetnének meg. A magyarázat abból a már említett jelenségből indulhat ki, hogy míg a versek gyakori nem funkciószavai jellemzően a hétköznapi gondolkodás számára jól ismert, általános kategóriákra vonatkoznak (vö. Horváth et al. 2022b, Horváth–Tátrai 2023), addig a ritkább szavak többnyire ezeknél specifikusabb kategóriákra (entitásokra, minőségekre, eseményekre) utalnak. A rövid versek magasabb lexikai választékossága talán azzal magyarázható, hogy mivel a megnyilatkozó lírai én kevesebb szóban közvetíti mondanivalóját, a gondolatok, érzések fókuszáltabb megfogalmazást kapnak, ami szükségessé teheti a specifikusabb jelentésű szavak nagyobb mértékű használatát.

Míg a rövid verseknél a fókuszáltabb megfogalmazás igénye okozhatja a specifikusabb jelentésű és ezért ritkább szavak használatát, addig a hosszabb versek esetében pont fordítva, a versekbe bekerülő részletező leíró részek adhatnak nagyobb teret a specifikusabb jelentésű ritkább szavak használatának. A hosszabb versekben a lírai énnak lehetősége van környezetének aprólékosabb megfigyelésére, gondolatainak, érzéseinek részletezőbb közvetítésére, illetve a megfigyelt külső világ részleteinek és saját belső világának kidolgozottabb metaforikus összekapcsolására. A lírai énnak ez a hosszabb versekben megjelenő, részletekre (is) fókuszáló attitűdje eredményezheti, hogy bizonyos szerzőknél a specifikusabb jelentésű ritkább szavak valamivel nagyobb mértékben jelennek meg a hosszú versekben, mint a közepes hosszúságú versekben.

4.2. Versforma és lexikai választékosság

A második vizsgálandó kérdés a versforma és a lexikai választékosság összefüggésére vonatkozik. A versformának természetesen nagyon sok jellemzője bevonható egy vizsgálatba a jellemzők kategorizálásának a legkülönbözőbb részletezettségével. Ez a vizsgálat csupán annak a kérdésnek a megválaszolására törekszik, hogy a magyar kanonikus költészetben a formailag egyszerű versek lexikai választékossága mutat-e valamilyen eltérést a többi vers lexikai választékosságától. Ennek vizsgálatához azt az egyszerű módszert választottam, hogy minden szerző esetében a verseket két csoportba soroltam. Az első csoportba kerültek a formailag kifejezetten egyszerű versek, azaz azok a versek, amelyek a lehető legteljesebb mértékben megfelelnek az egyféleség szabályának, azaz az izosabálynak (Horváth 1991: 152–170, 2008). Ezek a versek (1) izometrikusak, azaz a versszakok minden sora ugyanannyi szótagszámú, (2) izorímeseek, azaz a versszakok minden sora rímel minden másik sorral (bokorrím), és (3) izostrofikusak, azaz a versszakok ugyanolyanok mind a sorok számát mind a sorok szótagszámát tekintve. További feltételként szabtam meg, hogy a vers sorainak a száma minimum négy legyen, mivel a négyenél kevesebb sorból álló szövegek esetében nem igazán értelmezhető a formai egyszerűség kritériuma. A másik csoportba került az összes fennmaradó vers. Az itt követett megközelítésben tehát a formai egyszerűséget azonosítom a variabilitás nélküli ismétlődéssel, azaz a lehető legteljesebben érvényesülő formai homogenitással.

A 24 szerzőből 18 írt a fenti kritériumoknak megfelelő, „formailag egyszerű” verset, így a vizsgálatba csak ezeknek a szerzőknek a versei kerültek bele. Első lépésben mindegyik szerzőnél kiszámítottam mind a két verscsoport esetében a versek lexikai választékossági pontszámainak az átlagát. (Ezeket az értékeket, illetve az egyszerű és nem egyszerű versek számát lásd a Függelék 13. táblázatában.) Ezt követően mind a két verscsoport esetében átlagoltam az egyes szerzőkre kapott átlagértékeket. A két referenciakorpusz alapján kapott eredményeket az 5. táblázat mutatja be.

5. táblázat. A formailag egyszerű és nem egyszerű versek átlagos lexikai választékossága

	Egyszerű versek	Többi vers
Verskorpusz	1,879	1,932
Regénykorpusz	2,042	2,103

Az 5. táblázatból látható, hogy mind a két referenciakorpusz alapján azt az eredményt kapjuk, hogy a formailag egyszerű versek lexikai választékossága alacsonyabb a többi vers lexikai választékosságánál. A 6. táblázat mutatja be, hogy a 18 szerzőből hány szerző felel meg, illetve megy szembe ezzel a mintázattal. (A mediánértékeket lásd a Függelék 14. táblázatában.)

6. táblázat. A versforma és a lexikai választékosság összefüggésének mintázata megvalósító és nem megvalósító szerzők száma

	ÁTLAG		MEDIÁN	
	egyszerű < többi	egyszerű > többi	egyszerű < többi	egyszerű > többi
Verskorpusz	14	4	14	4
Regénykorpusz	13	5	14	4

A 6. táblázatból látható, hogy ha a Verskorpuszt használjuk referenciakorpuszként, akkor az átlagok és a mediánok alapján is a 18-ból 14 szerző, ha a Regénykorpuszt, akkor pedig 13, illetve 14 szerző esetében érvényesül az, hogy a formailag egyszerű versek lexikai választékossága alacsonyabb, mint a többi vers lexikai választékossága. Úgy tűnik tehát, hogy a magyar kanonikus költészetnek egy viszonylag erős mintázataról van szó. Ha az átlagértékeket nézzük, akkor a mintázattal szembe menő szerzők Csokonai, Arany, Reviczky, Ady és Tóth (az utóbbi csak a Regénykorpusz alapján). Az ő esetükben az egyszerű verseknek nagyobb az átlagos lexikai választékossága, mint a többi versnek. A mediánértékek esetében a mintázattal szembemenő négy szerző a Verskorpusz referenciakorpuszt használva Csokonai, Arany, Ady és Dsida, a Regénykorpuszt referenciakorpuszként használva pedig Csokonai, Arany, Petőfi és Ady.

A jelenség megmagyarázásához abból érdemes kiindulni, hogy a ritkábban előforduló szavak jellemzően nem pusztán specifikusabb jelentésűek, hanem szemantikailag komplexebbek, mint a gyakrabban előforduló

szavak. Szemantikai komplexitás alatt a következő jellemzőket értem: nem alapszintű kategóriák megnevezései; nem rendelkezünk velük kapcsolatban vizuális reprezentációval; a memóriából nehezebben előhívhatóak; illetve a nyelvelsajátítás során később tanuljuk meg őket (Rosch 1978; Schmid 2007: 122–125). Emellett sok esetben a gyakori szavaknál hosszabbak és morfológiailag összetettebbek (összetett szavak, képzett szavak), vagyis a szavak komplexebb fonológiai szerkezete ikonikusan leképezi a szavak komplexebb jelentését (lásd Ladányi 2017: 524–526).

A nyelv alapvetően ikonikus természete, miszerint a fonológiai szerkezetet a szemantikai szerkezet motiválja, azaz a forma valamilyen absztrakt módon leképezi a jelentést, természetesen a szavak morfológiai szerkezete mellett a nyelv számos egyéb területén is érvényesül (például lásd Haiman 1980; Pusch 2001; Langendonck 2007). A kapott eredmények azt mutatják, hogy nemcsak a morfológia, a grammatika vagy éppen a szövegstruktúra és a jelentés, hanem a versforma és a vers tartalma között is ikonikus összefüggés lehet: a vers szemantikailag kevésbé komplex szókinccsét ikonikusan leképezi a formailag egyszerű versforma. Azaz a versforma és a tartalomnak a szókinccsel megragadható dimenziója az egyszerűség szempontjából hasonlít egymásra, a kettő mintegy egymás tükörképe.

A versforma egyik leginkább kutatott jellemzője, a vers metrumba és a vers tartalma közti összefüggésre már az ókorban felhívták a figyelmet (például lásd Arisztotelész Poétikáját (1992: 46–47) vagy Horatius Ars poeticáját⁶). Az utóbbi időben több korpusznyelvészeti vizsgálat is született, amelyek nagy korpuszokon, kvantitatív eszközökkel igazolták a versmetrum és a vers tartalma közötti szisztematikus, azaz nem véletlenszerű összefüggést (Piperski 2017; Šeĵa et al. 2020, 2022). Ezt az összefüggést a metrum és a tartalom között ugyanakkor elsősorban kulturális okokkal magyarázzák. Az itt vizsgált esetben viszont feltételezhető, hogy a versforma és a vers tartalma közötti megfelelésnek egy olyan aspektusáról van szó, amely nem önkényes, azaz amely nemcsak szisztematikus, hanem egyben ikonikus is.

⁶ „Hogy mily versmérték illik hősök s fejedelmek / tetteihez s vészes harchoz, példázza Homerus. / Két sor más-más mértékű, első a panasz, majd / feltör a hálaadás, ha beteljesedett a kívánság. / Azt, hogy a könnyed elégia atya ki volt, a tudósok / nem döntötték el, vita dűl most is körülötte. / Archilocusnak a düh fegyverként adta a jambust, / ily versláb-
ra színészi cipő is jó, s a cothurnus, / mert dialógushoz sem rossz, s elnyomja a lármás /
nézőtér moráját, s megpörgeti jól a cselekményt.” (Bede Anna fordítása)

4.3. A vers belső szerkezete és lexikai választékossága – versszakok

A versek lexikai választékossága nem csupán a vers egészére vonatkozathatóan értelmezhető, hanem a vers egyes részeire is. Például felvethető a kérdés, hogy a versek elejének, közepének és végének a lexikai választékossága milyen módon változik, a változás kirajzol-e valamilyen tipikus mintázatot. A kérdés vizsgálatához a versek minden versszakának külön kiszámoltam a lexikai választékosságát, és azt néztem meg, hogy a versek első versszakainak, középső versszakainak és utolsó versszakainak a lexikai választékossága milyen mintázatot rajzol ki. Egy vers középső versszakainak a lexikai választékosságát az ide tartozó versszakok (azaz az első és az utolsó versszak közötti összes versszak) lexikai választékosságának az átlagolásával számítottam ki. Mivel a vizsgálat elvégzéséhez minimum három versszak szükséges (első, középső és utolsó versszak), a vizsgálatba csak a minimum három versszakos verseket vettem bele (összesen 8447 vers). A 7. táblázat a vizsgálatba bevett, 1000 szónál rövidebb, minimum három versszakból álló versek eloszlását mutatja be a versszakszámok alapján.

7. táblázat. A vizsgálatba bevont versek eloszlása versszakszám szerint

Versszak-szám	Versek száma
4	2122
3	1408
5	1230
6	962
7	607
8	499
8-nál több	1619

Első lépésben a versek három részére kapott lexikai választékossági pontszámoknak számítottam ki az átlagát minden szerző esetében (ezeket az értékeket lásd a Függelék 15. táblázatában), majd az egyes szerzőkre kapott értékeket átlagoltam. A kapott átlagokat a 8. táblázat mutatja be.

8. táblázat. A versek három szerkezeti egységének átlagos lexikai választékossága

	Első versszak	Középső versszak(ok)	Utolsó versszak
Verskorpusz	1,966	1,945	1,884
Regénykorpusz	2,149	2,120	2,062

A 8. táblázatból látható, hogy mind a Verskorpusz, mind a Regénykorpusz referenciakorpuszként való használata esetén kirajzolódik az a mintázat, miszerint a versek előrehaladásával a lexikai választékosság mértéke csökken. Ha a Verskorpuszt használjuk referenciakorpuszként, akkor ettől a mintázattól az átlagok alapján hét szerző, Kisfaludy, Kölcsey, Petőfi, Vajda, Komjáthy, Kosztolányi és József Attila tér el. Ha a Regénykorpuszt használjuk referenciakorpuszként, akkor csupán négy szerző, Kisfaludy, Petőfi, Gyulai és Vajda tér el a mintázattól.

Az átlagok mellett érdemes azt is megnézni, hogy a vizsgált 24 szerzőből hány olyat találunk, aki a lehetséges konstellációk közül a fent bemutatott mintázatnak megfelelő konstellációt valósítja meg a legtöbb versében. A 9. táblázat azt mutatja be, hogy a három szerkezeti egység lexikai választékosságának hat lehetséges mintázatából melyiket hány szerző valósítja meg a legtöbb és a második legtöbb versében. (Az egyes esetekhez tartozó versek számát szerzőkre lebontva lásd a Függelék 16. táblázatában.) A legnagyobb értékeket félkövérrel emeltem ki. A << és a >> szimbólum arra utal, hogy a szimbólum után szereplő utolsó versszak lexikai választékossága nemcsak a középső versszak(ok), hanem az első versszak lexikai választékosságánál is magasabb, illetve alacsonyabb. A második leggyakoribb mintázatok adataihoz nem számoltam hozzá azokat a szerzőket, akiknél a második helyen egyenél több mintázat is megjelenik egyenlő előfordulással.

9. táblázat. A szerkezeti egységek lexikai választékosságának hat lehetséges mintázata és az adott mintázatot a legnagyobb és második legnagyobb mértékben megvalósító szerzők száma

	Verskorpusz		Regénykorpusz	
	I	II.	I.	II.
Első versszak > középső versszak > utolsó versszak	23	1	23	0
Első versszak > középső versszak << utolsó versszak	0	0	0	0

	Verskorpusz		Regénykorpusz	
	I	II.	I.	II.
Első versszak > középső versszak < utolsó versszak	0	1	0	1
Első versszak < középső versszak < utolsó versszak	0	4	0	7
Első versszak < középső versszak >> utolsó versszak	1	16	1	12
Első versszak < középső versszak > utolsó versszak	0	0	1	2

A 9. táblázatból láthatjuk, hogy mind a Verskorpusz, mind a Regénykorpusz referenciakorpuszként való használata esetén a 24 szerzőből 23 azt az esetet valósítja meg a legtöbb versében, amikor a középső versszak(ok) lexikai választékossága alacsonyabb az első versszak lexikai választékosságánál, az utolsó versszak lexikai választékossága pedig alacsonyabb a középső versszak(ok) lexikai választékosságánál. Mind a két referenciakorpusz esetében Kisfaludy az egyetlen szerző, aki eltér ettől a mintától. (A Regénykorpusz referenciakorpusza esetén egyenlő számmal valósította meg az ötödik és a hatodik mintázatot.) A táblázatból az is látható, hogy a legtöbb szerző esetében a második leggyakrabban előforduló mintázat az az eset, amikor a vers középső versszakának/versszakainak a lexikális választékossága magasabb, mint az első versszak lexikai választékossága, az utolsó versszak lexikai választékossága viszont a leggyakoribb mintázathoz hasonlóan nemcsak a középső versszak(ok), hanem az első versszak lexikai választékosságánál is alacsonyabb. Úgy tűnik tehát, hogy a két mintázatból egy még általánosabb mintázat absztrahálható, amely csupán annyit specifikál, hogy az utolsó versszak lexikai választékossága alacsonyabb a vers első és középső versszakainak a lexikai választékosságánál.

A 8. és 9. táblázatból látható tehát, hogy a magyar kanonikus költészetben a legtipikusabb tendencia az, hogy a versek lexikai választékossága csökken a versek előrehaladásával.⁷ A lexikai választékosságnak ez a csökkenése

⁷ Megjegyzendő, hogy a használt módszer nem tudja kimutatni, hogy három versszaknál hosszabb versek esetében a lexikai választékosság változásának a fő tendenciája egy kvázi folyamatos csökkenés a vers előrehaladása során, vagy pedig a vers eleje, közepe és vége között van csupán eltérés a lexikai választékosságban, anélkül, hogy a középső versszakokon belül megjelenne egy csökkenő tendencia.

feltételezhetően a vers tartalmi változásának egy mérhető rétege. A lexikai választékosság az alacsony gyakoriságú szavak gyakoribb, illetve ritkább megjelenése következtében lesz magasabb vagy alacsonyabb. Feltételezve, hogy a gyakoribb szavak jellemzően általánosabb jelentésű fogalmakra vonatkoznak, a ritkábban előforduló szavak pedig speciálisabb kontextusokban használatos, specifikusabb jelentésű fogalmakra utalnak, a lexikai választékosság alakulását összefüggésbe hozhatjuk a tartalom általánosabb, illetve specifikusabb jellegével. A vers lexikai választékosságának a vers előrehaladásával való csökkenése ily módon a vers tartalmi/szemantikai általánosulásának a jele. Máshogyan fogalmazva a vers lexikai választékosságának a csökkenése a specifikustól az általános fele tartó tartalmi mozgással hozható összefüggésbe.

A (2) és a (3) példa két három versszakos verssel mutatja be a jelenséget. A szavak gyakorisági pontszámait szögletes zárójelben, a versszakok átlagos pontszámait pedig a versszakokat követően, félkövérrel adtam meg. A gyakorisági pontszámok kiszámításához a Verskorpuszt használtam referenciakorpuszként. Az Ady versben szereplő *ér* szó 1-es pontszáma jól mutatja, hogy az automatizált módszer nem képes homonímiák között különbséget tenni. Úgyszintén torzíthatja a pontszámokat, ha egy adott szó nem a szten-derd helyesírás szerint van leírva. A korrigált pontszámok esetében megadtam, hogy mennyi lenne a versszakok pontszáma, ha az *ér* szó 4 pontot, a nem sztenderd módon leírt *océán* szó pedig 3 pontot kapna.⁸ Mindazonáltal hangsúlyozandó, hogy mivel ezeknek a típusú hibáknak a versekben való megjelenési helye teljesen véletlenszerű, a kutatás nagy adatmennyiségen alapuló kvantitatív eredményeire nincsenek hatással.

⁸ Az *océán* és *oceán* lemmák együttes relatív gyakorisága egymillió szavanként 44,182.

(2) Dsida Jenő: Ki-ki magában Gyötrelmek [3] széles [3] országútja [3] ez [1]. Ki-ki [3] magában [1] szenved [2]. Széttört [3] kristálypoharak [4] szilánkjai [4] lépten [4] sebeznek [3]. [2,833] Vándordal [4] nem [1] születik [2]. Csak [1] a [1] kiszikkadt [4] torok [2] tátog [4]. Hiába [2] emelünk [2] kalapot [2] a [1] feszületnek [3], a [1] feszületről [3] sem [1] száll [1] le [2] senki [2]. [2,053] A [1] láthatáron [3] semmi [1] sincs [2]. A [1] kövekben [2] nincs [1] víz [2]. A [1] fák [1] alatt [1] sincs [2] árnyék [2]. Kőben [2], fűben [2], égben [1] nincs [1] irgalom [3]. [1,611]	(3) Ady Endre: Az Értől az Oceánig Az [1] Ér [1 (4)] nagy [1], álmos [2], furcsa [2] árok [3], Pocsolyás [4] víz [2], sás [3], káka [4] lakják [2]. De [1] Kraszna [4], Szamos [3], Tisza [2], Duna [2] Oceánig [4 (3)] hordják [2] a [1] habját [2]. [2,300, korrigált pontszám: 2,4] S [1] ha [1] rám [1] dől [2] a [1] szittya [3] magasság [3], Ha [1] száz [2] átok [2] fogja [1] a [1] vérem [1], Ha [1] gátat [3] túl [3] föl [2] ezer [2] vakond [4], Az [1] Oceánt [4 (3)] mégis [1] elérem [3]. [1,913, korrigált pontszám: 1,870] Akarom [2], mert [1] ez [1] bús [1] merészség [4], Akarom [2], mert [1] világ [1] csodája [2]: Valaki [2] az [1] Értől [1 (4)] indul [2] el [1] S [1] befut [3] a [1] szent [1], nagy [1] Oceánba [4 (3)]. [1,650, korrigált pontszám: 1,750]
---	--

A lexikai választékosság csökkenésének, illetve az ezzel összefüggésbe hozott tartalmi általánosulásnak a magyar kanonikus költészetben megfigyelt tendenciáját alapvetően két kiindulópontból magyarázhatjuk. Az egyik lehetőség az, hogy feltételezzük, hogy valamilyen, a magyar költészetben kulturálisan öröklődő mintázatról van szó. Azaz a tendenciát csupán azzal magyarázzuk, hogy az egyes szerzők nem feltétlenül tudatosan, de utánozzák a korábbi szerzőket. Ezzel szemben a másik lehetséges megközelítés nem a kulturális hatás

alapján, hanem a nyelvi tevékenységnek a lírai diskurzusokban is megnyilvánuló kognitív meghatározottsága alapján magyarázhatná a jelenséget. Persze egy ilyen magyarázat is sokféle irányba szétágazhat. Például helyezhetné a hangsúlyt a szövegtípus által képviselt általános megismerési módra, amelyről feltételezhetné, hogy annak inherensen része valamiféle, a specifikustól az általános felé haladó induktív logika. De legalább ennyire érvényes lehet egy olyan magyarázati kísérlet is, amely az alkotói folyamat pszichológiai aspektusaira helyezi a hangsúlyt, és a szerzői invenció, illetve figyelem változásával magyarázza a lexikai választékosság csökkenését és az ezzel összefüggésbe hozott tartalmi általánosulást. E kérdések részletes vizsgálata azonban túlmutat a tanulmány célkitűzésein.

4.4. A vers belső szerkezete és lexikai választékossága – verssorok

Végezetül érdemes azt is megvizsgálni, hogy a versszakon belüli sorok lexikai választékossága nem mutat-e valamilyen szabályos változást. Ennek vizsgálatához a minimum három versszakból álló versek első, második és utolsó versszakaiban szereplő első hat sor lexikai választékosságát néztem meg. A vizsgálatba a hatnál kevesebb sorból álló versszakok sorai is bekerültek, ugyanakkor kritériumként adtam meg, hogy a versszakoknak legalább két sorból kellett állniuk. A vizsgálat a versszakok vizsgálatához hasonló módon történt. Minden vers esetében kiszámítottam sorok lexikai választékosságát, majd az első, második és utolsó versszak első hat sorára kapott értékeket szerzőnként átlagoltam. A 10. táblázat a szerzőkre kapott átlagok átlagait mutatja be a vizsgált 18 sorra.

10. táblázat. A versek első, második és utolsó versszakai első hat sorának átlagos lexikai választékossága

	1. sor	2. sor	3. sor	4. sor	5. sor	6. sor
Első versszak						
Verskorpusz	2,006	2,045	2,033	2,049	2,010	2,049
Regénykorpusz	2,189	2,239	2,215	2,234	2,181	2,219
Második versszak						
Verskorpusz	1,977	2,057	2,008	2,050	2,034	2,032
Regénykorpusz	2,137	2,242	2,184	2,240	2,206	2,216

	1. sor	2. sor	3. sor	4. sor	5. sor	6. sor
Utolsó versszak						
Verskorpusz	1,902	1,970	1,950	1,974	1,950	1,983
Regénykorpusz	2,072	2,158	2,126	2,158	2,124	2,169

A 10. táblázat értékeiből egy szembetűnő mintázat rajzolódik ki: a versszakok páros sorainak a lexikai választékossága magasabb, mint a páratlan soroké. A lexikai választékosság változása tehát egy versszakon belüli szabályosan hullámzó mozgást mutat: a második, negyedik és hatodik sorban emelkedik az érték, míg a harmadik és ötödik sorban csökken. Ha a Verskorpuszt használjuk referenciakorpuszként, akkor ez a mintázat csak egy helyen, a második versszak hatodik sorában törik meg, a Regénykorpusz alapján kapott értékek esetében pedig nincsen mintatörés.

Felmerülhet, hogy a páratlan és páros sorok esetében talált mintázat összefügghet a magyar kanonikus költészet rímelésének azon egyszerű, egy korábbi vizsgálatban kimutatott jellemzőjével, miszerint eltérő szótagszámú rím��avak esetén jellemzőbb, hogy a rímpár második tagja a hosszabb (Horváth 2023). Ahogyan arra már utaltam, a hosszabb szavak jellemzően specifikusabb jelentésű és ezért ritkábban előforduló szavak. A páros rímű versszakok páros soraiban szereplő felelőrímek tehát felfelé húzhatják a páros sorok lexikai választékosságának az értékét. A rímelés hatásának vizsgálatához a 10. táblázat adatait eredményező számítást külön-külön elvégeztem azokra a versekre szűkítve, amelyekben a versszakok rímképlete páros rím, keresztrím, valamint félrím.⁹ Az egyes sorokra kapott pontszámokat a 11. táblázat mutatja be. A mintatöréseket, azaz amikor a páros sor lexikai választékossága alacsonyabb, mint a versszakban előtte lévő páratlan soré, illetve a páratlan sor lexikai választékossága magasabb, mint a versszakban előtte lévő páros soré, szürke cellaszínezéssel emeltem ki.

⁹ A vizsgálatba a következő rímképletű versszakok kerültek bele. Páros rím: aa, aabb, aabbaa, aabbcc, aabbaabb, aabbaacc, aabbccaa, aabbccbb, aabbccdd; keresztrím: ab, abab, ababab, abababab, abababcd, ababacaca, ababcbcb; félrím: abcb, abcbd, abcbdb, abcbdbe, abcbdbb, abcbdefe

11. táblázat. A versek első, második és utolsó versszakai első hat sorának átlagos lexikai választékossága páros rímű, keresztrímű és félrímű versszakok esetében

		1. sor	2. sor	3. sor	4. sor	5. sor	6. sor
Első versszak							
páros rím	Versk.	2,026	1,987	1,986	1,994	2,012	2,064
	Regényk.	2,217	2,159	2,174	2,150	2,212	2,245
keresztrím	Versk.	1,986	2,038	2,005	2,027	2,013	2,021
	Regényk.	2,166	2,245	2,209	2,235	2,232	2,205
félrím	Versk.	1,973	1,996	1,984	2,069	2,011	2,107
	Regényk.	2,177	2,222	2,138	2,238	2,158	2,215
Második versszak							
páros rím	Versk.	1,991	2,042	1,971	1,976	1,973	1,916
	Regényk.	2,165	2,231	2,186	2,164	2,102	2,074
keresztrím	Versk.	2,001	1,986	1,979	2,053	2,017	1,990
	Regényk.	2,174	2,185	2,145	2,255	2,238	2,274
félrím	Versk.	1,984	2,104	2,013	2,090	2,261	2,025
	Regényk.	2,175	2,301	2,188	2,277	2,363	2,134
Utolsó versszak							
páros rím	Versk.	1,880	1,910	1,940	1,916	2,067	1,968
	Regényk.	2,042	2,080	2,106	2,111	2,268	2,128
keresztrím	Versk.	1,883	1,935	1,906	1,937	1,938	1,968
	Regényk.	2,061	2,134	2,076	2,134	2,139	2,183
félrím	Versk.	1,927	1,988	1,922	1,937	1,791	1,869
	Regényk.	2,096	2,188	2,067	2,106	1,971	2,011

A 11. táblázatból láthatjuk, hogy szemben a kiinduló feltételezéssel, a páros rímű versszakok esetében szerepel a legtöbb mintatörés: ha a Verskorpust használjuk referenciakorpuszként, a mintatörések száma hét, ha a Regénykorpust, a mintatörések száma kilenc. Ezzel szemben a keresztrímű versszakok esetében a mintatörések száma három, illetve kettő, a félrímű versszakok esetében pedig mind a két referenciakorpusz esetében kettő. A 11. táblázat adatai tehát azt mutatják, hogy a 10. táblázat adatai alapján megállapított általános mintázat nem függ össze a rímpárok tagjainak a szótagszámbeli eltéréssel. A páros rímű versek kapcsán talált nagyszámú mintatörés magyarázatára e rész végén tér ki az elemzés.

Ha nem a szerzőkre kapott átlagértékek átlagait nézzük, hanem az egyes szerzőkre kapott átlagokat, akkor a 10. táblázathoz képest változatosabb kép bontakozik ki. A 12. táblázat azt mutatja be, hogy mely szerzők valósítják meg teljes mértékben a mintát, és mely szerzőknél jelenik meg egy, kettő, három stb. mintatörés.

12. táblázat. A lexikai választékoság páratlan és páros sorokra kapott mintájának a megvalósítása szerzőnként

	Verskorpusz		Regénykorpusz	
Minta-törések száma	Szerzők száma	Szerzők	Szerzők száma	Szerzők
0	2	Tompa, Petőfi	2	Tompa, Petőfi
1	1	Berzsenyi	0	
2	1	Arany	4	Berzsenyi, Kisfaludy, Vörösmarty, Tóth
3	4	Kisfaludy, Vörösmarty, Vajda, Ady	2	Kölcsey, Arany
4	3	Csokonai, Babits, Dsida	3	Csokonai, Ady, Babits
5	3	Reviczky, Tóth, József	5	Madách, Vajda, Somlyó, József, Dsida
6	4	Kölcsey, Madách, Somlyó, Kosztolányi	1	Reviczky
7	1	Kaffka	3	Gyulai, Kosztolányi, Radnóti
8	2	Gyulai, Radnóti	3	Komjáthy, Kaffka, Juhász
9	1	Juhász	1	Reményik
10	1	Komjáthy		
11	1	Reményik		

A 12. táblázatból láthatjuk, hogy Tompa és Petőfi valósítja meg a legszabályosabban a mintázatot, az ő esetükben egyik referenciakorpusz alapján sincsen mintatörés. Úgyisntén nagymértékben követi a mintát Berzsenyi, Kisfaludy, Vörösmarty és Arany. Az ő esetükben egyik referenciakorpusz

alapján sem több háromnál a mintatörések száma. Vannak olyan szerzők is, akik a lehetséges mintatörések több mint felében, azaz minimum nyolcszor megvalósítják a mintatörést mind a két referenciakorpusz alapján, ők tehát szembemennek a mintázattal. Ilyen szerző Reményik, Komjáthy és Juhász. Megjegyzendő, hogy a felsorolt hat nagymértékben mintakövető szerző mind a 19. század elején, illetve közepén alkotott, míg a három nagymértékben mintatörő szerző a 20. század első felében, illetve a 19. század végén. Az adatokból tehát úgy tűnik, hogy a magyar kanonikus költészetben kirajzolódik valamiféle időbeli tendencia, amely a mintakövetéstől a mintatörés felé mutat.

A (4) és az (5) Petőfi Sándor *Szomorú éj* és Tóth Árpád *Esti sugárkoszorú* című verseinek második versszakai alapján mutatja be a sorok páratlan és páros sorai kapcsán megállapított mintázatot. A szavak gyakorisági pontszámait a szavak Verskorpuszbeli gyakoriságai alapján adtam meg. A hibásan lemmatizált *tevék* ige és a nem sztenderd helyesírással leírt *szív* főnév esetében az automatizáltan megállapított pontszámok helyett a korrigált pontszámokat tüntettem fel. A sorokat követően adtam meg a sorok átlagos pontszámait.

(4)

Ez [1] hát [1] a [1] költő [2] sorsa [2], mindig [2] ez [1],	1,429
Hogy [1] örök [2] vészű [4] tengeren [2] evez [3]?	2,400
S [1] mit [1] ér [1], ha [1] őt [1] a [1] haragos [3] habokból [2]	1,375
Mentő [3] sajkával [3] partra [2] is [1] tévék [1],	2,000
Ha [1] itt [1] az [1] bántja [2], hogy [1] mi [1] lesz [1] azokból [1],	1,125
Kik [1] ott [1] maradtak [2] a [1] hajóba' [2] még [1]?	1,333
(Petőfi Sándor: <i>Szomorú éj</i> , második versszak)	

(5)

Illattá [2] s [1] csenddé [2]. Titkok [2] illata [2]	1,800
Fénylett [4] hajadban [2] s [1] béke [2] égi [2] csendje [2],	2,167
És [1] jó [1] volt [1] élni [1], mint [1] ahogy [2] soha [2],	1,286
S [1] a [1] fényt [1] szemem [1] beitta [4] a [1] szívembe [1]:	1,429
Nem [1] tudtam [1] többé [2], hogy [1] te [1] vagy [1] -e [1] te [1],	1,125
Vagy [1] áldott [2] csipkebokor [3] drága [2] tested [2],	2,000
Melyben [1] egy [1] isten [1] szállt [1] a [1] földre [1] le [2],	1,143
S [1] lombjából [2] felém [2] az [1] ő [1] lelke [1] reszket [2]?	1,429
(Tóth Árpád: <i>Esti sugárkoszorú</i> , második versszak)	

A lexikai választékosságnak a páros és páratlan sorokban megfigyelt emelkedő és ereszkedő fluktuációja feltételezhetően olyan jellemző, amely a sorok valamilyen általános szemantikai tulajdonságával is összefügghet. Amennyiben feltételezzük, hogy a lexikai választékosság fluktuációja és a sorok jelentése között összefüggés van, akkor azt is feltételeznünk kell, hogy a verssorok jellemzően párokba rendeződő nagyobb jelentésegységeket alkotnak, amelynek első eleme a páratlan sor jelentése, második eleme pedig a rá következő páros sor jelentése. A lexikai választékosság különbözőségét ebben a megközelítésben a páratlan és a páros sor által összekapcsolt két jelentésegység közti viszony, illetve a jelentésegységeknek az e viszonyból következő eltérő jellemzői magyarázhatják. A kérdés tehát úgy vethető fel, hogy mi az a tipikusan megjelenő absztrakt viszony a páratlan és a páros sor által kifejezett jelentésegységek között, amely a lexikai választékosságnak a megfigyelt fluktuációjához vezet.

Ez a viszony feltételezhetően az általános és a specifikus valamilyen összekapcsolásán alapszik, hiszen ahogyan arra már többször történt utalás, a lexikai választékosság magasabb értékét előidéző ritkább szavak jellemzően specifikusabb jelentésűek. A páros rímű versszakoknál talált mintatörések nagy száma is magyarázható azzal, hogy a páros rímelés motiválja és egyben ikonikusan le is képezi az egymás melletti páratlan és páros soroknak a szemantikailag párhuzamosabb szerkesztésmódját, ami megbontja az általános-specifikus viszonyon alapuló aszimmetrikus jelentésviszonyt. Ezzel szemben a keresztrímű és a félrímű versszakok egymás mellett lévő sorainak a fonológiai aszimmetriája motiválhatja, illetve ikonikusan leképezi a páratlan és páros soroknak az aszimmetrikus jelentésviszonyát (a rímelés jelentéskezdeményező szerepéről lásd Simon 2014).

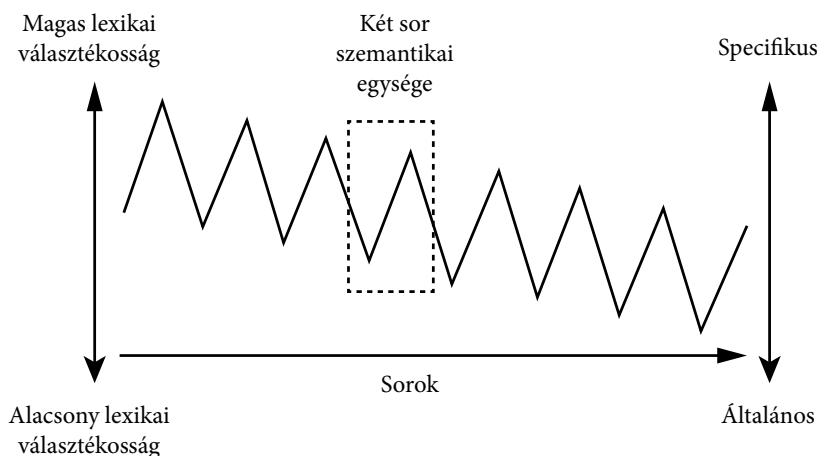
Az általánosabb és a specifikusabb jelentés ugyanakkor sokféle egymást nem kizáró logika mentén összekapcsolódhat. Ez a logika egyrészt alapulhat az állítás és a magyarázat, a következtetés és a megfigyelés vagy éppen a cselekvés és a belső motiváció kauzális viszonyán. De alapulhat egy általánosabb érzés, gondolat és az arról asszociált, arra hasonlító, illetve azt specifikáló tárgyi világ viszonyán is. Ha a költői nyelv metaforizáltsága szempontjából tekintjük a kérdést, akkor pedig alapulhat a metafora céltartománya és forrástartománya közötti viszonyon, ahol a céltartomány absztraktabb, érzékszervekkel nem megtapasztalható jelenségét teszi elgondolhatóvá a forrástartomány konkrétabb, érzékszervekkel megtapasztalható jelensége (Lakoff–Johnson 1980). E kérdés kvalitatív vizsgálata azonban már szétfeszítené ennek a kutatásnak keretét.

4.5. A prototipikus vers lexikai választékosságának és szemantikai specifikusságának kétszintű sémája a kanonikus magyar költészetben

Ahogy azt a 4.3. és 4.4. részből láthattuk, a magyar kanonikus költészet verseiben mind makroszinten, mind mikroszinten sikerült azonosítani a lexikai választékosságnak egy tipikus mintázatát. A makroszintű mintázatot az első, középső és utolsó versszakok lexikai választékosságának a mérésével sikerült absztrahálni. A versek eleje, közepe és vége esetében kapott lexikai választékossági pontszámok alapján az a fő tendencia rajzolódott ki, hogy a versek lexikai választékossága a vers előrehaladásával csökken, azaz a vers közepének a lexikai választékossága alacsonyabb, mint a vers elejének, a vers végének a lexikai választékossága pedig alacsonyabb, mint a vers közepének. Feltételeztem, hogy ez a mintázat a specifikustól az általános irányába haladó szemantikai mozgásnak a leképeződése, hiszen a magasabb lexikai választékosságot a ritkább és általában specifikusabb jelentésű szavak idézik elő.

A versszakok sorai esetében mért átlagos lexikai választékossági pontszámok pedig rámutattak a vers lexikai választékosságának a mikroszintű mintázatára. E mikroszintű mintázat lényege az, hogy a páros sorok lexikai választékossága magasabb, mint a páratlan sorok lexikai választékossága. Feltételezve azt, hogy a páratlan és páros sorokra kapott mintázat szemantikailag motivált, amellett érveltem, hogy a szomszédos páratlan és páros sorok egy nagyobb szemantikai egységként funkcionáló sorpárt alkotnak, amelyekben az alacsonyabb lexikai választékosságot képviselő első sor általánosabb és a magasabb lexikai választékosságot képviselő második sor specifikusabb jelentése kerül egymással viszonyba.

Látható tehát, hogy a lexikai választékosságnak a vers szerkezeti egységeinek makro- és mikroszintjén elkülönített két mintázata két szemantikai mintázat felszíni, mérhető rétegeként azonosítható. E szemantikai mintázatok a szemantikai specifikusság változásaként írhatók le. Mindezek alapján a kanonikus magyar költészet prototipikus versének vagy legalábbis egy kitüntetett verstípusának a lexikai választékosságához és a lexikai választékosságból levezetett szemantikai specifikusságához kapcsolódó kétszintű sémát az 1. ábra mutatja be.



1. ábra. A prototipikus vers lexikai választékosságának és szemantikai specifikusságának kétszintű sémája a kanonikus magyar költészetben

5. Összegzés

A tanulmány az idegen nyelv elsajátításának a vizsgálatában használt lexikai választékosság mérése alapján vizsgálta 24 magyar kanonikus költő verseit. A lexikai választékosság mértékét a szövegben szereplő szavak referenciakörpuszbeli gyakorisága alapján határoztam meg. A magyarázatok legfontosabb kiindulópontja az volt, hogy a lexikai választékosság és a szemantikai specifikusság mértéke között összefüggés van, mivel a ritkábban előforduló szavak jellemzően specifikusabb jelentésűek. A vizsgálat során a szavakat automatizáltan négy gyakorisági csoportba soroltam két referenciakörpusz, az ELTE Verskörpuszból kinyert kutatási körpusz, valamint az ELTE Regénykörpusz alapján, és a versek szavait a referenciakörpuszbeli gyakoriságuk alapján pontoztam egy egytől négyig terjedő skálán. A szavak pontszámait alapján kiszámítottam a versek, a versszakok és a verssorok lexikai választékosságát, majd ezeket a pontszámokat átlagoltam minden egyes szerző esetében a vizsgált kutatási kérdéseknek megfelelően.

A vizsgálat első kérdése az volt, hogy a versek hosszúsága és lexikai választékossága között van-e összefüggés. A három terjedelmi kategóriába sorolt

versek lexikai választékossági pontszámainak az átlagaiból kiderült, hogy a közepes, 50 és 199 szó közötti versek lexikai választékossága jellemzően valamivel alacsonyabb az 1 és 49 szó közötti rövid versek és a 200 és 999 szó közötti hosszú versek lexikai választékosságánál. A rövid és a hosszú versek magasabb lexikai választékosságának oka az lehet, hogy a rövid versekben a fókuszáltabb megfogalmazás miatt, a hosszú versekben pedig a részletesebb leírások miatt jelenhet meg valamivel nagyobb mértékben a specifikusabb jelentésű kevésbé gyakori szavak használata.

A vizsgálat második kérdése az volt, hogy a formailag egyszerű versek lexikai választékossága eltér-e a többi vers lexikai választékosságától. A vizsgálatból kiderült, hogy az egyszerre izometrikus, izorímes és izostrofikus, formailag egyszerű versek lexikai választékossága jellemzően alacsonyabb, mint a többi, formailag nem vagy kevésbé egyszerű vers lexikai választékossága. Ezt az összefüggést a versforma és a vers tartalma közötti ikonikus megfeleléssel magyaráztam, azaz a vers egyszerű formája és a vers jelentésének a szemantikailag kevésbé komplex szókincsben megnyilvánuló egyszerűsége leképezi egymást.

Ügyszintén vizsgáltam a vers belső szerkezetének a lexikai választékossággal való összefüggését. Ebben az esetben tehát nem arra kerestem a választ, hogy a bizonyos formai jellemzőket felmutató vers egésze és a lexikai választékosság között milyen összefüggés van, hanem arra, hogy a vers egymást követő szerkezeti egységeinek a lexikai választékossága mutat-e valamilyen tipikus mintázatot. A vizsgálat során a versek első, középső és utolsó versszakainak a lexikai választékosságát külön mérve bemutattam, hogy a versekben megjelenő legjellemzőbb tendencia az, hogy a lexikai választékosság mértéke a versek előrehaladásával csökken. A lexikai választékosság csökkenő tendenciája feltételezésem szerint a specifikustól az általános felé haladó tartalmi mozgásnak a jele.

Végezetül megvizsgáltam, hogy a lexikai választékosságnak a versszakon belüli változása mutat-e valamilyen szabályos mintázatot. A vizsgálatból kiderült, hogy míg a versszak páratlan sorainak alacsonyabb, addig a páros soroknak jellemzően magasabb a lexikai választékossága, vagyis a versszakon belül a lexikai választékosság egy fluktuáló mintázatot mutat, amelynek alapegysége a sorpár. A jelenséget megkísértem szemantikai okokra visszavezetni, feltételezve, hogy a sorpárok olyan jelentésszerkezetet alkotnak, amelyben az első sor által kifejezett első jelentésegység és a második sor által kifejezett második jelentésegység közötti viszony az általános-specifikus absztrakt viszonyra ragadható meg. A páros, kereszt-, és félrimű versszakok

vizsgálatából az is kiderült, hogy a sorok lexikai választékossága és a versszak rímképlete között van összefüggés: úgy tűnik, hogy a kereszt- és fél-rímű versszakok a megállapított mintázat megvalósulását ikonikusan leképezik, illetve motiválják, míg a páros rímű versszakok nem motiválják ezt.

A versek első, középső és záró versszakainak lexikai választékossága és a versszakok páratlan és páros sorainak lexikai választékossága alapján le lehetett írni a magyar kanonikus költészetre jellemző prototipikus vers lexikai választékosságának és az ebből levezetett szemantikai specifikusságának a kétszintű sémáját.

Irodalom

- Arisztotelész 1992. *Poétika*. Budapest: Kossuth Könyvkiadó.
- Bardel, Camilla – Gudmundson, Anna – Lindqvist, Christina 2012. Aspects of lexical sophistication in advanced learners' oral production: Vocabulary acquisition and use in L2 French and Italian. *Studies in Second Language Acquisition* 34: 269–290. <https://doi.org/10.1017/S0272263112000058>
- Covington, Michael A. – McFall, Joe D. 2010. Cutting the gordian knot: The moving-average type–token ratio (MATTR). *Journal of Quantitative Linguistics* 17 (2): 94–100. <https://doi.org/10.1080/09296171003643098>
- Cruse, Alan 1977. The pragmatics of lexical specificity. *Journal of Linguistics* 13: 153–166.
- Haiman, John 1980. The iconicity of grammar: Isomorphism and motivation. *Language* 56 (3): 515–540. <https://doi.org/10.2307/414448>
- Horváth Iván 1991. *A vers. Három megközelítés*. Budapest: Gondolat Kiadó.
- Horváth Iván 2008. A magyar vers a reneszánsz és a reformáció korában. 1536: Megjelenik két verskötet. In: Szegedy-Maszák Mihály (szerk.): *A magyar irodalom története I*. Budapest: Gondolat Kiadó. 236–249.
- Horváth Péter 2021. Két eljárás magyar nyelvű versek metrumának gépi felismertetéséhez. *Digitális Bölcsészet* 4: 79–103. <https://doi.org/10.31400/dh-hun.2021.4.2361>
- Horváth Péter 2023. Egy automatikusan generált rímshóttár fejlesztése és a magyar kanonikus költészet rímshóttáinak néhány jellemzője. In: Tick József – Kokas Károly – Holl András (szerk.): *Új technológiákkal, új tartalmakkal a jövő digitális transzformációja felé. Workshop 32: országos konferencia*. Budapest: HUNGARNET Egyesület. 77–84. <https://doi.org/10.31915/NWS.2023.12>

- Horváth Péter – Kundráth Péter – Indig Balázs – Fellegi Zsófia – Szlávich Eszter – Bajzát Tímea Borbála – Sárközi-Lindner Zsófia – Vida Bence – Karabulut Aslihan – Timári Mária – Palkó Gábor 2022a. ELTE Vers-korpusz – 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. <https://doi.org/10.31915/NWS.2022.35>
- Horváth Péter – Kundráth Péter – Palkó Gábor 2022b. ELTE Népdalkorpusz – magyar népdalok gépileg annotált adatbázisa. In: Tick József – Kokas Károly – Holl András (szerk.): *Valós térben – az online térért. Workshop 31: országos konferencia*. Budapest: HUNGARNET Egyesület. 276–283. <https://doi.org/10.31915/NWS.2022.35>
- Horváth Péter – Tátrai Szilárd 2023. A személy, az idő és a mód jelölésének igei mintázatai József Attila verseiben. In: Lőrincz Csongor – Halász Hajnalka (szerk.): *A lírai hang túloldalai. József Attila-olvasatok*. Budapest: Prae Kiadó. 207–237.
- Indig Balázs – Sass Bálint – Simon Eszter – Mittelholcz Iván – Kundráth Péter – Vadász Noémi 2019. emtsv – Egy formátum mind felett. In: Berend Gábor – Gosztolya Gábor – Vincze Veronika (szerk.): *XV. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: SZTE Informatikai Intézet. 235–247.
- Jarvis, Scott 2013. Capturing the diversity in lexical diversity. *Language Learning* 63 (1): 87–106. <https://doi.org/10.1111/j.1467-9922.2012.00739.x>
- Johnson, Wendell 1944. Studies in language behavior: I. A program of research. *Psychological Monographs* 56 (2): 1–15. <https://doi.org/10.1037/h0093508>
- Kyle, Kristopher – Crossley, Scott A. 2015. Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly* 49 (4): 757–786. <https://doi.org/10.1002/tesq.194>
- Ladányi Mária 2017. Alaktan. In: Tolcsvai Nagy Gábor (szerk.): *Nyelvtan*. Budapest: Osiris Kiadó. 501–660.
- Lakoff, George – Johnson, Mark 1980. *Metaphors we live by*. Chicago: The University of Chicago Press.
- Langacker, Ronald W. 2008. *Cognitive Grammar: A basic introduction*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195331967.001.0001>

- Langendonck, Willy van 2007. *Iconicity*. In: Geeraerts, Dirk – Cuyckens, Hubert (eds.): *The Oxford handbook of Cognitive Linguistics*. Oxford: Oxford University Press. 394–418.
- Laufer, Batia – Nation, Paul 1995. Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics* 16 (3): 307–322. <https://doi.org/10.1093/applin/16.3.307>
- Malvern, David – Richards, Brian 2002. Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing* 19 (1): 85–104. <https://doi.org/10.1191/0265532202lt221oa>
- Naismith, Ben – Han, Na-Rae – Juffs, Alan – Hill, Brianna – Zheng, Daniel 2018. Accurate measurement of lexical sophistication in ESL with reference to learner data. In: *Proceedings of the 11th International Conference on Educational Data Mining*. Buffalo, NY (USA): International Educational Data Mining Society (IEDMS). 259–265.
- Novák Attila – Rebrus Péter – Ludányi Zsófia, 2017. Az emMorph morfológiai elemző annotációs formalizmusa. In: Vincze Veronika (szerk.): *XIII. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi Tudományegyetem, Informatikai Intézet. 70–78.
- Piperski, A[leksander] Ch. 2017. Semantic halo of a meter: A keyword-based approach. In: *Kompiuternaia lingvistika i intelektualnyie tekhnologii*. Vol. 2. *Kompiuternaia lingvistika: lingvisticheskie issledovaniia*. Moscow: RGGU. 342–354.
- Pusch, Claus D. 2001. Ikonizität. In: Haspelmath, Martin – König, Ekkehard – Oesterreicher, Wulf – Raible, Wolfgang (eds.): *Language typology and language universals. Handbücher zur Sprach- und Kommunikationswissenschaft*. Band 20.1. Berlin, New York: Walter de Gruyter. 369–384.
- Rosch, Eleanor 1978. Principles of categorization. In: Rosch, Eleanor – Lloyd, Barbara B. (eds.): *Cognition and categorization*. Hillsdale: Lawrence Erlbaum. 27–48. <https://doi.org/10.4324/9781032633275-4>
- Schmid, Hans-Jürgen 2007. Entrenchment, salience, and basic levels. In: Geeraerts, Dirk – Cuyckens, Hubert (eds.): *The Oxford handbook of Cognitive Linguistics*. Oxford: Oxford University Press. 117–138.
- Scott, Mike 2007. *Oxford WordSmith tools: Version 4.0*. Oxford: Oxford University Press.
- ŠeĽa, Artjoms – Orekhov, Boris – Leibov, Roman 2020. Weak genres: Modeling association between poetic meter and meaning in Russian poetry. In: Karsdorp, Folger – McGillivray, Barbara – Nergheš, Adina – Wevers, Melvin (eds.): *CHR 2020: Workshop on Computational Humanities Research*. Amsterdam: CEUR-WS. 12–31.

- Šeļa, Artjoms – Plecháč, Petr – Lassche, Alie 2022. Semantics of European poetry is shaped by conservative forces: The relationship between poetic meter and meaning in accentual-syllabic verse. *PLoS ONE* 17 (4): e0266556. <https://doi.org/10.1371/journal.pone.0266556>
- Simon Gábor 2014. *Egy kognitív poétikai rímelmélet megalapozása*. Budapest: Tinta Könyvkiadó.
- Tátrai Szilárd – Ballagó Júlia 2020. A stílustulajdonítás szociokulturális szituáltsága. Funkcionális kognitív pragmatikai megközelítés. *Magyar Nyelvőr* 144 (1): 1–43.
- Tolcsvai Nagy Gábor 1996. *A magyar nyelv stilisztikája*. Budapest: Nemzeti Tankönyvkiadó.
- Tweedie, Fiona J. – Baayen, R. Harald. 1998. How variable may a constant be? Measures of lexical richness in perspective. *Computers and the Humanities* 32 (5): 323–352. <https://doi.org/10.1023/A:1001749303137>
- Váradí Tamás – Simon Eszter – Sass Bálint – Gerőcs Mátyás – Mittelholcz Iván – Novák Attila – Indig Balázs – Prószéky Gábor – Farkas Richárd – Vincze Veronika 2017. Az e-magyar digitális nyelvfeldolgozó rendszer. In: Vincze Veronika (szerk.): *XIII. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: SZTE Informatikai Intézet, 49–60.



Patterns of lexical sophistication and semantic specificity in Hungarian canonical poetry

This paper investigates the relationship between vocabulary and text size, vocabulary and verse form, and vocabulary and structural units in the poems of 24 Hungarian canonical authors, measuring lexical sophistication through the frequency of words in a reference corpus, which is a method developed in L2 research. Firstly, it shows that the lexical sophistication of medium-length poems is typically slightly lower than the lexical sophistication of short and long poems. Secondly, it highlights that the lexical sophistication of poems with a simple verse form is usually lower than that of the others, which can be explained on the basis of an iconic relationship between verse form and content. Thirdly, it also demonstrates that the lexical sophistication of poems typically decreases as the poem progresses, which can be related to the generalization of the content of the

poems. Finally, the paper presents a general pattern regarding odd and even lines in the stanzas: the lexical sophistication of even lines is higher than the lexical sophistication of odd lines. This pattern suggests that the line pairs form semantic units in which the more general meaning of the first line and the more specific meaning of the second line are connected in an asymmetric semantic relation. Based on the quantitative results, a two-level scheme of lexical sophistication and semantic specificity of the prototypical poem in Hungarian canonical poetry can be described.

Keywords: type-token ratio, lexical sophistication, semantic specificity, ELTE Poetry Corpus, verse form, iconicity, stanza, poem line

Horváth Péter

Eötvös Loránd Tudományegyetem

horvath.peter@btk.elte.hu

ORCID: 0000-0002-3517-5623

Függelék

13. táblázat. Az egyes terjedelmi csoportokba és az egyszerű (Izo.), illetve nem egyszerű (Hetero.) csoportokba tartozó versek száma szerzőnként, a Verskorpusz (Átlag V.) és a Regénykorpusz (Átlag R.) alapján számított lexikai választékossági értékek átlagaival

Szerző	0–49			50–199			200–999			Izo.			Hetero		
	db	Átlag V.	Átlag R.	db	Átlag V.	Átlag R.	db	Átlag V.	Átlag R.	db	Átlag V.	Átlag R.	db	Átlag V.	Átlag R.
Csokonai	70	2,053	2,227	179	2,018	2,209	127	2,023	2,161	3	2,443	2,627	369	2,026	2,196
Berzsenyi	16	2,183	2,302	92	2,076	2,295	28	2,055	2,253	0	no	no	130	2,078	2,283
KisfaludyK	20	2,011	2,180	67	1,915	2,100	32	2,030	2,217	0	no	no	116	1,959	2,145
Kölcsey	38	2,014	2,313	92	1,981	2,310	20	2,025	2,282	0	no	no	146	1,996	2,309
Vörösmarty	287	1,987	2,123	247	1,935	2,116	111	1,934	2,111	0	no	no	527	1,944	2,112
AranyJ	130	2,049	2,155	140	1,974	2,099	118	2,010	2,113	5	2,160	2,275	359	2,003	2,116
Tompa	29	1,873	2,051	152	1,922	2,136	291	1,945	2,107	4	1,821	2,063	467	1,935	2,114
Petőfi	166	1,831	1,979	528	1,832	1,969	135	1,870	1,984	9	1,755	1,945	813	1,839	1,973
Madách	88	1,967	2,044	121	1,853	2,067	104	1,888	2,079	1	1,704	1,741	277	1,892	2,078
Gyulai	12	1,912	2,119	111	1,818	2,023	31	1,868	2,060	0	no	no	154	1,835	2,038
Vajda	14	1,904	2,039	104	1,864	1,992	68	1,853	1,980	3	1,614	1,773	181	1,863	1,991
Reviczky	33	1,852	2,011	236	1,858	2,035	65	1,891	2,036	5	1,908	2,098	329	1,863	2,032
Komjáthy	32	1,878	2,069	110	1,906	2,149	104	1,900	2,129	10	1,839	2,072	230	1,912	2,144
Ady	116	1,864	1,992	918	1,863	2,004	82	1,880	2,030	21	1,893	2,040	1095	1,864	2,004
Kaffka	3	2,083	2,238	63	1,976	2,123	33	1,986	2,088	0	no	no	99	1,983	2,115
Somlyó	17	1,754	1,923	326	1,897	2,051	33	1,943	2,070	19	1,821	1,952	357	1,899	2,052
Juhász	156	1,825	2,052	1086	1,829	2,061	36	1,908	2,097	93	1,768	2,008	1171	1,834	2,063
Babits	18	1,827	2,004	363	1,933	2,107	129	1,933	2,079	8	1,756	1,898	502	1,932	2,099
Kosztolányi	74	1,907	2,070	467	1,919	2,085	85	1,961	2,130	14	1,878	2,063	602	1,924	2,090
Tóth	79	1,977	2,230	279	1,976	2,153	93	2,023	2,161	3	1,964	2,213	439	1,987	2,167
Reményik	61	1,949	2,104	525	1,914	2,046	81	1,902	2,019	5	1,812	1,921	662	1,917	2,049
József	84	2,037	2,121	467	1,965	2,089	46	1,966	2,062	8	1,824	1,893	578	1,977	2,094
Dsida	147	1,984	2,149	839	1,943	2,102	69	1,976	2,126	8	1,941	2,076	1045	1,951	2,110
Radnóti	37	1,958	2,108	225	1,974	2,118	35	1,916	2,054	1	1,916	2,096	296	1,965	2,109

14. táblázat. Az egyes terjedelmi csoportokba és az egyszerű (Izo.), illetve nem egyszerű (Hetero.) csoportokba tartozó versek száma szerzőnként, a Verskorpusz (Med. V.) és a Regénykorpusz (Med. R.) alapján számított lexikai választékossági értékek mediánjaival

Szerző	0–49			50–199			200–999			Izo.			Hetero.		
	db	Med. V.	Med. R.	db	Med. V.	Med. R.	db	Med. V.	Med. R.	db	Med. V.	Med. R.	db	Med. V.	Med. R.
Csokonai	70	2,025	2,165	179	2,000	2,200	127	2,008	2,146	3	2,378	2,538	369	2,008	2,176
Berzsenyi	16	2,218	2,327	92	2,068	2,297	28	2,064	2,287	0	no	no	130	2,071	2,297
KisfaludyK	20	2,025	2,156	67	1,937	02,09	32	2,013	2,169	0	no	no	116	1,982	2,145
Kölcsey	38	2,025	2,261	92	1,954	2,291	20	2,036	2,270	0	no	no	146	1,974	2,289
Vörösmarty	287	1,975	2,108	247	1,936	2,110	111	1,912	2,097	0	no	no	527	1,933	2,100
AranyJ	130	2,0	2,101	140	1,948	2,051	118	2,008	2,086	5	2,109	2,327	359	1,986	2,083
Tompa	29	1,865	2,050	152	1,919	2,133	291	1,945	2,104	4	1,878	2,082	467	1,938	2,117
Petőfi	166	1,806	1,974	528	1,813	1,960	135	1,859	1,961	9	1,765	1,983	813	1,821	1,962
Madách	88	1,947	2,028	121	1,847	2,049	104	1,885	2,069	1	1,704	1,741	277	1,874	2,062
Gyulai	12	1,874	2,091	111	1,814	2,013	31	1,895	2,079	0	no	no	154	1,838	2,035
Vajda	14	1,841	2,042	104	1,833	1,962	68	1,850	1,987	3	1,545	1,667	181	1,846	1,974
Reviczky	33	1,824	1,967	236	1,847	2,036	65	1,899	2,023	5	1,691	1,897	329	1,852	2,032
Komjáthy	32	1,853	2,079	110	1,869	2,133	104	1,901	2,103	10	1,836	2,116	230	1,902	2,124
Ady	116	1,858	1,972	918	1,850	2,000	82	1,873	2,023	21	1,907	2,024	1095	1,851	2,000
Kaffka	3	2,263	2,439	63	1,936	2,092	33	1,988	2,120	0	no	no	99	1,965	2,116
Somlyó	17	1,783	1,953	326	1,885	2,044	33	1,940	2,087	19	1,811	1,957	357	1,890	2,051
Juhász	156	1,783	2,034	1086	1,810	2,054	36	1,878	2,081	93	1,764	2,012	1171	1,812	2,056
Babits	18	1,818	1,996	363	1,909	2,097	129	1,941	2,087	8	1,809	1,934	502	1,923	2,089
Kosztolányi	74	1,897	2,051	467	1,914	2,087	85	1,943	2,123	14	1,901	2,056	602	1,925	2,093
Tóth	79	1,977	2,167	279	1,984	2,157	93	2,017	2,154	3	1,929	2,142	439	1,988	2,163
Reményik	61	1,914	2,098	525	1,909	2,041	81	1,890	2,004	5	1,745	1,823	662	1,909	2,041
József	84	2,019	2,132	467	1,941	2,058	46	1,984	2,065	8	1,767	1,837	578	1,957	2,070
Dsida	147	2,000	2,167	839	1,938	2,098	69	2,005	2,110	8	1,948	2,045	1045	1,944	2,103
Radnóti	37	1,979	2,125	225	1,959	2,101	35	1,908	2,069	1	1,916	2,096	296	1,953	2,105

15. táblázat. A minimum három versszakos versek száma, valamint az első, középső és utolsó versszakokra kapott lexikai választékossági értékek átlagai szerzőnként a Verskorpusz (Átlag V.) és a Regénykorpusz (Átlag R.) alapján

Szerző	Versek száma	Első versszak		Középső versszakok		Utolsó versszak	
		Átlag V.	Átlag R.	Átlag V.	Átlag R.	Átlag V.	Átlag R.
Csokonai	186	2,063	2,277	2,025	2,206	1,954	2,145
Berzsenyi	95	2,123	2,379	2,087	2,304	2,021	2,222
KisfaludyK	81	1,927	2,13	1,966	2,164	1,908	2,070
Kölcsey	90	1,986	2,314	1,997	2,307	1,958	2,267
Vörösmarty	215	1,943	2,146	1,921	2,098	1,862	2,047
AranyJ	237	2,025	2,164	2,014	2,121	1,969	2,101
Tompa	438	1,986	2,185	1,958	2,134	1,899	2,082
Petőfi	585	1,856	1,991	1,872	2,001	1,824	1,963
Madách	187	1,920	2,134	1,898	2,111	1,786	2,015
Gyulai	130	1,878	2,080	1,841	2,031	1,807	2,041
Vajda	168	1,859	1,990	1,899	2,026	1,845	1,979
Reviczky	296	1,920	2,083	1,880	2,058	1,832	2,008
Komjáthy	210	1,904	2,156	1,911	2,153	1,826	2,066
Ady	992	1,957	2,091	1,879	2,019	1,822	1,965
Kaffka	75	2,086	2,233	2,003	2,14	1,917	2,056
Somlyó	349	1,954	2,107	1,917	2,074	1,851	1,997
Juhász	1001	1,894	2,127	1,846	2,078	1,821	2,055
Babits	397	1,958	2,126	1,947	2,113	1,895	2,069
Kosztolányi	423	1,938	2,128	1,946	2,125	1,859	2,032
Tóth	336	2,033	2,197	2,009	2,181	1,961	2,132
Reményik	469	1,966	2,107	1,928	2,063	1,851	1,984
József	451	2,002	2,142	2,003	2,123	1,920	2,044
Dsida	842	1,979	2,135	1,969	2,129	1,884	2,049
Radnóti	194	2,015	2,163	1,971	2,120	1,942	2,104

16. táblázat. A minimum három versszakos versek száma az első (E), középső (K) és utolsó (U) versszakok lexikai választékossági értékeinek különböző lehetséges együttállásaiban, szerzőnként, a Verskorpusz (V) és a Regénykorpusz (R) alapján.

Pl.: $E > K > U$: az első versszak lexikai választékossága nagyobb, mint a középső versszak lexikai választékossága, a középső versszak lexikai választékossága pedig nagyobb, mint az utolsó versszak lexikai választékossága; $E > K << U$: az első versszak lexikai választékossága nagyobb, mint a középső versszak lexikai választékossága, az utolsó versszak lexikai választékossága pedig nagyobb, mint a középső és az első versszak lexikai választékossága; $E > K < U$: az első versszak lexikai választékossága nagyobb, mint a középső versszak lexikai választékossága, az utolsó versszak lexikai választékossága pedig nagyobb, mint a középső lexikai választékossága, de kisebb, mint az első versszak lexikai választékossága.

(A táblázatban nincsenek feltüntetve azok az esetek, amikor a vizsgált három szerkezeti egységből kettőnek vagy mind a háromnak ugyanakkora a lexikai választékossága.)

	Összes	$E > K > U$		$E > K << U$		$E > K < U$		$E < K < U$		$E < K >> U$		$E < K > U$	
		V	R	V	R	V	R	V	R	V	R	V	R
Csokonai	186	49	51	18	17	25	34	26	18	37	36	27	25
Berzsenyi	95	32	33	4	8	15	14	16	13	14	13	12	13
KisfaludyK	81	16	16	11	7	6	12	14	11	19	17	13	17
Kölcsey	90	20	25	7	9	15	15	17	16	17	15	11	9
Vörösmarty	215	53	66	27	24	25	23	34	42	36	38	32	21
AranyJ	237	68	60	21	28	29	42	46	44	37	33	32	29
Tompa	438	136	139	50	55	56	55	65	62	68	56	54	63
Petőfi	585	136	125	52	59	80	77	113	105	91	104	92	93
Madách	187	63	56	9	12	20	24	22	27	40	31	28	34
Gyulai	130	34	32	9	21	21	16	25	24	24	10	14	21
Vajda	168	46	42	16	17	18	18	30	38	26	25	30	26
Reviczky	296	83	78	33	37	48	43	37	35	53	51	32	48
Komjáthy	210	65	49	15	20	19	25	34	28	39	48	26	27
Ady	992	285	271	107	107	158	162	112	123	166	171	98	99
Kaffka	75	25	28	6	4	12	12	8	8	16	16	8	6
Somlyó	349	93	99	28	27	59	57	56	53	72	61	34	43
Juhász	1001	264	250	130	144	164	155	152	155	155	152	104	123

	Összes	E > K > U		E > K << U		E > K < U		E < K < U		E < K >> U		E < K > U	
		V	R	V	R	V	R	V	R	V	R	V	R
Babits	397	103	105	37	35	52	64	64	70	71	61	61	56
Kosztolányi	423	110	115	47	44	44	43	59	62	82	86	67	50
Tóth	336	83	81	48	53	44	47	41	41	68	65	43	46
Reményik	469	128	135	43	53	70	67	59	52	89	82	65	65
József	451	104	101	48	49	70	69	60	66	92	91	66	66
Dsida	842	219	202	87	89	118	122	113	126	165	161	120	123
Radnóti	194	53	55	28	27	31	24	22	23	35	34	18	25

SZEMES BOTOND

**Hozzászólás A lexikai választékosság és szemantikai
specifikusság mintázatai a magyar kanonikus
költészetben című tanulmányhoz**

Horváth Péter tanulmánya egyszerre inspiratív irodalomtörténeti, poétikai, módszertani és nyelvelméleti szempontból is. Rövid hozzászólásomban csak ez utóbbira szorítkozom, mivel úgy gondolom, a digitális bölcsészet, illetve a korpusznyelvészet kvantitatív eredményei fontos tanulságokkal szolgálnak a nyelvi működés egészének megértéséhez, új fénytörésbe helyezve a korábbi belátásokat – mégis kevés szó esik ezekről és a belőlük levonható következtetésekről. Egyre több kutatás támasztja ugyanis alá, hogy a nyelvi jel önkényességén túl olyan kognitív meghatározottságok is szervezik megnyilatkozásainkat, amelyek nemcsak vizsgálhatók statisztikai módon, hanem gyakran maguk is implicit statisztikai tudáson alapulnak. Hozzászólásomat motiválja továbbá, hogy Horváth Péter szintén ilyen irányú értelmezéseket vet fel szövegében, hangsúlyosabban korábbi, hasonló kutatásaihoz képest (ezek közül kiemelendők: A szereplői tudat megjelenítésének kvantitatív vizsgálata Ady, Babits és Kosztolányi összes versében [2019], Zárt fogalmi tartományok a magyar kanonikus költészetben – egy korpusznyelvészeti megközelítés [2022], A személy, az idő és a mód jelölésének igei mintázatai József Attila verseiben [2023], A versbeszélő nyelvi jelölésének összefüggései a versek szerzőjével, keletkezési idejével, hosszúságával és metrumával [2024].)

Mindenekelőtt érdemes felidézni néhány példát a kvantitatív kutatások nyelvelméleti jelentőségéről. A szerzőazonosítás bevett eljárásai kimutatták, hogy egy korszak, műfaj vagy egyén szövegei elkülöníthetők másoktól a leggyakrabban használt, elsősorban tartalmas jelentés nélküli funkciószavak eloszlását vizsgálva. Ez a jelenség arra mutathat rá, hogy egy nyelvhasználó megnyilatkozásainak nagy részét olyan szerkezetek (szavak, szókapcsolatok, nyelvi sémák) automatikus aktiválásával hozza létre, amelyek éppen gyakori használatuk miatt „készen kapva”, kevés mentális erőfeszítéssel

hozzáférhető a procedurális emlékezetben (Nini 2023). Vagyis nem elvont nyelvtani szabályok, hanem e sémák bevésoódése és gyakori alkalmazása tekinthető a nyelvhasználat alapjának. Az egyéni sajátosságok mellett ugyanakkor közösségi szempontok is hatással vannak a gyakoriságok alakulására: a természetes nyelvekre jellemző Zipf-féle eloszlás például a nyelvelsajátítás során játszik fontos szerepet, amennyiben a minél jobban elkülönülő, ismétlődő sémák a könnyebb tanulhatósághoz járulnak hozzá (Arnon–Kirby 2024). Ugyanezt a célt szolgálhatják az egyes nyelvosztályok formai megvalósulásában megfigyelhető szisztematikus egyezések is (például több nyelvben, így a magyarban is, a közelre mutató névmásokban tendenciózan magas hangrendű magánhangzók, míg a távolra mutatókban mélyek szerepelnek, vö. Dingemanse et. a. 2015). Számszerűsíthető továbbá az is, hogy a nyelvi egységek milyen gyakorisággal fordulnak elő egymás kontextusában; ezen a statisztikai tudáson alapul a nagy nyelvmodellek (LLM) teljesítménye – és feltételezhető, hogy az emberi beszélők is (többek között) erre hagyatkoznak a jelentések létrehozásakor.

Horváth Péter tanulmánya szintén szavak gyakoriságából indul ki, hogy megragadja azok előfordulásának valószínűségét, így a rájuk jellemző lexikai választékosság mértékét. Mindezt nem elsősorban a kifejezések stíláriis jelöltségének, hanem szemantikai komplexitásuk és specifikusságuk kvantifikálása érdekében teszi. Ezáltal az általánosság-specifikusság, illetve az egyszerűség-komplexitás tengelyén tudja elhelyezni a verseket és részeit, amin keresztül a szövegek általános mintázatait képes azonosítani. Kimutatja, hogy még a leginkább kreatív és reflektív nyelvhasználat, a lírai szöveg esetében is tetten érhető kognitív meghatározottságok, mint például a specifikustól az általános felé tartó induktív logika előnyben részesítése. Mindez összefüggésben állhat a versek hatásmechanizmusával és poétikai felépítésével is, amennyiben a váratlan, specifikus jelentésektől az absztraktabb tartományokig vezető út egyrészt a figyelem megragadásához, másrészt ahhoz járul hozzá, hogy versolvasás során az egyediben ismerhessük fel az univerzalist (ami rokonságot mutathat a költői képeknek a dolgozatban is megidézett, kognitív nyelvészeti értelmezésével).

Látni kell ugyanakkor, hogy gyakoriak az ellenpéldák is a vizsgált korpuszban és a művek szervezőoódése sok esetben nem felel meg az azonosított sémáknak. Ez persze nem a statisztikai eredmények hitelességét kérdőjelezi meg (azt csak erősíti), hanem a lírai megszólalás alapvető sajátosságára hívja fel a figyelmet: ez ugyanis éppen a hétköznapi, sematikus nyelvhasználat megtörésében, kibillentésében érdekelt. Azaz a tanulmány eredményei felől

úgy is fogalmazhatnánk, hogy a költői szöveg az alapvető kognitív sémák hatásos kidolgozásának és felülírásának dinamikájában alakulnak. Ez az alakulás az egyes versekben eltérően realizálódik; a tanulmányban idézett, szemléletes példák felvetik a szavakhoz rendelt lexikai változatosságból kiinduló verselemzések lehetőségét is.

A vázolt nyelvelméleti belátások felől a versek formája és tartalma között feltárt összefüggés válik különösen érdekessé. Eszerint a versek hossza és formai összetettsége, valamint szemantikai komplexitása között ikonikus kapcsolat áll fenn: minél egyszerűbb, vagyis szabályosabb ismétlődéséken alapul, illetve minél átlagosabb hosszúságú egy költemény, annál inkább gyakoriak benne a mentálisan könnyebben hozzáférhető kifejezések. Ez az ikonikus kapcsolatnak a bevett nyelvészeti (a grammatikai-morfológiai és a szemantikai szintek megfelelésből kiinduló) értelmezéshez képest egy általánosabb érvényű, statisztikailag megragadható vizsgálatát teszi lehetővé. A kvantitatív kutatás ez esetben is olyan mintázatok feltárását teszi lehetővé, amelyek az eseti megvalósulásokban, „szabad szemmel” nem mindig érzékelhető mélystruktúrákhoz való hozzáférést biztosítják. Figyelemre méltó, hogy mindezt lírai művek kapcsán mutatja be a tanulmány, hozzájárulva ahhoz, hogy a Roman Jakobson által leírt poétikai funkciót, azaz egy nyelvi közlés formai és tartalmi dimenziójának ikonikus megfelelését, ne kizárólag a költői lelemény eredményeként ismerjük fel, hanem mint alapvető, általában öntudatlanul működtetett eljárást. Megnyilatkozásainkat erősen befolyásolja a szavakra és egyéb szerkezetekre vonatkozó implicit statisztikai tudásunk – még a költészet birodalmában is.

Irodalom

- Arnon, Inbal – Kirby, Simon 2024. Cultural evolution creates the statistical structure of language. *Scientific Reports* 14: 5255. <https://doi.org/10.1038/s41598-024-56152-9>
- Dingemanse, Mark – Blas, Damián E. – Lupyan, Gary – Christiansen, Morten H. – Monaghan, Padraic 2015. Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences* 19 (10): 603–615. <https://doi.org/10.1016/j.tics.2015.07.013>
- Nini, Andrea 2023. *A theory of linguistic individuality for authorship analysis*. Elements in forensic linguistics. Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/9781108974851>

Szemes Botond

Bölcsészettudományi Kutatóközpont; Digitális Bölcsészet Nemzeti Labora-
tórium

szemes.botond@abtk.hu

ORCID: 0000-0002-0637-6776