

**Beférkőző igék kollostrukciós vizsgálata.  
Az igekötő és főnévi igenév közé ékelődő igék  
kolexémaelemzése az MNSZ2-UD korpuszban<sup>1</sup>**

**Kivonat**

Jelen tanulmány az igekötő és főnévi igenév közé való beférkőzés kollostrukciós vizsgálatát mutatja be. A kolexémaelemzés 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

---

<sup>1</sup> 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 kolecéma-elemzés

## 1. Beférkőzés és segédigeség

Jelen tanulmány a segédigeséggel igen gyakran asszociált konstrukció, az ige-kötő 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.<sup>2</sup> 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 kolecémaelemzést – a disztinktív és a kovariációs kolecémaelemzés 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

<sup>2</sup> 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 infinitívusos beférkőző szerkezetben a kevésbé variábilis tag a beférkőző ige. A tág 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 grammatikalizációja (vö. Kalivoda 2021; lásd az Imrényi 2013: 292 lapról származó 1a-b példákat) és a frazeológiai szókapcsolatok fokozatos lexikalizációja (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)<sup>3</sup>

A keretet az általunk vizsgált esetben tehát egy infinitívusos 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 infinitívusos keretek megállapításánál csak az igekötőként annotált első tagokat vettük figyelembe, de a jelenség lehetséges tágabb értelmezésére reflektálunk az 5. pontban is.<sup>4</sup>

<sup>3</sup> 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.

<sup>4</sup> 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 sokféleknek 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.<sup>5</sup> 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.

<sup>5</sup> 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 ígére. 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 segédigeség 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).<sup>6</sup> A beférkőzés tulajdonsága ez alapján tekinthető egyfajta grammatikalizációs markernek, amely a klitizáció speciális eseteként a kötöttség, vagyis szintagmatikai kohézió (boundedness, Fügungsenge, vö. Diewald 1999, Dér 2008) ikonikus<sup>7</sup> 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

<sup>6</sup> 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.

<sup>7</sup> 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 segédigéség 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 infinitívusban (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 segédigéség 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. Érdeemes közelebről megnézni, hogy milyen forrásokból származnak a kritériumok, felállításukkor található-e elméleti alapot.

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 kieleézve 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*<sup>8</sup> 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

<sup>8</sup> 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 információszerkezet befolyásolja az adott setting (alap). A nagyméretű korpuszok a teljes nyelvhasználatot kívánják reprezentálni úgy, ahogy az a nyelvhasználók elméjében is érvényesítheti gyakorisági hatásait.<sup>9</sup> 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 segédigeiként 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 kolexémaelemzést. A módszerről általában lásd Stefanowitsch-Gries (2003), Gries (2015), Stefanowitsch (2013),

<sup>9</sup> 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, 'igekötő<sup>10</sup>+ 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

<sup>10</sup> 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 nyelvtechnológiai 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. kontingenciatábla (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 gyakoriságával<sup>11</sup> (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 pusztán 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 pusztán 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

<sup>11</sup> 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 előfordulású adatokra, ezeknek kiemelt erősséget tulajdonít (például a  $\chi^2$ -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 kolecémák 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 erősségéért 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.<sup>12</sup> 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 adattisztító lépéseket iktattunk be – gépi és kézi módszereket egyaránt –, hogy

<sup>12</sup> A kollostruációs 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 kolecémák 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 legújabban 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.<sup>13</sup>

### 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 méretegologozott 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).<sup>14</sup> 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.

<sup>13</sup> 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.

<sup>14</sup> 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 áll, 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*, *telí* → *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 adathalható 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 *szere*t lemmához tartozó találatokat két részre bontottuk: *szere*t illetve *szere*tne lemmák találataira. Ezt a döntést az indokolta, hogy mind az intuíciónk, mind a szakirodalomból ismert korábbi megfigyelések alapján meglehetősen különböző eloszlási mintákra számítottunk a *szere*t, illetve a *szere*tne 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 amellet, 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),<sup>15</sup> valamint az

<sup>15</sup> 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)<sup>16</sup>, és egyesítettük azokat a lemmákat, melyek gyakori gépelési hiba (például *akad>akar*; *kell>lett*, *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 kolxémákat 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 kolxéma 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,<sup>17</sup> „ns” jelöli a nem szignifikáns eredményt).

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

<sup>17</sup> \*\*\*\* = p<0,00001, \*\*\* = p<0,0001, \*\* = p<0,001, \* = p<0,01, ns = 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	****



Kolexéma	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 kolexémaként, 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 *kell*-ről pedig nem állapítható meg, hogy szignifikáns kolexémája a beférkőzésnek, noha az adatok az ige asszociáltságát mutatják. A *van* lexéma egyaránt tartalmazza a *lehet*, *van* és *volna* töveket is.<sup>18</sup>

<sup>18</sup> Az adatelemzés utolsó fázisában észleltük, hogy ebben az esetben is érdemes lett volna újraellematizálni 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 *kell* *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 kolecémákat 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 kolecémaelemzés lényegi vonása, hogy nem az abszolút gyakoriságot veszi figyelembe annak megítélésakor, 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 kolecxémavizsgálata kizárólag a következő igéket találta szignifikánsan beférkőzőnek: *kell*, *tud*, *akar*, *fog*, *mer* és *talál*, legutóbbi csak alacsonyabb szignifikanciaküszöbvel. A legerősebben vonzott kolecxémák 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 ige<sup>19</sup> 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 kolecxémák alacsony száma és a kisszámú vonzott kolecxéma 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

<sup>19</sup> 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 infinitívussal 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 nullhipotézie a lexémák random megjelenése a konstrukciókban. Azonban egyrészt a legtöbb ige nem rendelkezik beférkőző használattal, másrészt a beférkőzőként is használható igék asszociálnak 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 szignifikanciaszűrés 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 kontrollcsoporthoz 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 tolneni* 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.<sup>20</sup>
- (8) El kellene döntened, hogy *engedsz letölteni* fájlokat vagy sem. Mert egyszerre engedni, meg nem engedni, az nem megy.<sup>21</sup>

### 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 igtől 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 igeekötős főnévi igenév esetében sem.

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

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

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

- (10) kérdezze meg az urakat, *mikor óhajtták 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 használatlalt 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 igékkal: *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 vizsgálattal 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 szignifikanciasztek 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 asszociálnak 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 mondatsztintú 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](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](http://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](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. *Igekötős 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 igekötőzés 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 – Prószték Gábor 2023. Infinitívuszi szerkezetek a magyarban: adatvezérelt megközelítés. *Nyelvtudományi Közlemények* 119: 223–245.
- Kalivoda, Ágnes – Prószték, 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 – Nádasy Ádám – Prószték 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 magmondattal 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. Igekötő-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 kérdése 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