

VIZKELETY ANDRÁS 2009b. Ómagyar Mária-siralom. In: MADAS EDIT szerk., „*Látjátok feleim...*”. *Magyar nyelvemlékek a kezdetektől a 16. század elejéig*. Országos Széchényi Könyvtár, Budapest. 226–227.

Could we gain new insight into the discovery of the Leuven Codex?

The Rosenthal and Olschki families in the 19th–21st centuries

This paper aims to promote further inquiry into the following issue: Where did the Leuven Codex, containing the earliest surviving Hungarian poem, the Old Hungarian Lament of Mary, hide before its discovery in the 20th century? The first well-established event in the history of this codex is that, in 1910, it was purchased in Toscana by Jacques Rosenthal, an antiquarian from Munich. In 1922, the book was transferred to the library of the University of Leuven (hence the name of the codex). Finally, since 1982, it has been stored in the National Széchényi Library in Budapest, under catalogue number MNy 79. – A new avenue of research may be opened by a monograph presenting the history of the Rosenthal family (ANGERMAIR et al. 2002). It describes the extraordinary antiquarian activities of Jacques Rosenthal and his descendants, including friendly and familial relations (dating back to the beginning of the 20th century) between Jacques Rosenthal and Leo Olschki, an antiquarian from Florence. This Florence link undoubtedly deserves close inspection for gaining new insight into the history of the codex. By sharing key information about the Rosenthal and Olschki families, the paper aims to give new impetus to research on the codex's history prior to 1910.

Keywords: Leuven Codex, Lament of Mary, Jacques Rosenthal, Leo S. Olschki.

KOROMPAY KLÁRA
ELTE Eötvös Loránd Tudományegyetem

SZEMLE

Korpusznyelvészeti módszerek a kognitív nyelvészeti kutatásban

Szerk. KUGLER NÓRA, MODRIÁN-HORVÁTH BERNADETT, SIMON GÁBOR. DiAGram Könyvek 10. Eötvös Loránd Tudományegyetem Bölcsészettudományi Kar, Budapest, 2025. 336 oldal

1. Adatok és elméletek metszéspontján. Mint az a címből kiderül, a kötet a kognitív nyelvészet és a korpusznyelvészeti elemzések metszéspontján helyezkedik el. A tanulmányok olyan esettanulmányokat mutatnak be, amelyek a magyar nyelv szintaktikai, szemantikai és morfológiai szerkezeteit, valamint nyelvhasználati mintázatait vizsgálják. A kötet egyik legfőbb érdeme, hogy innovatív adatfeldolgozási és elemzési módszereket vonultat fel, amivel hozzájárul a hazai nyelvészeti kutatások metodológiai megújulásához. A kognitív nyelvészeti keret összekapcsolása a nagyméretű korpuszok elemzésével új meglátásokat eredményez: a nyelvi jelenségeket mind elméleti, mind empirikus szempontból mélyebben érthetjük meg.

Szintén figyelemre méltó a kötet felépítése, hiszen nem csupán a tanulmányokat tartalmazza, hanem az egyes írásokhoz fűzött lektori hozzászólásokat is. Ez a szerkesztői megoldás lehetővé teszi az egyes témák alaposabb megvitatását és a különböző nézőpontok ütköztetését. A hozzászólások változatos hangvételűek: egyesek méltatják a tanulmányok újdonságértékét és szakmai megalapozottságát, míg mások óvatos kritikákat fogalmaznak meg, elsősorban a módszertani megközelítések finomhangolását és az értelmezési keretek pontosítását javasolva. Összességében e reflexiók kritikus, de konstruktív módon járulnak hozzá a tudományos párbeszédhez, tovább emelve a kötet értékét és koherenciáját.

2. Az egyes fejezetek ismertetése. A kötet bevezető tanulmányát a szerkesztők, KUGLER NÓRA, MODRIÁN-HORVÁTH BERNADETT és SIMON GÁBOR írták, célja a kötet koncepcionális és elméleti kereteinek megrajzolása. A szerzők bemutatják, hogyan kapcsolódhat egymáshoz a korpusznyelvészet és a kognitív nyelvészet mint két eltérő, de egyre inkább egymást kiegészítő kutatási irányvonal. A bevezető részletesen tárgyalja a korpusznyelvészet kognitív nyelvészeti integrációjának elméleti hátterét, a kezdeti szkepsztől az együttműködés előnyei felé mutató tendenciákig. Rávilágítanak arra, hogy a korpusznyelvészet nem csupán módszertani segédeszköz lehet, hanem a kognitív modellek empirikus tesztelésének és finomításának eszköze is. A szerkesztők kiemelik, hogy a kötet tanulmányai ezt az integratív megközelítést példázzák, különös figyelmet fordítva a magyar igei konstrukciók vizsgálatára. Emellett a bevezetés ismerteti a fentebb említett sajátos szerkesztési elvet is: a tanulmányokhoz nevesített lektori hozzászólások és szerzői válaszok is tartoznak, ezzel elősegítve a nyílt tudományos diskurzust és a tudás közös alakítását.

2.1. Szimultán igekötőzés és igeképzés. KALIVODA ÁGNES és PALÁGYI LÁSZLÓ tanulmánya Szimultán igekötőzés és igeképzés – konstrukcióegyesítés a magyarban címmel a magyar igeragozás egyik különleges jelenségét, a prefixumok és az igei képzők együttes előfordulását vizsgálja, amelyet a szerzők konstrukcióegyesítésként értelmeznek. A kutatás abból a hipotézisből indul ki, hogy ezen szerkezetek mögött a morfológiai konstrukciók hálózatos szerveződése áll, és ezek működését a nyelvi sémák és az analógia alapvetően befolyásolják. Két passzív állapotváltozást kifejező sémára fókuszálnak: a *-(s)Ul* és a *-(s)Odik* képzős konstrukciókra, különös tekintettel azok igekötős megvalósulásaira.

A tanulmány empirikus alapját a Magyar Nemzeti Szövegtár 2.0 UD (MNSZ2-UD) adja. A szerzők ebből kinyerték az összes releváns adatot: az *-(s)ódik* és *-(s)ül* képzős igék előfordulásait, igekötőkkel együtt, kontextusban és metaadatokkal ellátva. Az így létrehozott, több mint 1,1 millió adatpontot tartalmazó, nyíltan elérhető adattábla nemcsak a kutatás transzparens adatvezérelt szemléletét jelzi, hanem egyben értékes erőforrás a további vizsgálatok számára is.

A szerzők statisztikai és hálózatelemző módszerekkel térképezik fel az igekötős és képzős igealakok használati mintázatait. A gyakorisági és produktivitási mutatók elemzésével kimutatják, hogy az ilyen szerkezetek nemcsak gyakoriak, hanem rendszerszerűen szerveződnek a magyar nyelvben. Az egyik legfontosabb eredmény, hogy az igekötő megléte növeli a képző produktivitását: míg például a *-(s)Ul* képző hagyományosan nem számított produktívnak, a vizsgálat alapján igekötővel együtt egyértelműen azzá válik.

A tanulmány külön érdeme, hogy összekapcsolja a nyelvi sémák elméleti keretét a gyakorlati, korpusz alapú elemzéssel, és új megvilágításba helyezi a magyar morfológia produktivitásával kapcsolatos kérdéseket. Tematikusan is szoros kapcsolatban áll a kötet

későbbi passzív konstrukciókat elemző fejezetével, így a két tanulmány együtt komplex képet ad a magyar passzív szerkezetek használatáról és működéséről.

Lektori hozzászólások: INDIG BALÁZS méltatta a vizsgálat módszertani alaposágát, ugyanakkor véleménye szerint a cikk az interdiszciplináris jellegéből adódóan jobban hajlik a leíró nyelvészeti irány kibontására, ami több helyen a számítógépes szempontok kidolgozásának rovására történik. LADÁNYI MÁRIA szintén méltató megjegyzésekkel kezd, jelezve, hogy a tanulmány új perspektívákat kínál a konstrukcióegyesítés jelenségének megértéséhez, és jelentős mértékben járul hozzá a magyar morfológia leírásához azáltal, hogy egy eddig csak részben értelmezett jelenségre kínál használat alapú, empirikusan megalapozott magyarázatot, miközben sikeresen ötvözi a hálózatelméleti és nyelvtechnológiai módszereket, ugyanakkor kiemeli, hogy a használat alapú megközelítés szellemében további mélyebb adatelemzésre lenne szükség a jelenség teljesebb feltárásához.

2.2. Appozitív határozók korpusz alapú vizsgálata. INDIG BALÁZS, KURUCZ RÉKA, ERNHÖFFER ALIZ és KUGLER NÓRA tanulmánya az appozitív határozók – vagyis a főnevek utáni, esetragos bővítményekkel megvalósuló, értelmező funkciójú szerkezetek – korpusz alapú vizsgálatát végzi el, különös figyelemmel azok azonosítási nehézségeire és kognitív nyelvészeti értelmezésére. A szerzők kiindulópontja, hogy ezek a posztnominatív struktúrák (pl. *visszaélés a hatalommal, kilábalás a válságból*) finom átmenetet képeznek a határozói és a jelzői funkciók között, és ezért különösen izgalmas kérdés, hogy milyen nyelvi mintázatokkal rokoníthatók. A vizsgálat során az appozitív határozói konstrukciókat a *való* kötőszóval jelölt szerkezetekkel (*válságból való kilábalás*) és a „szomszédos” szórendű, de *való* nélküli formákkal (*válságból kilábalás*) vetik össze, e három konstrukciót egységes hálózatos keretben értelmezve. A kutatás során a Magyar Nemzeti Szövegtár adataira támaszkodva fokozatosan fejlesztették ki keresési módszertanukat, mely végül lehetővé tette a különféle szerkezetek automatizált azonosítását és összevetését. A kvantitatív elemzések alapján kirajzolódik, hogy bár az appozitív és a *való*-s szerkezetek funkcionálisan átfedést mutatnak, szórásuk és természetességük nem azonos minden nyelvi környezetben. A tanulmány jelentős módszertani és elméleti tanulságokat hordoz: egyrészt bemutatja, hogyan azonosíthatók a finom grammatikai alternációk nagy korpuszokban, másrészt hozzájárul a kognitív szemantikai elemzések eszköztárának gazdagításához az appozitív szerkezetek vizsgálatán keresztül.

Lektori hozzászólások: KALIVODA ÁGNES és PALÁGYI LÁSZLÓ elismerik a kutatás összetettségét a konstrukció automatikus azonosításában, ám megjegyzik, az elemzés sok esetben prozódiai vagy kontextuális diszambiguálást igényel, amit a gépi elemzők jelenleg nem tudnak megbízhatóan kezelni. A kutatás azonban jelen formájában is átgondolt, jól dokumentált és reprodukálható, eredményei pedig jelentős elméleti és nyelvtechnológiai értékkel bírnak.

2.3. Beférkőző igék kollostrukciós vizsgálata. KALIVODA ÁGNES, MODRIÁN-HORVÁTH BERNADETT és SIMON GÁBOR közös munkája a beférkőző igék jelenségét elemzi kollostrukciós módszerekkel. A beférkőzésen azt a szerkezeti megoldást értjük, amikor egy ige úgy ékelődik be, hogy elválasztja az igekötőt a főnévi igenévtől – például az olyan szerkezetekben, mint *le tudja írni*, ahol a *leírni* igekötős szerkezetbe a *tudja* ige ékelődik. A szakirodalom ezt a jelenséget az auxiliáris (segédigei) funkció formailag jól megragadható jelenségeként tartja számon. A tanulmány ugyan nem törekszik a magyar segédigék kategóriájának meghatározására, mégis értékes ebből a szempontból is, mivel lehetőséget

ad az igék kvantitatív jellemzésére a beférkőzés – mint potenciálisan segédigei viselkedés – és a konstrukcióhoz való kapcsolódásuk erőssége alapján.

A kutatás az MNSZ2-UD korpusz adataira támaszkodva kolexémaelemzést (collostructional analysis) alkalmazott. Ez a módszer lehetővé teszi annak számszerűsítését, hogy az egyes igék a gyakoriságukhoz képest milyen erősen asszociálódnak a vizsgált szerkezettel. Az eredmények összesen hét darab statisztikailag szignifikánsan vonzónak mutatózó ígét azonosítottak, amelyek jelentése jobbra modális (módbeli) vagy temporális (időbeli) jellegű. Ezek a segédige-gyanús igék – nem meglepő módon – szinte pontosan megegyeznek azokkal a központi segédigékkel, amelyeket TOLCSVAI NAGY GÁBOR (2009) elméleti alapon már korábban kiemelt. A kapott lista megerősíti, hogy a magyar nyelvben szűk körű, jól körülhatárolható csoportja van az ilyen grammatikai szerepű igéknek (mint például a *fog, tud, szokott* stb.). A szerzők kitérnek néhány olyan ige példájára is, amelyek előzetes várakozás alapján segédigeként funkcionálhatnának, de a korpuszadatok alapján mégsem sorolódtak a legszignifikánsabbak közé – ezeket az eredményeket is értelmezik, jelezve, hogy a segédigeiség fokozati kérdés, és bizonyos igék átmeneti státuszúak lehetnek. Összességében a tanulmány fontos empirikus bizonyítékokkal szolgál a magyar segédigék körének meghatározásához, és módszertani példát nyújt arra, hogyan tárható fel egy jelenség rejtett lexikai szintű struktúrája a korpusz asszociációs mintázatai alapján. E fejezet eredményei szorosan illeszkednek a kötet többi, igei szerkezetekkel foglalkozó tanulmányához (különösen az 1. és 4. fejezethez), hiszen mind a prefixumok szerepét, mind a grammatikalizálódó igék körét a használati statisztikák felől közelítik meg.

Lektori hozzászólások: DÖMÖTÖR ANDREA kiválóan ítélte a tanulmányban alkalmazott statisztikai-nyelvészeti módszertant, de úgy vélte, hogy az osztályozás további finomítása még árnyaltabbá tehetné az eredményeket. Egy további, a korpuszra vonatkozó megjegyzése, hogy a beférkőző igék listáján szerepel a *szabad* is, „igeként” annotálva. Bár ennek részletes vizsgálata valószínűleg túlmutat a tanulmány keretein, érdekes és releváns kérdés lenne feltárni, hogy pontosan milyen esetekben jelenik meg a *szabad* igeként a korpuszban, és mennyire következetes ez az annotáció. IMRÉNYI ANDRÁS szintén dicsérte az elméleti háttér alapos kifejtését, azonban megjegyzi, hogy a korpuszelemzés viszonylag kevés lexémát azonosított szignifikánsan beférkőző igeként, miközben több, látszólag központi ige – mint például *szokott, tetszik* vagy akár *szeretne* – nem került be a szignifikáns eredmények közé. Ez arra utalhat, hogy az alkalmazott módszer további finomításra és fejlesztésre szorul, melyhez a lektor számos új szempontot javasol.

2.4. A magmondat szerepű, passzív jelentésű sémák produktivitása. MODRIÁN-HORVÁTH BERNADETT, KALIVODA ÁGNES és TANOS BÁLINT tanulmánya négy, különböző mértékben passzív jelentésű igei konstrukció (*V-ódik/-ődik, V-(t)atik/-(t)etik, V-va/-ve van, V-va/-ve lesz*) globális produktivitását vizsgálja korpusz alapú módszerekkel. A kutatás az MNSZ2 és az újraannotált MNSZ2-UD korpuszok gépi és kézi feldolgozásán alapul, és ötvözi a kvantitatív mutatókon alapuló produktivitási indexet (hapaxok, típus- és példánygyakoriság) a kvalitatív elemzéssel, különösen az igenem szerinti besorolással. A szerzők egyik központi kérdése az volt, hogy összehasonlítható-e a szintetikus és analitikus konstrukciók produktivitása, amire a *V-ni fog* kontrollkonstrukció elemzése alapján nemleges választ kaptak. Fontos kiemelni, hogy a szerzők számára nem az abszolút gyakoriságok, hanem a passzív jelentésű konstrukciók produktivitási mutatóinak összehasonlítása volt lényeges, mivel a produktivitást – mint az általánosan elfogadottnak tekinthető – fokozatos

jelenségként értelmezik. Ennek vizsgálatához kontrollként bevonták az analitikus *V-ni fog* konstrukciót is, hogy teszteljék, vajon az analitikus szerkezetek nagyobb szintaktikai kombinálhatósága önmagában növeli-e a produktivitást, valamint azért is, mert ez a konstrukció az inflexiók tartományhoz közel áll, amelyet korábbi kutatások (pl. GAETA 2007) jellemzően magas produktivitással hoztak összefüggésbe.

Az eredmények szerint a *V-va/-ve van* séma rendelkezik a legnagyobb globális produktivitással, a legtöbb típust a *V-ódik/-ődik* konstrukció mutatja, míg a *V-va/-ve lesz* séma bizonyult a leginkább potenciálisan produktívnak. A hapaxok elemzése alapján a *V-(t)atik/-(t)etik* konstrukció a legstabilabban passzív jelentésű, amit szorosan követ a *V-va/-ve lesz*, míg a *V-ódik/-ődik* esetében a passzív jelentés csak a hapaxok kevesebb mint felében volt megfigyelhető. A tanulmány részletesen vizsgálja az alkorpuszok szerinti eltéréseket is: például a „személyes” stílusú szövegekben a *V-va/-ve lesz* konstrukció kiemelkedően produktívnak és passzívnak bizonyult, így ez tekinthető a dinamikus passzív jelentéstartalom legfőbb hordozójának.

A kutatás külön érdeme, hogy nemcsak a nyelvi adatok típusgyakoriságát és előfordulási mintáit térképezi fel, hanem figyelembe veszi a sémák közötti paradigmaticus és jelentésbeli viszonyokat is (pl. *V-va/-ve volt – V-va/-ve lett*). A tanulmány új meglátásokat nyújt a passzív konstrukciók működéséről, miközben módszertanilag is hozzájárul a kognitív korpusznyelvészet fejlődéséhez azáltal, hogy részletesen feltárja az eddig kevésbé vizsgált variabilitást a magyar nyelv igei sémáiban.

Az eredmények több szempontból is izgalmasak. A kontrollként bevont *V-ni fog* szerkezet (amely a jövő idejű analitikus alak) elemzése azt mutatta, hogy a globális produktivitas – meglepő módon – nincs összefüggésben azzal, hogy egy szerkezet szintetikus vagy analitikus jellegű-e. Mind a négy vizsgált séma esetében más és más mutató mentén volt kimagasló a produktivitas: a *V-va/-ve van* szerkezet bizonyult a legnagyobb összproduktivitásúnak, a *V-ódik/-ődik* séma rendelkezett a legnagyobb típusgazdagsággal (vagyis ebben jelent meg a legtöbb különböző ige), míg a *V-va/-ve lesz* mutatta a legnagyobb potenciális produktivitást (tehát elvben sok új alak képezhető vele). A régebbi keletű *-(t)atik/-(t)etik* képzős forma ugyanakkor alacsony gyakoriságú, viszont szinte kizárólag passzív jelentésben fordul elő – e konstrukció tehát ritkasága ellenére a legstabilabb passzív séma. Hasonlóan erős passzív jelentésű a *V-va/-ve lesz*, míg a *V-va/-ve van* túlnyomórészt, de nem kizárólag passzív tartományú. A *V-ódik/-ődik* esetében derül ki a leginkább, hogy bár viszonylag gyakoribb, sok esetben nem tisztán passzív (hanem mediális vagy reflexív) jelentésben használatos. Ezek az eredmények együttesen arra utalnak, hogy a nyelvben a passzív kifejezésére szolgáló formák különböző funkcionális szerepeket töltenek be: van amelyik ritka, de egyértelmű jelentésű, míg mások gyakoribbak, de jelentésük tágabb. A tanulmány fontos tanulsága, hogy a passzív szerkezetek produktivitása és szemantikai specifikitása nem választható el egymástól, és a pusztán formai kategorizálás (szintetikus vs. analitikus) nem jelzi előre a használati gyakoriságot. A fejezet tematikusan kapcsolódik a szimultán igeikötözésről szóló korábbi tanulmányhoz (2.1.), hiszen mindkettő érinti a passzív *-ódik/-ődik* képzőt: míg a 2.1. fejezet a prefixummal együtt való előfordulására koncentrált, a jelen fejezet a képző produktivitását tágabb összefüggésben elemzi.

A szerzők a korpusz mérete miatt elsősorban gépi elemzést alkalmaztak, de bizonyos jelenségek – például a hapaxok jelentésének vagy a *V-va/-ve van* konstrukciók aspektuális sajátosságainak vizsgálata – kézi elemzést igényeltek. A manuális ellenőrzés során derült ki,

hogy több kizárt szerkezet (pl. *lebegtetve van*) jelentős részét képezi a séma használatának, és az igekötős alakok túlnyomó része hamis pozitív találat. Ez rávilágít arra, milyen fontos a gépi eredmények alapos kézi kontrollja.

Lektori hozzászólások: KUGLER NÓRA elismerően nyilatkozott a kutatás módszertani alaposságáról, legfőbb újtásként kiemelve, hogy a tanulmány kidolgozza a szintetikus (pl. *betömetik*) és analitikus (pl. *be lett tömve*) igealakok produktivitásának összevethető mérését, gépi és kézi elemzések kombinálásával. A lektor további érdemként említi meg, hogy a szerzők figyelembe vették a szórendi változatosság funkcionális jelentőségét, és az *emPreverb* eszközzel megoldották az igekötős igék egységes kezelését, emellett a metaadatok alapján nagy adatmennyiség időbeli rendezését is elvégezték. PALÁGYI LÁSZLÓ hangsúlyozta az adatvezérelt megközelítés előnyeit és a korpuszmunka értékét, kiemeli a tanulmány már említett legfontosabb eredményét: az analitikus és szintetikus szerkezetek globális produktivitási indexe összehasonlítható, így a grammatikai kontinuum különböző pontjain elhelyezkedő sémák termékenysége egységes mérési keretben vizsgálható. Továbbá hangsúlyozza, hogy a *V-va/-ve van* séma használati köre jóval tágabbnak bizonyult, mint amit a korábbi, főként introspektív elemzések sugalltak, produktivitása pedig megközelíti a kontrollként vizsgált *fog V-ni* konstrukciót. A lektor kritikai észrevételként javasolta a kvalitatív szempontok és értelmezések bővebb integrálását, annak érdekében, hogy a statisztikai eredmények értelmezési kerete még gazdagabb lehessen.

2.5. A tagmondatkapcsolattól a komplex birtokos szerkezetig. BAJZÁT TÍMEA BORBÁLA, SZEMES BOTOND és SZLÁVICH ESZTER tanulmánya az irodalmi nyelvhasználat világába vezeti az olvasót: különböző hasonlattípusok azonosítását és vizsgálatát végzik el nagy méretű irodalmi korpuszban. A kutatás hármas célt követ: egyrészt a hasonlatok számitógépes azonosítását, másrészt e hasonlattípusok kognitív szemantikai elemzését, harmadrészt pedig kvantitatív vizsgálatát a kanonikus magyar prózairodalom (regények) anyagában. A szerzők 100 kanonikus magyar regényt tartalmazó korpuszon dolgoztak, amely 173 évnyi időszakot ölel fel – ez impozáns történeti távlatot is ad a vizsgálatnak. A hasonlatok automatikus felismerése nem triviális feladat, mivel sokféle nyelvi alakban fordulhatnak elő. A kutatók egy többlépcsős, reguláris kifejezésekre és a szerkezetek komponenseinek morfológiai-szemantikai tulajdonságaira épülő módszert dolgoztak ki. A kutatás célja ötféle, eltérő komplexitású hasonlatkifejező konstrukció – köztük a prototipikus két tagmondatból álló mintázat és négy sűrített forma (mint a *-ként* esetragos, *-szerű* utótagos, *gyanánt* névutós és birtokos szerkezetbe ágyazott hasonlatok) – automatikus azonosítása, valamint ezek kognitív szemantikai jellemzése. A szerzők a funkcionális nyelvészeti kontinuum-elv alapján részletes jelentéshálózati leírást nyújtanak, és a konstrukciók felismerésére egy több lépcsőből álló, morfológiai és szemantikai jellemzőket is figyelembe vevő algoritmikus megközelítést dolgoznak ki. A 100 kanonikus magyar regényt tartalmazó ELTE DH Regénykorpusz felhasználásával végzett korpuszvizsgálat lehetőséget ad a különféle hasonlatok gyakoriságának, szerkezeti sajátosságainak és stílusbeli eloszlásának empirikus feltérképezésére.

A kutatás újdonsága abban rejlik, hogy nemcsak a nyelvi hasonlósághoz kapcsolódó formákat vizsgálja, hanem a kevésbé feltűnő, sűrített konstrukciók azonosítására is megoldást kínál, új utakat nyitva ezzel a digitális szövegelemzés és a stílustörténeti kutatások számára. A hasonlat mint kognitív művelet itt nem csupán nyelvi szerkezetként jelenik meg, hanem tudásképző mechanizmusként is: az entitások közti viszonyok révén

új jelentésszintűk tárulnak fel a befogadó előtt. A gépi elemzési módszerek – például token alapú keresés, reguláris kifejezések és kollokáció alapú stop-szólísták – nemcsak pontosítják a hasonlattípusok formai és gyakorisági jellemzőit, hanem lehetőséget kínálnak hosszabb távú, például történeti-pragmatikai vagy stilometriai irányú kutatásokhoz is. A módszertani újítás abban is áll, hogy a szerzők azonosítják a nehezen elkülöníthető hasonlatszerkezeteket is, melyekre külön keresési feltételeket alakítanak ki. A konstrukció alapú megközelítés – amely a nyelvi szerkezeteket jelentésükkel együtt vizsgálja – biztosítja, hogy az azonosítás nem csupán formai, hanem tartalmi alapon is releváns legyen.

A sikeres azonosítási eljárás lehetővé tette a hasonlatok kvantitatív elemzését a korpuszban. A szerzők megvizsgálták például, hogy a különböző típusú hasonlatok milyen gyakorisággal fordulnak elő a 19. és 20. századi magyar regényirodalomban, és hogyan változik használatuk az időben. A kognitív szemantikai elemzés révén kategorizálták a hasonlatokat aszerint, hogy milyen jelentésbeli kapcsolatot teremtenek (például metaforikus, metonimikus vagy más kognitív összefüggés alapján működő hasonlatok). Az így kapott eredményeket összevetették a hasonlító összetett mondatok hagyományos gyakorisági adataival, és pontosították azok értelmezését. Kiderült, hogy bizonyos sűrített hasonlatok (ahol a hasonlítás grammatikailag nem expliciten jelenik meg, hanem például jelzői formában) gyakoriságukat tekintve alul voltak eddig becsülve: az új módszerrel kimutathatóvá váltak és számszerűsíthetőek lettek ezek az esetek is. A fejezet egyik fontos tanulsága, hogy a korpusznyelvészeti és stilometriai módszerek ötvöztetésével az irodalmi szövegekben rejlő nyelvi mintázatok is feltárhatók, ezáltal a hagyományosan inkább kvalitatív irodalomtudományi kérdések kvantitatív megközelítéssel is vizsgálhatók. Érdekes megfigyelés, hogy a *mintha* kötőszó, amely a feltételes jelentésáramlással szubjektívizálja a hasonlítás műveletét, egyúttal diskurzusjelölőként is működhet, további kifejtést igényel azonban ennek mibenléte és a diskurzusjelölők kritériális jegyeinek (vö. pl. FURKÓ 2020) megfeleltethetősége a *mintha* egyes tokenjei esetében.

A kutatómódszertan alaposságát mutatja, hogy a szerzők táblázatos formában is összefoglalják a módszertani lépéseket. A szerzők kiemelik, hogy a szerkezet típusok nagy formai és funkcionális variabilitása miatt az azonosítás többlépcsős, kvalitatív elemeket is integráló folyamatot igényelt. Eredményeik szerint egyszerű módszerekkel is jelentősen csökkenthető a fals pozitív találatok aránya. A folytatásban gold standard korpusz létrehozását és részletesebb szemantikai vizsgálatot terveznek. E tanulmány és a következő fejezet (2.6.) együtt jól mutatja, miként használhatók a korszerű nyelvtechnológiai eszközök a szépirodalmi szövegek nyelvészeti elemzésében: az előbbi a hasonlatok, az utóbbi a lexikai választék kapcsán nyújt betekintést a kanonikus művek világába.

Lektori hozzászólások: HORVÁTH PÉTER kiemeli, hogy a tanulmány elméletileg és módszertanilag is megalapozott, és fontos hozzájárulást nyújt a hasonlatok konstrukciós leírásához és irodalmi funkcióik jobb megértéséhez. Hangsúlyozza, hogy a kutatás példászerűen mutatja be, hogy a formailag és funkcionálisan összetett nyelvi jelenségek is eredményesen azonosíthatók szabály alapú módszerekkel. A lektor szerint a szabály alapú megközelítés előnye, hogy pontosabb és nyelvészeti szempontból relevánsabb jellemzést tesz lehetővé, különösen reguláris kifejezések és stop-szó listák alkalmazásával, következésképpen érdemes megvizsgálni a szabály alapú azonosítás lehetőségét még a gépi tanuló módszerek alkalmazása előtt, mindemellett javasolja a stilisztikai elemzések mélyebb kifejtését, hogy a hasonlatok formai megoldásai és stílusi hatásai még

részletesebb értékelést nyerjenek. – SIMON GÁBOR kiemeli, hogy a tanulmány kognitív korpusznyelvészeti megközelítése korszerűen ötvözi a konstrukciós nyelvleírás eszközeit és szemléletmódját. Hangsúlyozza, hogy a szerzők figyelemre méltó kísérletet tesznek a magyar nyelvű hasonlatok formai sokféleségének és funkcionális változatainak feltérképezésére, különös tekintettel a hasonlítás mint jelentéskonstruálás skaláris értelmezésére. Ugyanakkor a lektor rámutat arra is, hogy – mint a figuratív nyelvhasználat kutatásaiban gyakran – a formai eltérések azonosítása jobban sikerül, mint a funkciók pontos elkülönítése. Elismerve a tanulmány módszertani alaposságát, kritikai megjegyzésként jelzi a lektor, hogy a szöveg nem mindig következetesen különíti el a korpusz alapú és korpuszvezérelt elemzési szemléletmódokat, holott ezek eltérő teoretikus és módszertani kiindulópontokat képviselnek. A lektor szerint a funkcióból kiinduló formaértelmezések a korpusz alapú megközelítéshez állnak közelebb, míg a nyelvi mintázatokból kiinduló, lekérdezéseken alapuló elemzési lépések inkább a korpuszvezérelt irányt képviselik. A kettő összehangolása ugyan gazdagíthatja a kutatást, ám a tanulmányban ezek együttes alkalmazása nem mindig reflektált módon történik. SIMON GÁBOR azt is hangsúlyozza, hogy bizonyos szerkezetek – például referenciális azonosságon alapuló kifejezések – nem feltétlenül hasonlatként, hanem inkább jellemzők előtérbe helyezéseként értelmezhetők. A lektor szerint az analógián alapuló hasonlítás azokban a birtokos szerkezetekben a legindokoltabb, ahol a birtokos típusként jelenik meg, és a hasonlítás referenciapont–célpont viszonya világosan rekonstruálható.

2.6. A lexikai választékosság és szemantikai specifikusság mintázatai. HORVÁTH PÉTER zárótanulmánya a magyar kanonikus költészet nyelvezetét vizsgálja korpusznyelvészeti eszközökkel: a lexikai választékosság (szókincsbeli gazdagság) és a szemantikai specifikusság mintázatait elemzi. A kutatás alapját 24 magyar klasszikus költő (Csokonai-tól Radnótiig) életművének verskorpusza adja, amelyet a Digitális Irodalmi Archívum adatbázisából merítettek. E versek szókészletének jellemzéséhez a szerző egy olyan mérőszámot alkalmaz, amelyet eredetileg idegennyelv-elsajátítási vizsgálatokban használnak a nyelvi választékosság kvantifikálására. Lényege, hogy a szövegben található szavak ritkaságát veszi alapul: minél ritkább egy szó egy referencia-korpuszban (jelen esetben a Magyar Nemzeti Korpusz irodalmi anyagában vagy egy külön verskorpuszban), annál nagyobb súlyt kap a választékosság indexében. A szerző két referencia-korpusz (egy verses és egy prózai korpusz) alapján minden szót besorolt négy gyakorisági kategóriába, majd minden egyes versre, versszakra és sorra kiszámította a lexikai választékosság pontszámát. Emellett feltételezte, hogy a ritkább szavak egyben specifikusabb jelentésűek, így a lexikai választékosság tulajdonképpen a szemantikai specifikusság mérőszámának is tekinthető. A vizsgálat több szinten folyt: nemcsak az egész életművek, hanem az egyes költemények belső szerkezeti egységei (pl. első vs. utolsó versszak) között is összehasonlításokat tett.

A tanulmány eredményei rámutatnak néhány meglepő összefüggésre a költői nyelvhasználatban. Az egyik fő kérdés az volt, hogy a vers hosszúsága és lexikai választékossága közt van-e kapcsolat. A korpusz elemzése azt mutatta, hogy a közepes hosszúságú versek (50–199 szavasak) lexikai választékossága általában alacsonyabb, mint a nagyon rövid (1–49 szavas) vagy a hosszú (200 szónál hosszabb) verseké. A magyarázat szerint a rövid versekben a fókuszált megfogalmazás, a hosszú versekben pedig a részletező leírás eredményez gyakoribb ritka, specifikus szavakat, míg a közepes terjedelmű művek nyelve átlagosabb. A másik vizsgált szempont a versforma komplexitása volt: kiderült, hogy

a formailag egyszerű versek (ahol a versszakok, sorok hossza és rímképlete egységes, azaz például izometrikus és izorímes a vers) lexikai választékossága általában alacsonyabb, mint a bonyolultabb, változatos formájú verseké. Ezt a szerző egyfajta ikonicitással magyarázza: az egyszerű forma egyszerűbb, kevésbé változatos szókincset tükröz. Ugyanakkor a versek belső szerkezetét tekintve is talált mintázatot: tipikusan a versek elején magasabb a választékosság mértéke, majd a mű előrehaladtával csökken. Az első versszakok gyakran tele vannak egyedi, konkrét képekkel és szavakkal, míg a záró szakaszok inkább általánosabb nyelvezetűek – ezt a szerző akként értelmezi, hogy a versek jelentése a specifikustól az általános felé halad. Ezek az eredmények újfajta, számszerű nézőpontból világítják meg a kanonikus költők stílusát, és alkalmat adnak az egyes szerzők közti különbségek további vizsgálatára is.

A fejezet módszertani jelentősége abban áll, hogy bebizonyítja: a stilisztikai sajátosságok is elemezhetők kvantitatív eszközökkel, a korpusznyelvészet így az irodalomtudomány szempontjaira is reflektálni tud. E tanulmány szorosan összecseng a megelőző fejezettel (2.5.), hiszen mindkettő a kanonikus irodalom nyelvi mintázatait vizsgálja nagy adathalmazokon.

Lektori hozzászólások: SZEMES BOTOND méltatta a kutatás innovativitását és azt, hogy a szerző hatékonyan alkalmazta a digitális szövegbányászati eszközöket az irodalmi szövegek elemzésére. A lektor kiemeli, hogy a tanulmány innovatív módon közelíti meg a lírai szövegek vizsgálatát: a szavak előfordulási valószínűségéből kiindulva kvantitatívan méri azok lexikai választékosságát, nem stilisztikai, hanem szemantikai komplexitásuk és specifikusságuk alapján. Hangsúlyozza, hogy ezáltal a versek és versrészletek az általánosság–specifikusság, illetve az egyszerűség–komplexitás tengelyén helyezhetők el, ami lehetővé teszi a költői szövegek mögötti általános mintázatok feltárását. A lektor a kutatás különösen értékes felismerésének tartja, hogy még a lírai, reflektív nyelvhasználatban is kimutathatók kognitív mintázatok, mint például az induktív logika dominanciája. A lektor ugyanakkor rámutat arra is, hogy a vizsgált versek korpuszában számos ellenpélda is megfigyelhető, amelyek nem illeszkednek az azonosított sémákhoz. Ez nem a módszer gyengesége, hanem éppen a lírai megszólalás sajátosságát tükrözi: azt, hogy a költői nyelv gyakran épp a szokásos kognitív és nyelvi sémák kibillentésére törekszik. A lektor szerint a tanulmány eredményei ennek a dinamikának – az alapsémák kidolgozásának és felülírásának – érzékeny nyelvészeti megragadását teszik lehetővé, és felvetik a szavak lexikai változatosságán alapuló, új típusú verselemzés további lehetőségeit is.

3. Általános összefoglalás. Összegzőképpen, a Korpusznyelvészeti módszerek a kognitív nyelvészeti kutatásban című kötet jelentős értéket képvisel a magyar nyelv korpusz alapú vizsgálatának terén. A szerzők alapos és ötletes elemzései nemcsak megerősítik a magyar nyelvtani szerkezetek és nyelvhasználati jelenségek kutatásában eddig elért eredményeket, hanem új perspektívákat is nyújtanak. A különböző fejezetek – bár látszólag változatos témákat ölelnek fel a grammatikai konstrukcióktól az irodalmi stílusig – szorosan kapcsolódnak egymáshoz, és módszertani szempontból is kiegyensúlyozott, egységes egészet alkotnak. A statisztikai modellezések, a kvalitatív elemzések és a korpusz alapú megközelítés kombinációja minden tanulmányban központi szerepet játszik. Ez az egységes szemlélet biztosítja, hogy a sokféle vizsgált jelenség egy koherens tudományos keretbe illeszkedjen: a nyelvhasználati adatok szisztematikus elemzése teremti meg a hidat a fejezetek között.

Különösen figyelemre méltó, hogy a kötet szerzői a magyar nyelv olyan aspektusait vizsgálják, amelyek a szintaktikai és morfológiai sajátosságok mellett a nyelvhasználati kontextusokat és a jelentésbeli tényezőket is fókuszba helyezik. Ennek köszönhetően a tanulmányok eredményei számos új meglátással gazdagítják a nyelvészeti diskurzust, és több esetben kiindulópontként szolgálhatnak a további kutatások számára. A lektori hozzászólásokban megfogalmazott észrevételek tovább növelik a kötet értékét: rámutatnak a vizsgálatok finomhangolási lehetőségeire, és arra ösztönzik a szerzőket, hogy jövőbeli kutatásaikat újabb aspektusok bevonásával mélyítsék el. Emellett a könyv nem csupán az elméleti nyelvészet számára szolgáltat releváns tanulságokat, hanem a nyelvtechnológia és az alkalmazott nyelvészet területén dolgozók is hasznosíthatják az itt bemutatott eredményeket és módszereket, többek között a korpuszokra épülő nyelvmodellek fejlesztésében vagy az oktatásban. Összességében a kötet tanulmányainak magas szakmai színvonala, az adatok körültekintő feldolgozása és a módszertani újítások együttesen teszik ezt a munkát meghatározóvá a magyar nyelv korpusz alapú kutatásának jelenkori palettáján.

Hivatkozott irodalom

- FURKÓ BÁLINT PÉTER 2020. *Discourse Markers and Beyond: Descriptive and Critical Perspectives on Discourse-Pragmatic Devices across Genres and Languages*. Springer International Publishing, Cham.
- GAETA, LIVIO 2007. On the double nature of productivity in inflectional morphology. *Morphology* 17/2: 181–205.
- TOLCSVAI NAGY GÁBOR 2009. A magyar segédige + igenév szerkezet szemantikája. *Magyar Nyelvőr* 133: 373–393.

FURKÓ PÉTER
Károli Gáspár Református Egyetem

Pusztay János, Egy nyelvcsalád tündöklése és ...?

Magyar Esszék, Nap Kiadó, Budapest, 2023. 483 lap

Előttünk egy vaskos könyv. A borítóján található térképen nyelvrokonaink elhelyezkedése látható (egy nagy hibával és több kis pontatlansággal). Az ismertetendő nagyszabású munka kérdéses címét írója a mű előszavában (5–6) kiegészíti. Könyvét a finnugor (uráli) nyelvcsalád közelmúltjának és főleg jelenének bemutatása során elsősorban e népek nyelvéről ad összképet, és a nyelvcsalád megmentését tekinti írása céljának.

Az alkotás szerkezetét jól visszaadják 11 fejezetének (plusz irodalomjegyzék) és ezen belül számos alfejezetének címei. Véleményem szerint a címek szerint rendezve áttekinthetjük a kitűnő felépítésű és szerkezetű munkát.

1. Fakuló nyelvi tarkaság. A világ nyelvi képe (7–34). E fejezet számos olyan kérdést vet fel, amelyek az embert, gyakran a szakembert is foglalkoztatják. Mitől nyelv egy idióma? A szerző az ismert skandináv és szláv nyelvi kölcsönös érthetőség,