DOI: 10.15774/PPKE.BTK.2025.020

Dömötör Andrea

Hány VAN nincs?
A zéró létigék jellemzői és számítógépes elemzési lehetőségei

Doktori (PhD) értekezés

Pázmány Péter Katolikus Egyetem

Bölcsészet- és Társadalomtudományi Kar

Nyelvtudományi Doktori Iskola

Vezetője: **Dr. Surányi Balázs**

egyetemi tanár, az MTA doktora

Nyelvtechnológia Műhely

Témavezető

Dr. Prószéky Gábor

egyetemi tanár, az MTA doktora

Budapest, 2025

DOI: 10.15774/PPKE.BTK.2025.020

"Minden forrásom belőled fakad"

Zsolt 87,7

DOI: 10.15774/PPKE.BTK.2025.020

Tartalomjegyzék

Köszönetnyilvánítás **1**

1. Bevezetés **2**

1.1. Terminológia 2

1.2. Hány VAN van? 4

1.3. Van semmi, vagy nincs semmi? 10

1.3.1. Zéró elemek az elméletben 10

1.3.2. Az elhagyhatóság szempontjai 12

1.4. A dolgozat felépítése 14

2. Módszertan **16**

2.1. Korpuszalapú és korpuszvezérelt módszerek 16

2.2. Számítógépes nyelvfeldolgozás 17

2.3. Függőségi elemzés 18

2.3.1. Zéró elemek a korpuszokban és a számítógépes elemzésben 20

2.4. Felhasznált korpuszok 21

2.5. Felhasznált nyelvfeldolgozó eszközök és nyelvmodellek 23

3. Zéró kopulák a számítógépes elemzésben **26**

3.1. Kutatási cél és módszer 28

3.2. Szabályalapú bináris osztályozás 31

3.2.1. Az osztályozási feladat 31

3.2.2. Címkézés párhuzamos korpuszon 32

3.3. A zéró kopulák beszúrása 37

3.3.1. Tanító adat 37

3.3.2. Eredmények 39

3.3.3. Hibaelemzés 42

3.3.4. Összegzés 44

DOI: 10.15774/PPKE.BTK.2025.020

4. Zéró létigék a korpuszokban **45**

4.1. Ingadozó bővítmények 49

4.2. Ott van, tehát van 55

4.2.1. Az *ott* mint igekötő a szakirodalomban 58

4.2.2. Korpuszalapú vizsgálatok 61

4.2.3. Következtetések 65

4.3. *Nincs*, ami nincs 67

4.4. Zéró *lehet* 69

4.5. Összefoglalás 76

5. A kopulás mondat szerkezete **78**

5.1. A testes és zéró kopulás mondatok szerkezete 78

5.1.1. Nem alapértelmezett esetet jelző tagmondatok 78

5.1.2. Az alapértelmezett esetet kizáró lexikális elemek 79

5.1.3. Hiányzó névszói állítmány 80

5.1.4. Igekötős igéből képzett névszói állítmány 81

5.1.5. Tagadó mondatok szórendje 81

5.2. Fókusz vagy igemódosító? 83

5.2.1. Távolságok és szórendek a Szeged Treebankben 84

5.2.2. Elméleti keret, hipotézisek 90

5.2.3. Kérdőíves vizsgálat 93

5.2.4. Korpuszalapú vizsgálat 95

5.2.5. Következtetések, összegzés 99

6. A predikatív névszó (segéd)igei tulajdonságai **100**

6.1. Igére jellemző szórend 100

6.1.1. Igéből képzett névszók és igekötő-leválás 100

6.1.2. Az *-e* kérdő partikula helye 105

6.2. A predikatív névszók vonzatai 108

6.2.1. Infinitívuszi vonzatok 110

6.2.2. Az igévé válás útján: *szabad* 112

DOI: 10.15774/PPKE.BTK.2025.020

6.3. Összefoglalás 115

7. Összegzés **117**

Hivatkozások **120**

DOI: 10.15774/PPKE.BTK.2025.020

Köszönetnyilvánítás

Mindenek előtt köszönöm mindenkinek a türelmet.

Köszönöm a témavezetőmnek, Prószéky Gábornak, hogy rám bízta ezt a témát, és hogy a kezdetektől figyelemmel kísért és biztatott. Köszönöm, hogy a kutatásomhoz biztos hátteret adott, és része lehettem a "százpí" szobának.

Nagy szeretettel és hálával gondolok vissza a pázmányos éveimre: a támogató, inspiráló és otthonos közegre, amit a BTK és az ITK campusain megtapasztaltam, a "digibölcsi" csoport összetartására, tanáraim kedves közvetlenségére. Külön köszönet illeti Ligeti-Nagy Noémit és Tornai Kálmánt, akiktől programozni tanultam és É. Kiss Katalint, aki megmutatta, hogyan legyek kutató.

Köszönöm az opponenseim, É. Kiss Katalin és Héja Enikő alapos munkáját, valamint a munkahelyi vita bizottságának, Surányi Balázsnak, Sass Bálintnak és Halm Tamásnak a hasznos hozzászólásokat. Észrevételeik és javaslataik sokat segítettek a disszertáció végleges formájának kidolgozásában.

Hálás vagyok a csoporttársaimért, az "Ágikért", amiért együtt élhettük át a doktori évek összes örömét, bánatát, bosszúságát és izgalmát.

Köszönöm a volt és jelenlegi munkatársaimnak a sok támogatást, együttgondolkodást és közös munkát. Külön köszönöm Yang Zijian Győzőnek és Novák Attilának a 3. fejezetben való közreműködésüket.

Köszönöm a családomnak, hogy mellettem álltak, elsősorban Balázsnak, hogy a hosszas időhúzásaim alatt sem lankadt a belém vetett bizalma, és kitartó "szurkolóm" volt.

Végül, de elsősorban hálát adok a mennyei Atyának, akiben mindenre van erőm, és akitől mindent kaptam. A disszertáció első valamelyest épkézláb változata a Budapesten rendezett 52. Nemzetközi Eucharisztikus Kongresszus idején készült el, ezért is választottam ennek az eseménynek a mottóját a dolgozat mottójául.

DOI: 10.15774/PPKE.BTK.2025.020

1. Bevezetés

A dolgozat a magyar létigés mondatok azon eseteit vizsgálja, ahol a morfológiailag jelöletlen esetben (azaz kijelentő mód, jelen idő, egyes szám 3. személyben) a létige nem, vagy nem kötelezően jelenik meg a mondatban. Ide tartozik az úgynevezett zéró kopula jelensége, illetve azok a létigét tartalmazó nem nominális mondatok, ahol a *van* igei elem elhagyható (pl. *Kulcs a lábtörlő alatt.*). Előbbi már jól ismert és leírt jelenség, utóbbiról azonban még viszonylag kevés említést találunk a szakirodalomban.

A dolgozat elsősorban számítógépes és adatvezérelt szemléletű. A fő elméleti kérdése az, hogy mely esetekben opcionális a létige a nem nominális mondatokban, és erre leginkább empirikus úton, a korpuszadatok elemzésével keres választ. A "klasszikus" zéró kopula kérdéskörében a számítógépes nyelvfeldolgozás problémáira helyezi a hangsúlyt, az ige nélküli mondatok ugyanis komoly kihívást jelentenek a gépi szövegelemzés számára.

Kutatásaimnak tehát kettős célja van. Az elméleti nyelvészet szempontjából a létigék típusainak és az elhagyhatóságuk szempontjainak jobb megértését tűztem ki célul. Másrészt szem előtt tartom a nyelvtechnológia igényeit is: munkám során a számítógépes elemzést segítő algoritmusok kidolgozására és a gépi tanuláshoz használható feladatspecifikus korpuszok létrehozására is figyelmet szenteltem.

Bevezetésként fontos kitérni két általános, a szakirodalomban sokat tárgyalt kérdésre: a létige (*van* szótő) poliszémiájára és a "zéró vagy semmi" kérdésére. A létigék kategorizálása azért lényeges kérdés, mert a különböző bővítményekkel rendelkező létigés mondattípusokra más szabályok érvényesek a létige elhagyhatóságának tekintetében. Másrészről az is vita tárgyát képezheti, hogy a vizsgált mondatokban valóban a létige elhagyásáról van-e szó, vagy ezekben is feltételezünk egy zéró hangalakú igei elemet.

1.1. Terminológia

A létigékkel és kopulákkal kapcsolatos terminológia nem egységes a szakirodalomban, ezért az egyes elméletek ismertetése előtt ebben a fejezetben tisztázom a dolgozatban használt fogalmak jelentését.

DOI: 10.15774/PPKE.BTK.2025.020

- **kopula**

  A magyar hagyományt követve a kopula megnevezés a dolgozatban kizárólag a nominális mondatok igei elemére vonatkozik. Ide tartoznak a *van* és a *lesz* szótő alakjai, előbbi esetén kivétel a kijelentő mód, jelen idő, egyes szám 3. személyű (*van*) alak. A kopula legfőbb tulajdonsága, amely megkülönbözteti ezen szótövek más típusba tartozó előfordulásaitól éppen a *van* alak hiánya. Az angolszász szakirodalmi hagyománytól eltérően a lokatív és egyéb határozós mondatok *van/lesz* igéjét nem tekintem kopulának.

- **zéró kopula**

  A zéró kopula létéről és szükségességéről megoszlanak a vélemények a szakirodalomban (l. 1.3. fejezet). A számítógépes elemzés számára azonban mindenképpen hasznosak a zéró elemek, így a dolgozat ilyen motiváltságú részeiben használni fogom a klasszikus leíró nyelvészetben szokásos zéró kopula kifejezést. Ebben az esetben a nominális mondatokban egy zéró kopula elemet feltételezünk ott, ahol nem alapértelmezett esetben a testes kopula lenne.

- **alapértelmezett eset**

  Alapértelmezett esetnek nevezzük az igei paradigma azon tagját, amely nem visel testes toldalékokat. A magyarban ez a kijelentő mód, jelen idő, egyes szám 3. személy. A kopula paradigmájából (a *van* tő esetén) az alapértelmezett eset hiányzik.

- **zéró kopulás és nominális mondat**

  Nominális mondat alatt azokat a mondatokat értem, amelyekben a logikai állítmány szerepét ragtalan (predikatív) névszó vagy névszói csoport tölti be. Ezen belül zéró kopulásnak nevezzük azokat, amelyekben nincs testes igei elem (kopula).

- **létige**

  Létigének nevezem a *van* és a *lesz* szótő mindazon előfordulásait, amelyek nem kopulák, illetve nem a *lehet* modális segédige előfordulásai. Ezen kívül szintén a létige kategóriába sorolom még a *nincs,sincs* tagadó alakokat is. A létigék osztályán belül Clark (1978) nyomán további alcsoportokról is fogok beszélni. A létigét lokatívnak tekintem, ha a mondat egy specifikus alanyról állítja, hogy az valahol van (*A növények az erkélyen vannak.*). Az egzisztenciális mondat ezzel szemben valaminek a meglétét vagy hiányát állítja, ebből kifolyólag

DOI: 10.15774/PPKE.BTK.2025.020

az alanya nem lehet specifikus. A lokatív és egzisztenciális létigék megkülönböztetésében a helyre utaló elem megléte vagy hiánya nem releváns, a fő szempont az alany specifikussága. (Egzisztenciális például: *Vannak növények (az erkélyen).*) Végül, a birtokos létige azon mondatok igéjét jelenti, ahol egy nem specifikus alanyról állítjuk, hogy valakinek vagy valaminek a birtokában van, létezik (*Balázsnak vannak növényei (az erkélyen).*). Az előbbi, Clark (1978) által definiált kategóriák egyikébe sem sorolható létigei alakokra a MGr. nyomán "egyéb határozós létige" néven fogok hivatkozni, ilyenek például az időre (*10-kor lesz a megbeszélés.*) vagy okra (*Ez azért van, mert nem figyeltél.*) vonatkozó létigés mondatok.

- **specifikus** (NP)

  Az egzisztenciális létige egyik fő jellemzője a specifikussági megszorítás, miszerint a létezést jelentő ige alanya nem lehet specifikus. A határozott főnévi kifejezések (DP) mellett specifikusnak nevezzük azt a határozatlan főnévi kifejezést, amely úgy azonosít egy meghatározatlan individuumot, hogy besorolja egy, a beszélő és a hallgató számára is ismert, meghatározott halmazba. Például: *A beteggel egy egész orvoscsapat foglalkozott.* ***Egy orvos*** *megmérte a vérnyomását.* (É. Kiss 2003: 73.)

- **zéró létige**

  Zéró létige alatt azokat az eseteket értem, amikor a mondatból hiányzik a létige. A megnevezés nem jelent elméleti állásfoglalást a zéró igei elemek mellett, csupán a létige hiányára utal.

- **funkció, funkcionális**

  Funkció alatt azokat a nyelvtani kategóriákat (idő, mód, szám, személy) értem, amelyeket a kopula mint segédige hordoz (Lengyel, 2000b). A kopuláról mint funkcióelemről akkor beszélhetünk, ha ezeken a funkciókon kívül nem bír más jelentéssel. A *funkcionális* jelző tehát lényegében a *lexikális* ellentéte.

1.2. Hány VAN van?

A *van* szótőt a magyar nyelvészeti szakirodalom gyakran poliszémnek feltételezi. Az ÉrtSz. 9 fő jelentéscsoportot, azon belül pedig 65 jelentést különböztet meg. Az egyes jelentések megkülönböztetésének létjogosultságát igazolja az, hogy más nyelvekben külön lexémákat használnak a

DOI: 10.15774/PPKE.BTK.2025.020

VAN különböző jelentéseire. Részletes tipológiába itt nem bocsátkoznék, csak néhány széles körben ismert példát említek. Ilyen például a birtoklásige, amely számos nyelvben (pl. angol, német és általában az indoeurópai nyelvekben) eltér a létigétől és a kopulától. Az is gyakori jelenség, hogy a létezést a nominális mondat segédigéjétől eltérő szerkezettel (pl. angol *there is*, német *es gibt*) vagy igével (spanyol *hay*) fejezik ki a nyelvek, és ezeknek jellemzően hiányos a paradigmája: csak 3. személyű alakjuk van. A spanyol ezen kívül különböző kopulát használ az úgynevezett individuumszintű (individual-level) és helyzetképszintű (stage-level) predikátumok (Kratzer (1995), Mejías-Bikandi (1993)) esetén. A *ser* az állandó tulajdonságokhoz kapcsolódik (1a), míg az *estar* az állapotot jelölő névszói állítmányokkal (1c) és a lokatív mondatokban (1b) fordul elő.

(1) a. *Ana María **es** profesora de baile.*

’Annamária tánctanár.’

b. *Mi hermana **está** en casa.*

’A húgom otthon van.’

c. *Tomás **está** cansado.*

’Tamás fáradt.’

A *van* szótövet poliszémnek feltételező magyar nyelvleírások is hasonló jelentéskategóriákat különítenek el, mint amelyekre más nyelvekben külön lexémákat találunk. A magyarban ezek a jelentéskategóriák leginkább szintaktikai szempontok alapján írhatók le: a létige bővítményeinek morfológiai tulajdonságaival, az alanyuk specifikusságával, az elhagyhatósággal és a szórenddel.

A Magyar grammatika (Lengyel (2000a), továbbiakban MGr.) szerint a létige lehet tartalmas ige és kopula. Előbbit jelentései és bővítményei szerint 5 altípusra osztja fel: létezést jelentő, birtokos, partitívuszi jelentésű, határozói bővítménnyel társuló és "sajátos többtagú szerkezettípus" részét képező létigére. A "csupán létezést jelentő" VAN-t (2) tartja a legteljesebbnek tartalmasság szempontjából. A típus formai jellemzője, hogy az alanyon kívül semmilyen más bővítményt nem kíván meg. (2b)-t és (2c)-t a típus "sajátos előfordulásainak" tartja, melyek természeti vagy környezeti állapotokat fejeznek ki.

DOI: 10.15774/PPKE.BTK.2025.020

(2) a. Van igazság (a földön).

b. Meleg van.

c. Hétfő van. / Csend van.

A birtoklásigét a MGr. az előzőektől külön kategóriába sorolja a kötelező részeshatározói bővítménye miatt, a jelentés tekintetében azonban megjegyzi, hogy ez a létigetípus a birtoklás mellett létezést is kifejez. A partitívuszi jelentésű létige (3) szintén csak az argumentumszerkezetében különbözik az egyszerű létezést jelentőtől. A negyedik csoportba sorolt, "kötelezően határozói bővítménnyel társuló" létigéknek (4) Lengyel (2000a) szerint gyengül a tartalmas jelentése, azaz ezek már inkább a kopulához hasonlítanak. "Sajátos többtagú szerkezettípusnak" nevezi a Magyar grammatika azokat a szerkezeteket, ahol az alaptag létezést kifejező létige, alanyi bővítménye egy főnévi igenév, amely kötelezően kiegészül egy tárgyi vagy határozói bővítménnyel, amit egy kérdő formájú névmás tölt be. (5)

(3) Van még szőlő. (= Van még a szőlőből.)

(4) Az edzés ötkor van. / A pénz aranyból van. stb.

(5) Van mit olvasni. / Van hol aludnod?

A MGr. felosztása tehát alapvetően a létige bővítményeire alapoz, a kopula és létige közti különbséget pedig leginkább a "lexikális tartalmasság" mentén igyekszik megragadni, ennek meghatározásához azonban nem ad egzakt, mérhető támpontokat, az egyes létigetípusok tartalmasságának megítélése erősen szubjektív marad.

A Magyar Grammatika jelentéscsoportjait gondolja újra Kádár (2004). Szabolcsi (1992)-re alapozva a birtoklást kifejező létigét is az egzisztenciális jelentésűek közé sorolja, mivel az egyszerű létigéhez hasonlóan a birtoklásige is egyargumentumú, és erre a [+birt] jeggyel ellátott alanyi argumentumra ugyanazok a határozottsági megszorítások vonatkoznak, mint az egzisztenciális létigére. (6a és b) Szabolcsi (1992) példái.

DOI: 10.15774/PPKE.BTK.2025.020

(6) a. Van kalap. / Péternek van kalapja.

b. *Van a kalap. / *Péternek van a kalapja.

Kádár (2004) a partitívuszi jelentésű létige külön osztályba sorolását is elveti, a (3)-típusú mondatok VAN-ját ugyanolyan egyargumentumú létigének tartja, mint az egzisztenciális mondatokét. Szintén az egzisztenciális mondatok közé sorolja az (5)-ben látott "sajátos többtagú szerkezeteket" is, ezek esetében a VAN egyetlen argumentuma egy NP értékű kérdő kifejezés. A Magyar Grammatika határozós létige csoportjával kapcsolatban Kádár (2004) megjegyzi, hogy a lokatív bővítmény másodlagos állítmányi szerepet is betölthet, és bizonyos korlátozásokkal a VAN elhagyható, hasonlóan a nominális mondatok kopuláihoz. A Kádár (2004) által csak megemlített jelenséget a 4.2. fejezet részletesen tárgyalja majd.

Kádár külön tanulmányban (Kádár, 2011a) foglalkozik a *Meleg van.* (2b-c) típusú mondatokkal. Ebben az időjárásra és egyéb környezeti tényezőkre vonatkozó NP+van szerkezeteket környezeti kopulatív szerkezeteknek (environmental copula construction) nevezi, ugyanakkor amellett érvel, hogy a mondattípus NP tagja – predikatív tartalma ellenére – a szintaktikai alany szerepét tölti be. A névszói predikátumként való elemzés ellen többek között a jelen idejű, egyes szám 3. személyű kopula (*van*) meglétét hozza fel érvként.

Kádár (2004) és Kádár (2011a) tehát szintén a bővítményekre helyezi a hangsúlyt a létige típusainak leírásakor, de azok morfológiai tulajdonságai (szófaj, esetragok) mellett az alany határozottságát is szempontként említi. Hasonlóképpen az Új Magyar Nyelvtan (É. Kiss et al., 2003) is az alany határozatlanságát emeli ki a létige fő ismérveként. Szerinte a valóban egzisztenciális jelentésű létige (és általában a levést, létrejövést jelentő, igekötő nélküli igék) fő ismérve, hogy alanya nem-specifikus határozatlan NP kell, hogy legyen, hiszen már ismert dolognak nem állíthatjuk a létezését. Ezeket csak úgy tudjuk egyéníteni, hogy egy határozatlan NP-vel kategóriába soroljuk őket:

(7) a. Volt egyszer egy király.

b. *Volt egyszer a király.

Ha azonban a létige igekötős, akkor már lehet határozott NP az alanya, hiszen ekkor már a létrejövés bevégzettségét állítjuk:

DOI: 10.15774/PPKE.BTK.2025.020

(8) Megvan a király.

((7) és (8) forrása: É. Kiss 2003: 74-75.)

Az összes eddig említett leírásban közös, hogy az alany az egyetlen argumentum, amelyik minden létigetípusnál kötelező. Az, hogy a létige létezést jelent-e, vagy valamilyen (grammatikai vagy logikai) viszonyt jelöl, jelentős részben függ az alany határozottságától, vagy pontosabban specifikáltságától. A határozottság és határozatlanság meghatározása formálisan jól leírható. Határozatlan NP-nek tekintjük a névelőtlen, határozatlan névelős, illetve a határozott vagy határozatlan számneves főneveket vagy főnévi kifejezéseket. Határozott NP-k (más néven DP-k) a tulajdonnevek és a determinánst tartalmazó főneves kifejezések. Determinánsok lehetnek kvantorok (*bármelyik, mindegyik, valamelyik, melyik*), mutató névmások (*e, eme, ezen*) vagy a határozott névelő (*a/az*). Határozott NP-nek számít a birtokos szerkezet is (*János könyve*), hiszen aminek a birtokosát ismerjük, annak magának is ismertnek kell lennie. (É. Kiss 2003: 78-84.)

Az egzisztenciális (Kálmán (2001) még pontosabb megnevezése szerint individuumegzisztenciális) mondat sajátossága, hogy új diskurzusszereplőt vezet be, így ebben a jelentésben a létige nem használható határozott (vagyis már ismert individuumot jelölő) névszói csoporttal. Ha a létige alanya határozott, akkor azt valamilyen igevivőnek (igekötő vagy vonzat) vagy a kommenten belül kontrasztba állított elemnek kell megelőznie.

(9) a. Van egy könyv az asztalon.

b. *Van a könyv az asztalon.

c. Ott van a könyv az asztalon.

d. A könyv van az asztalon (nem a szótár).

(Kálmán 2001: 18.)

Ugyanakkor az egzisztenciális jelentés elvesztéséhez nem feltétlenül szükséges, hogy az alany határozott legyen, elég ha specifikus, azaz olyan individuumot jelöl, amelynek létezését a beszélő előfeltételezi. A következő példák azt mutatják, hogy az *egy macska* határozatlan NP specifikus és nem-specifikus témavonzat szerepét is betöltheti:

DOI: 10.15774/PPKE.BTK.2025.020

(10) a. Van egy macska az udvaron. - Nincs egy macska se az udvaron.

b. Egy macska az udvaron van. - Egy macska nincs az udvaron.

Mint a tagadásból is látszik, (10b)-ben a határozatlan névelő ellenére konkrét macskáról van szó, akinek léteznie kell, hogy a mondat igaz lehessen, és a tagadó változat nem jelenti, hogy nincs macska az udvaron. (Kálmán 2001: 58.)

A fenti jelenséget, miszerint bizonyos igék megkövetelik, hogy valamelyik argumentumuk határozatlan, vagy legalábbis specifikálatlan legyen, először Milsark (1977) írta le Definiteness Effect néven. Nála a határozatlansági megszorítás a létezést jelentő *there is, there are* szerkezetre vonatkozott. Szabolcsi (1986) kiterjesztette az effektust minden olyan igére, amelynek van létezést, létrejövést vagy létrehozást kifejező jelentéskomponense. Emellett a definiteness megszorítást specificity megszorításra pontosította, azaz azt mondta ki, hogy az említett igéknek nem lehet specifikus alanya. Mindebből az következik, hogy az egzisztenciális és a lokatív (illetve egyéb határozós) létigéss mondatok megkülönböztetésénél elsősorban az alany specifikusságát kell vizsgálni.

Összegezve tehát a magyar nyelvészeti szakirodalom a *van* lexéma többértelműségét leginkább az argumentumszerkezetek különbségeiből vezeti le. Az elhagyhatóság kérdésének szempontjából fontos kiemelni az egyes nyevleírásokban explicit (pl. MGr.) vagy implicit (pl. É. Kiss et al. (2003)) módon megjelenő megkülönböztetést az úgynevezett tartalmas létigé(k) és a kopula között. Előbbi jellemzője, hogy "valóban létezést állít", és emiatt nem lehet specifikus alanya, utóbbinak viszont pusztán grammatikai szerepe van (személy, mód és idő kifejezése), "tartalmas jelentése" nincs.

Ezzel összhangban Heggie (1988) a kopulát igei operátorként (verbal operator), azaz nem lexikális, hanem a szemantikai kiüresedés magas fokát mutató funkcionális elemként írja le. Mivel a mondatok kanonikus predikátuma a VP, így a kopula olyan funkcionális elem, amely verbális predikátumot hoz létre bármely más, nem igei predikátumból. Erre azért van szükség, mert a predikatív NP-nek nincs esetosztási képessége, így az alany csak a kopula által hordozott igei inflexióból tud esethez jutni. A kopula más tekintetben is mutat igei tulajdonságokat (a magyarban is): idő- és egyeztetőmorfémákat hordoz, vannak non-finit alakjai (*lenni, lévén*), és a tagadó mondatokban az igékhez hasonló szórendet mutat (erről bővebben l. 5.2. fejezet).

A magyar leíró nyelvészeti hagyomány abban tér el Heggie (1988), és általában az angolszász

DOI: 10.15774/PPKE.BTK.2025.020

leíró nyelvészet definíciójától, hogy a kopula fogalmát szűkebben értelmezi: csak a nominális mondatok segédigéjét sorolja ide, a lokatív mondatok igéjét már nem. Mivel a kopula–lexikális ige megkülönböztetésnek lesz még szerepe a dolgozatban, ezért fontos tisztázni, hogy a kopula fogalma alatt a magyar hagyomány szerinti definíciót értem, azaz a lokatív igéket már a lexikális igék közé sorolom. Ez az elkülönítés leginkább az elhagyhatósággal, és ezzel összefüggésben a paradigmával indokolható: a *van* szótő azon előfordulásait tekintem kopulának, amelyeknek nincs kijelentő módú, jelen idejű, 3. személyű alakja (*van, vannak*).

Mivel a dolgozatban a létigetípusok megkülönböztetése elsősorban az elhagyhatóság kritériumainak meghatározásában játszik szerepet, meg kell még említeni Clark (1978) tipológiai munkáját, melyben azt vizsgálta, mennyire gyakori a zérós szerkesztésmód a világ nyelveiben a különböző létigetípusok esetén. Az általa felállított univerzális hierarchia a következő (példák tőlem):

1. predikatív (*Eszti okos.*)
2. azonosító (*Eszti az első unoka.*)
3. birtoklásmondat2 – be-possessive (*Ez a cipő az enyém.*)
4. lokatív (*A macska a hűtő mögött van.*)
5. egzisztenciális (*Van elég liszt.*)
6. birtoklásmondat1 – have-possessive (*Esztinek rengeteg játéka van.*)

Az első három kategória az általam (és a magyar leíró nyelvészeti hagyomány által) használt kopula fogalmába esik. A többi létigét tekintve a dolgozat további részében, elsősorban Clark (1978) nyomán egizsztenciális, birtokos és lokatív létigéket fogok megkülönböztetni, ezek ugyanis, mint látni fogjuk, a magyarban is eltérő tulajdonságokat mutatnak az elhagyhatóság tekintetében.

1.3. Van semmi, vagy nincs semmi?

1.3.1. Zéró elemek az elméletben

Az úgynevezett zéró kopula jelenségének értelmezése szintén nem egységes a szakirodalomban: bizonyos leírások a jelen idő 3. személyű nominális mondatokban feltételeznek egy "testetlen",

DOI: 10.15774/PPKE.BTK.2025.020

"fonetikailag zéró", "üres" igei elemet, míg mások szerint ezekben az esetekben egyszerűen nincs kopula a mondatokban.

Pete (2004) a zéró kategóriák elméletét Bloomfieldtől és a strukturalizmustól eredezteti, ahol a zéró elemek célja a nyelvi aszimmetria kiküszöbölése. A zéró tehát nem egyszerűen valaminek a hiánya, hanem a rendszerben funkcionális vagy szemantikai oppozícióban van egy nem zéró nyelvi elemmel. (Funkcionális oppozíció alatt valamilyen nyelvtani kategória, szemantikai oppozíció alatt pedig valamilyen jelentéstartalom eltérő értékeit értjük.)

Pete (2004) a zéró elemek használatát tekintve maximalista és mérsékelt felfogásról beszél. Előbbire szerinte "a zérók számának és típusainak logikus, de a nyelvleírást bonyolultabbá tevő, mértéken felüli szaporítása" jellemző, melynek célja az aszimmetria teljes megszüntetése. A Pete (2004) által is követett mérsékelt felfogás ezzel szemben természetesnek tekinti a nyelvi aszimmetriát, nem próbálja azt mindenáron kiküszöbölni, ugyanakkor elfogadja a zéró elemek létét is, de csak azokban az esetekben, ahol a zéró elem szembeállítható egy azonos jelentésű vagy funkciójú jelölt elemmel. Eszerint a magyar főnevek (névszók) egyes szám alanyesetében egyetlen zérót sem tételezhetünk fel, mivel az alanyesetet és az egyes számot egyetlen főnéven sem jelöli soha külön morféma. Az igeragozásban azonban feltételezhetünk zéró E/3. személyragot, amely szembeállítható testes személyragos alakokkal: *vár-0, várt-0, várna-0 vkit – esz-ik húst, vár-ja a barátját*.

Mivel Pete (2004) morfémákkal foglalkozik, a kopulát nem említi a kérdéskörben, ám az előbbiek alapján elméletében a zéró kopula léte attól függ, hogy van-e olyan mondat, ahol a jelen idejű 3. személyű kopula megjelenik testes alakban. Erre nehéz egyértelmű választ adni, hiszen, mint az előző fejezetekben utaltunk rá, maga a kopula fogalma sem egységes a nyelvészeti szakirodalomban. A dolgozatban (elsősorban 4.1. fejezet) azonban fogunk majd találkozni Pete (2004) példáihoz hasonló *van*–semmi váltakozásokkal.

A Pete (2004) által maximalista felfogásúnak nevezett MGr. értelmezése szerint a kopulának nincs "tartalmas jelentése", a szerepe annyi, hogy hordozza azokat az igei inflexiós jegyeket, amelyeket a névszói állítmány nem tud hordozni. Abban az esetben, amikor ezek a jegyek nem jelöltek, zéró kopulát feltételez, hasonlóképpen ahhoz, ahogyan feltételezi az alapértelemezett eseteket (időt, módot, személyt) "jelölő" zéró morfémákat is az igék esetén.

A zéró kopula mellett foglal állást Szécsényi (2000) és Alberti és Medve (2002) is. Előbbi szerint a nominális mondatokban van egy fonológiailag testetlen létige, amelynek van mondattani

DOI: 10.15774/PPKE.BTK.2025.020

szempontból alanya, szemantikailag azonban ez az alany a predikatív főnév argumentuma. Szerkezetileg a predikatív főnév igemódosító szerepet tölt be a mondatban. Alberti és Medve (2002) szerint is szükség van egy szemantikailag üres elemre a jelen idő 3. személyű nominális mondatokban, mert a mondatnak időjelöltnek kell lennie, és az időt jelölő inflexiót a zéró kopula tudja magára venni ezekben az esetekben.

Ezzel szemben a Strukturális Magyar Nyelvtan (Kiefer (1992), továbbiakban: SMNy.) szerint a kopula nem zéró hangalakú az inflexió nélküli esetekben, hanem egyszerűen nincs kijelentő mód, jelen idő, 3. személyű alakja. Ennek az alaknak a személy- és időjegyei az állítmány névszói részén jelennek meg a kopula nélküli mondatokban, és ezekkel a jegyekkel egy produktív lexikai szabály látja el az állítmányi névszókat (tehát eszerint ige nélküli mondatszerkezet is generálható).

Hasonlóan vélekedik É. Kiss (2002), aki szerint a kopula csak akkor van jelen a mondatban, ha azt valamilyen morfofonológiai megszorítás megkívánja. Ezt szórendi példákkal támasztja alá: egyrészt azzal, hogy a predikatív névszó nem lehet kontrasztív topik, másrészt hogy a fókuszos mondatokban a kopula a fókusz után áll, ha azonban nincs kopula a mondatban, akkor a predikatív névszó kerül a fókusz utáni pozícióba, azaz lényegében átveszi az ige helyét.

Mivel a dolgozat leíró nyelvészeti fejezetei az elméletfüggetlenségre törekednek, ezekben a "zéró vagy semmi" kérdése általában nem releváns. A 4. fejezetben ismertetett jelenségek zéró kopulával (létigével) és a kopula (létige) elhagyásával is magyarázhatók. A kopulás mondatok szórendjével foglalkozó 5.1. fejezet egyes (É. Kiss (2002)-höz hasonló) megállapításai viszont inkább a kopula teljes hiányára utalnak. Az elméleti fejezetekben ezért a létige (vagy kopula) hiányára elhagyásként fogok hivatkozni.

A számítógépes nyelvfeldolgozásban azonban mindenképpen lényeges gyakorlati kérdés, hogy megjelenítsük-e a szövegben a zéró kopulát mint feldolgozható, feldolgozandó elemet. Ennek részletesebb szempontjairól a 2.3.1. és a 3. fejezetben lesz szó.

1.3.2. Az elhagyhatóság szempontjai

A zéró kopula jelensége és a kérdés, hogy mitől függ az elhagyhatóság nem csak a magyar nyelvben ismert, a szakirodalomban tipológiai keretben is találkozhatunk vele. A Stassen (1994) által megfogalmazott "zero copula question" így hangzik:

What (if any) is the principal on the basis of which the cross-linguistic distribution of full and

DOI: 10.15774/PPKE.BTK.2025.020

zero copula encoding can be predicted? (Stassen 1994: 109)

Kutatásában 385 nyelvet vizsgált meg, és ezek alapján három fő szempontot állapított meg, amelyek meghatározhatják, hogy mely esetekben szükséges a testes kopula a mondatokban:

- Permanency Parameter (állandósági paraméter)
- Present Parameter (jelen idő paraméter)
- 3rd Person Parameter (3. személy paraméter)

A Permanency Parameter szerint ha a névszói predikátum állandó, lényegi tulajdonságra utal, akkor zéró kopulás lesz a mondat, átmeneti, esetleges tulajdonság esetén viszont testes. Ezt az elvet követik például a dravida kannada nyelv kopulái. A Present Parameter jellemző például az oroszra, ahol minden jelen idejű nominális mondatban zéró kopulával találkozunk. A 3rd Person Parameterre Stassen (1994) éppen a magyart hozza példának: a magyarban a jelen idő mellett a 3. személy is feltétele a zéró kopulának. A magyar tehát a zéró kopula lehetőségét tekintve a legszigorúbb nyelvek közé tartozik. A magyarra jellemző feltételeket Stassen (1994) a zéró kopula minimális disztribúciójának is nevezi:

The minimal distribution of zero copulas: If a language allows a zero copula at all, it will minimally select this option for predicate nominal sentences in the Present Tense with a Third Person Subject. (Stassen 1994: 111)

Ugyanakkor Stassen (1994) úgy véli, az a tény, hogy például a magyarban a zéró létige használata nem terjed ki a lokatív mondatokra, problémássá teszi azt az általános vélekedést, miszerint a kopula egy szemantikailag üres, igei inflexióhordozó elem. Szerinte ugyanis a jelen idő, egyes szám 3. személyű lokatív mondatokban a *van* kopula nem indokolt, hiszen nem hordoz inflexiót, és nincs lexikális jelentése sem.

A MGr. ezt a kérdést úgy oldja fel, hogy a lokatív mondatok *van*-ját már nem kopulának, hanem tartalmas igének tekinti. A MGr. a tartalmasság szempontjából egy skálát állít fel a létigék között:

- létezést jelentő ige
- birtoklásige
- határozói bővítménnyel társuló létige

DOI: 10.15774/PPKE.BTK.2025.020

- kopula

Eszerint a kopula a *van* szótő egyetlen olyan megvalósulása, amelynek egyáltalán nincs lexikális jelentése, ezért zéró alakú abban az esetben, ha nem hordoz inflexiót. Azonban, mint említettük, a "tartalmasság" fogalma nincs megfelelően definiálva a MGr.-ban.

A MGr.-hoz hasonló skálát állít fel a már szintén említett Clark (1978), aki azt vizsgálta, mennyire gyakori a zérós szerkesztésmód a világ nyelveiben a különböző létigetípusok esetén. Itt megismétlem az általa felállított, az 1.2. fejezetben már ismertetett univerzális hierarchiát:

1. predikatív (*Eszti okos.*)
2. azonosító (*Eszti az első unoka.*)
3. birtoklásmondat2 – be-possessive (*Ez a cipő az enyém.*)
4. lokatív (*A macska a hűtő mögött van.*)
5. egzisztenciális (*Van elég liszt.*)
6. birtoklásmondat1 – have-possessive (*Esztinek rengeteg játéka van.*)

Az 1-3. kategória a MGr. kopula fogalmába esik, a lokatív ige a határozói bővítménnyel társuló létige csoportjába sorolható, míg az egzisztenciális és birtoklásmondat a létezést jelentő és a birtoklásigének feleltethető meg. A Magyar grammatika "tartalmasági" és Clark (1978) "elhagyhatósági" hierarchiája tehát közel azonos, így a "lexikális tartalmasság" és az elhagyhatóság között valóban fennállhat az összefüggés. Ennek részletes vizsgálatával foglalkozik a 4. fejezet.

1.4. A dolgozat felépítése

A dolgozat elsődleges vizsgálati területe a létigék elhagyhatóságának szempontjai, melyet valós, korpuszokból mért nyelvi adatokra alapoz. Ezen kívül kitér a nominális mondatok néhány szerkezeti tulajdonságára, és bemutat egy automatikus zérókopula-beillesztő eszköz kidolgozására tett kísérletet.

A 2. fejezet a dolgozatban használt módszereket és elméleti megközelítéseket foglalja össze. Ide tartoznak a korpuszalapú és korpuszvezérelt módszerek jellemzői, a gépi tanulás, a kiértékelési

DOI: 10.15774/PPKE.BTK.2025.020

módszerek, valamint a számítógépes mondatelemzés, azon belül elsősorban a függőségi elemzés megközelítései. Szintén ez a fejezet mutatja be a kutatásaimhoz használt korpuszokat és nyelvfeldolgozó eszközöket.

A 3. fejezet egy olyan számítógépes nyelvfeldolgozó eszköz fejlesztését írja le, amely alkalmas a zéró kopulás mondatok automatikus felismerésére, sőt képes a zéró kopulát a mondatok megfelelő helyére beilleszteni. A rendszer alapja egy párhuzamos korpusz, ahol a zéró kopulás mondatokban a célnyelvi oldalon szerepel a kopula helyét jelző címke. Ezt tanítóanyagként felhasználva egy gépi fordítót tanítottunk be arra, hogy a nominális mondatokba beillessze a zéró kopula címkét. Bár az eszköz pontossága egyelőre nem éri el azt a szintet, hogy érdemes legyen korpuszépítésre használni, a fejlesztését célzó kutatás mégis számos tanulsággal szolgált. A dolgozat természetesen ezekre is kitér.

A 4. fejezet azokat a "zéró létigéket" vizsgálja a korpuszokban, amelyek nem a nominális mondatok körébe tartoznak. A fejezet három jelenségkört elemez részletesen. A 4.2. fejezet az *Ott az ajtó.*-típusú mondatokra fókuszál, és arra keres választ, miért támogatja a helyre utaló demonstratív elemek (*itt, ott, hol*) jelenléte a létige elhagyását. A 4.3. fejezetből az derül ki, hogy nem csak a *van* létige hagyható el bizonyos mondattípusokból, hanem annak tagadó változata, a *nincs/sincs* is. Ez az alfejezet a *Semmi gond.*-típusú mondatok lexikális sajátosságait vizsgálja. Végül, a 4.4. fejezet a *lehet* modális segédige elhagyhatóságának kérdését helyezi középpontba, azaz hogy mely igék esetén lehetséges az *Innen jól látni.*-típusú segédige nélküli szerkesztésmód.

Az 5. fejezet a kopulás mondatok néhány szerkezeti (szórendi) kérdésével foglalkozik. Az 5.1. fejezet, a zérókopula-beillesztő program fejlesztésének tanulságaira alapozva, olyan eseteket mutat be, ahol a zéró és a testes kopulás mondatok szerkezete eltér egymástól, és emiatt a kopula elhagyása hibás mondatot eredményez. Az 5.2. fejezet pedig a predikatív névszók szerkezeti helyének kérdését járja körül: mely esetekben jellemző ezekre a fókuszálás, és mikor preferált a semleges (igemódosítós) szórend.

Végül, a 6. fejezet a predikatív névszók (segéd)igei tulajdonságait, illetve a kopula "tartalmas jelentésének" gyengülését vizsgálja két szórendi kérdés (az igekötő-leválás és az *-e* kérdő partikula helye), illetve az igékre jellemző határozói és infinitívuszi vonzatok tükrében. Utóbbiaknál a fejezet kitér a segédigévé válás útjára lépett predikatív névszókra (*kell, muszáj, szabad*) is.

DOI: 10.15774/PPKE.BTK.2025.020

2. Módszertan

2.1. Korpuszalapú és korpuszvezérelt módszerek

A dolgozat mind módszereiben, mind céljaiban nagy hangsúlyt fektet a korpusznyelvészet szempontjaira. Az egyre nagyobb változatosságot mutató és egyre nagyobb méretű elérhető korpuszok megjelenése, valamint a számítógépes módszerek alkalmazása korábban elképzelhetetlen mennyiségű szöveges adat elemzését teszi lehetővé, ami segítheti a nyelvi jelenségek jobb megértését, vagy éppen új vagy eddig nem vizsgált jelenségek felfedezését. A korpuszok használata mára az elméleti nyelvészetben is általánossá vált (Teubert, 2005), a gépi tanulásra alapozott nyelvtechnológiai fejlesztésekben pedig egyenesen elengedhetetlen.

A korpusznyelvészetben a szöveges adatok felhasználásának két fő módszere van. Korpuszalapúnak (corpus-based) nevezzük azt a megközelítést, amikor valamilyen eleve meglévő hipotézist tesztelünk az adatokon, például gyakoriságok vagy disztribúciók vizsgálatával. A korpuszvezérelt (corpus-driven) megközelítés esetén viszont csak minimális előzetes feltevéseink vannak, minden következtetést a korpuszadatokból vonunk le (Tognini-Bonelli, 2001). A dolgozatban ismertetett kutatásokra nagyrészt az előbbi megközelítés jellemző. Az általam végzett korpuszalapú mérések a zéró létigék előfordulásait, a létigés mondatok lexikális és disztribúciós jellemzőit, valamint a kopulás mondatok szerkezetét (szórendi változatait) vizsgálják, jellemzően meglévő elméletekre alapozva.

A korpuszokra alapozott kutatások legfőbb mérőszáma a gyakoriság (és az ebből következtetett valószínűség), amit más módszerekkel nem igen lehet hatékonyan vizsgálni. A gyakorisági értékeket tekintve megkülönböztetünk abszolút és relatív gyakoriságot. Abszolút gyakoriság alatt a keresett jelenség előfordulásainak számát értjük, a relatív gyakoriság pedig a jelenség arányát jelenti a korpuszmérethez viszonyítva. Utóbbi általában informatívabb, hiszen egy abszolút számérték, viszonyítási alap nélkül keveset mond el arról, hogy az adott jelenség mennyire tekinthető gyakorinak vagy ritkának. Ezen kívül a relatív gyakoriság lehetővé teszi a gyakorisági értékek összehasonlítását különböző korpuszokban, akkor is, ha a vizsgált korpuszok mérete nem azonos. A kutatásaim során használt korpuszadatok értékelésénél ezért legtöbbször a relatív gyakoriságot használtam.

DOI: 10.15774/PPKE.BTK.2025.020

2.2. Számítógépes nyelvfeldolgozás

A dolgozat az elméleti nyelvészeti kutatások mellett a számítógépes nyelvfeldolgozással is foglalkozik, azaz a számítógépes módszerekre nem csak a kutatás eszközeként tekint, hanem részben céljaként is. A 3. fejezet teljes mértékben nyelvtechnológiai motiváltságú: egy automatikus zérókopula-felismerő eszköz kidolgozását mutatja be.

Az eszköz egyik része egy szabályalapú algoritmus implementációja. A szabályalapú számítógépes nyelvfeldolgozó módszerek Chomsky formális nyelvtanáig nyúlnak vissza. Az ilyen módszert alkalmazó eszközök kézzel írt, nyelvészetileg megalapozott szabályok és szótárak segítségével oldanak meg nyelvi feladatokat, esetünkben a kopulák és a létigék megkülönböztetését.

A számítógépes nyelvfeldolgozás másik (mára elsődleges) módszere a gépi tanulás. Mitchell (1997) a gépi tanulást a következőképpen definiálta: *"Egy számítógépes programról azt mondjuk, hogy T feladattípust és P teljesítménymutatót tekintve tanul az E tapasztalatból, ha a P-vel mért teljesítménye T feladatokban az E tapasztalattal javul."*[1] Vagyis explicit szabályok megfogalmazása helyett nagy mennyiségű tanítóadatot (tapasztalatot) biztosítunk a számítógépes programnak, amely ebből a mintázatok felismerésével képes "tanulni". A gépi tanuláson alapuló módszereknél tehát ismét jelentős szerepe van a korpuszoknak, ezek szolgáltatják ugyanis a tanítóadatot.

Az általam fejlesztett zérókopula-beillesztő eszköz esetén a program által elvégzendő feladat a zéró kopulák beillesztése a bemenetül kapott mondatokba. A tapasztalat ehhez egy szabályalapú módszerrel készített zéró kopula címkéket tartalmazó tanítókorpusz. A kiértékeléshez a klasszifikálási feladatokban széles körben használt pontosság (precision) és fedés (recall), illetve az f-mérték (f-score) értékeket használtam (Kent et al., 1955). A pontosság azt fejezi ki, hogy a kapott találatok közül hány volt helyes (tehát amit az eszköz kopulaként címkézett, az valóban kopula-e), más szavakkal ez a mérőszám a fals pozitív találatokat bünteti. Ezzel szemben a fedés a fals negatív találatok számáról ad információt: azt adja meg, hogy az elvárt találatokból (kopulákból) hányat talált meg az eszköz. Az f-mérték a pontosságot és a fedést is figyelembe véve egy átlagos mérőszámot ad az algoritmus teljesítményéről. A pontosság-, fedés- és f-mérték értékek képletei a következők:

[1] "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E." – fordítás tőlem

DOI: 10.15774/PPKE.BTK.2025.020

$precision = \frac{TP}{(TP+FP)}$
$recall = \frac{TP}{(TP+FN)}$
ahol:

- TP: valódi pozitív találatok száma
- FP: fals pozitív találatok száma
- FN: fals negatív találatok száma

$F = \frac{2*precision*recall}{(precision+recall)}$

2.3. Függőségi elemzés

A számítógépes szövegfeldolgozásban a mondatszerkezet reprezentálására a függőségi elemzés terjedt el, számos nyelven léteznek ilyen annotációt tartalmazó korpuszok, illetve ezeken betanított szintaktikai elemző eszközök. A modern függőségi nyelvtant Tesnière (1959) alapozta meg, majd Mel'čuk (1988, 2003) dolgozta ki a részleteit a szintaxisra. A függőségi elemzési fa az összetevős elemzéssel szemben nem összetevőket ábrázol, hanem a mondat minden szava a fa egy csomópontjának felel meg. A fa egy virtuális gyökércsomópontból indul ki, a szavakat reprezentáló csomópontok ebből ágaznak el, szigorúan hierarchikus struktúrában, azaz egymásnak alárendelve. Az egymásnak alárendelt szavak között különbözö, általában a szintaktikai szerepeket jelölő relációk állhatnak fenn. Ezek mindig egy az egyhez megfeleltetések, azaz minden szónak csak egyetlen "feje" lehet. Ennek az összetevős elemzéssel összehasonlítva hátránya, hogy nem határozza meg az azonos fejhez tartozó összetevők közötti hierarchiát, és nem (vagy csak nagyon nehézkesen) lehet vele frázisokat kezelni. Előny viszont, hogy a szavak között definiált relációk pontosabb képet adnak az egyes vonzatok funkciójáról. Ezen kívül a számítógépes feldolgozás számára előnyös, hogy az egy-egy megfeleltetésnek köszönhetően a mondat szerkezete fa struktúrában ábrázolható.

Mel'čuk (1988, 2003) 12 relációtípust határoz meg, melyek közül 6 az ige és vonzatai közötti kapcsolatokat írja le, a maradék 6-ot pedig a koordinációkra és a különféle módosítókra alkalmazza. Mel'čuk függőségi nyelvtana a mellérendelést is egyfajta alárendelésként értelmezi: a kötőszót és az egymás mellé rendelt elemeket a COORD reláció köti össze. Másik fontos sajátossága a nyelvtannak, hogy megenged zéró elemeket, ilyen például az orosz (és a magyar) kopula.

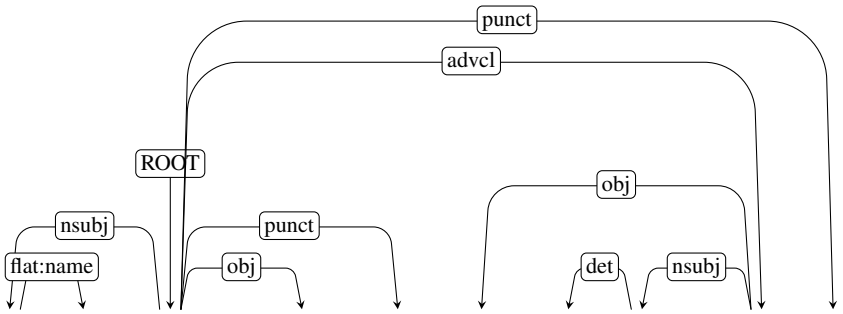
DOI: 10.15774/PPKE.BTK.2025.020

A függőségi elemzést is tartalmazó korpuszok (treebankek) növekvő száma miatt szükségessé vált az annotációs sémák egységesítése. Erre a célra jött létre a Universal Dependencies projekt[2], melyhez jelenleg 138 nyelv 244 korpusza tartozik. A Universal Dependencies által meghatározott legfontosabb relációk a következők:

- `nsubj`: alany
- `obj`: tárgy
- `iobj`: indirekt tárgy
- `cop`: kopula
- `obl`: határozói vonzatok
- `aux`: segédige
- `conj`: konjunkció
- `nmod`: NP-módosító
- `det`: determináns

Magyar nyelvre eddig egy nagyobb, függőségi annotációt is tartalmazó korpusz létezik: a Szeged Treebank (bővebben ld. 2.4. fejezet), amelynek egy részére a Universal Dependencies elveit követő annotáció is készült[3]. (11)-ben egy példát láthatunk a Szeged Treebank függőségi elemzésére.

(11)



[2]https://universaldependencies.org/

[3]https://universaldependencies.org/treebanks/hu_szeged/index.html

DOI: 10.15774/PPKE.BTK.2025.020

A dolgozat számítógépes elemzésre irányuló részei (3. fejezet) elsősorban a függőségi elemzés szempontjait veszik figyelembe. A leíró részek (4 – 5. fejezet) azonban igyekeznek elméletfüggetlenek maradni.

2.3.1. Zéró elemek a korpuszokban és a számítógépes elemzésben

A zéró elemek feltételezése sérti a függőségi elemzés azon alapelvét, miszerint minden szónak pontosan egy csomópont felel meg a függőségi fában. Ennek ellenére több nemzetközi és magyar függőségi elemzést tartalmazó korpusz (treebank) mégis alkalmaz üres csomópontokat, például az ellipszis vagy éppen a zéró kopulák kezelésére.

Mel'čuk (1988) az orosz nominális mondatokban zéró kopulát feltételez jelen időben, ezzel teljessé teszi a múlt időben testesen is megjelenő kopula paradigmáját. Ennek megfelelően a függőségi elemzésben amellett érvel, hogy a zéró kopulás mondatok feje egy üres kopula csomópont legyen, így az elemzési fa szerkezete azonos lesz a múlt idejű nominális mondatokéval.

Ezt az elvet követi a Szeged Treebank, ahol a függőségi elemzésben üres igei fejek helyettesítik a zéró kopulákat és az egyéb elliptált igéket. A korpusz elemzési sémája tehát (Mel'čuk (1988)-hoz hasonlóan) mindig a ragozott igét tekinti a mondat fejének, így ha az valamilyen okból (nominális mondat vagy ellipszis) hiányzik, zéró elemmel kell jelölni a helyét, hogy a mondat elemezhető legyen.

Hasonlóképpen a KorKor pilotkorpusz (Vadász, 2020) is zéró elemekkel jelöli a hiányzó igéket (kopulákat és egyéb elliptált igéket is). Ezen kívül a KorKor korpusz a "testetlen névmásokat" is jelöli, ide sorolva a finit igék jelöletlen alanyát, határozott ragozású igék jelöletlen tárgyát, a jelöletlen birtokost és az infinitívuszok alanyát.

A Universal Dependencies alapszabályai azonban sem a morfológiai, sem a szintaktikai elemzésben nem engedik meg a zéró elemeket (előbbinél ez nehezen is lenne értelmezhető, hiszen a UD morfológiai annotációja nem jelöli a szóelemeket). Ennek a megközelítésnek az a fő előnye, hogy a szövegfeldolgozás során csak a felszíni alakok (és "látható" szavak) elemzésével kell foglalkozni, nem nehezítjük az elemzési feladatot zéró elemek megtalálásával és beillesztésével. Hátrány viszont, hogy a ragozott igét nem tartalmazó mondatok esetén így nehezebb a mondat fejének meghatározása. (A testes kopulát nem tartalmazó mondatok számítógépes elemzési lehetőségeiről a 3. fejezet bevezetőjében lesz részletesebben szó.) A Universal Dependencies szigorúsága nem

DOI: 10.15774/PPKE.BTK.2025.020

csak a zéró elemek kapcsán vet fel kérdéseket, bizonyos nyelvi jelenségek esetén a címkekészlet korlátozottsága és az egy-egy relációs fa struktúra is jelentősen nehezíti az "ésszerű" ábrázolást. Éppen ezért az alap UD függőségi elemzés mellett kidolgoztak egy megengedőbb, úgynevezett "kiterjesztett" UD elemzési sémát, amely a függőségi fa helyett gráf reprezentációt alkalmaz (az-az egy szónak egynél több feje is lehet), és megengedi a zéró elemeket az ige elliptálása esetén (Schuster és Manning, 2016).

A zéró elemek korpuszokban való megjelenítésének kérdésénél érdemes még megemlíteni a korpuszfelhasználó nyelvészek szempontjait is. A nominális mondatokat elméleti szempontból vizsgáló nyelvész számára a zéró kopula megjelenítése mindenképpen hasznos, enélkül már ott nehézségbe ütközik, amikor adatot keres a kutatásához. (Nem tud arra keresni, ami nincs ott.) A disszertáció 3. fejezetét ilyen jellegű személyes tapaszatalat is motiválta.

2.4. Felhasznált korpuszok

- **MNSz2** (Oravecz et al., 2014)

  A Magyar Nemzeti Szövegtár v2.0.4. verziója 1,04 milliárd szövegszót tartalmaz 6 stílusrétegből és 5 regionális nyelvváltozatból. Alkorpuszai a sajtó, a szépirodalom, a tudományos, a hivatalos, a személyes és a beszéltnyelvi szövegtípusokat ölelik fel, utóbbi kettő különböző internetes fórumok és a közösségi média szövegeit, illetve országgyűlési jegyzőkönyvek és rádióinterjúk átiratait jelenti. A szövegeket egy tokenizáló-szegmentáló, morfológiai elemző és POS-tagger modulokat tartalmazó pipeline dolgozta fel. A tokenizálást a Huntoken eszköz végezte, a morfológiai elemzések pedig a Humor (Novák, 2014) elemzővel készültek. A korpusz weben elérhető grafikus keresőfelülete a Sketch Engine keretrendszerbe tartozik. Ennek fő funkciói a szóalak, szótő és címke szerinti keresés, a gyakorisági listák készítése, valamint a kollokációkereső. A korpusz anyaga (vagy annak keresett részletei) szöveges és xml formátumban is letölthető.

- **Szeged Treebank** (Vincze et al., 2010)

  A Szeged Treebank az első magyar nyelvű, kézzel annotált, függőségi elemzést is tartalmazó korpusz. Az eredeti Szeged Treebank 82000 mondatból (1,2 millió szó) áll, melyek 6 különböző domainből származnak: irodalmi szövegek, 14-16 évesek fogalmazásai, újságcikkek,

DOI: 10.15774/PPKE.BTK.2025.020

informatikai témájú szövegek, jogi szövegek és üzleti-pénzügyi témájú hírek. Ehhez eredetileg egy közvetlen összetevős elemzést reprezentáló szintaktikai annotáció tartozott, később azonban elkészült a korpusz függőségi elemzése is.

A korpusz nagyobb része nem a Universal Dependencies annotációs sémáit követi, a fő elemzési szempontok azonban azonosak. Sajátos, és a dolgozat témájának szempontjából fontos elemzési megoldás, hogy a korpusz annotációja a kopula és a predikatív névszó közötti viszonyt a `PRED` relációval jelöli, testes kopula hiányában pedig egy virtuális csomóponthoz köti a predikatív névszót.

A korpuszban használt főbb relációtípusok a következők[4]:

- `ATT` – főnevek és attributív melléknevek, főnevek és névutók, illetve főnevek és főnévi módosítók közötti viszony
- `INF` – infinitívusz és főige közötti viszony
- `LOCY` – *hol?* kérdésre válaszoló határozószó, esetragos vagy névutós névszó és feje közötti viszony
- `MODE` – "egyéb" (nem időre vagy helyre vonatkozó) határozószó, esetragos vagy névutós névszó és feje közötti viszony
- `OBL` – az ige és annak határozói vonzatai közötti viszony
- `PRED` – predikatív névszó és kopula közötti viszony
- `SUBJ` – ige és alanya közötti viszony
- `TLOCY` – *mikor?* kérdésre válaszoló határozószó, esetragos vagy névutós névszó és feje közötti viszony

A korpusz formátuma a CoNLL 2009 Shared Task (Hajič et al., 2009) normáit követi, és kutatási célokra szabadon hozzáférhető.

- **OPUS OpenSubtitles** (Lison és Tiedemann, 2016)
  Az OPUS OpenSubtitles egy 62 nyelven elérhető, összesen 22,1 milliárd szövegszót tartalmazó, filmfeliratokból álló párhuzamos korpusz. Ennek magyar-angol részéből készült

[4] A teljes listát ld. Vincze et al. (2010):1856–1857

DOI: 10.15774/PPKE.BTK.2025.020

a 3. fejezetben felhasznált lemmatizált, morfológiailag elemzett és egyértelműsített, szó szinten megfeleltetett korpusz (Novák et al., 2019), amely 644,5 millió tokennyi angol szövegből és ezek magyar fordításából áll. Az angol oldal elemzése a morpha lemmatizálóval (Minnen et al., 2001) és a Stanford taggerrel (Toutanova et al., 2003) történt. A magyar oldalon a PurePos (Orosz és Novák, 2013) és a Humor eszközök végezték el a morfológiai elemzést. Az elemzett szövegekben mindkét oldalon minden eredeti tokent két token reprezentál: a szótő és a fő szófajcímke, illetve az egyéb morfológiai címkék. (12)-ban láthatunk példát az előfeldolgozott mondatokra (*Szeretlek, kedvesem. - I love you, dear*).

(12) a. `szeret[IGE] [Ie1] ,[PUNCT] kedves[FN] [PSe1][NOM]`

b. `I#PRP love#VB [P] you#PRP ,#, dear#RB`

Az előfeldolgozott mondatokon a fast align program (Dyer et al., 2013) segítségével szó szintű megfeleltetések készültek. Ehhez a morfológiai címkék külön tokenként ábrázolása előnyös, mert így lehetőség van arra, hogy bizonyos, lexikális megfelelővel nem rendelkező szavakat a morfológiai címkékkel kössünk össze, például az angol prepozíciókat a magyar oldalon az esetragokat jelző címkékkel.

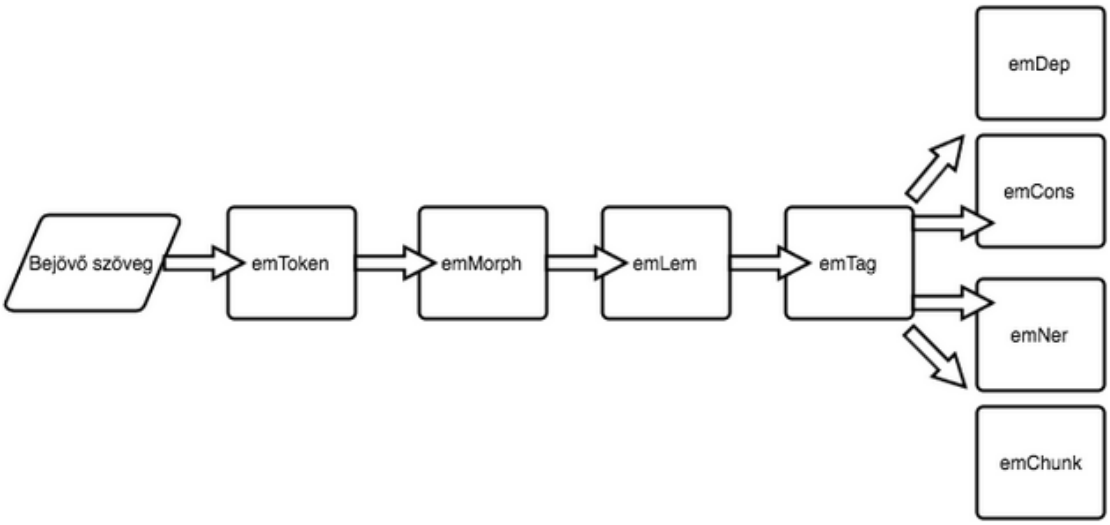
2.5. Felhasznált nyelvfeldolgozó eszközök és nyelvmodellek

- **e-magyar (emtsv)** (Váradi et al., 2018; Indig et al., 2019)

  Az `e-magyar.hu` Digitális Nyelvfeldolgozó Rendszer különböző szövegfeldolgozó eszközöket gyűjt össze egy egységes, szabványos keretben működő láncba. Az egyes elemzési lépéseket különböző modulok végzik el. A számítógépes szövegfeldolgozásnak alapfeltétele a tokenizálás, azaz a szöveget alapegységekre (szavakra és írásjelekre) kell bontani. Ezt az emToken nevű tokenizáló eszköz végzi el. Ezután a morfológiai elemzés, szótövesítés és szófaji egyértelműsítés következik az emMorph, emLem és emTag eszközök segítségével. A rendszerben kétféle szintaktikai elemzésre: közvetlen összetevős (emCons) és függőségi elemzésre (emDep) van lehetőség. A rendszer (többek között) rendelkezik még főnévicsoport-azonosító (emChunk) és névelem-felismerő (emNer) modulokkal is.

DOI: 10.15774/PPKE.BTK.2025.020

Az e-magyar rendszer működését az 1. ábra foglalja össze[5].



1. ábra. Az `e-magyar` elemzőrendszer működése

A kiinduló fő elemzési modulokon túl az emtsv rendszer folyamatosan bővül további eszközök beépítésével, így például már elérhetők rajta keresztül más modellekre épülő elemzők (pl. UDPipe, Stanza, BERT) és egyéb nyelvfeldolgozó (pl. igekötő-kereső, tagmondathatár-meghatározó) eszközök[6].

- **Marian NMTT**[7] (Junczys-Dowmunt et al., 2018)

  A neurális gépi fordítás (NMT – Neural Machine Translation) a gépi fordítás egy újszerű megközelítése, amely, mint neve is mutatja, mesterséges neurális hálón alapszik. A neurális háló alapú nyelvmodellek a disztribúciós szemantikából indulnak ki amely szerint a szavak jelentése szorosan összefügg azzal, hogy milyen kontextusban használjuk őket. Ennek megfelelően a vektoros szómodellek az egyes szavak fix méretű környezetét veszik figyelembe, ez alapján nyernek ki a nyers szövegkorpuszból szemantikai információt. A célszó környezetében előforduló szavak vektorai egy neurális hálózat bemenetét képezik. A háló feladata az, hogy a bemeneti vektorok összegéből megjósolja a célszót. (Ebben az esetben a CBOW

[5] Az ábra forrása az e-magyar.hu

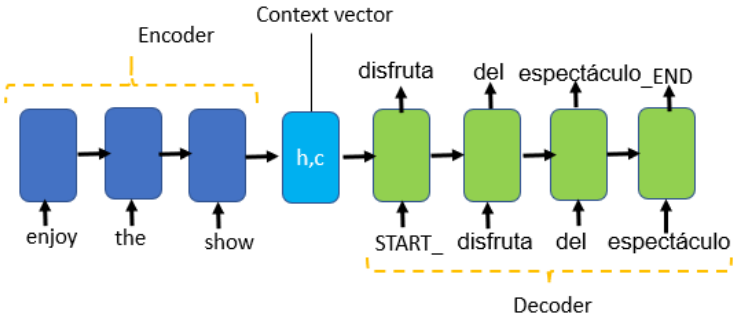
[6] Teljes lista: https://github.com/nytud/emtsv/tree/master

[7] https://marian-nmt.github.io/

DOI: 10.15774/PPKE.BTK.2025.020

– Continous Bag of Words módszerről beszélünk. Egy másik lehetőség, amikor a célszó képezi a bemenetet, és ebből próbáljuk megjósolni a szó környezetét. Ez utóbbi módszert skip-gram modellnek nevezzük.) A tanítás során a hiba visszaterjesztésével és a bemeneti vektorok frissítésével létrejön a célszót reprezentáló vektor. A vektoros ábrázolás célja, hogy a jelentésükben vagy/és morfológiájukban hasonló szavak egymáshoz közel helyezkedjenek el a vektortérben, így például két vektor (vektoralgebrai művelettel mért) távolsága meghatározza a két elem szemantikai hasonlóságának mértékét.

A neurális gépi fordítás lényege az úgynevezett enkóder–dekóder algoritmus, amely köztes nyelvmodell és fordítási modell nélkül, csupán a vektoros reprezentáció segítségével képezi le a forrásnyelvi mondatot a célnyelvire. Az enkódert és a dekódert rekurrens neurális hálók alkotják, melyek feladata a mondatjelentés és a vektorreprezentáció közötti megfeleltetés. A rekurrens neurális háló lényege, hogy az input szöveget szavanként olvassa be, így minden egyes szóval új információ adódik hozzá a mögöttes vektorhoz. A mondat végére érve így a vektor a teljes mondat jelentését fogja tárolni. A dekóder ebből a rejtett vektorból állítja elő a célnyelvi mondatot, az enkóderhez hasonlóan szekvenciálisan, mindig a következő szót megjósolva. A 2. ábrán látható egy példa az enkóder–dekóder működésére[8].



2. ábra. A neurális gépi fordító működése

A Marian NM egy C++ nyelven íródott, szabadon hozzáférhető neurális gépi fordító programcsomag. A 3. fejezetben ismertetett kutatásban a gépi fordítás módszerét és a Marian NMT-t egy nem klasszikus értelemben vett fordítási feladatra: a zéró kopulák beillesztésére használtuk.

[8]Az ábra forrása: https://towardsdatascience.com/implementing-neural-machine-translation-using-keras-8312e4844eb8

DOI: 10.15774/PPKE.BTK.2025.020

3. Zéró kopulák a számítógépes elemzésben

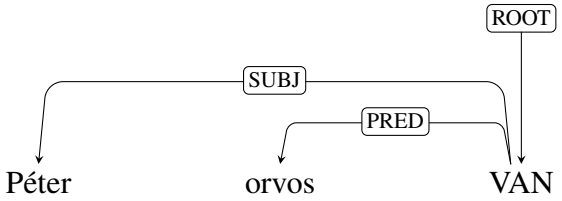
A fejezet bevezetőjében a zéró kopulák függőségi elemzésének lehetséges megközelítéseit foglalom össze, majd ismertetem egy jelölt zéró kopulákat tartalmazó korpusz építésének módszertanát. Végül, a fejezet egy automatikus zérókopula-beillesztő program fejlesztését célzó kutatást mutat be.

Mint a 2.3. fejezetben említettük, a függőségi elemzés, a szintaktikai reprezentációk nagy részéhez hasonlóan, meglehetősen igeközpontú. A mondat feje jellemzően a ragozott ige, és a relációk nagy része is az ige és argumentumainak viszonyát írja le. Ez problémákat vet fel az olyan nyelvekben, ahol a nominális mondatok nem minden esetben tartalmaznak testes kopulát, amilyen például a magyar nyelv is. Ha ragaszkodunk ahhoz, hogy a mondatok feje igei elem legyen, szükséges a zéró (lét)igék bevezetése. A zéró elemeket azonban nem minden elemzési séma engedi meg, így ezekben a nominális mondatok esetén le kell mondani az igeközpontúságról.

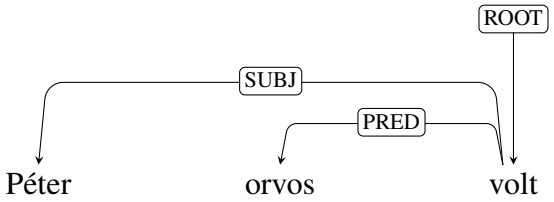
A számítógépes nyelvfeldolgozás számára az elemzési séma kiválasztásánál fontos szempont, hogy a séma alkalmas legyen hatékony gépi tanulásra. Simkó és Vincze (2017) a zéró kopulás mondatok függőségi elemzésében háromféle megközelítéssel kísérletezik: a funkciófejes, a tartalmas fejes és a komplex címkés elemzéssel. A funkciófej-elemzést követi a Szeged Treebank, ahol a predikatív névszó egy üres igei (VAN) fejhez kapcsolódik. Ennek előnye, hogy a zéró kopulás mondatok így a testes kopulásokkal analóg elemzést kaphatnak (13), cserébe viszont a zéró kopulás mondatok feje egy üres csomópont lesz. A tartalmasfej-elemzés ezzel szemben nem enged meg üres fejeket, ezért a kopulás mondatok feje minden esetben az állítmányi névszó, amihez az esetleges testes kopula COP éllel kapcsolódik (14). Ez egyébként megfelel a Universal Dependencies elveinek. A zéró kopulás mondatok elemzése így az egyargumentumú igét tartalmazó mondatokéhoz hasonló. Ez abból a szempontból előnyös, hogy a mondat feje az úgynevezett logikai állítmány lehet, cserébe viszont az elemzőnek meg kell tanulnia megkülönböztetni a morfológiailag azonos címkézésű alanyt és névszói állítmányt, illetve azt is, hogy a *van* szótő nem jelen idejű alakjai az igei fej (14c) és a névszói fejnek alárendelt kopula (14b) szerepeit is betölthetik. Végül, a komplex címkés elemzésben a testes kopulás mondatok elemzése megegyezik a funkciófej-elemzéssel, a kopula hiányát pedig a predikatív névszó speciális ROOT-VAN-PRED címkéje jelzi (15). Ez tehát megőrzi az elemzés igeközpontúságát, de elkerüli azt, hogy üres csomópont legyen a mondat feje.

DOI: 10.15774/PPKE.BTK.2025.020

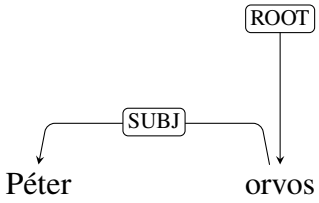
(13) a.



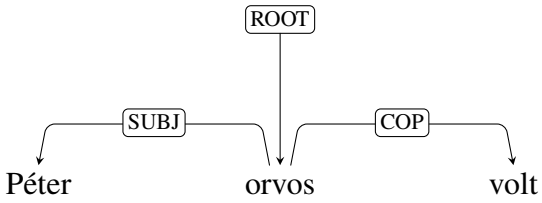
b.



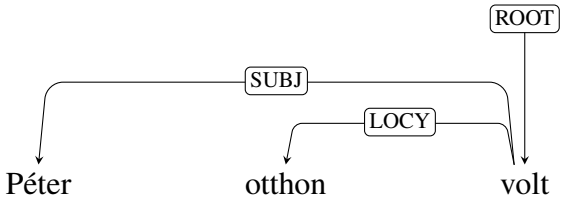
(14) a.



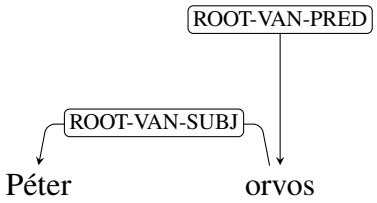
b.



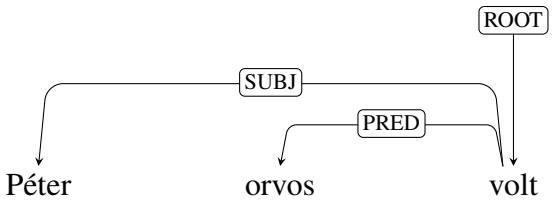
c.



(15) a.



b.



DOI: 10.15774/PPKE.BTK.2025.020

A tanulmány kísérletei során a Bohnet parsert (Bohnet, 2010) tanították be a szerzők a három elemzési módszerre. A betanított elemző teljesítményét külön értékelték ki egzisztenciális, testes kopulás és zéró kopulás mondatokon. Az 1. táblázatban az adott elemzési módszerrel elért helyes mondatok aránya látható.

	Funkciófej	**Tartalmas fej**	**Komplex fej**
Egzisztenciális	78%	80%	80%
Testes kopula	62%	42%	52%
Zéró kopula	70%	68%	30%
Összesen	70%	63%	54%

1. táblázat. Simkó és Vincze (2017) eredményei a zéró kopula különböző elemzéseire

Az eredmények azt mutatják, hogy a zéró kopulák beszúrása a tanítókorpuszba (funkciófej-elemzés) összességében kevesebb hibát eredményez a létigés és kopulás mondatok elemzésében, mint a másik két elemzési módszer. Az eredmények részletes ismertetéséből kiderül, hogy a leggyakoribb hiba mindhárom elemzési módszernél az alany és a predikatív névszó felcserélése volt. Ezen kívül a tartalmasfej-elemzésnél nehézségnek bizonyult az, hogy ebben az elemzési sémában a testes kopulás és az egzisztenciális mondatok elemzése eltérő: előbbinél a predikatív névszó, utóbbinál a létige a mondat feje. A tartalmasfej-elemzés gyenge teljesítménye (42%) a testes kopulás mondatokon abból fakadt, hogy az elemző nem tudta megkülönböztetni a kopulát az egzisztenciális mondat létigéjétől.

Simkó és Vincze (2017) eredményei szerint tehát a Universal Dependencies térhódítása ellenére is van létjogosultsága a zéró kopulás függőségi elemzési sémáknak. Az ilyen elemzést tartalmazó korpuszok létrehozásához szükség van automatikus zérókopula-beillesztésre. Erre tesz kísérletet a következőkben ismertetett kutatás.

3.1. Kutatási cél és módszer

A fejezetben ismertetett kutatásban Yang Zijian Győző és Novák Attila működött közre (Dömötör et al. (2020a), Dömötör et al. (2020b)). A kutatás célja egy olyan eszköz kidolgozása, amely képes a nominális mondatok megfelelő helyére automatikusan beilleszteni a zéró kopulát. "Megfelelő

DOI: 10.15774/PPKE.BTK.2025.020

hely" alatt azt a pozíciót kell érteni, ahol nem alapértelmezett esetben (azaz nem kijelentő mód, jelen idő, 3. személyben) a testes kopula lenne. A rendszer elvárt működése tehát a következő:

(16) Bemenet: *Fábián elvtárs is vámpír.*
Kimenet: *Fábián elvtárs is vámpír Ø.*

Bár, amint a későbbiekben látni fogjuk, a zéró-jelenség nem feltétlenül korlátozódik a "klasszikus" értelemben vett kopulára, jelen kutatásban csak azokat tekintjük kopulás mondatnak, ahol az adott alapértelmezett esetben soha nem fordul elő testes ige. Nem soroljuk ide a (17)-típusú létige–semmi váltakozásokat (a példák forrása az MNSz2). Más szóval jelen fejezet csak a nominális mondatok zérókopula-jelenségével foglalkozik, és nem terjed ki az opcionálisan elhagyható egyéb létigékre (17a,-b,) vagy a címek szerkezeti sajátosságaira (17c).

(17) a. Ott (van) a csodarendszer, hát alkalmazzák!

b. Ennek semmi értelme (nincs), csak eszembe jutott.

c. Veszélyben (van) az olajellátás

A cél, a már említett függőségi elemzés mellett,a korpusznyelvészeti kutatások támogatása egy olyan eszköz kifejlesztésével, amelynek segítségével a nominális mondatok kvantitatív vizsgálatára alkalmas, nagy méretű korpusz hozható létre.

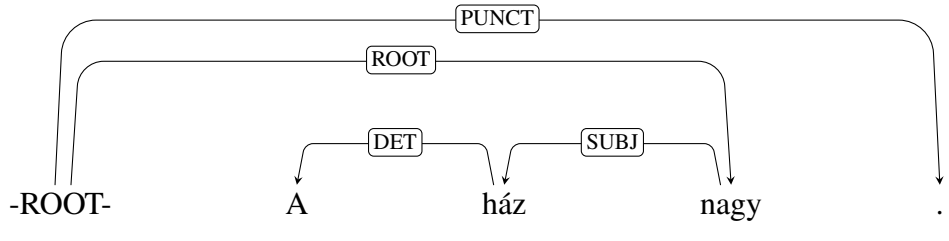
Hasonló, üres funkciófejek automatikus beillesztésére irányuló kísérletek már vannak a szakirodalomban. Seeker et al. (2012) kutatásának célja a zéró szóalakok (ellipszis miatt hiányzó szavak vagy zéró kopulák) megjóslása a függőségi elemzés során magyar és német nyelvre. Erre a szerzők szerint azért van szükség, mert az igei ellipszis (ide értve a zéró kopulát) jelensége gyakori a magyar nyelvben, és az ige hiánya esetén gondot okoz az ige függőségeinek elhelyezése az elemzési fában. A zéró elemek beillesztésére tehát azért van szükség, hogy az elliptált ige argumentumainak (függőségeinek) legyen feje. A cikk három módszert mutat be: az elsőben az üres fejeket a parser illeszti be az elemzés során, a másodiknál az üres fejek a címkekészletben vannak kódolva, míg a harmadik esetén az üres fejek szükségességéről egy osztályozó dönt az elemzés előtt. Számunkra ez utóbbi módszer a leginkább érdekes, hiszen itt a miénkhez hasonló feladatról van szó. Seeker et al. (2012) első lépésben egy tagmondat-detektálót implementált, amely a magyar nyelvre 86.8-as F-mértéket ért el. Az osztályozó modul ennek a kimenetén dönti el, hogy mely tagmondatokba

DOI: 10.15774/PPKE.BTK.2025.020

kell zéró igei elemet beszúrni. Ehhez felhasználja a szavak morfológiai elemzését és a tagmondat-hierarchiát. Az eszköz célja - a miénkkel szemben - tehát csak osztályozás, nem tesz kísérletet a zéró kopula helyének megtalálására, a zéró elemeket minden esetben a tagmondat elejére illeszti be. Az osztályozó 70.1-es F-mértéket ért el a magyar tesztadatokon. A saját módszerünkre is készítettünk ezzel összevethető, csak az osztályozás sikerességét mérő kiértékelést.

Tudomásom szerint a zéró kopula felszíni szerkezetbe való beillesztésére alkalmas eszköz nem áll rendelkezésre a magyar nyelvre. Az elérhető szintaktikai elemzők sem alkalmazhatók erre a feladatra, ezek ugyanis, a Universal Dependencies elveit követve, nem illesztenek zéró kopulát az elemzéseikbe. Ennek megfelelően a magyar nyelvtechnológiában state-of-the-artnak tekinthető e-magyar elemző a zéró kopulás mondatok esetén a tartalmasfej-elemzést alkalmazza (18. példa).

(18)



A korpuszokat tekintve említettem, hogy a Szeged Treebankben és a KorKor pilotkorpuszban vannak zéró kopulát pótló üres funkcióelemek. A KorKor korpusz azonban túl kicsi ahhoz, hogy bármilyen gépi tanulásos algoritmushoz felhasználható legyen, a Szeged Treebank "zéró kopulái" pedig nem valódi mondatrészként, csak virtuális csomópontokként szolgálnak, így a mondatbeli pozícióknak nem tulajdonítottak jelentőséget a korpusz készítői. A Szeged Treebank üres fejeket jelölő szimbólumai gyakorlatilag véletlenszerű helyeken jelennek meg a felszíni szerkezetben, ezért ezeket nem lehet közvetlenül zéró kopulás tanítóanyagként felhasználni az általunk kitűzött célra. Előnye azonban a Szeged Treebanknek, hogy bináris osztályozási feladatra (zéró kopulás-e a mondat vagy nem) gold standard adatként rendelkezésre áll. A korpusz 16003 darab zéró kopulás mondatot tartalmaz, ami a teljes méretének nagyjából 17%-a. Ezt az empirikus arányt használtuk fel a tanító- és tesztanyagaink összeállításában.

A zéró kopulák beszúrására a gépi fordítás módszerét alkalmaztuk, melynek lényege, hogy transzformációt képez tetszőleges forrás- és célnyelvi mondatpár között. A rendszer betanításához egy kétnyelvű párhuzamos korpuszra van szükség. Ebben az esetben a "forrásnyelv" a mondat felszíni alakja, a "célnyelv" pedig a jelölt zéró kopulás alak. A legtöbb mondat "fordítása" tehát

DOI: 10.15774/PPKE.BTK.2025.020

önmaga, kivéve a zéró kopulás mondatpárokat, melyekben a célnyelvi mondatban a zéró kopula helyén egy <zerokop> címke áll.

Munkánk során a Marian NMT (Junczys-Dowmunt et al., 2018) nevű keretrendszert használtuk. Az NMT tanítását és futtatását Yang Zijian Győző végezte a jelenleg state-of-the-art Transformer (Vaswani et al., 2017) modell és Sentece Piece (Kudo és Richardson, 2018) tokenizáló segítségével. A rendszer beállításai és paraméterei a következők voltak:

- Sentence Piece: szótárméret: 16000; egy szótár a forrás- és egy a célnyelvi korpusznak; karakter lefedettség a teszt korpuszon: 100%
- Transformer modell: enkóder és dekóder rétegeinek száma: 6; transformer-dropout: 0.1;
- Learning rate: 0,0003; lr-warmup: 16000; lr-decay-inv-sqrt: 16000;
- Optimizer-params: 0,9 0,98 1e-09; beam-size: 6; normalize: 0,6
- Label-smoothing: 0.1; exponential-smoothing

A fordító betanításához olyan korpuszra volt szükség, ahol a zéró kopulák jelölve vannak. Ennek előállításához egy szabály alapú bináris osztályozót dolgoztam ki, amely képes eldönteni a *volt, voltak* szóalakokról, hogy azok a mondatban kopulák-e. Az osztályozó által kopulának címkézett mondatokban lecseréltük az igei elemet egy <zerokop> címkére, ez adta a fordító zéró kopulás tanító anyagát.

3.2. Szabályalapú bináris osztályozás

3.2.1. Az osztályozási feladat

Az osztályozó célja a *van* szótő testes megvalósulásainak (továbbiakban: VOLT-oknak) bináris osztályoza. Az osztályozás szempontja az, hogy jelen idő harmadik személyben is testes-e a kérdéses létige. A továbbiakban a két osztályt a *kopula* és a *lexikális VAN* nevekkel említem. Kopulának nevezem (a magyar szakirodalom hagyományához hasonlóan) a jelen idő harmadik személyben zéró alakú, predikatív névszót kiegészítő segédigét, míg a teljes paradigmával, azaz testes "van" alakkal is rendelkező létigéket (így a lokatív, egzisztenciális és birtokos mondatok igéjét) egységesen a *lexikális VAN* osztályába sorolom.

DOI: 10.15774/PPKE.BTK.2025.020

Összehasonlításképpen kipróbáltam, milyen hatékonysággal oldja meg ezt a feladatot az e-magyar automatikus szövegelemző rendszere. A VOLT-ok osztályozásához elkészíttettem az elemzővel egy 1000 mondatból álló teszthalmaz függőségi elemzését, és a kapott elemzés szerint osztályoztam a kérdéses létigéket. A tesztmondatokat véletlenszerűen választottam ki az OPUS OpenSubtitles magyar nyelvű morfológiailag elemzett alkorpuszának *van* szótőt tartalmazó mondataiból. Ezekben a kopulák és lexikális VOLT-ok eloszlása nagyjából egyenletes volt: 598 mondatban szerepelt kopula, és 402-ben lexikális ige. Ha a VOLT-hoz tartozott vele `PRED` relációban álló névszó, akkor a VOLT *kopula* címkét kapott, egyéb esetben pedig *lexikális*at. Az osztálycímkék kézi ellenőrzésének (*kopula* címkék szerinti) eredményeit a 2. táblázat tartalmazza.

Pontosság	87,8%
Fedés	81,0%

2. táblázat. Az e-magyar osztályozási eredményei 1000 mondaton

Az eredmények szerint a különböző létigetípusok megfelelő elemzésének kiválasztása nem egyszerű feladat egy automatikus elemzőrendszer számára. A nehézség egyik oka az lehet, hogy a magyar nem konfigurációs nyelv, így a szórend nem segít abban, hogy megállapítsuk az egyes szavak szintaktikai szerepét. A magyar nyelv másik, automatikus feldolgozást nehezítő tulajdonsága a pro-drop, emiatt nem lehet a létigetípusok megkülönböztetését az esetrag nélküli névszók számára alapozni, hiszen ha csak egyetlen ilyen is van, akkor is könnyen lehetséges, hogy ez az egyetlen névszó predikatív, az alany pedig nincs jelölve.

A magyar nyelv nehéz elemezhetősége miatt a címkézéshez egy angol-magyar párhuzamos korpuszt használtam, az angol nyelv konfigurációs jellegének köszönhetően ugyanis az angol mondatokon sokkal könnyebb lokális információk alapján, szabályok segítségével szintaktikai döntéseket meghozni. Vagyis a magyar mondatok angol megfelelői segítenek abban, hogy meghatározzuk a VOLT aktuális mondatbeli funkcióját (típusát).

3.2.2. Címkézés párhuzamos korpuszon

Az adatgyűjtéshez, a tesztmondatok kiválasztásához hasonlóan, a 2.4. fejezetben ismertetett OPUS OpenSubtitles párhuzamos korpuszt használtam. Ebből először is kinyertem azokat a mondatokat,

DOI: 10.15774/PPKE.BTK.2025.020

ahol a magyar oldalon VOLT szerepel, majd ezeket egy Python3-ban implementált szabályalapú algoritmussal címkéztem.

A címkéző algoritmus első lépésben megnézi a VOLT-nak megfeleltetett angol tokeneket. Ha ezek között szerepel nem segédigei *have* vagy expletív *there*, a mondat `lexikális` címkét kap. Előbbiek ugyanis nyilvánvalóan birtokos mondatok, utóbbiak pedig egzisztenciálisak, mert az angol *there is/are* létezést kifejező szerkezetek mindkét tagja a létigével van megfeleltetve a magyar oldalon. Ha a VOLT *be*-vel van megfeleltetve, akkor a mondat lexikális igés és kopulás egyaránt lehet, ekkor tehát további vizsgálati lépésekre van szükség. Ha az előbbi tokenek (*have, there, be*) egyike sem szerepelt a VOLT-nak megfeleltetett angol tokenek között, akkor a mondat kikerült a címkézendő anyagból, ekkor ugyanis feltehetően az eredetitől nagyon eltérő fordításról lehet szó, így a mondat angol megfelelője nem megbízható kiindulópontja a címkézésnek.

Ha a VOLT az angol oldalon *be*-nek felel meg, akkor a program megkeresi az angol mondatban azt a "kulcsszót", amely alapján a címkézés megtörténik. A kulcsszó feltételezésem szerint vagy egy predikatív névszó(i csoport), vagy egy nem nominatívuszi argumentum (illetve ezek része). Az algoritmus tehát ezeknek az elemeknek a kanonikus pozícióját keresi az angol mondatban (ami a magyar mondat esetén a szórend kiszámíthatatlansága miatt lehetetlen lenne).

A kulcsszó kiválasztásához először is meg kell állapítani, hogy kijelentő vagy kérdő szórendű mondatról van-e szó. Ezt a kérdőszó vagy a *do* segédige megléte alapján ellenőrzi az algoritmus. Kijelentő mondatok esetén a kulcsszó a *be*-t követő első olyan token, amely nem tagadószó vagy NP-t módosító elem (például: *very, more*). A kulcsszónak való "alkalmasság" megállapítása elsősorban a szófajcímkék alapján történik. (19. példa) Eldöntendő kérdés vagy a *what*, *who*, *whose*, *which*, *how* és *why* kérdőszavak esetén a program a (tag)mondat utolsó szavát választja ki. (20) Az egyéb kérdószóval (pl. *where*, *when* stb.) bevezetett mondatok `lexikális` címkét kapnak.

(19) a. *Régen ez egy kvalitás volt.*

It used to be **a** quality.

b. *Nem volt otthon.*

He was not **at** home.

DOI: 10.15774/PPKE.BTK.2025.020

(20) a. *Mi volt ez a zaj?*

What was that **noise**?

b. *Miről volt szó?*

What was it **about**?

A kulcsszó kiválasztása után az algoritmus megnézi a kulcsszóhoz rendelt magyar tokeneket, és ezek szófaj- illetve morfológiai címkéi alapján megállapítja a kérdéses létige típusát. Ha a kulcsszóhoz tartozik egy nem nominatívuszi esetet jelölő morfológiai címke, akkor a mondat a `lexikális` címkét kapja. Ha a kulcsszónak névelő vagy nominatívuszi névszó felel meg a magyar oldalon, akkor a mondat címkéje *kopula* lesz. A korpusz kulcsszavaihoz tartozó szófaj- és morfológiai címkéket, valamint az algoritmus által hozzájuk rendelt létigetípust a 3. táblázat tartalmazza.

Némely esetben az algoritmus speciális lexikális szabályokat is tartalmaz, a POS-tagek ugyanis félrevezetőek lehetnek, például az időre, időjárásra vonatkozó mondatok esetén . Ezeknél a kulcsszókeresés rossz eredményt ad, ezek ugyanis az angolban kopulás mondatok, a magyarban viszont (szerkezetileg) egzisztenciálisak (21).

(21) ***Sötét*** *volt és köd.*

It was **dark** and foggy.

Ezek a szerkezetek ezért lexikális kivételként vannak kezelve egy névszólista alapján, amely az MNSZ2 kollokációkeresőjével készült, a *van* szótővel leggyakrabban kollokációt alkotó névszók alapján.

Az időjárásra utaló szerkezetek mellett van még néhány olyan speciális eset, ahol a kulcsszókeresés félrevezető lehet, ezeket leginkább "állandó fordítási különbségnek" nevezhetnénk. Ez alatt azokat a szerkezeteket értjük, amelyek angolul kopulásak, magyarra viszont lexikális VAN-nal fordítjuk őket. Ennek leggyakoribb esete a *being right* 'igaza van' mondat, de ez alá a speciális lexikális szabály alá soroltam a *being lucky* 'szerencséje van', *being necessary* 'szükség van' és a *being ready* 'kész van' szerkezeteket is. Ez utóbbi különösen érdekes, valódi kivételnek tűnik, erre a 4.1. fejezetben térünk vissza.

DOI: 10.15774/PPKE.BTK.2025.020

lexikális		kopula	
HA	határozószó	DET	determináns
HA|NM	határozói névmás	DET|NM	determinatív névmás
NU	névutó	MN	melléknév
INE	inesszívusz	MN|NM	melléknévi névmás
SUP	superesszívusz	FOK	középfokú melléknév
ELA	elatívusz	FF	felsőfokú melléknév
ADE	adesszívusz	SZN	számnév
ESSMOD	módhatározó	SZN|NM	számnévi névmás
ILL	illatívusz	FN	főnév
ALL	allatívusz	FN|NM	főnévi névmás
SUB	szublatívusz	PS	birtokos névszó
CAU	okhatározó (-ért)	OKEP	folyamatos melléknévi igenév
ABL	ablatívusz	MIB	befejezett melléknévi igenév
HIN	határozói igenév		
INS	instrumentális		
DEL	delatívusz		
DAT	datívusz		
TER	terminatívusz		
TEM	temporális		
ESSNUM	számnévi esszívusz (’valahányan’)		

3. táblázat. A kulcsszavaknak megfeleltetett morfológiai címkék és a hozzájuk tartozó létigetípusok

A címkéző algoritmust a 4. táblázat foglalja össze.

Az OPUS korpuszon futtatva a címkéző algoritmus kimenete 791130 címkézett mondat volt, ezek közül 458270 kapott `kopula` címkét, és 332860-ban szerepelt (az algoritmus szerint) lexikális ige. Az algoritmus teljesítményét ugyanazon az 1000 mondatos mintán értékeltem ki, amelyeket az e-magyar elemzéséből kiinduló előzetes teszthez is használtam. Az eredményeket az 5. táblázat tartalmazza.

Mint látható, az algoritmus mind pontosságban, mind fedésben jobban teljesített az `e-magyar`nál. A hibák elemzése során kiderült, hogy a hibás címkék nagy része nem az algoritmusból, hanem valamilyen "külső körülményből" ered. Tipikusan ilyenek például a hibás POS-tagelés vagy szómegfeleltetés. Szintén gyakori külső hibaforrás volt az angol eredetitől jelentősen eltérő magyar fordítás. Bár az algoritmus igyekszik az ebből származó problémákat kikü-

DOI: 10.15774/PPKE.BTK.2025.020

1. lépés: A *volt*-tal megfeleltetett tokenek ellenőrzése	
have	`lexikális`
there	`lexikális`
be	2. lépés
egyéb	mondat figyelmen kívül hagyva
2. lépés: Speciális lexikális szabályok	
környezeti kopulás szerkezetek	`lexikális`
right, lucky, necessary, ready	`lexikális`
egyéb	3. lépés
3. lépés: Kulcsszóválasztás	
kijelentő mondat	a *be*-t követő token
eldőntendő kérdés	tagmondat utolsó tokenje
what, who, whose, which, how, why	tagmondat utolsó tokenje
egyéb wh-kérdés	`lexikális`
4. lépés: Címke kiosztása a kulcsszó szerint	

4. táblázat. A címkéző algoritmus összefoglalása

Pontosság	90,8%
Fedés	91,1%

5. táblázat. A címkéző algoritmus kiértékelésének eredménye 1000 mondaton

szöbölni azzal, hogy figyelmen kívül hagyja az olyan mondatokat, ahol a VOLT-hoz sem *be*, sem *have* nem volt hozzárendelve, ez a megszorítás azonban még mindig sok olyan mondatot átenged a szűrőn, ahol a mondatszerkezeti vagy akár jelentésbeli különbségek megnehezítik a címkézést. (22a)-ban az angol verzióban *have* szerepel, ezért a magyar (kopulás) mondat `lexikális` címkét kapott. (22b) magyar fordítása lokatív mondat, az angol eredeti viszont kopulás, emiatt a címkéző algoritmus itt is hibázott.

(22) a. *Egy rossz álom volt.*

You had a bad dream.

b. *Ők voltak itt először.*

They were the first ones here.

DOI: 10.15774/PPKE.BTK.2025.020

Ezeket a címkézési hibákat nehezen lehetne elkerülni. Egy "szigorúbb" fordítású korpuszban, mint például az Európai Uniós dokumentumok, valószínűleg kevesebb ilyen adódna, ennek azonban az erősen korlátozott domain lenne a hátránya. Mindennek ellenére elmondható, hogy a párhuzamos korpuszok jól alkalmazhatók egynyelvű adatok gyűjtésére is minden olyan esetben, ahol a keresett szintaktikai információ egy másik nyelven explicitebb a magyarnál.

3.3. A zéró kopulák beszúrása

3.3.1. Tanító adat

A tanítóanyag létrehozásához a bináris osztályozó kimenetét használtuk fel. A program által kopulásnak címkézett mondatokban a testes kopulákat egy `<zerokop>` jelre cseréltük, ezek adták a tanítóanyag pozitív példáit (318843 mondat). Itt meg kell jegyezni, hogy a testes kopulák zéróra cserélése nem mindig eredményez helyes mondatot, a jelen időt kizáró lexikai elemek vagy a mondatszerkezet miatt. Ezekről az 5.1. fejezetben lesz részletesen szó. Az így létrehozott "mesterséges" zéró kopulás mondatokhoz hozzáadtunk az OPUS korpuszból 1 millió random mondatot (természetesen ügyelve arra, hogy ne legyen átfedés a zérókopulás mondatokkal), és ebből tanítottuk be az alapmodellt. Ez elég jó pontosságot, viszont gyenge fedést produkált (3.3.1. táblázat), ami nem meglepő, hiszen a negatív példának szánt 1 millió mondatban valószínűleg sok valódi, jelöletlen zéró kopula volt. Ennek kiküszöbölésére az OPUS korpuszból kiszűrtük azokat a mondatokat, ahol a *be* valamilyen jelen idejű alakja szerepel nem segédigei funkcióban az angol oldalon harmadik vagy második személyű alannyal (az utóbbinak az esetek nagy részében harmadik személyű a magyar fordítása). Ugyan az angol kopulás mondatok fordítása nem szükségszerűen kopulás a magyar oldalon és fordítva, ezzel a módszerrel viszonylag hatékonyan ki tudtuk szűrni a hamis negatív példák nagy részét a tanítóanyagból. (A szűrést Novák Attila végezte.) További szűrésképpen a maradékon lefuttattuk az alapmodellt. A továbbiakban azokat a mondatokat (illetve ezek egy részhalmazát) használtuk negatív példának, amelyekbe a modell nem illesztett be zéró kopulát. Végül, az alapmodellt lefuttattuk a kiszűrt (azaz az angol oldalon *be*-t tartalmazó) mondatokon is, így összesen további 161223 darab zéró kopulás mondatot nyertünk.

Az így rendelkezésre álló adatokból a következő modelleket építettük fel (a modellek tanítását és finomhangolását Yang Zijian Győző végezte):

DOI: 10.15774/PPKE.BTK.2025.020

Pontosság	89,57%
Fedés	58,19%

6. táblázat. Az alapmodell eredményei az OPUS tesztkorpuszon

- **Eredeti szűrt**: ez a modell a címkéző algoritmus segítségével előállított 318843 zéró kopulás, és 1 millió random kiválasztott nem zéró kopulás mondatot tartalmaz. A tanítóanyagba nem kerültek egyszavas, illetve speciális karaktereket tartalmazó mondatok. Ennek a szűrésnek a zajcsökkentés volt a célja.

- **Bővített szűrt**: A pozitív példákhoz hozzáadtuk az alapmodellel előállított zéró kopulás mondatokat (így összesen 477082 zérót tartalmazó mondatunk lett), a negatív példákat pedig újabb 1 millió random nem zéró kopulás mondattal bővítettük.

- **Eredeti javított**: Az eredeti szűrt modell tanítóanyagába visszakerültek az egyszavas mondatok és néhány olyan speciális karakter, amely a Szeged Treebankben gyakorinak bizonyult (pl. §). Ezen kívül Novák Attila javított néhány, a tesztadatok átnézése során feltűnő könnyen kiszűrhető hibát, amelynek altípusait a (23) alatti példákkal szemléltetjük. A javító algoritmus az eredeti modell által zéró kopulásnak jelölt mondatok közül kiszűrte azokat, ahol a magyar oldalon az eredeti mondatban szereplő *volt(ak)* szóalak valamilyen minden esetben hangsúlytalan (de nem enklitikus) összetevő: kötőszó (23a), vonatkozó névmás (23b) vagy névelő (23d, 23e) után következik. Az előbbieket csak a mindig hangsúlyos lexikális (egzisztenciális (23a) vagy birtokos (23c)) *volt* vagy *voltak* igealak követheti, az utóbbiak pedig szinte biztosan az ‘ex’ jelentésű szintén hangsúlyos *volt* melléknév hibás annotációjával keletkeztek. Lényegében az első eset speciális változata ezen kívül a vesszőt követő vagy mondat eleji *volt* (23f, 23g). Ebben az esetben ugyanúgy nagyrészt kizárt a zéró kopulával való helyettesítés (kivéve a beágyazott vonatkozó mellékmondatok esetét). Ahogy a (23) alatti példák is mutatják, az eredeti algoritmus nagyrészt akkor hibázott így, ha az eredeti mondat múlt idejű névszói állítmányt és hangsúlyos (lexikális vagy melléknévi) *volt*-ot is tartalmazott, és a szóösszerendelési modell az angol kopulát hibásan hozzákapcsolta az utóbbihoz (is). Emellett az is egy érdekes eset, amikor ugyan kopuláról van szó, de ugyanakkor ellipszis is van a mondatban (23h), ezért múlt időben a *volt* a fókuszos kontrasztív szerkezet miatt

DOI: 10.15774/PPKE.BTK.2025.020

kötelezően hangsúlyos, azonban jelen időben muszáj lenne egy másik elemnek megjelennie, hogy legyen, amit fókuszálni lehet. Ez a modell így 314607 zéró kopulás és 1515204 nem zéró kopulás mondatot tartalmaz. (Azaz, ez a tanítóanyag már megfelel a Szeged Treebankből megállapított empirikus aránynak.)

- **Bővített javított**: A bővített szűrt modell anyagához hozzáadtuk az alapmodellel generált zéró kopulás mondatokat, és a negatív példákat is kibővítettük a 17%-os aránynak megfelelően, így ez a modell 475830 zéró kopulás, és 2574207 nem zéró kopulás mondatot tartalmaz.

(23) a. De igen, a legtöbb az volt, de **volt** fehér is, valamint ilyen bébifosbarna is.

b. Az a belépőkártya... volt minden, amim **volt**.

c. Házas volt, és **volt** egy fia.

d. Megértjük, hogy a **volt** férje és Ő üzleti társak voltak.

e. A különleges osztag egy **volt** kihallgatótisztje csinálta.

f. Egy szinttel lejjebb voltak a testőrök, **volt** ejtőernyősök vagy idegenlégiósok.

g. Ahhoz képest, hogy fiú vagy, **voltak** jó válaszok.

h. Nem olyan, mint **volt**.

3.3.2. Eredmények

A kiértékeléshez kétféle teszthalmazt használtunk. Az egyik ugyanabból a korpuszból készült, mint a tanítóanyag (OPUS), a másik pedig a Szeged Treebankből. Ez utóbbi azt a célt szolgálja, hogy kipróbáljuk, hogy működik a rendszer más domaineken. Mint említettem, a Szeged Treebank alkalmas arra, hogy zéró kopulás mondatokat keressünk benne, hiszen ezekben egy üres igei csomópont alkotja a mondat fejét.

Mindkét tesztkorpusz 2000 mondatot tartalmaz, ebből, a 17%-os empirikus aránynak megfelelően, 340 zéró kopulás. A tesztmondatokat ellenőriztem és kézzel javítottam, ahol szükséges volt, illetve a Szeged Treebank mondatainak zéró kopuláit szintén kézzel kellett a helyükre illeszteni. A végeredmény tehát két, méretben és arányban megegyező, de különböző forrásból származó, és ezáltal különböző szövegtípusokat tartalmazó gold standard tesztkorpusz lett.

DOI: 10.15774/PPKE.BTK.2025.020

A Szeged Treebank mondatain tesztelve azonban váratlan nehézségekbe ütköztünk. Egyrészt kiderült, hogy ezek a mondatok olyan karaktereket is tartalmaznak, amelyek nem szerepelnek az első két modell tanítóanyagában (pl. §), és ezeket a fordító nem tudta kezelni. Emiatt a javított modellekbe visszakerültek a zajcsökkentés céljával kizárt, speciális karaktereket tartalmazó mondatok. Másrészt gondot okoztak a hosszú mondatok is, az OPUS korpuszból készült tanítóanyag mondatai ugyanis – műfaji sajátosságuknál fogva – jellemzően rövidek voltak. Erre a problémára valószínűleg a tanítóanyag hosszú mondatokkal való kiegészítése lehetne megoldás, de a Szeged Treebankből készült tesztanyagon sajnos nem értünk el olyan eredményt, ami lehetővé tette volna, hogy a program outputját tanítóanyagként használjuk.

Mindegyik modellt lefuttatuk mindkét teszthalmazon, és a kimeneten kétféle kiértékelést végeztünk. Az egyik esetén osztályozási feladatnak tekintettük a zéró kopulák beillesztését, azaz csak azt mértük, hogy hány esetben találja el a program, hogy kell-e zéró kopula a mondatba, azt nem értékeltük, hogy jó helyre illeszti-e be azt. A filmfeliratkorpusz viszonylag egyszerű mondatai esetében egyébként szinte minden esetben, ahol nem a referenciával azonos helyre szúrta be a kopulát az algoritmus, helyes az általa javasolt megoldás is. (9. táblázat – A méréseket Novák Attila készítette.) A másik kiértékelésnél ("Beszúrás" oszlopok) már csak azokat a zéró kopulákat tekintettük jó találatnak, amelyek a mondat referenciával azonos pozíciójába kerültek. Az eredményeket a 7. és 8. táblázatok tartalmazzák. A táblázat "P" oszlopai a pontosságot, az "R" oszlopok a fedést, az "F1" oszlopok pedig az f-mértéket mutatják.

	Osztályozás			**Beszúrás**		
	P	**R**	**F1**	**P**	**R**	**F1**
eredeti szűrt	95,5	84,2	89,5	89,7	79,1	84,1
bővített szűrt	94,9	78,3	85,8	88,7	73,2	80,2
eredeti javított	93,6	82,8	87,9	85,6	75,7	80,4
szűrt javított	94,1	76,0	84,1	86,4	69,8	77,2

7. táblázat. Az OPUS tesztkorpusz eredményei a különböző modellekkel

Az OPUS tesztkorpusz esetén az osztályozási feladaton mindegyik modell magas pontosságot ért el. A fedés tekintetében érdekes módon a korpusz bővítése jelentős visszaesést eredményezett. Ez arra utalhat, hogy a bővítés során a szűrések ellenére is megnövekedhetett a fals negatív

DOI: 10.15774/PPKE.BTK.2025.020

minták aránya. A beszúrási feladatnál mind pontosságban, mind fedésben az eredeti szűrt modell volt a legjobb, 90, illetve 80%-ot közelítő eredménnyel. Összességében elmondható, hogy ezen a tesztkorpuszon mindegyik modell lényegesen jobban teljesített a Seeker et al. (2012) által mért eredményeknél, még a nehezebb beszúrási feladaton is.

	Osztályozás			**Beszúrás**		
	P	**R**	**F1**	**P**	**R**	**F1**
eredeti szűrt	52,9	29,5	37,8	37,8	21,0	27,0
bővített szűrt	59,6	28,5	38,5	45,1	21,5	29,2
eredeti javított	59,4	25,7	35,9	42,3	18,3	25,6
bővített javított	70,5	29,0	41,1	52,4	21,5	30,5

8. táblázat. A Szeged tesztkorpusz eredményei a különböző modellekkel

Más szövegtípusokon (Szeged Treebank) azonban már egyik modell sem volt ilyen sikeres. Bár az osztályozási feladat pontosságán sokat javított a korpusz bővítése és javítása (részben talán a speciális karakterek visszakerülésének köszönhetően), az eredmény még így is csak 70% körüli. A beszúrást is a bővített javított modell oldotta meg legjobban, 52%-os eredménnyel. A fedés láthatóan mindkét feladatnál és mindegyik modellnél gyenge, a 30%-ot sem éri el. Ennek egyik oka lehet a fentebb említett technikai problémák miatti "veszteség". Másrészt a Szeged tesztkorpusz különböző műfajú szövegei sok olyan mondattípust tartalmaznak, amelyek a modellek számára ismeretlenek, mert a filmfeliratok műfajára egyáltalán nem jellemzők.

	OPUS			**Szeged**		
Modell	**helyes**	**valószínűtlen**	**helytelen**	**helyes**	**valószínűtlen**	**helytelen**
Eredeti szűrt	53.9%	7.7%	38.5%	22.2%	22.2%	55.6%
Bővített szűrt	66.7%	6.7%	26.7%	36.4%	9.1%	54.6%
Eredeti javított	63.6%	4.6%	31.8%	50.0%	6.3%	43.8%
Bővített javított	70.0%	10.0%	20.0%	38.9%	11.1%	50.0%

9. táblázat. Helyes-e a zéró kopula modell által javasolt helye, ahol az nem egyezik a referenciával?

DOI: 10.15774/PPKE.BTK.2025.020

3.3.3. Hibaelemzés

A Szeged tesztkorpusz hibáinak egy részét a már említett technikai problémák okozták. Egyrészt a hosszú mondatoknál előfordult, hogy a fordítás során a mondat egy része egyszerűen elveszett, ami nagy mértékben rontotta a fedést. Másrészt problémásak voltak a tanítóanyagban csak ritkán előforduló karakterek, ezek időnként, a hosszú mondatokhoz hasonlóan, hiányos outputot eredményeztek, más esetekben a fordító lecserélte a problémás karaktereket más, nem odaillőekre. A hibák arányát az egyes modelleknél a 10. táblázat tartalmazza. (A hibaelemzést Novák Attila készítette.) A táblázatban jelölt értékek csak a Szeged tesztkorpuszra vonatkoznak, az OPUS esetén egyáltalán nem fordultak elő ilyen hibák. Mint látható, a modelleken végzett javítások eredményesen oldották meg a szimbólumcsere problémáját, és a hiányos outputok száma is jelentősen csökkent.

Modell	**Szimbólumcsere**	**Hiányos output**
eredeti szűrt	6,80%	10,15%
bővített szűrt	8,25%	6,85%
eredeti javított	0,0%	3,80%
bővített javított	0,0%	3,75%

10. táblázat. A hibás outputok aránya a Szeged tesztkorpuszon

A hibásan beillesztett zéró kopulák áttekintésekor feltűnt néhány mondattípus, amelyeket a program látszólag következetesen (vagy legalábbis gyakran) elrontott. Ilyenek voltak például az egyszavas mondatok, ezekbe mindegyik modell hajlamos volt hibás zéró kopulákat beszúrni: **Mióta Ø?; *Elnézést Ø!; *Értem Ø.* stb. Ez érthető azoknál a modelleknél, ahol a tanítóanyagból kiszűrtük az egyszavas mondatokat, de ugyanezeket az outputokat kaptuk az egyszavas szűrő kikapcsolása után is. Úgy tűnik tehát, hogy a rendszernek mindenképpen nehézséget okoz a nagyon rövid mondatok kezelése. Ennek részben az is oka lehet, hogy a tanítóanyagban sok olyan egyszavas mondat szerepel, ami valóban zéró kopulás (*Jó.; Olcsó.;* stb.).

Hasonlóan általános hibajelenségnek bizonyultak a vonatkozó névmással vagy *mint* kötőszóval összekapcsolt összetett mondatok is (24a és 24b). Ennek valószínűleg a zéró kopulák hasonló helyzetben való gyakorisága lehet az oka. Ugyanígy jellemző hiba, hogy a rendszerek minden esetben zéró kopulát szúrnak be az *ez a(z)...* típusú szerkezetekbe, feltehetően szintén azért, mert ezek

DOI: 10.15774/PPKE.BTK.2025.020

gyakran valóban zéró kopulásak (24c). A gyakoriság problémakörébe tartozik még a *tagadószó + NP* (24d) és a birtokos szerkezetek (24e) hibás zéró kopulával való ellátása is.

(24) a. *Hallottam, hogy megtiltottad a belépést a szentélybe, és a boszorkány Ø, aki magához vette a herceget, eltűnt.

b. *Erzsi híreinek valóságtartalma Ø, mint valami méreg, szívódott fel a szervezetébe.

c. *...mert ez Ø az út a hegy másik felére vezetett el...

d. *A mű nem Ø üzletszerű többszörözése és terjesztése a szabad felhasználás körébe tartozik.

e. *Kedden a tokiói tőzsde vezető részvényindexe Ø, a Nikkei 225 mintegy 280 pont, azaz 2,7 százalékos erősödést jelzett.

Más forrású lehet, de szintén nagyon gyakori hiba a halmozott jelzős szerkezetek hibás értelmezése is (25). Ezekbe mindegyik modell hajlamos az első jelző után zéró kopulát beilleszteni. Ennek az a valószínű oka, hogy az OPUS korpuszra az ilyen leírások nem jellemzőek, így a rendszernek nem volt esélye ezek helyes kezelését megtanulni.

(25) a. *A bennszülött bivalyszerű Ø, fekete nyakizmai kidagadtak az erőfeszítéstől.

b. *A fordulat a második Ø, 1994-es választás után következett be.

Az előbbihez valamelyest hasonló az a hibatípus, amelyet akár pszicholingvisztikai motiváltságúnak is nevezhetnénk, ezek a mondatok ugyanis a szekvenciális feldolgozás egy pontján (vagy csak egy adott tagmondatot tekintve) valóban zéró kopulásnak tűnhetnek. Tipikusan ilyenek az ellipszisek és a koordinált névszói állítmányok (26a-26c).

(26) a. *És mi tudna ezen változtatni,... ha nem egy újabb vihar Ø.

b. *Nem szeretnék valami hatalmas villában vagy panellakásban élni, mert az előbbi túl nagy Ø, az utóbbi túl kicsi lenne a család számára.

c. *A Pénzügyi Szervezetek Állami Felügyelete engedélyezte, hogy Szobonya Csaba Zoltán, az OTP Lakástakarékpénztár Rt. igazgatósági tagja Ø, egyben vezérigazgatója legyen.

DOI: 10.15774/PPKE.BTK.2025.020

Végül meg kell említeni a csak a bővített modellekre jellemző hibajelenségeket. Ezekben előfordult a megszólítások zéró kopulás értelmezése, ami az eredeti modelleknél nem merült fel. Továbbá ezeknél a modelleknél találtuk a legtöbb "indokolatlan" hibát is, a rendszer hajlamos volt akár ragozott igék után is zéró kopulát beilleszteni. Ez arra utal, hogy a több iterációs tanítóanyaggyártásnál fennáll a hibák terjedésének és halmozódásának veszélye.

(27) a. *A kiképzésetek még nem ért véget, ifjú barátok Ø.

b. *Először sokat ittak Ø, aztán csókolóztak...

3.3.4. Összegzés

A kutatás során létrehoztunk egy eszközt, amely a zéró kopulás mondatok automatikus felismerésére és a zéró kopulák mondatba való beillesztésére alkalmas. A tanítóanyaggal megyegyező forrásból származó tesztkorpusz esetén az eszköz közel 90%-os pontossággal tudta megfelelő helyre beilleszteni a zéró kopulát. A kutatást kiterjesztettük a Szeged Treebankre, amely a főleg egyszerű beszélt nyelvi szövegekből álló tanítóanyagunktól lényegesen különböző, és jóval bonyolultabb szerkezeteket tartalmazó jogi, irodalmi, illetve sajtószövegeket is tartalmaz. Ennek következtében ezen a korpuszon jóval gyengébb teljesítményt mértünk.

Bár az eredmények még nem elég meggyőzőek ahhoz, hogy az eszközt korpuszépítésre használjuk, a modellek előállítása mégis érdekes tanulságokkal szolgált. Bemutatta például a párhuzamos korpuszok egynyelvű adatgyűjtésre való alkalmasságát. Az osztályozó algoritmus lexikális kivételei arra világítanak rá, hogy a mondat felszíni szerkezete nem mindig utal egyértelműen arra, hogy elhagyható-e belőle a létige az alapértelmezett esetben. Ezen kívül a beszúró eszköz hibáiból azt is láttuk, hogy a zéró kopuláknak vannak bizonyos "gyakori környezetei", amikre a rendszer túltanult. A kutatás további, a kopulás szórendjével kapcsolatos tanulságaira az 5.1. fejezetben térek vissza.

DOI: 10.15774/PPKE.BTK.2025.020

4. Zéró létigék a korpuszokban

A fejezet a nem nominális mondatok "zéró létigéit" vizsgálja a főbb magyar nyelvű elemzett korpuszok (az MNSz2 és a Szeged Treebank) adatain. A cél annak feltérképezése, hogy a leíró nyelvtanokból ismert zéró kopulán kívül milyen esetekben lehetséges még a létige elhagyása.

Bár a 3.2. fejezetben használt bináris (kopula – lexikális VAN) osztályozás a számítógépes szövegfeldolgozás szempontjából kétségtelenül előnyös, fontos látni, hogy a létige elhagyhatóságának kérdése a magyarban valójában nem bináris kérdés. Nem arra kell tehát választ keresni, hogy elhagyható-e egy bizonyos létige, hanem arra, hogy mennyire valószínű az elhagyása.

Megfigyelhető, hogy a létige-zéró váltakozás a magyarban nem korlátozódik csupán a nominatívuszi névszói predikátumokra. Az ige hiánya lokatív állítmányú mondatokra is jellemző, például: *Hol a könyved?, Kulcs a lábtörlő alatt*. A fejezetben ismertetett kutatások azt vizsgálják, hogy a "hagyományos értelemben vett" zéró kopulás mondatokon kívül milyen mondatokra és milyen mértékben jellemző még a létige elhagyása.

A zéró létigék kezdeti feltérképezésére a Szeged Treebanket használtam, ebben ugyanis minden nem hangzó funkcionális elem jelölt. A zéró kopulás mondatokban az üres VAN fejhez PRED éllel kapcsolódik az állítmányi névszó, így az egyéb zéró létigék kereséséhez azokra a mondatokra kellett szűrni, amelyekben van üres VAN csomópont, de ehhez nem tartozik PRED él. Az automatikus kereséssel kapott találatokat kézzel is át kellett nézni, mert sok esetben nem az okozta a PRED él hiányát, hogy a mondat nem volt nominális. Az egyik probléma az volt, hogy ha az alany nem szerepel a mondatban, a Szeged Treebank annotációja sokszor az állítmányi névszót jelöli annak (28a). Más esetekben a létige azért hiányzik a nem nominális mondatrészből, mert egy másik (testes) létigével van koordinálva (28b). Előfordult az is, hogy bár az annotáció zéró létigeként címkézte az üres csomópontot, valójában azonban nem egyértelmű, hogy a hiányzó elem létige (28c). Ehhez hasonló probléma, hogy a Szeged Treebank zéró létigét jelöl a *mint*tel bevezetett hasonlító mellékmondatokban is (28d). Végül szűrni kellett a hiányos (28e) és hibás (28f) mondatokat. Az áttekintés után 227 mondat maradt, amelyeknél egyértelműsíthető, hogy valamelyik létige (*van, lesz, nincs*) valamilyen alakja hiányzik a mondatból. További 7 esetben a segédigei *lehet* elhagyásával találkozunk, ezzel a későbbiekben külön alfejezetben foglalkozom.

DOI: 10.15774/PPKE.BTK.2025.020

(28) a. Jó tényező Ø például, hogy sok haverja lehet az embernek, számtalan kapcsolatot alakíthat ki.

b. Péntek volt és nagy hőség Ø már korán reggel.

c. Talán csak azért Ø, mert a gondolataimmal messze jártam.

d. Bent az ember olyan kicsinek érzi magát, mint egy kisegér Ø az elefánt mellett.

e. Erzsi már nem is egészen Ø...

f. *Sokat szurkoltunk és a meccs végeredménye 4:1 Ø arányban a bikák nyerték.

A kapott zéró létigés mondatokat formai és szemantikai tulajdonságaik szerint a következő csoportokba osztottam (gyakoriság szerinti sorrendben):

1. itt/ott/hol
2. leírás
3. gyakori kifejezések
4. semmi/se/sehol/mi
5. eseményleltár
6. lokatív mondatok
7. időre utaló mondatok
8. ritka kifejezések
9. passzív szerkezet

Az **itt/ott/hol** típus olyan lokatív vagy egzisztenciális mondatokat jelöl, amelyek tartalmaznak valamilyen helyre utaló demonstratív elemet (*itt,ott,hol*). Ezek közül az *itt* volt a leggyakoribb 25 időre (29a) és 19 helyre (29b) utaló példával. Az *ott* 15 (29c), a *hol* 14 mondatban szerepelt (29d).

(29) a. De nemsokára itt a nyár és a vakáció.

b. Itt a bőrtok, a fésű és a plakett.

c. Most se félek, csak magának ott az a rengeteg esze...

d. Hol az az izé, az a tetem?

DOI: 10.15774/PPKE.BTK.2025.020

A **leírás** kategóriába jellemzően egzisztenciális mondatok tartoznak, amelyek valamilyen helyszínt vagy személyt írnak le nominális stílusban (30). Ez főleg az irodalmi szövegekre jellemző, de az iskolai fogalmazásokban is előfordul.

(30) a. A hátán zöld műanyag kabát, a feje tetején toll forma kopaszodás.

b. Odabent piros dzsúdóruhás személyzet, kedvesen invitálnak ijesztően amatőr pincérek.

A **gyakori testes létige nélküli kifejezések** közé tartozik a *vége* 14, a *rendben* 4, a *hiába* és a valamennyi ideje (például *éve*) 3-3, az *ideje*, a *miért* és az *ahány-annyi* 2-2, és a *tele, ezért, rajta a sor, késő, láthatáron, ahol-ott* 1-1 példával (31).

(31) a. - A verekedésnek ezennel vége! - jelentette ki határozottan.

b. Minderről azonban a kisbefektetők nem tudhattak, kívülről úgy tűnt, minden rendben a bank körül.

c. Mert hát hiába az erő és gyorsaság, amivel a busz után iramodtam, a technika csodájával én sem vehetem fel a versenyt.

d. Annak már néhány éve, azt hiszem, hogy itt volt Gubbióban.

e. Majdnem reggel volt már, mikor úgy gondoltuk, ideje aludnunk egy kicsit.

f. Miért ez a hirtelen elhatározás?

g. Ellenben felvetődik a kérdés, hogy számomra mit jelent a jó pap, ugyanis ahány ember, annyi elképzelés.

h. a város tele boltokkal és játéktermekkel.

i. Ezért ez a komédia az olasz lánnyal, be kell várni a végét.

j. Hamarosan, nagyon hamar, talán öt perc múlva, talán már most, azt fogja jelenteni a csizmák trappolása, hogy rajta a sor.

k. Persze már késő mert az ember mindig utólag okos!

l. Újabb autószerelő a láthatáron.

m. Az SPÖ szerint azért is fontos az osztrák állami vállalatok központját Ausztriában tartani, mert ahol a centrum, ott folyik a kutatás-fejlesztés

DOI: 10.15774/PPKE.BTK.2025.020

A **semmi/se/sehol/mi** szavakkal jellemzett csoportba többnyire a *nincs* létigét elhagyó mondatok tartoznak (a *mi* kivételével). Ezek formálisan egzisztenciális vagy birtokos mondatok. A *semmi* 20 (32a), a *mi* 7 (32b), a *sehol* 4 (32c), a *se-se* pedig 2 példában (32d) fordult elő.

(32) a. A sietségre azonban semmi ok.

b. Mi dolgod erre?

c. Végre elértem célomat: bent voltam a teremben és a tanárnő még sehol.

d. A Honolulu-Star reggel indul, és se fűtő, se pincér, mert maga leütötte őket!

Az **eseményleltár** elnevezés azokra a nominális stílusú mondatokra utal, ahol a megnyilatkozó leltárszerű felsorolással meséli el a történéseket vagy a napirendjét (33). Ez leginkább az iskolások fogalmazásaira volt jellemző.

(33) Délelőtt iskola, dolgozat, felelés, a hatodik óra után rohanás a buszhoz, mivel kollégista vagyok.

A létige elhagyása az itt/ott/hol csoportba nem tartozó **lokatív mondatok** (34) esetén 12-szer fordult elő. Különösen érdekes (34b), ahol az *-e* kérdő partikula a lokatív kifejezéshez kapcsolódik. Olyan példát is találunk ebben a csoportban, ahol a mondat csak formálisan lokatív, valójában azonban időre utal (34c).

(34) a. Nem látja, hogy kezemben a cipőm?!

b. Hogyha baj van, kissé mozgatja a vállát, és akkor a hóna aljával megérzi, hogy helyén-e az ólmos.

c. Már nagyon közel a megegyezés a hitelező pénzintézetekkel a rövid lejáratú hitelek átütemezéséről.

Az **időre utaló kifejezések** jellemzője, hogy a létige hiányában általában jövőre utalnak, azaz itt valójában nem a *van* hiányzik a mondatokból, hanem a *lesz*. Ezt mindig valamilyen jövő időre utaló kifejezés (*mindjárt, hamarosan*) erősíti meg, ami nélkül nem is lehetséges a létige elhagyása (35).

DOI: 10.15774/PPKE.BTK.2025.020

(35) Mindjárt éjfél, nekem a templomba kell mennem.

Két esetben fordult elő a létige nélküli szerkesztésmód határozói igenevekkel, úgynevezett **passzív mondatok**ban (36). A fejezetben látott példák közül ez a típus áll legközelebb a nominális mondatokhoz, hiszen az ilyen típusú mondatok létigéjét a MGr. is segédigének tartja.

(36) Belülről óriási helyiség és az egész kivilágítva gyertyákkal, egyszerűen gyönyörű volt.

Végül, (37) az előbbiek közül sehova sem sorolható, és a létige elhagyása furcsán is hat ebben az esetben. Ennek megfelelően neveztem **ritka zéró létigés kifejezések**nek az általa alkotott "csoportot".

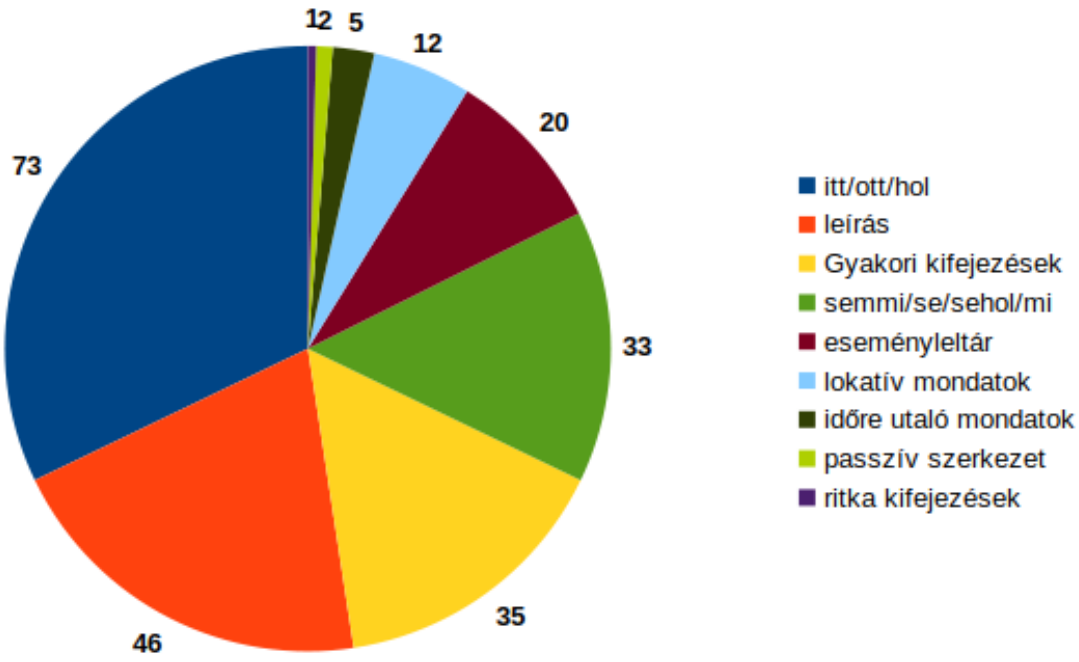
(37) a. A másikkal a szélében horgásztam, és szép piros gilisztával (sügéreket) kb. 15-20 dkg-os sügéreket lehet fogni, ha csendben az ember.

Az egyes mondattípusok gyakoriságát a 3. ábra szemlélteti. Mindebből azt láthatjuk, hogy a nem nominális mondatok létigéjének elhagyása nem ritka jelenség. A Szeged Treebankből kapott példák alapján a továbbiakban a különböző zéró létigék produktívnak tűnő, nem műfajspecifikus (szépirodalmi) eseteit vizsgáltam meg részletesebben. A 4.1. fejezet a létige azon predikatív bővítményeit járja körül, amelyek előfordulnak a korpuszban testes és zéró *van* létigével is. A 4.2. fejezet a *van* létige elhagyásával foglalkozik a helyre utaló demonstratív elemet (*itt/ott/hol*) tartalmazó mondatokban. A 4.3. fejezet a tagadó formájú létigék (*nincs, sincs*) elhagyhatóságát vizsgálja. Végül a 4.4. fejezet kitér a segédigei *lehet* elhagyására.

4.1. Ingadozó bővítmények

A hagyományos leíró nyelvtanból ismert, hogy a predikatív szerepet betöltő, ragtalan névszók (vagy névszói csoportok) mellől minden esetben elhagyjuk a létigét az alapértelmezett esetben. Azonban láttuk, hogy vannak ezen kívül opcionálisan elhagyható létigék is. Ezek egyik csoportját "gyakori testes létige nélküli kifejezések"-nek neveztem, úgy tűnik ugyanis, hogy a létige bizonyos bővítményei megengedőbbek a létige elhagyását tekintve. A leggyakoribb ide sorolt kifejezések (*vége, rendben*) valamilyen helyzetre utalnak, így ezekben a *van* valóban kopulaszerűnek tekinthető. A kopulaszerű *van* további előfordulásainak vizsgálatához az előző keresés fordítottjára van

DOI: 10.15774/PPKE.BTK.2025.020



3. ábra. A zéró létigét tartalmazó nem nominális mondatok gyakorisága a Szeged Treebankben

szükség: amíg a fejezet bevezetőjében olyan zéró létigéket kerestem, amelyeknek vannak nem predikatív bővítményei, addig ez a fejezet éppen azokat a predikatív bővítményeket keresi, amelyek előfordulnak testes *van*-nal is.

A Szeged Treebank függőségi elemzési annotációját felhasználva tehát azokat a szavakat gyűjtöttem össze, amelyek előfordulnak PRED relációban zéró kopulával (üres igei fejjel) és testes *van, vannak* jelen idejű, 3. személyű igealakokkal is. A korpuszban 54 különböző PRED éllel jelölt névszó fordult elő a *van, vannak* igealakokkal, ami összesen 111 előfordulást jelent. A 11. táblázat azokat a predikatív névszókat tartalmazza, amelyek legalább háromszor előfordultak testes kopulával. A táblázat "Testes ige" oszlopa jelzi, hányszor fordult elő az adott névszó a *van, vannak* igealakokkal, míg a "Zéró ige" oszlop az üres igei fejjel való előfordulás gyakoriságát mutatja. A "Példa" oszlop a testes kopulás esetre ad példát. A táblázatba nem kerültek bele azok a példák, ahol a PRED éllel jelölt elem nehezen értelmezhető a létige predikatív vonazatként. Ilyenek voltak az *olyan* és a *szó*. Az *olyan* PRED-nek van jelölve a (38a)-típusú egzisztenciális mondatokban, ez azonban nyilvánvalóan nem a predikatív *olyan* testes létigés változata. A *szó* esetén a *szó van valamiről* kifejezések (38b) esetén kapott a *szó* predikatív címkét, ám ez sem sorolható egyértelműen

DOI: 10.15774/PPKE.BTK.2025.020

a testes–zéró váltakozás témakörébe (bár közelít a környezeti kopulás szerkezetekhez).

(38) a. **Vannak** viszont **olyan** tanárok, akik pont az ellenkezőjét csinálják.

b. ugyanis egy általa írt darabról **van szó**

Névszó	Testes ige	Zéró ige	Példa
késő	9	7	*Nagyon **késő van** már.*
tele	7	2	*Az élet **tele van** megokolatlan koincidenciákkal.*
meleg	6	3	*S mindenekelőtt: ott **meleg van**.*
több	5	50	*És vérrokon **több van**, mint egypetéjű ikertestvér.*
óra	4	8	*Öt **óra van**, most már menni kell.*
sok	3	51	*Hát ilyen elég **sok van**.*
kevés	3	25	*Rendőr pedig azért **van kevés**, mert aki teheti, jobban fizető helyet keres.*
hideg	3	4	*— Öltözzenek melegen, mert bent nagyon **hideg van**.*
szombat	3	1	***Szombat van**, ma van a 50. házassági évfordulójuk mamáéknak.*

11. táblázat. Testes és zéró kopulával is előforduló predikatív névszók a Szeged Treebankben

A 11. táblázatban látható névszók némelyike (*késő, meleg, hideg*) egyértelműen a Kádár (2004) által vizsgált "környezeti kopulás szerkezetek" közé sorolhatók. Amennyiben ezek valóban időjárásra vagy valamilyen "környezeti helyzet" jellemzésére utalnak, a mondat formálisan egzisztenciális lesz, azaz kötelezően testes létigével bír, és az alany névelőtlen. A Szeged Treebank annotációja azonban jogosan jelöli ezeket a névszókat predikatív relációval, hiszen valójában egy nem meghatározott, de odaérthető alanyról tesznek állítást. Amikor ez az alany explicit, a mondat zéró kopulás lesz. (39) és (40) példái szintén a Szeged Treebankből származnak.

(39) a. Az idő egy kicsit **meleg**, de megéri a látványért.

b. Nagyapám arra hivatkozott, hogy nagyon **hideg** a víz.

(40) a. Mert előre közölte a magyar tanárunk, hogy erre tanulni már **késő**, aki eddig sem tanulta meg az anyagot, az már úgysem fogja tudni megtanulni.

b. Szent Isten, milyen **késő** van!

DOI: 10.15774/PPKE.BTK.2025.020

(40) esete különösen érdekes, mert itt jelentéskülönbség is megfigyelhető a testes igés és a zéró kopulás szerkesztésmód között. A testes igével bíró *késő van* szerkezet általában a napszakra vonatkozik: már kései óra, este van. Ezzel szemben a zéró kopulás *késő* arra utal, hogy valaminek az ideje már elmúlt, elkéstünk vele, és emiatt általában infinitívuszi vagy igéből képzett névszói alannyal jár együtt. (Előfordul a korpuszban egy kivételes példa, ahol ez utóbbi jelentésben jelenik meg a testes *van*: *Az igazság az, hogy a Xerox sohasem fog felzárkózni, ehhez túl késő van*.)

A gyakorisági listában második helyen álló *tele* esetén azonban semmilyen jelentéskülönbség nincs a testes *van*-os és a zérós szerkesztés között, ez esetben egyszerűen ingadozik a *van* jelenléte és elhagyása (41). A korpuszban ritkább volt, ezért nem került a táblázatba, de hasonló ehhez a *kész* helyzete is (42). Ezekben az esetekben az lehet a váltakozás oka, hogy a *tele* és a *kész* szófaját tekintve is többértelmű: ha névszónak tekintjük őket, akkor a nominális, azaz zéró kopulás mondatszerkesztés a "szabályos", ha azonban határozószóként értelmezzük őket, akkor az egyéb határozós mondatokhoz hasonlóan kaphatnak testes létigét.

(41) a. Az élet **tele** van megokolatlan koincidenciákkal.

b. Bár a felnőttek világa elég borzasztó, **tele** gondokkal, küszködésekkel és persze ellentétekkel.

(42) a. — **Kész** van az államtanácshoz?

b. Úgy gondolja, hogy másképp nem menekülhet meg, és igenis **kész** rá, hogy így meneküljön meg.

A mennyiséget jelentő kifejezéseknél (*sok, több, kevés*) váltakozik az egisztenciális és a nominális mondatokra jellemző szerkesztésmód. Mivel ezek a kifejezések a névszók közé sorolhatók, alkalmasak a névszói állítmányi szerep betöltésére, ekkor az alany kötelezően specifikus, és nem jelenik meg hangzó kopula az alapértelmezett esetben (43). A funkciójukat tekintve azonban ezek a mondatok valaminek a (valamilyen mértékű, mennyiségű) létezését fejezik ki, ezért előfordulhatnak egzisztenciális mondatokban, puszta főnévi alannyal. Ilyenek voltak a 11. táblázat példái.

DOI: 10.15774/PPKE.BTK.2025.020

(43) a. Szerintem az iskolának sokkal **több** a pozitív tulajdonsága, mint a negatív, és vajon mennyivel jobb dolgozni, mint tanulni?

b. **Sok** az összeütközés közöttünk, de azt hiszem jó osztály vagyunk.

c. De szerintem ennek ellenére **kevés** az őszinte ember.

Érdekes az aktuális időre vonatkozó kifejezések (*óra, szombat*) esete. Ezek a környezeti kopulás szerkezetekhez hasonlóan általában testes létigével állnak együtt. Akadnak azonban zérós példák is (44). (Az *órá*val több hasonló is található a korpuszban, a *szombat* egyetlen zéró kopulás példája azonban más jelentésű, nem aktuális időre vonatkozik: *A második nap szombat.*) Ez azért különleges, mert itt nem a *van* váltakozik zéróval, hanem a *lesz*: a korpusz mindegyik (44)-hez hasonló példája jövő időre utal.

(44) a. Ebben a pillanatban O'Brien a karórájára pillantott, látta, hogy mindjárt tizenegy **óra**, s nyilvánvalóan úgy döntött, hogy az irattári osztályon marad, amíg a Két Perc Gyűlölet tart.

A 11. táblázatban nem szereplő, ritkább példák közül még a *rendben* és a *vége* esetét érdemes említeni, ezek ugyanis a zéró létigék nem predikatív bővítményeiként is megtalálhatók a korpuszban. (Ez arra utal, hogy a Szeged Treebank nem következetes ezek annotációjában: néhol `PRED`, máshol `OBL` címkét kapnak.) A *rendben* előfordul a Szeged Treebankben testes létigés (45a-b) és zéró kopulás (45c) mondatban is, a zéró és a testes létige azonban nem minden példánál cserélhető fel egymással (46) ugyanolyan természetességgel. Az ige nélküli szerkesztés különösen furcsán hat (46b) esetén, ahol a *rendben* predikátum módosítót kapott.

(45) a. De most már **rendben** van, minden **rendben** van, a küzdelemnek vége.

b. Az ördög tudja miért, de úgy látszik, ez nagyon **rendben** van, hogy Mr. Irving parancsoljon az ilyen embereknek, mint amilyen ő.

c. Mire a válaszom az volt, hogy **rendben**.

DOI: 10.15774/PPKE.BTK.2025.020

(46) a. De most már **rendben** ?(van), minden **rendben** (van), a küzdelemnek vége.

b. Az ördög tudja miért, de úgy látszik, ez nagyon **rendben** #(van), hogy Mr. Irving parancsoljon az ilyen embereknek, mint amilyen ő.

c. Mire a válaszom az volt, hogy **rendben** (van).

A *vége van valaminek* kifejezés formálisan birtokos mondat, de már idiomatikussá vált: (47a) és (47b) jelentése nem azonos. Míg (47b) valódi birtokos mondat, (47a) inkább helyzetre utal. A zérós szerkesztésű (48) egyértelműen (47a)-val azonos jelentésű, nem lehet (47b) változataként értelmezni. A *vége van* szerkezet létigéje tehát abban az esetben hagyható el, ha a mondat "itt van a vége" jelentésű. (Az *itt van a vége* mondat esetén egyébként szintén elhagyható a létige.)

(47) a. Ennek a napnak is vége van.

b. Ennek a napnak is van vége.

(48) Ennek a napnak is vége.

Az eddigiekből azt láthattuk tehát, hogy a létige elhagyhatósága nem csak annak funkciójával hozható összefüggésbe, hanem a "predikátumjelölt" szófaji és lexikális tulajdonságaival is. Ez az összefüggés a Szeged Treebank példái alapján a következőkben foglalható össze:

1. **Ragtalan névszók**

   Ha a mondatban ragtalan névszó tölt be predikatív szerepet, akkor a mondat nominális, és jelen idő, egyes szám harmadik személyben nem tartalmaz igei elemet. Ez alól kivételt képeznek az időjárásra vagy más "környezeti helyzetre" utaló névszók, amelyek formálisan nem névszói predikátumot képeznek, hanem egzisztenciális mondat alanyát. Ingadozás figyelhető meg a számneveknél is: ezek mind állítmányi névszóként, mind egzisztenciális mondat alanyát módosító jelzőként előfordulhatnak. Váltakozhat az ige nélküli és igés szerkesztésmód azokban az esetekben is, ahol a predikatív elem szófaji besorolása nem egyértelmű, ilyenek a *tele* és a *kész*.

2. **Határozóragos névszók, határozószók**

   A határozóragos névszók esetén általában szükséges a *van* igei elem jelenléte. Bizonyos, 4. fejezetben látott esetekben a határozós mondatokban is opcionális lehet a létige.

DOI: 10.15774/PPKE.BTK.2025.020

3. **Birtokos névszók**

   Birtokos névszók esetén csak speciális esetekben, elsősorban idiomatikus kifejezésekben megengedett a létige a elhagyása.

4.2. Ott van, tehát van

A Szeged Treebank példáin láttuk, hogy a *van* létige lokatív és egzisztenciális mondatokból is elhagyható lehet bizonyos esetekben. Ilyen jellemző esetek a helyre utaló demonstratív elemet (*itt, ott, hol*) tartalmazó mondatok. Ebben a mondattípusban a létige jelenléte opcionálisnak tűnik. Az alany lehet határozott vagy határozatlan névelős (49), viszont nem lehet névelőtlen (50) vagy többes számú (51), bár ez utóbbira volt egy megkérdőjelezhető példa a korpuszban: *?Időre hazamenni, nem lehet este egy-egy moziba menni, mert ott a gyerekek.*

(49) a. Itt a kulcs. / Ott a kulcs. / Hol a kulcs?

b. Itt egy kulcs. / Ott egy kulcs. / Hol egy kulcs?

(50) a. *Itt hely. / *Ott élet. / *Hol szabad asztal?

b. *Itt János pénze. / *Ott Péter. / *Hol Punxsutawney?

c. *Itt minden hozzávaló. / *Ott két kutya. / *Hol elég filctoll?

(51) a. *Itt a könyvek. / *Ott a könyvek. / *Hol a könyvek?

b. *Itt állatok./ *Ott állatok. / *Hol állatok?

Az ige nélküli szerkesztésmód az *itt, ott* esetén csak semleges mondatban lehetséges. Fókuszos mondatban (52) és kontrasztív topik esetén (53) kötelező megjelennie az igének. Tagadó mondatban sem lehetséges a létige elhagyása, de ennek az is oka lehet, hogy ezeknek a mondatoknak a tagadása nem a tagadószóval, hanem a *nincs* tagadó létigével történik (54), kivéve ha a tagadás hatóköre maga a helyre utaló névmás. Ám az ige nélküli szerkesztésmód ez utóbbi esetben is szokatlan (55).

(52) *MIÓTA itt a nyomtató? / *A TE MACSKÁD ott.

DOI: 10.15774/PPKE.BTK.2025.020

(53) *A könyv itt/ott, de a füzetemet nem találom.

(54) Nincs itt/ott a gekkó. → *Nem itt/ott a gekkó.

(55) Nem itt/ott van a gekkó. → #Nem itt/ott a gekkó.

Érdekes kérdés lehet még, hogy az *itt/ott* demonstratív elemekhez kapcsolódhat-e *-e* kérdő partikula a vizsgált mondattípusokban (például: *Megkérdeztem, ott-e a könyv.*) Az MNSz-ben az *ott* esetén 13, az *itt* esetén pedig 4 példát találtam erre. Ez arra utal, hogy *-e* kérdő partikulás mondatok esetén a létige elhagyása lehetséges, de nem preferált.

Összefoglalva elmondhatjuk, hogy az *itt/ott/hol* demonstratív elemeket tartalmazó lokatív és egzisztenciális mondatok létigéje, a kopulával ellentétben, csak egyes számban hagyható el. Az ige nélküli szerkesztés ezen kívül csak akkor lehetséges, ha az alany valamilyen determinánst visel (*a, az, egy*), és a mondat semleges, az azonban nem feltétel, hogy az alany határozott névelős legyen (56). Ezen kívül azt is észre lehet venni, hogy (56b) esetén egyáltalán nem szükséges, hogy a demonstratív elem konkrét helyet nevezzen meg. A b. mondat jelentése is inkább egzisztenciális, mint lokatív: a szótár létezik, rendelkezésre áll. (57) példái azt mutatják, hogy az igemódosító *ott* lehetővé teszi az egzisztenciális mondatra jellemző szórendet határozott névelős alany esetén is.

(56) a. Itt egy konnektor.

b. Ha nem értesz valamit, ott a szótár (a polcon).

(57) a. Van egy szótár (a polcon).

b. *Van a szótár a polcon.

c. Ott van a szótár a polcon.

Mindez azt az újabb kérdést veti fel, hogy milyen prozódiai vagy/és szemantikai funkciója van az igemódosító *itt/ott*-nak? Más igék esetén is megfigyelhető a kérdéses demonstratív elemeknek egy olyan sajátos esete, amikor úgy tűnik, hogy ezek rámutató funkciója gyengül, vagy redundánssá válik.

DOI: 10.15774/PPKE.BTK.2025.020

(58) a. Az oroszok már itt vannak a spájzban.

b. A kocsi ott áll a ház előtt.

(58) példáiban az *itt/ott*-on kívül is van tartalmas lokatív bővítmény a mondatokban, de nem teljesen egyértelmű, hogy a demonstratív elemek rájuk mutatnak-e. Elképzelhető, hogy deiktikusak, és a lokatív bővítmény csak specifikálja a jelentésüket. Ugyanakkor (58) mindkét példájában érezhető némi jelentésváltozás az *itt/ott* nélküli változathoz képest:

(59) a. Az oroszok már a spájzban vannak.

b. A kocsi a ház előtt áll.

Amint (59) példái mutatják, a helyre mutatás erősebb, ha a tartalmas lokatív bővítmény megelőzi az igét. Ekkor az állítás valóban az, hogy az oroszok egy adott helyen vannak, illetve a kocsi egy adott helyen áll. Ezzel szemben (58)-ban inkább a „jelen van", „rendelkezésre áll" jelentések tűnnek elsődlegesnek, a helyet meghatározó lokatív bővítmények másodlagosak. Azt gondolhatjuk tehát, hogy az *itt/ott* demonstratív elemek jelentésének (hasonlóan az angolhoz) valamilyen köze van a létezéshez.

Mindezzel kapcsolatban két kérdést tehetünk fel:

1. Milyen szemantikai funkciót tölt be az *itt* és az *ott* (58) példáiban látott, közvetlenül az igét megelőző pozícióban?

2. Miért hagyható el a létige (56) példáihoz hasonló lokatív és egzisztenciális mondatokban?

A kérdésekre empirikus vizsgálattal, korpuszadatok elemzésével keresek választ. A fejezetben amellett fogok érvelni, hogy az *itt/ott* a vizsgált esetekben igemódosító (azaz igekötőszerű) szerepben van, és az igéhez adott jelentéstartalma az, hogy „a cselekvés/esemény közben az alany egy adott helyen jelen van, tehát létezik". Amennyiben (egzisztenciális) létigét előz meg, az ige szemantikájának módosításával lehetővé teszi, hogy az egyébként úgynevezett definiteness effect igének specifikus alanya legyen. Másrészt mivel az *itt/ott maga* is hordoz létezés jelentést, a létige tartalmassága gyengül, és a kopulához hasonló inflexióhordozóvá válik. Emiatt jelöletlen esetben lehet zéró hangalakú.

DOI: 10.15774/PPKE.BTK.2025.020

4.2.1. Az *ott* mint igekötő a szakirodalomban

A vizsgált demonstratív elemek szintaktikai státuszával és jelentéstartalmával kapcsolatos kérdésekre már mások is felfigyeltek. Elsőként érdemes megemlíteni, hogy egyes szófajokkal vagy helyesírással foglalkozó művekben az *ott* mint igekötő is szóba jön. Komlósy (1992) összefoglaló táblázata alapján Temesi et al. (1955) egyértelműen igekötőnek minősíti az *ott*-ot, továbbá A magyar nyelv értelmező szótára is megemlíti, hogy bizonyos igék felvehetik igekötőként. A táblázatban felsorolt 7 műből azonban 4 határozottan nem sorolja az *ott*-ot az igekötők közé. (Az *itt* egyáltalán nem merül fel a kérdéskörben.)

A preverbális *ott* státuszával először Deme (1959) foglalkozott részletesen. Empirikus kutatásában az *ott* olyan előfordulásait kereste, amelyek nem annyira rámutatások, határozatlanok, indokolatlannak tűnnek. (60) az ő példája.

(60) A bérkocsi ott áll a ház előtt.

Megfigyelése szerint az *ott* leginkább tárgyatlan igéket előz meg, közülük is leggyakrabban a létigét vagy egyéb helyzetet, tartózkodást jelentő igéket (*ül, áll, fekszik, lakik* stb.). A jelenség okát a prozódiában látja, szerinte ugyanis a preverbális *ott* azért kell ezekben az esetekben, mert másképp nem oldható meg, hogy nyomatéktalan legyen a mondat. Az igekötőszerű *ott* lehetővé teszi, hogy az állítmány egyenes szórendben állhasson, és némileg hangsúlyos legyen, illetve hogy más mondatrész ne kaphasson hangsúlyt. (61b) nehezen fogadható el semleges mondatként.

(61) a. A bérkocsi ott áll a ház előtt.

b. %A bérkocsi áll a ház előtt.

Ez a megfigyelés az igék hangsúlykerülő tulajdonságával függhet össze. (61a)-ban a preverbális *ott*-nak köszönhetően maradhat hangsúlytalan az ige. Kálmán (2001) ezt a funkciót igevivőnek nevezi, és Deme (1959)-hez hasonlóan úgy véli, hogy a hangsúlykerülő igék azért hajlamosak egy igekötőként viselkedő névmást felvenni, hogy elkerüljék a hangsúlyosságot. (Kálmán 2001: 56-57.) Komlósy (1992) megfogalmazásában a hangsúlykerülés „elsősorban a 'valahol levés' jelentéselemét tartalmazó" igékre jellemző (Komlósy 1992:316). Ennek megfelelően a Deme (1959) által említett igék szemantikája is a hangsúlykerülésre utal.

DOI: 10.15774/PPKE.BTK.2025.020

Deme (1959) tehát egyfajta nem lexikalizálódott igekötőnek tekinti a preverbális *ott*-ot. A kérdéskörben még több fontos, de kevéssé magyarázott észrevétele is volt, ezekről majd a fejezet későbbi pontjain teszek említést.

Deme (1959) példájából kiindulva É. Kiss (2004) is egy igekötő-elmélet kereteibe illeszti bele az *ott* + ige kapcsolatok magyarázatát. Szerinte a magyarban lehetséges egy létezést vagy térbeli helyzetet kifejező elsődleges predikátum és egy helyet kifejező másodlagos predikátum egy tagmondatba való sűrítése (62).

(62) a. A bérkocsi ott áll a ház előtt.

b. A bérkocsi a ház előtt áll.

c. A bérkocsi ott áll.

„Az (a) mondatváltozatban egy hol kérdésre felelő proadverbium szerepel lokatív másodlagos predikátumként, mely fogalmi egységet alkot egy ige utáni tartalmas lokatív főnévi kifejezéssel. A (b) változatokban a tartalmas lokatív kifejezés került a másodlagos predikátum helyére. A (c) változatokban a proadverbium egyedül képviseli a lokatív predikátumot." (É. Kiss 2004: 38.) A (62a)-féle lokatív másodlagos predikátum főleg a létigére vagy egyéb térbeli helyzetet jelentő igékre jellemző. Ezek az igék kétértelműnek tekinthetők: van egy agentív, „adott pozíciót (szándékosan) fenntart", és egy non-agentív, létezést kifejező jelentésük. A preverbális ott az utóbbi jelentést hívja elő, hatására az ige agentív jelentéstartalma háttérbe szorul, és az ige a „van", illetve a „meghatározott térbeli helyzetet foglal el" sajátos jelentéselemekkel dúsított szinonimájává válik.

É. Kiss (2004) elemzése szerint tehát az *ott* „olyan lokatív másodlagos predikátum, melynek alanya a létezést vagy térbeli helyzetet kifejező elsődleges predikátum jellemzett szerepű belső argumentuma." (É. Kiss 2004: 39)

Surányi (2005) szintén igemódosítónak tekinti a preverbális *ott*-ot, szerinte azonban az az ige utáni lokatív bővítménnyel egy argumentumot képez, és jegymozgatással kerül az igemódosító pozícióba. Az igemódosító *ott* szemantikailag inkorporálódik az igébe, és egy „adott helyen végzi a cselekvést" jelentéstartalommal bővíti azt. A létige kapcsán megemlíti, hogy bár ez hagyományosan definiteness effect ige, az *ott*-tal módosítva mégis megenged specifikus bővítményt, hiszen a szemantikája megváltozott. (l. (57) példái)

Az ige előtti "redundáns" *ott* kérdéskörével részletesen foglalkozik Kocsány (2021). Szerinte

DOI: 10.15774/PPKE.BTK.2025.020

a preverbális *ott* akkor tekinthető redundánsnak, ha a mondatban a nominális helymeghatározás is jelen van (pl. (62a)). A jelenség magyarázatához É. Kiss (2004) lokális igekötő fogalmából indul ki. Az ige előtti *ott*-ot mint *in statu nacendi* igekötőt írja le. Azoknál az igéknél, ahol az *ott* igekötő már lexikalizálódott (pl. *otthagy, ottfelejt, ottmarad, ottvész, ottfog, ott talál, ottreked, ottragad, ott terem, ott pusztul*) egyszerre fejezi ki az esemény kezdőpontját és a valahol levést. Az *ott* igekötő lexikalizálódásának egyik fő ismérve (az egybeírás mellett) az, hogy akkor is megmarad(hat) a mondatban az *ott*, amikor a nominális helyhatározó fókuszba kerül (63a). Ez a *van*-ra és a többi helymeghatározást kifejező hangsúlykerülő igére nem igaz (63b), Kocsány (2021) szerint azonban ezeknél is megfigyelhető az igekötővé válás folyamata, méghozzá elsősorban éppen a *van* esetén. A *van* előtti *ott* lokális jelentése halványul, az ige jelentése pedig "megvan, adva van, előfordul" jelentéssé módosul. Ha megjelenik a mondatban nominális helymeghatározás, akkor az *ott* szórendileg az igekötő szabályait követi. Összefoglalva, az *ott* mint igekötő eseményszerkezetét Kocsány (2021) egy kezdőpont mint határpont, ezzel együtt impliciten egy esemény bekövetkezése, egy helyzet beállta, ÉS maga a helyhez kötött folyamat/fennmaradás nyomatékosításaként írja le.

(63) a. Az ASZTALON hagyta (ott) a kulcsot.

b. Az ASZTALON van (*ott) a kulcs.

Az eddigi megközelítések tehát egyrészt a prozódiában, másrészt a szemantikában keresik az igemódosító *ott* magyarázatát. Az, hogy a preverbális *ott* jelenléte nem csak prozódiai eredetű, azzal támasztható alá, hogy olyankor is megjelenik, amikor a hangsúlykerülés miatt ez nem lenne indokolt. (64a) és (65a) korpuszbeli (MNSz2) példák az igemódosító nélkül is grammatikusak, ám a jelentések nem teljesen azonosak az a, és b, változatok esetén.

(64) a. 3,6 millió gépen volt ott a vírus.

b. 3,6 millió gépen volt a vírus.

(65) a. Timóteust nem találták ott azon a vidéken.

b. Timóteust nem találták azon a vidéken.

(64a) csak disztributívan érthető, azaz 3,6 millió gép volt vírusos, (64b) viszont megengedi a kollektív olvasatot is, miszerint a vírusnak akkora memóriaigénye van, hogy 3,6 millió gép kell a

DOI: 10.15774/PPKE.BTK.2025.020

tárolásához. (65a) azt implikálja, hogy Timóteus nem volt azon a vidéken, mikor keresték, (65b) viszont csak azt állítja egyértelműen, hogy nem találták meg, ám ettől még lehetséges, hogy ott volt. Azt mondhatjuk tehát, hogy az igemódosító itt/ott nem csak prozódiai jelenség, hanem szemantikai funkciója is van.

Mindezek alapján a fejezet bevezetőjében is megfogalmazott hipotéziseim a következők:

1. A preverbális *itt/ott* olyan igemódosító, amely az ige eredeti jelentéstartalmát „a cselekvés/esemény közben az alany egy adott helyen jelen van" elemmel egészíti ki.

2. Az általa módosított igék eredeti jelentéstartalma ezáltal gyengül, és a létige valamilyen specifikusabb szinonimáivá válnak.

3. Amennyiben az *itt/ott*-tal módosított létige zéró inflexiót hordoz, a kopulához hasonlóan elhagyható, mert szemantikailag redundáns, a létezés vagy helyben levés jelentését (1. szerint) az igemódosító is tartalmazza.

4.2.2. Korpuszalapú vizsgálatok

Az előbbiekben megfogalmazott hipotéziseket korpusz alapú empirikus vizsgálattal teszteltem. Az adatok forrása az MNSz2 volt, ahol a következő formájú kereséseket végeztem el:

```
[lemma="itt"] [msd="IGE\.\S+[123]"]
```

A keresést természetesen elvégeztem az *ott*-ra, illetve a fordított szórendre is. A hipotézisek szempontjából releváns, és számítógépes programmal mérhető kérdések a következők:

1. Milyen szórend jellemző a tagadó mondatokra?

2. Milyen igéket módosít az *ott* és az *itt*?

3. Melyik igékkel szokás egybeírni az *itt/ott*-ot?

Az 1. kérdés arról adhat információt, hogy igemódosító pozícióban van-e az *itt/ott*. A 2. és 3. kérdés az igemódosító *itt/ott* szemantikáját vizsgálja. Mit adhatnak hozzá ezek az elemek az ige eredeti jelentéséhez? Milyen esetekben látszik az *itt/ott* igekötőként lexikalizálódni?

Az itt megfogalmazott kérdések egyébként nagyon hasonlóak azokhoz, amelyeket Deme (1959) vizsgált, szintén írott szöveges adatokon. A különbség egyrészt, hogy nála mind az adatgyűjtés,

DOI: 10.15774/PPKE.BTK.2025.020

mind azok feldolgozása manuálisan történt, így a vizsgált adathalmaz természetesen jóval kisebb volt. Másrészt, jelen korpusz alapú kutatással szemben, az ő módszere inkább korpuszvezéreltnek mondható, azaz kezdeti hipotézisek nélkül fordult az adatokhoz, és csak azokból próbált következtetéseket levonni. Mindemellett az eredményeinkben is sok hasonlóság látható.

A szórend kérdésével kapcsolatban Deme (1959) azt figyelte meg, hogy a preverbális ott bizonyos esetekben hajlamos a fordított szórendre, például folyamatosság jelzése (*volt ott...*) vagy tagadás (*nem volt ott*) esetén, illetve főhangsúlyos rész, mutatószó vagy kérdőszó mellett (*Úgy ült ott..., Miért volt ott?*). Felsorolásából a tagadó mondatok szórendjét vizsgáltam meg a korpuszadatokon. A mérés következő szerkezetek gyakoriságára terjedt ki:

- tagadószó + *itt/ott* + ige – egyenes szórend
- tagadószó + ige + *itt/ott* – fordított szórend

A vizsgált szórendek általános gyakoirsági eloszlását a 12. táblázat és a 4. ábra mutatja. Látható, hogy a tagadó mondatokban mindkét demonstratív elem jellemzően az igekötőkhöz hasonló módon viselkedik, azaz fordított szórendben áll. Ez alapján úgy tűnik, hogy a vizsgált elemek valóban az igemódosító szerepét töltik be.

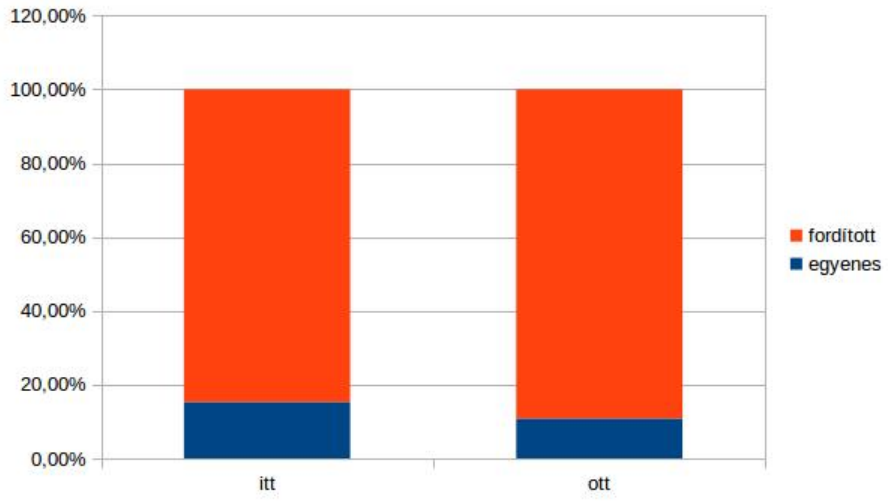
	Egyenes	**Fordított**
Itt	409	2265
Ott	448	3686

12. táblázat. Az *itt/ott* demonstratív elemeket tartalmazó tagadó mondatok szórendje

Az igekötőszerű *itt/ott* szemantikáját az általuk gyakran módosított igéken keresztül vizsgáltam. A korpuszadatokon azoknak az igéknek a gyakoriságát mértem, amelyeket közvetlenül megelőz egy demonstratív elem (*itt* vagy *ott*). Az 5. ábrán megjelenített igék listáját úgy állítottam össze, hogy az *itt* és *ott* külön-külön mért gyakorisági listája szerint összegeztem az igék helyezését, és a 10 legalacsonyabb pontszámú igét vettem fel az ábrába. Az igék az *ott*-tal való előfordulásuk gyakorisága szerint vannak rendezve.

Az adatok alapján a listát egyértelműen a *van* vezeti. A leggyakoribb igék között szerepelnek még más létigék (*lesz, nincs*), illetve „térbeli pozíciót elfoglal” jelentéstartalommal bíró, egyargumentumú igék, mint amilyen az *áll, ül, marad, él*. A tárgyas igék, a *lát*, a *talál*, a *tart* és a *hagy*

DOI: 10.15774/PPKE.BTK.2025.020



4. ábra. Az *itt/ott* demonstratív elemeket tartalmazó tagadó mondatok szórendje

esetén az "adott helyen van" jelentéstartalom a témára vonatkozik, azaz ezek az igék is "valahol levést" eredményeznek (66).

(66) a. Egyszer aztán, nagy későre, ott látták, akit vártak. → Ott volt, akit vártak.

b. Ott találták őket lefejezetten. → Ott voltak lefejezetten.

c. De valami ott tartotta a félhomályos szobában. → Ott maradt/ott volt a szobában.

d. Másnap elmentek, ajándékként ott hagytak egy parancsnokot. → Ott maradt/ott volt egy parancsnok.

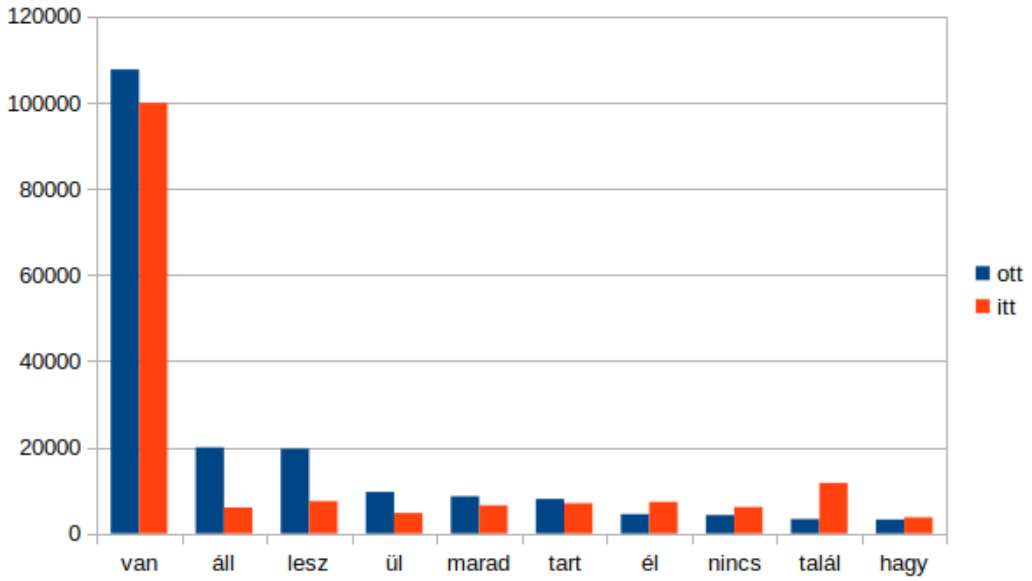
Összességében elmondható, hogy az 5. ábrán látható igék jellemzően statikusak (azaz nem fejeznek ki mozgást), időbeli kiterjedésük van, és a jelentésük valamelyest adott helyen levésre utal. Ezen kívül ezek jellemzően igekötő nélküli, hangsúlykerülő igék.

A demonstratív elemek szemantikáját vizsgáló másik kérdés az volt, hogy mely igékkel szokás egybeírni az *itt/ott*-ot. Bár a jelenlegi akadémiai helyesírás egyiket sem sorolja az igekötők közé, a gyakorlatban mind az *itt*, mind az *ott* előfordul az igével egybeírt, sőt, kettőzött formában is. Ez utóbbira összesen három példa fordult elő az MNSz2-ben, mindegyik a *hagy* igével:

(67) a. A rektor most már tanításidőben is ott-otthagyta az osztályt s kötést cseréltetett.

b. Ezt is ott-otthagyom , hogy megírjam , elvégezzek valamit

DOI: 10.15774/PPKE.BTK.2025.020



5. ábra. A leggyakoribb igekötőszerű *itt/ott* által módosított igék

c. Valamiképpen itt-itthagyjuk az emberre valló önmagunkat!

Az egybeírás és a kettőzés is arra utal, hogy az anyanyelvi intuíció a szabályzattal ellentétben igekötőnek tekinti a demonstratív elemeket, valószínűleg azért, mert valamilyen jelentésváltozást érzékel ezekben az esetekben. Az *ott* kapcsán már Deme (1959) is megemlíti, hogy néhány ige (*marad, ragad, hagy*) esetén az egyébként pusztán „mondatdinamikai" kapcsolat az *ott* és az ige között lexikalizálódott, ezért gyakori az egybeírt alak. Kocsány (2021) Kiefer és Ladányi (2000) és Kiefer (2003) alapján négy olyan igét említ, amelyeknél az *ott* már lexikalizált igekötőnek tekinthető (és ennek megfelelően jellemző rájuk az egybeírt alak): *otthagy, ottfelejt, ottmarad, ottvész*. Az egybeírt alakokat a következő keresőkifejezéssel kerestem:

```
[lemma="ott.+" & msd="(IK.)?IGE\.\S+[123]"]
```

A keresést elvégeztem az *itt*-re is, ám erre nem kaptam megfelelő találatot. Az *ott*-ra 12 különböző igetövet találtam (nem különbözőnek számítva a hatóképzős és műveltető alakokat). Ezek eloszlását a 13. táblázat mutatja.

Látható, hogy az *ott*-tal való egybeírás valóban a Kocsány (2021) által említett igék és azok képzett változatai esetén a leggyakoribb. Ezeken kívül még az *ottfog* és az *ottreked* tűnik fel a korpuszadatok között igekötősige-jelöltként. Ezek jelentése annyiban hasonló a Kocsány (2021) által felsorolt igékhez, hogy ezek esetében is megfigyelhetünk egyfajta szándékolatlanságot a jelentés-

DOI: 10.15774/PPKE.BTK.2025.020

Igető	Gyakoriság
hagy	12793
marad	883
felejt	519
veszik	51
fog	33
reked	32
felejtődik	23
felejtkezik	14
feled	13
feledkezik	8
hagyatik	4
veszt	2

13. táblázat. Az *ott*-tal egybeírt igetövek gyakorisága

tartalomban. Az *ott*-tal egybeírt igéknek jellemzően nem ágense, inkább experiense van, az ágens alanyú igéknél pedig a páciensnek tartozik az akaratlanság és a „térbeli helyet elfoglal” jelentés. Ez alátámasztja Deme (1959) azon észrevételét, miszerint az egybeírást okozó jelentéstömbösödés éppen a dinamika, sőt általánosabban nézve a szándékos cselekvés hiányára utal.

4.2.3. Következtetések

A mérések eredményei alapján az alábbi következtetéseket vonhatjuk le:

1. Az *ott* „igekötőszerűbb”, mint az *itt*

   Mind a szórendi elemzésből, mind az igével való egybeírás gyakoriságából az derült ki, hogy az *itt*-re kevésbé jellemző az igemódosító funkció, mint az *ott*-ra. Ennek szemantikai oka lehet: míg az *ott* anaforikus szerepet is betölthet, az *itt* jelentésénél fogva mindig deiktikus. Mivel értelmezése mindig az aktuális beszédhelyzettől függ, nem tud olyan absztrakt módon helyre utalni, mint azt az *ott* esetén (56b)-ben láttuk.

2. Az *itt/ott* által módosított igék többnyire „specifikus VAN-ok”

DOI: 10.15774/PPKE.BTK.2025.020

A korpuszadatok világosan mutatják, hogy a kérdéses demonstratív elemek többnyire egyfajta helyben levést jelentő, statikus igéket módosítanak. Ezek jelentése tehát a lokatív *van*-hoz hasonló, csak specifikusabb annál: az „udvaron áll" jelentése nagyjából úgy írható le, hogy „az udvaron van álló helyzetben". É. Kiss (2004) feltevése szerint a preverbális *ott* ezt a specifikusságot gyengíti, és az ige eredeti lexikális jelentése helyett annak non-agentív, létezést kifejező jelentését hozza előtérbe. Ugyanakkor ebből még nem egyértelmű, hogy az igék specifikus jelentésének gyengülése valóban az igemódosító *itt/ott*-tal függ-e össze vagy a hangsúlytalansággal (amit az igemódosító tesz lehetővé). A vizsgált pozícióban gyakran előforduló igékről ugyanis az is kiderült, hogy többnyire hangsúlykerülők, és mint korábban említettem, ebbe a kategóriába is elsősorban a 'valahol levés' jelentéselemet tartalmazó igék tartoznak, mint például *húzódik, elterül, elhelyezkedik*. Ezekről pedig szintén elmondható, hogy a jelentésük általában megegyezik a lokatív *van*-nal, vagy legalábbis nagyon közeli ahhoz. Az egybeírás és a kettőzés szokására vonatkozó eredmények azonban arra utalnak, hogy az igemódosító *itt/ott*-nak (különösen az *ott*-nak) a prozódiai funkción kívül sajátos szemantikája is van.

3. Az „eggyel kevésbé specifikált" VAN zéró

   Ha az előzőek szerint a létige lexikális jelentése is gyengül, ha *itt/ott* igemódosító előzi meg (és hangsúlytalan), akkor ez azt eredményezi, hogy elveszíti a lexikális jelentését, és a kopulához hasonló funkcióelemmé válik. Az, hogy más lokatív igemódosítók mellől kevésbé jellemző a létige elhagyása, az *ott* szemantikájával magyarázható: mivel ennek van a legkevésbé specifikus jelentése, így ez a legalkalmasabb arra, hogy az igék 'adott helyen van' jelentéselemét előhívja.

A korpuszadatok elemzése tehát empirikus bizonyítékokkal szolgált arra vonatkozóan, hogy az *(itt/)ott* igemódosító pozícióban is állhat, és ez esetben „adott helyen (jelen) van" jelentéstartalmat hordoz, más szóval statikusságot, és némely esetben szándékolatlanságot is kifejez.

Az igemódosító *itt/ott* és a vele gyakran előforduló igék kapcsolatából az derült ki, hogy a vizsgált proadverbiumok gyengítik a módosított ige deskriptív tartalmát, így azok valójában a létige (valamivel specifikusabb) szinonímáivá válnak. Ezért ha a magához létigéhez kapcsolódnak, az szemantikailag redundánssá válik, tartalmas jelentése gyengül, a funkciója inkább kopulaszerű inf-

DOI: 10.15774/PPKE.BTK.2025.020

lexióhordozóhoz lesz hasonló. Emiatt ha a mondat jelen idő, egyes szám 3. személyben áll, azaz nincs jelölt inflexió, a létige lehet zéró hangalakú is. Ez elméleti magyarázattal szolgál arra a jelenségre, miszerint opcionálisan a lokatív és egzisztenciális létige is lehet testetlen, ha valamilyen, lehetőleg minél általánosabb jelentésű helyre utaló névmás (*itt/ott/hol*) előzi meg.

4.3. *Nincs*, ami nincs

A tagadó formájú (*nincs, sincs*) létigék hiánya összesen 26-szor fordult elő a Szeged Treebankben. Ezek mindegyike tartalmazott valamilyen egyéb tagadó kifejezést. Ezek eloszlását a 14. táblázat tartalmazza.

Tagadó kifejezés	Gyakoriság	Példa
semmi	20	*A sietségre azonban semmi ok.*
sehol	4	*A rend őrei persze sehol.*
se	2	*A Honolulu-Star reggel indul, és se fűtő, se pincér...*

14. táblázat. Az ige nélküli tagadó mondatok tagadó kifejezés szerinti eloszlása

A kezdeti eredmények alapján a *semmi* + főnév szerkezetek további korpuszbeli keresésre és vizsgálatra érdemesnek tűntek. Ezek formálisan egzisztenciális (68a) és birtokos (68b) mondatokra oszthatók.

(68) a. Angyali gyerekarcán **semmi gőg**, és mégis felette állt mindig azoknak, akikhez szólt.

b. Az angol orvos megint eljött hozzá, megvizsgálta, és közölte vele, hogy **semmi baja**, néhány nap múlva felkelhet.

A kétféle mondattípusra külön-külön kerestem az MNSz2-ben az alábbi keresőkifejezésekkel:

- `[lemma="[Ss]emmi"] [msd="FN.NOM"]`
- `[lemma="[Ss]emmi"][msd="FN.PS[et]\d.NOM"]`

Emellett a korpusz szűrő funkcióját alkalmaztam a finit igét tartalmazó találatok szűrésére. Így 14444 létige nélküli egzisztenciális, és 17331 birtokos mondatot kaptam. Ezután a kereséseket megismételtem úgy, hogy a keresőkifejezésekbe beillesztettem a hiányzó *nincs/sincs* létigét (kétféle

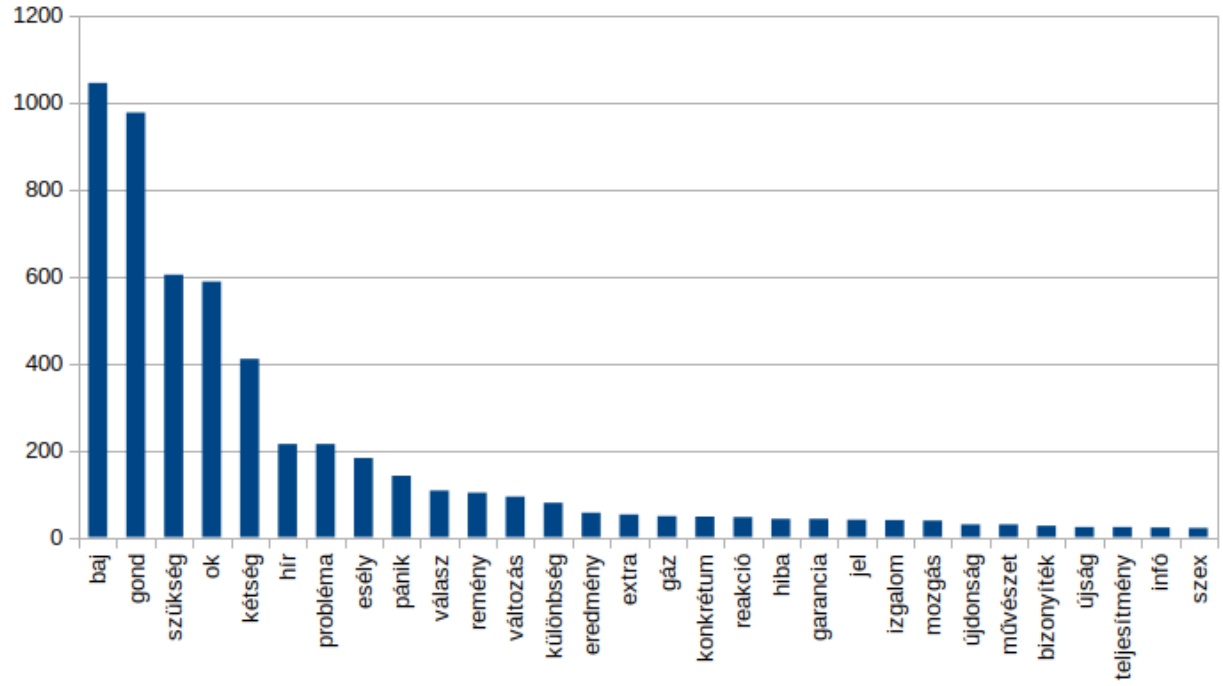
DOI: 10.15774/PPKE.BTK.2025.020

szórenddel: *nincs semmi FN*, illetve: *semmi FN nincs*), és az így kapott találatok számát vetettem össze az azonos szerkezetű, ige nélküli mondatok számával. Ennek eredményeit a 15. táblázat tartalmazza.

Típus	Van létige	Nincs létige	Milyen arányban nincs létige?
egzisztenciális	11090	14444	56,57%
birtokos	14673	17331	54,15%

15. táblázat. A testes és zéró létigék eloszlása a vizsgált mondattípusokban

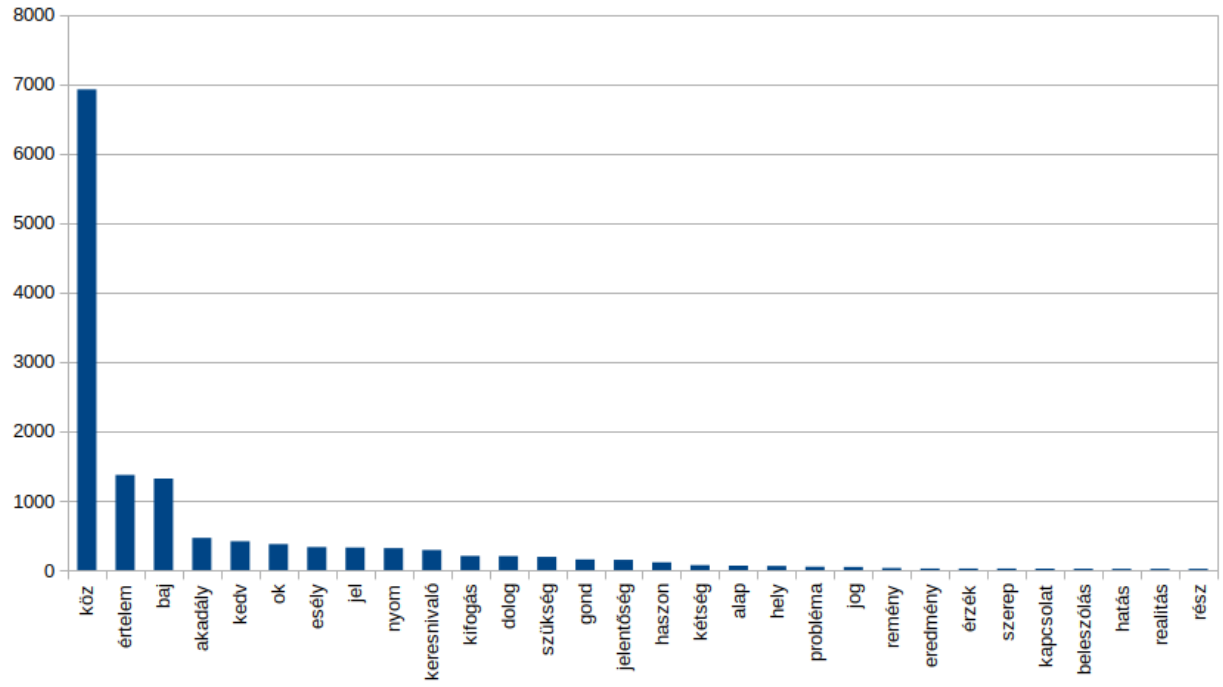
Az eredmények meglepőek, úgy tűnik ugyanis, hogy a vizsgált mondattípusokból az esetek felében hiányzik a létige. A jelenség okának felderítésére elvégeztem ezen szerkezetek névszói részének (alanyának) gyakorisági vizsgálatát, azaz azt mértem, mely névszók esetén milyen gyakorisággal jellemző a *nincs* létige elhagyása. Az eredmények grafikonos szemléltetését a 6. és 7. ábrák tartalmazzák.



6. ábra. Az egzisztenciális mondatok alanyainak gyakorisági megoszlása (30 leggyakoribb szótő)

A vizsgált létige nélküli egzisztenciális mondatokban összesen 2496-féle szótő fordul elő az alany helyén, a birtokos mondatoknál ezzel szemben mindössze 418, holott a birtokosok eleve

DOI: 10.15774/PPKE.BTK.2025.020



7. ábra. A birtokos mondatok alanyainak gyakorisági megoszlása (30 leggyakoribb szótő)

majdnem harmincezerrel többen voltak az egzisztenciálisoknál. Amint a 7. ábrán is látszik, a birtokos mondatoknál az összes gyakoriság nagy részét néhány nagyon gyakori elem teszi ki, az összes előfordulás 90,5%-át lefedi a 15 leggyakoribb szótő (*köz, értelem, baj, akadály, kedv, ok, esély, jel, nyom, keresnivaló, kifogás, dolog, szükség, gond, jelentőség*). Ezzel szemben az egzisztenciálisok első 1500 leggyakoribb szótöve (amelyeknek jelentős része már a legritkább, 1 gyakoriságúak közül kerül ki) is csak az összes eset 89%-ának felel meg. Úgy tűnik tehát, hogy az egzisztenciális *semmi* + főnév szerkezet jóval produktívabb a birtokosnál. Mindemellett mindkét esetről elmondható, hogy bár a *semmi* kvantor jelenléte általában is támogatja a tagadó formájú *nincs/sincs* létige elhagyását, bizonyos gyakori *semmi* + főnév kifejezések erre különösen hajlamosak.

4.4. Zéró *lehet*

A Szeged Treebank "zéró létigés" mondatai között van egy speciális mondattípus, ahol a létige hatóképzős alakjával azonos alakú segédige (*lehet*) hiányzik a szerkezetből. A korpuszban 7 darab (69)-hez hasonló példa fordul elő.

DOI: 10.15774/PPKE.BTK.2025.020

(69) Rendőrök posztolnak a brit és a spanyol nagykövetség előtt is, **soha nem tudni**, hogy néhány tucat megszállott tüntetőnek mikor szottyan ismét kedve egy kis tojásdobálásra.

Ez a mondattípus azért különleges, mert a hiányzó igei alak itt már nem csak idő- és számjelölő, hanem modalitást is kifejez, meglepő tehát, hogy ennek ellenére is elhagyható a mondatból bizonyos esetekben, a jelentés sérülése nélkül. Ráadásul ez az elhagyhatóság csak a *lehet* segédigei funkciójú előfordulásaira igaz, ha a hatóképzős létigét vagy kopulát hagyjuk el, akkor a modális jelentés megszűnik (70).

(70) a. Minden nyelv a maga természete szerint lehet szép és tiszteletreméltó.

b. Minden nyelv a maga természete szerint szép és tiszteletreméltó.

c. Ősztől már a boltokban lehet az új verzió.

d. Ősztől már a boltokban az új verzió.

Úgy tűnik tehát, hogy a *lehet* elhagyhatósága csak az infinitívuszi alanyra korlátozódik, de azokra sem minden esetben, (71a)-ból például rossz mondatot eredményez a létige elhagyása.

(71) a. Ezekben a házakban nem lehet aludni.

b. *Ezekben a házakban nem aludni.

A különleges mondattípusra már 1890-ben felfigyelt Balogh (1890). Leírása szerint a (69) példához hasonló mondatok elliptikusak, a *lehet* elhagyásával keletkeztek. A problémával részletesebben foglalkozik Nádas (2002), aki az elliptált *lehet* mellett az igenevekbe inkorporálódott *lehet* jelenségére is kitér, elsősorban a *mondhatni* alak kérdéskörét vizsgálva. Vizsgálatai szerint a *-hatni/-hetni* végződésű főnévi igenevek 177-szer fordulnak elő a Magyar Történeti Szövegtárban[9], de internetes keresések alapján a *mondhatni* szóalakot a mai nyelvben is "a legkülönfélébb életkorú és foglalkozású emberek használják" (Nádas (2002):330), az eredmények részleteire és a keresés pontos módszerére azonban nem tér ki a tanulmány. A *mondhatni* szóalak Nádas (2002) szerint többféle funkciót is betölthet a mondatban: lehet főtag, módosítószó, partikula, kötőszó, sőt írásjelpótló is. Számunkra most csak az első szerepe érdekes, amikor főtagként a *lehet mondani* vagy

[9]http://www.nytud.hu/hhc

DOI: 10.15774/PPKE.BTK.2025.020

a *mondhatnánk* alakokat helyettesíti. Hasonlóképpen az ige nélkül álló, (69)-típusú főnévi igeneveket is a *lehet + INF* szerkezetből eredezteti a szerző: szerinte ebben az esetben jelentéstapadás történik, azaz az igenév átveszi a segédige modális jelentését. A segédige elhagyását megengedő igenevek szemantikáját tekintve Nádas (2002) "hozzájutás" igékről beszél, Lengyel (2002) és nem részletezett saját gyűjtése alapján a következő példákat sorolja fel: *jutni, tapasztalni/érezni, látni, hallani, fogni, sejteni, észrevenni, megvallani (vallani meg), találni, kapni, megismerni, tudni, tapintani*.

Ezek alapján azt tűztem ki kutatásom fő kérdéséül, hogy mely igék esetén lehetséges a fejezetben tárgyalt segédige nélküli infinitívuszos mondattípus. Ennek megválaszolásához ismét az MNSz2-ből kerestem példákat a kérdéses szerkezetre. Ehhez az infinitívuszt tartalmazó mondatokra kerestem rá (`< s/> containing ([msd="(IK\.)?IGE\.INF"])`), és a kapott találatokból kiszűrtem azokat, amelyekben az infinitívusz közelében (5 szavas ablakban) van finit ige. Az így maradt 117875 találatból egy script segítségével eltávolítottam azokat, amelyekben gyakori infinitívuszi alanyú névszói állítmányok (*érdemes, képes* stb.) szerepelnek. Az így maradt mondatokból kézzel átnézve választottam ki 1000-et elemzésre.

Külön korpuszkeresést végeztem a Nádas (2002) által vizsgált hatóképzős főnévi igenevekre (72a). A *mondhatni* előfordulásait leszámítva 1631 találatot kaptam. A *mondhatni* egyedül 10515 találatot hozott, ám ezt most azért hagyom figyelmen kívül, mert ez gyakran nem főtagként, hanem valamilyen egyéb, Nádas (2002) által említett szerepben fordul elő (72b).

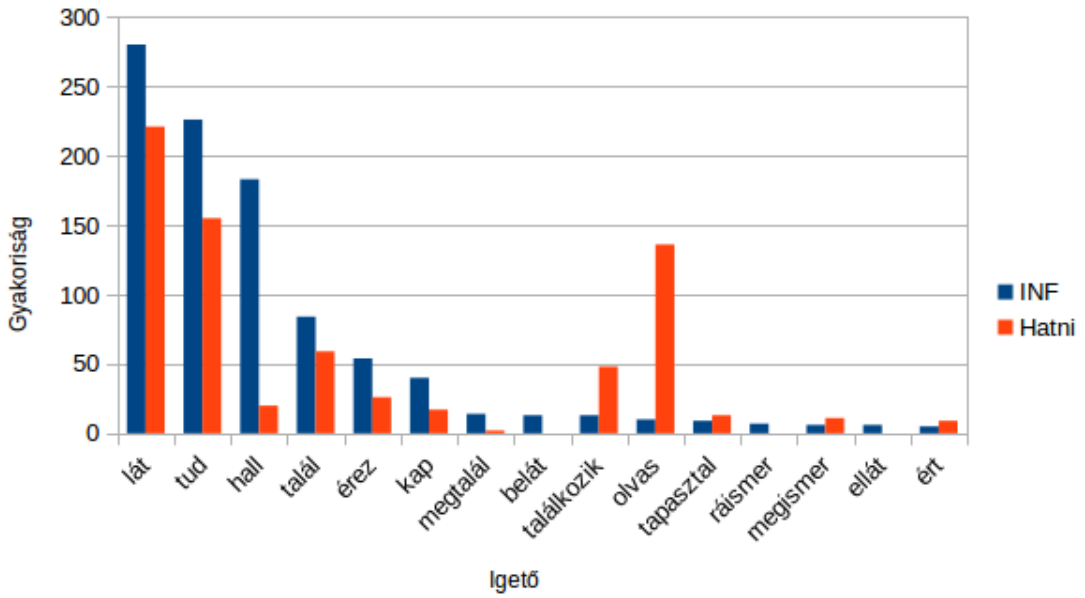
(72) a. Nem tudhatni, hogy melyik bolond társaság akarja így hívni fel magára a figyelmet.

b. Apádat régóta ismerem, közöttünk mindig elvtársias, mondhatni baráti volt a viszony.

A két kereséssel kapott találatokban azt vizsgáltam, hogy mely igék hajlamosak leginkább infinitívuszi főtagot alkotni, ezeknek milyen szintaktikai (tárgyas vagy tárgyatlan, igekötős vagy nem) illetve lexikális tulajdonságai vannak. Az igék gyakorisági eloszlását a 8. ábra mutatja.

A főtagként funkcionáló infinitívuszok leggyakrabban tárgyas igék (*lát, tud, hall*), esetleg valamilyen határozói bővítménnyel rendelkező igék (*ellát (valameddig), találkozik (valakivel)*). Egyargumentumú ige nem szerepel a találatok között. Azt is észre lehet venni, hogy a kapott gyakori igék általában nem ágens-, hanem inkább experiens-alanyúak. Az egyetlen ágens-alanyúnak tűnő ige a gyakori találatok között az *olvas*, ám a példákban ez a "könyvben, szövegben (meg)talál

DOI: 10.15774/PPKE.BTK.2025.020



8. ábra. A 15 leggyakoribb igető ige nélküli infinitívuszos mondatban

valamit" szinonímájaként értelmezhető (73). Az experiens-alanyúsággal összefügg az a lexikális sajátosság, hogy a *lehet* elhagyása a kérdéses mondattípusból leginkább az észlelést, érzékelést jelentő infinitívuszokra jellemző. Ez alól viszonylag gyakori kivétel a *kapni*, ami sima infinitívuszi alakban 40-szer, *kaphatni* alakban 17-szer fordul elő. A ritkább találatok között előfordulnak ágens-alanyúak is, például: *kiügyeskedik, elkerül, nyit* (74).

(73) a. Hát ilyen gyűlöletkeltő dumát, mint a tied, még itt is ritkán olvasni.

b. Teljesebb kifejtésében a kései SZOFISTÁBAN olvashatni...

(74) a. Hogyan kiügyeskedni majd mégis valami pihenést?

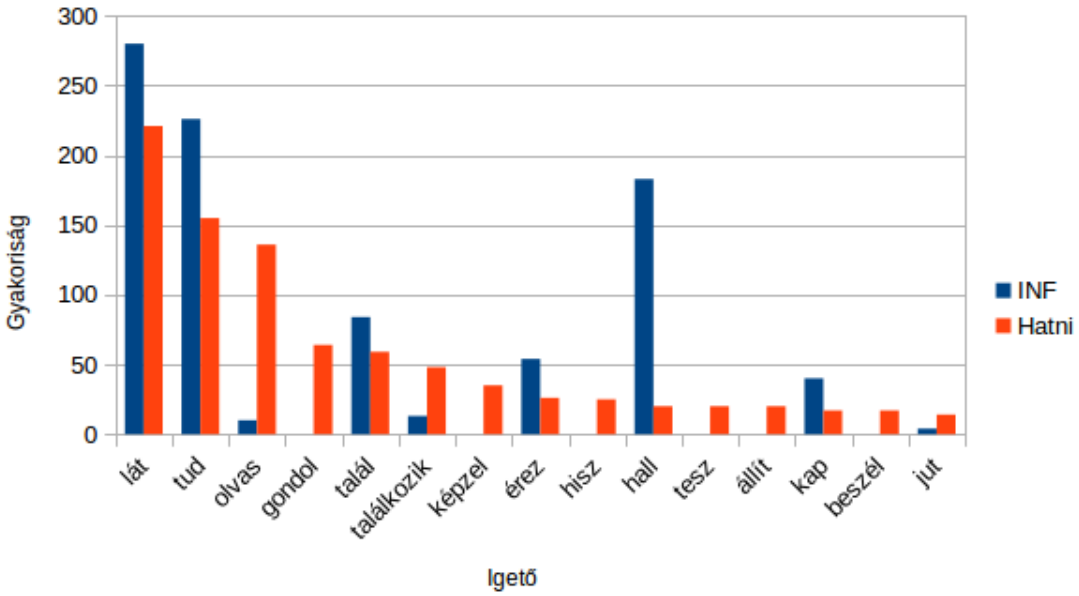
b. Hogyan elkerülni a szemérmetes magunk-kelletését?

c. Belülről nyitni kulccsal vagy gombnyomással.

A hatóképzős alakok eloszlása némileg különböző, itt több olyan gyakori igealak is elfordul, amely a sima infinitívuszokra jóval kevésbé, vagy egyáltalán nem volt jellemző, például: *olvashatni, gondolhatni, képzelhetni, hihetni, tehetni, állíthatni, várhatni*. Az itt előforduló 15 leggyakoribb

DOI: 10.15774/PPKE.BTK.2025.020

igetövet a 9. ábra mutatja. A különbségek oka valószínűleg a műfaji sajátosságokban keresendő, ennek felderítésére összehasonlítottam a jelenség gyakoriságát különböző műfajú szövegekben.

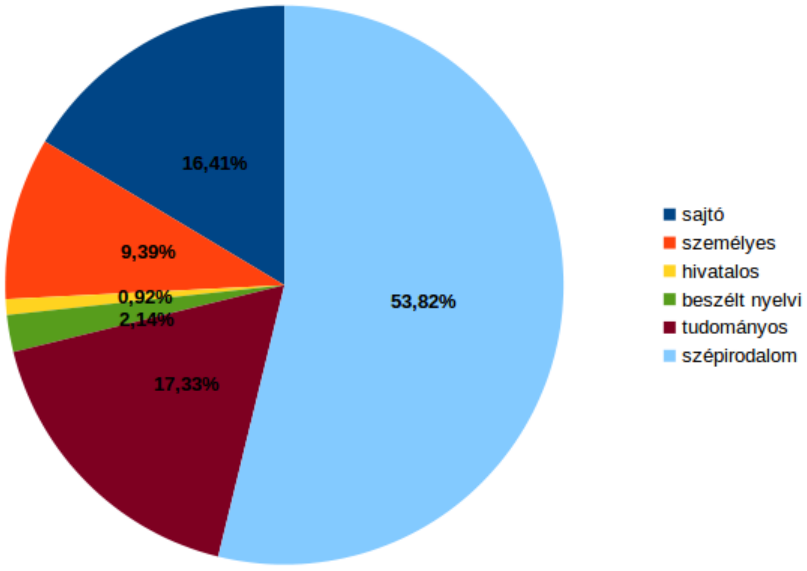


9. ábra. A 15 leggyakoribb igető ige nélküli hatóképzős infinitívuszos mondatban

A hatóképzős infinitívuszok előfordulásának szövegtípussal való összefüggését az MNSz2 alkorpuszainak segítségével vizsláltam meg. A 10. ábrán látható, hogy a kérdéses alak milyen arányban fordul elő az egyes alkorpuszokban. Látható, hogy a jelenség leginkább az írott műfajokra, azon belül is elsősorban a szépirodalomra jellemző. Legritkábban a hivatalos és a beszélt nyelvi szövegekben találkozunk a vizsgált szerkesztésmóddal. Ezzel szemben – bár a sok zaj miatt erre nehéz lenne mérést végezni, de naív megfigyelésem szerint – a sima infinitívuszos ige nélküli mondattípus nagyon is jellemző a beszélt nyelvre. (Meg kell jegyezni azonban, hogy az MNSz2 beszéltnyelvi alkorpusza nem feltétlenül reprezentatív mintája a hétköznapi beszélt nyelvnek, ennél ugyanis valószínűleg mind az országgyűlési jegyzőkönyvek, mind a rádióadások szövegei jellemzően szerkesztettebbek.)

Mindebből arra lehet következtetni, hogy a (69)-típusú mondatok nem csak a (72a)-típusúakkal mutatnak rokonságot. Ezek sokszor megfogalmazhatók hatóképzős névszói állítmánnyal is (75), és ebben az esetben úgy tűnik, hogy a *lehet* modális segédige a névszóba inkorporálódik. Elképzelhető tehát, hogy a *lehet* elhagyhatósága a vizsgált esetekben összefüggésben van a hatóképzős névszói

DOI: 10.15774/PPKE.BTK.2025.020



10. ábra. Az ige nélküli hatóképzős infinitívuszok megoszlása az MNSz2 alkorpuszaiban

állítmányok gyakoriságával.

(75) a. Az erkélyünkről jól látni a főteret.

b. Az erkélyünkről jól látható a főtér.

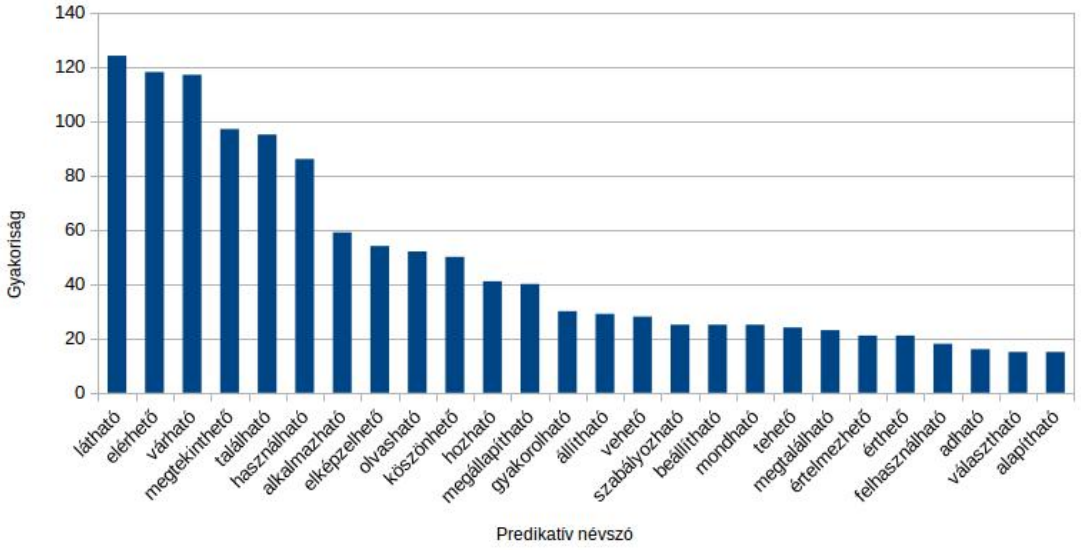
A kérdés vizsgálatához a Szeged Treebank Vincze et al. (2010) adatait használtam fel. Ez a korpusz összesen 30289 darab (5820 különböző) állítmányként funkcionáló (azaz `PRED` relációval jelölt) névszót tartalmaz, melyből 2700 hatóképzővel alkotott melléknév (507 különböző). Ez az összes állítmányi névszó 8,91%-a, elmondható tehát, hogy a hatóképzős alakok gyakran fordulnak elő állítmányi szerepben.

A 15 előfordulásnál gyakoribb hatóképzős névszói állítmányok gyakoriságát a 11. ábra mutatja. Akárcsak az ige nélkül álló infinitívuszok esetén, itt is a *lát* igető bizonyult a leggyakoribbnak, a további gyakorisági adatok azonban jelentős eltéréseket mutatnak az infinitívuszokhoz képest. A *látható*val közel azonos gyakoriságú *elérhető* és *várható* igetövei például egyáltalán nem szerepelnek a vizsgált infinitívuszok között, bár a *lehet* elhagyása ezekkel sem tűnik lehetetlennek. (76)-ben a Szeged Treebank hatóképzős példáit alakítottam át infinitívuszosra. Ezek nem eredményeztek agrammatikus mondatokat, inkább stílusbeli változás történt: az infinitívuszos verzió kevésbé hivatalosnak hat, mint a hatóképzős.

DOI: 10.15774/PPKE.BTK.2025.020

(76) a. A szoftverek számos különféle eszközről elérhetők a weben keresztül.
A szoftvereket számos különféle eszközről elérni a weben keresztül.

b. A ThinkPad sorozatból jövőre várható az A22-es megjelenése.
A ThinkPad sorozatból jövőre várni az A22-es megjelenését.



11. ábra. A leggyakoribb hatóképzős névszói állítmányok a Szeged Treebankben

A további gyakori hatóképzős találatok közül még a *található* és a *megtalálható* hozható összefüggésbe az infinitívuszok gyakorisági listájával, a *talál* és a *megtalál* ott is a 10 leggyakoribb igető között szerepelt. Ezen kívül az *elképzelhető* érdemes még említésre, a *képzel* tő ugyanis a hatóképzős infinitívuszok között bizonyult gyakorinak.

Összefoglalásképpen elmondható, hogy a modalitást magába sűrítő, segédige nélküli infinitívuszos szerkesztésmód az experiens-alanyú, többargumentumú (elsősorban tranzitív) igékre jellemző leginkább, közülük is leggyakrabban az észlelést-érzékelést kifejező igékre. A hatóképzős infinitívuszok gyakorisági eloszlása részben hasonló a sima segédige nélküli infinitívuszokéhoz, a különbségek (például az *olvas* szótő kiugró gyakorisága a hatóképzős adatok között) a műfaji sajátosságoknak tudhatók be, a *-hatni* alak ugyanis a korpuszadatok szerint elsősorban az írott nyelvre jellemző, a szépirodalmi és tudományos szövegekben pedig mások a gyakori igék, mint a beszélt nyelvben. Bár a segédige nélküli infinitívuszos mondatok jelentése hasonló a hatóképzős névszói állítmányos mondatokéhoz, a két mondattípus között az igető-gyakoriságra koncentráló kutatásaim

DOI: 10.15774/PPKE.BTK.2025.020

Elhagyandó		
Típus	**Feltételek**	**Példa**
nominális mondat segédigéje	kijelentő mód, jelen idő, egyes szám 3. személy; a predikatív névszó nem határozószerű	*A ház nagy.*
Elhagyható		
határozószerű névszói predikátum	kijelentő mód, jelen idő, egyes szám 3. személy; a predikatív névszó határozóként is értelmezhető	*Kész (van) az ebéd.*
itt/ott/hol + (van)	kijelentő mód, jelen idő, egyes szám 3. személy; az alany nem névelőtlen	*Ott (van) az ajtó.*
semmi/sehol + (nincs)	kijelentő mód, jelen idő, egyes szám 3. személy; bizonyos alanyoknál gyakori, a többinél ritka	*Erre semmi szükség (nincs).*
jövőre utaló határozószók + (lesz)	kijelentő mód, egyes szám 3. személy	*Mindjárt 11 óra (lesz).*

16. táblázat. Az elhagyandó és elhagyható létigetípusok összefoglaló táblázata

nem mutattak ki egyértelmű rokonságot. Megfigyelhető azonban, hogy az infinitívuszos és hatóképzős névszói állítmányos szerkesztésmód között stiláris különbség lehet (76).

4.5. Összefoglalás

A fejezetben tárgyalt megfigyelések összegzéseként elmondható, hogy a létige elhagyhatósága inkább skaláris mint bináris tulajdonságnak tekinthető. Az elhagyhatósági skálát 3 fő kategóriára lehet osztani: beszélhetünk elhagyandó, elhagyható és kötelező létigékről. Az első két kategória vizsgált típusait a 16. táblázat foglalja össze. A táblázat nem kimerítő, csak azokat az eseteket tartalmazza, ahol a korpuszadatok alapján kimutatható volt a létigék szisztematikus elhagyhatósága. Más mondattípusoknál is előfordultak szórványos példák a létige elhagyására, ezek azonban nem alkotnak jól definiálható, produktív kategóriá(ka)t.

A határozószerű névszói predikátumok a nominális mondatok olyan speciális esetei, ahol az egyébként elhagyandó kopula opcionálisan mégis megjelenhet a szófaji kétértelműség miatt. A többi elhagyható létigetípusnál az elhagyást az teszi lehetővé, hogy a mondat valamely másik szava kifejezi a létige által hordozott jelentéstartalmat.[10] Az *itt/ott/hol* típus esetén ez a jelentéstartalom a helyben levés, a *semmi/sehol* típusnál a tagadás, a jövő idejű határozóknál pedig a jövő idő.

Láttuk, hogy a magyar esetén a fenti létige-osztályozást használva nem tudunk egyértelmű

[10]Ezt a gondolatot Sass Bálint vetette fel a disszertáció munkahelyi vitáján.

DOI: 10.15774/PPKE.BTK.2025.020

határvonalat húzni az elhagyható és nem elhagyható kategóriák közé. Az elhagyhatóság kérdését inkább egy Clark (1978) hierarchiájához hasonló skálával érdemes vizsgálni. A 4. fejezet adatai szerint az elhagyhatóság szempontjából nagyjából 3 fő csoportra oszthatjuk a létigéket:

- alapértelmezett esetben kötelezően zéró
- alapértelmezett esetben (bizonyos feltételek esetén) opcionálisan elhagyható
- semmilyen esetben nem elhagyható

Ezt Clark (1978) és a MGr. típusaival összevetve elmondhatjuk, hogy a "klasszikus értelemben vett" kopula a kötelezően zéró kategóriába esik. A lokatív mondatokról azt állapítottuk meg, hogy bizonyos feltételek mellett a létige elhagyható belőle. Ezen feltételek közül a legfontosabb, a kopulától eltérő megszorítás az, hogy az ige nélküli szerkesztésmód csak egyes számban lehetséges. Nem szükségszerű, de támogatja a létige elhagyásának lehetőségét, ha valamilyen helyre utaló demonstratív elem (*itt, ott, hol*) szerepel a mondatban.

Ezzel szemben az egzisztenciális mondatok *van*-ja, az irodalmi szövegek leírásait leszámítva, csak az előbb említett demonstratív elemek megléte esetén tűnt elhagyhatónak. Érdekes jelenség viszont ennél a típusnál, hogy sok esetben a tagadó formájú létige (*nincs*) hiányzik a mondatokból. Ez akkor lehetséges, ha van a mondatban más tagadó értelmű kifejezés (*semmi, sehol*).

Végül, a birtokos mondatokra csak tagadás esetén jellemző a létige elhagyása, és ekkor is erősen függ az alany lexikális tulajdonságaitól. A gyakori esetek (*semmi köze; semmi értelme; semmi baja*) lényegében idiómáknak tekinthetők.

DOI: 10.15774/PPKE.BTK.2025.020

5. A kopulás mondat szerkezete

Ez a fejezet a kopulás mondat szerkezeti kérdéseivel foglalkozik, ezen belül két fő témakört ölel fel: egyrészt azt vizsgálja, mely esetekben mutatnak szórendi eltérést a testes kopulát tartalmazó illetve nem tartalmazó nominális mondatok, másrészt azt, hogy mi motiválja a predikatív névszó fókuszálását jelentésüket tekintve semleges mondatokban. Az 5.1. fejezet azokat a speciális eseteket mutatja be, amelyek a 3. fejezetben ismertetett zérókopula-beszúró eszköz fejlesztése során merültek fel. Az 5.2. fejezet a nominális mondatok szórendi változatainak gyakoriságát méri fel kérdőíves és korpuszalapú vizsgálattal.

5.1. A testes és zéró kopulás mondatok szerkezete

A 3. fejezetben ismertetett zérókopula-detektáló rendszer tanítóanyagához úgy állítottuk elő a pozitív példákat, hogy a testes kopulákat `<zerokop>` címkére cseréltük a mondatokban. Ez abból az alapfeltevésből indul ki, miszerint a testes és zéró kopulás mondatok szerkezete megegyezik, tehát bármely testes kopulás mondatból jó zéró kopulás mondatot kapunk, ha a testes kopulát zéróra cseréljük. Ez azonban nem minden esetben bizonyult igaznak. Az elkészült tanítóanyag kézi áttekintése során több olyan jelenség is felszínre került, amely nem tette lehetővé a testes és zéró kopulák teljes megfeleltetését. A fejezet ezeket foglalja össze. (A példák a program tanító- és tesztadataiból kerültek ki.)

5.1.1. Nem alapértelmezett esetet jelző tagmondatok

A legegyszerűbb (éppen ezért legkevésbé érdekes) probléma az, amikor a kopula összetett mondatban fordul elő, és más tagmondat(ok) igéje jelzi, hogy a mondat nem jelen idejű (77).

(77) a. Mire Han ideért, már halott volt/*Ø.

b. Emlékeznek, milyen volt/*Ø mielőtt Emily idejött?

Meg kell azonban jegyezni, hogy a múlt idejű tagmondat jelenléte sem zárja ki minden esetben a zérós szerkesztés lehetőségét. A vonatkozó mellékmondatok például sokszor jelen idejű főmondattal is értelmezhetők (78a). Hasonlóan jól működhet a kopula zéróra cserélése azokban az

DOI: 10.15774/PPKE.BTK.2025.020

esetekben is, ahol a múlt idejű tagmondat a névszói állítmányban jelölt állapot okára utal (78b). Megengedett lehet továbbá a zérós szerkesztés a közlést vagy érzékelést jelentő igét tartalmazó főmondatok esetén is (78c, 78d).

(78) a. Kiindulási alapunk az a holttest (volt), amit ön eltemetett.

b. Annyira dühös (volt), hogy a kocsiban kellett várnia.

c. Azt mondta, csak egy kis kitérő (volt) és én hittem neki.

d. És a családja úgy érezte, hogy ön és dr. Justin Wyatt, meg a kórház gondatlanok (voltak).

Ezek alapján valószínűsíthető, hogy a múlt idejű tagmondatot tartalmazó példák szűrése a fedés nagymértékű visszaesésével járt volna.

5.1.2. Az alapértelmezett esetet kizáró lexikális elemek

A testes kopula zéróra cserélése hibás mondatot eredményezhet olyan esetekben is, ahol valamilyen – általában időre vonatkozó – lexikális elem zárja ki a jelen idejű értelmezést. Ilyenek például az egyértelműen múlt időre utaló időhatározók (79).

(79) a. Ön volt/*Ø a Fehér Ház személyzeti főnöke évekkel ezelőtt.

b. Hát hogyne, hiszen kiskorotokban is ő volt/*Ø kettőtök közül a lükébb.

Problémásak lehetnek az intervallumot jelölő időhatározók is, mint például az *-ig* ragos, az *át* névutós névszók, vagy a *soha(sem), sosem* időre vonatkozó,negatív értelmű univerzális kvantorok. Ezekre jellemző, hogy bizonyos predikátumok esetén jelen időben furcsa mondatokat eredményeznek. (80) példái nehezen fogadhatók el valamilyen, múlt vagy jövő idejű, kopula nélkül. Más predikátumok azonban jelen időben is megengedik az *-ig*/*sosem*-típusú határozókat (81).

(80) a. Ő volt/lesz/#Ø Delgado tolmácsa két hétig.

b. Sohasem volt/lesz/#Ø nagy színész.

(81) a. Egy ideig jó (volt/lesz).

b. Tudod az én apám sosem (volt/lesz) ilyen.

DOI: 10.15774/PPKE.BTK.2025.020

Ez a jelenség arra vezethető vissza, hogy az individuum szintű predikátumoknak nem lehet spatio-temporális módosítója (Kratzer, 1995). Ha mégis ilyen módosítót kapnak, azzal helyzetkép szintűvé válnak. (80b)-ben például, bár a *nagy színész*t alapértelmezetten individuum szintűnek tekintjük, a kopula nélküli *Sohasem nagy színész* mondatot mégis csak "sosem tud jól színészkedni (amikor kéne)" jelentésben, azaz helyzetkép szintű predikátummal tudjuk értelmezni. Ha azonban a mondat időjelölt, azaz tartalmaz valamilyen kopulát, a predikátum visszanyeri az individuum szintű értelmezését. Ekkor ugyanis a *sohasem* módosító már nem a predikátumra, hanem az időre vonatkozik: "sohasem volt olyan időpont/időszak, amikor a nagy színészség tulajdonsággal rendelkezett". Ugyanígy (81b) is csak helyzetkép szintűként értelmezhető a kopula nélkül: *sosem ilyen*, azaz *sosem viselkedik így*.

Összegezve tehát a *soha(sem)/sosem* időre vonatkozó univerzális kvantorok akkor engedik meg a zéró kopulát, ha a predikátum helyzetkép szintű, vagy az adott kontextusban annak értelmezhető.

5.1.3. Hiányzó névszói állítmány

Bizonyos esetekben maga a predikatív névszó is elhagyható a mondatokból, ha az állítás explicit vagy implicit vonatkozó mellékmondatban van kifejtve. Ekkor az állítmányt kibontó mellékmondatot az *az* utalószóval helyettesítjük, ami testes kopula jelenléte esetén elhagyható. Ha azonban zéró kopulás a mondat, az utalószó elhagyása a vonatkozó mellékmondat előtt hiányosnak, vagy legalábbis régiesnek (vö. Szózat: *Ez a föld, melyen annyiszor atyáink vére folyt*) hathat (82).

(82) a. Ez a kormány volt/*Ø, amelyik ultimátumot adott.

b. Ez a kormány (volt) az, amelyik ultimátumot adott.

Amennyiben nincs explicit vonatkozói mellékmondatunk (csak a kontextus alapján értjük oda), az *az* elhagyása még határozottabban tiltott zéró kopula esetén, (83a) egyértelműen hiányos mondat a testes kopula vagy az utalószó nélkül.

(83) a. Ezt követően megtudtuk, hogy nem az oroszok voltak/*Ø .

b. Ezt követően megtudtuk, hogy nem az oroszok (voltak) azok.

DOI: 10.15774/PPKE.BTK.2025.020

5.1.4. Igekötős igéből képzett névszói állítmány

A predikatív névszók gyakran igéből képzett, hatóképzős alakok. A Szeged Treebankben található állítmányi névszók 8,9%-a ilyen. Ezeknél megfigyelhető, hogy fókuszos és tagadó mondatokban, zéró kopula esetén leválik róluk az igekötő, azaz a hatóképzős névszói állítmány testes kopula hiányában az ige helyére kerül és az igei tulajdonságait megőrizve viselkedik. Emiatt (84a)-ban nem tudjuk meghatározni a zéró kopula helyét, hiszen a mondatba csak a szórend megváltoztatásával illeszthető be a testes kopula (84b).

(84) a. Nyilvánvaló, hogy a találkozás ma reggeli módja nem ismételhető meg.

b. Nyilvánvaló, hogy a találkozás ma reggeli módja nem (volt) megismételhető.

A zérókopula-beillesztő program tesztadatai között is előfordult ilyen példa, ezekbe hatóképzős névszó elé illesztettem be a zéró kopula címkét (*Nyilvánvaló, hogy a találkozás ma reggeli módja nem Ø ismételhető meg.*), bár a tanítóanyagban biztosan nem szerepel ilyen mondat, a modell legfeljebb (84b) zéró kopulás változatával találkozhatott. Hogy ez mennyiben okozott nehézséget a programnak, azt nehéz megmondani, mert a Szeged tesztkorpusz összességében is nehéznek bizonyult. Azt mindenesetre fontos megjegyezni, hogy a (84a)-típusú mondatok esetén a zéró kopula szabály alapon is nagyon könnyen azonosítható, más esetben ugyanis az igekötő nem válhat le a hatóképzős névszóról, hiszen ekkor a testes kopula tölti be a fókuszált mondatrész (ige) szerepét.

5.1.5. Tagadó mondatok szórendje

Az állítmányi névszó a kopula mellett mind a SMNy. (Kiefer (1992):236-237), mind az Új Magyar Nyelvtan (É. Kiss et al. (2003):38) szerint igemódosítói szerepet tölt be. Ennek megfelelően tagadás esetén a kopula mögé kerül (85) a tagadószó–ige szomszédossági szabály miatt (É. Kiss et al. (2003):66). Nem kell azonban közvetlenül követnie a kopulát, közéjük ékelődhet az állítmányi névszó vonzata (86a) vagy akár az alany is (87a). Ezen mondatok jelen idejű változatánál azonban szórendcserét figyelhetünk meg: az állítmányi névszónak kötelezően követnie kell a tagadószót, azaz a predikatív névszó (igemódosító) tulajdonképpen átveszi az ige helyét (86b, 87b). Más hangsúllyal lehetséges a (87a)-hoz hasonló zéró kopulás mondat is, ez azonban csak fókuszosként értelmezhető, így a múlt idejű párjában máshol lenne a kopula (87c). (87b) zéró kopulás

DOI: 10.15774/PPKE.BTK.2025.020

változata (a hangsúlyviszonyok ismerete nélkül) kétértelmű: ha a kopula a tagadószó és a predikatív névszó között van, akkor semleges a mondat, azonban értelmezhető fókuszos mondatként is, ekkor a predikatív névszó után van a kopula helye (87d). A zérókopula-beillesztő program kiértékelésénél többek között az ilyen esetekre utaltunk, amikor a program által beillesztett kopula helye nem egyezett a referenciával, de az általa javasolt megoldás is helyes volt (9. táblázat).

(85) a. Ez **érdekes** volt.

b. Ez nem volt **érdekes**.

(86) a. Anyám nem volt/*Ø rá **képes**.

b. Anyám nem Ø/volt **képes** rá.

(87) a. Nem volt/*Ø Seger ügynök ítélőképessége **zavart**?

b. Nem Ø/volt **zavart** Seger ügynök ítélőképessége?

c. Nem SEGER ÜGYNÖK ÍTÉLŐKÉPESSÉGE Ø/volt **zavart**?

d. Nem **ZAVART** Ø/volt Seger ügynök ítélőképessége?

A tagadás tekintetében külön érdekesség az állítmányi névszó *sem*-mel való tagadása. Igei állítmány esetén ez ismétléssel oldható meg (88). A nominális mondatoknál ha van testes kopula, akkor ez kerül a tagadószó mögé, míg az állítmányi névszó megelőzi azt (89a). A zéró kopulás mondatban azonban kötelező az igéshez hasonló ismétlés (89b), azaz a tagadószó után nem állhat önmagában zéró.

(88) Nem találkoztam vele, és **látni sem láttam**.

(89) a. Talán még **legális sem volt**.

b. Talán még **legálisnak sem Ø legális**. (De: *Talán még legális sem Ø.)

Ezek alapján elmondhatjuk, hogy testes kopula hiányában a tagadószó utáni pozíciót (semleges mondatban) minden esetben a predikatív névszónak kell betölteni. Ha azonban testes kopula szerepel a modatban, akkor a szórend kérdése már nem ilyen egyértelmű. A következő fejezet ezt vizsgálja meg részletesen.

DOI: 10.15774/PPKE.BTK.2025.020

5.2. Fókusz vagy igemódosító?

Mint említettük, a predikatív névszók a testes kopula mellett általában igemódosító pozíciót töltenek be. A tagadó mondatok szórendi variációiból azonban kiderül, hogy a predikatív névszó bizonyos esetekben fókusz pozícióban is állhat, akkor is, ha jelentését tekintve a fókuszos szórendű változat ugyanolyan semlegesnek hat, mint az igemódosítós. Ez megfigyelhető a múlt idejű eldöntendő kérdésekre adott tagadó válaszok szórendi változataiban (90, 91).

(90) – Petőfi kékszemű volt?

a. – Petőfi nem volt kékszemű.

b. – Petőfi nem KÉKSZEMŰ volt.

(91) – A bíró kopasz volt?

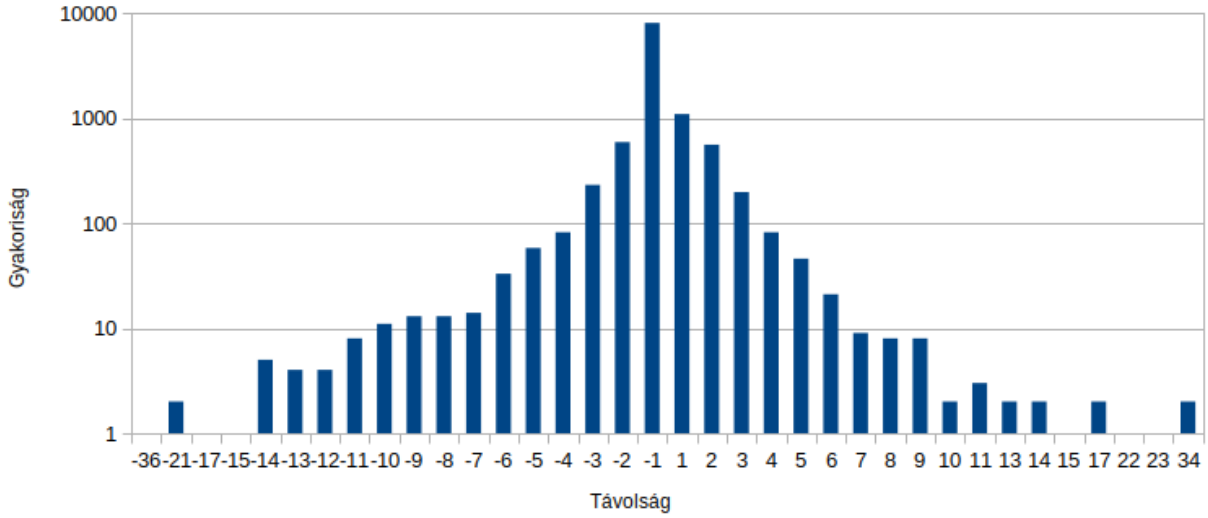
a. – A bíró nem volt kopasz.

b. – A bíró nem KOPASZ volt.

A jelenség több, eddig megválaszolatlan kérdést is felvet. Egyrészt nem világos, miért hajlamosak ezek az elemek fókuszpozícióba mozogni, hiszen úgy tűnik, erre nincs szemantikai-pragmatikai motiváció. Másrészt – és jelen fejezetnek ez az elsődleges kérdése – a fókuszmozgatás a névszói predikátumok esetén mennyire mondható opcionálisnak, más szóval van-e bármilyen tényező, amely meghatározza, hogy az adott mondatban megengedett-e, és ha igen, mennyire preferált a fókuszálás?

(90) és (91) válaszai mind a semleges, mind a fókuszos szórenddel grammatikusak, ugyanakkor észrevehetjük, hogy a két szerkezet mégsem egyformán természetes. Míg (90) kérdésére a fókuszos és az igemódosítós mondatváltozat is ugyanannyira jó, és lényegében egyenértékű válasz, addig (91b)-ben a *kopasz* fókuszálása esetén bizonyos hiányérzetünk van: ha nem kopasz volt, akkor milyen? Úgy tűnik tehát, hogy a *kopasz* névszói predikátum fókuszálása esetén preferált a kontrasztív olvasat, ezzel szemben a *kékszemű* esetén nincs ilyen elvárásunk. Ezek alapján azt feltételezhetjük, hogy az egyes szórendek elfogadhatósága közti különbség a jelzők valamilyen szemantikai tulajdonságából adódik.

DOI: 10.15774/PPKE.BTK.2025.020



12. ábra. A kopula és a predikatív névszó távolsága a Szeged Treebankben

A kérdés tehát az, hogy mi motiválhatja egyes névszói predikátumok fókuszálását, és hogy a fókuszpozíció preferálása a predikátum mely szemantikai tulajdonságával függhet össze. Ennek kezdeti feltérképezésére a Szeged Treebanket használtam.

5.2.1. Távolságok és szórendek a Szeged Treebankben

A Szeged Treebank alkalmas a szórendek vizsgálatára, mert ott a predikatív névszók a függőségi elemzési annotációban jelölve vannak. A korpuszból két, a kopulás mondatok szerkezetére vonatkozó mérést végeztem. Egyrészt azt vizsgáltam, jellemzően milyen távolság van a kopula és a predikatív névszó között, másrészt azt, hogy milyen szórendekkel fordulnak elő a korpuszban az előző példákhoz hasonló tagadó mondatok.

A távolság mérésénél egyszerűen a `PRED` címkével jelölt elem és annak (nem üres) feje közti tokenek számát vettem figyelembe. A távolság értéke negatív, ha a predikatív névszó a kopula előtt áll, -1, ha közvetlenül megelőzi, -2, ha egy szó van köztük, stb. A kopulát közvetlenül követő predikatív névszó távolsága pedig értelemszerűen 1. A korpuszban ilyen módon mért távolságokat a 12. ábra mutatja (a gyakoriságot logaritmikus skálán ábrázoltam). Ezen jól látható, hogy a predikatív névszók jelentős része közvetlenül a kopula előtt, azaz az igemódosító kanonikus helyén áll. A távolság növekedésével a gyakoriság mindkét irányba meredeken csökken, a nagyon távoli névszók skálája azonban meglepően szélesnek tűnik.

DOI: 10.15774/PPKE.BTK.2025.020

Az eredmények pontosabb értelmezéséhez áttekintettem azokat a mondatokat, ahol a kopula és a predikatív névszó távolsága 5 tokennél nagyobb. Ebből az derült ki, hogy a nagy távolság a legtöbb esetben csak látszólagos. Előfordult, hogy a predikatív névszó azért tűnt a kopulától távolinak, mert egy hosszú NP része volt (92a). Még nagyobbnak tűnik a távolság, ha a predikátum koordinálva van (92b), ekkor ugyanis a Szeged Treebank annotációja a mellérendelő szerkezet első tagját jelöli meg fejként (azaz PRED-ként). Más esetekben azt láttam, hogy valójában csak az alany áll a predikatív névszó és a kopula között, azonban ez az alany egy hosszú NP (92c). (93) tartalmazza azokat a példákat, ahol a predikatív névszó és a kopula közti nagy távolságot nem a predikatív vagy alanyi NP hossza okozza, hanem valóban távol vannak egymástól. A távolságértékeknél az alanyi NP egy szónak minősül.

(92) a. Ezeket a napokat akkor sem felejtjük el, mikor másik iskolába kerülünk, mert mi nekünk ez **volt** az utolsó ilyen szép három **nap**.

b. A táj, a növényzet viszont **gyönyörű**, szépséges szép, kellemes levegőjű, csodaszép **volt**.

c. Kicserélték a Konzumbank vezető tisztségviselőit is: az igazgatóság új **elnöke** a lemondott Pázmándi Gyula helyett Baranyai László, az MFB ügyvezető igazgatója **lett**

(93) a. -8: **Alapító** a Gt. (új) 3.§ (1) bekezdése szerint bel- és külföldi jogi személy, jogi személyiséggel nem rendelkező gazdasági társaság és természetes személy **lehet**.

b. -7: A legtöbb nő, akárcsak a legtöbb férfi, könnyen válhat középszerű villamosvezetővé, ám **anyák** – legyenek is bármennyire középszerűek –csak a nők **lehetnek**.

c. -7: Az egyesülés alapítója és **tagja** - csakúgy mint a gazdasági társaságoké - külföldi és belföldi természetes személy, jogi személy és jogi személyiség nélküli gazdasági társaság **lehet**.

d. -7: Ezért kamatot a társaság a Gt. (új) 188.§ (2) bekezdése szerint csak akkor fizethet, ha van nyeresége és már az adóját is kifizetette, mert a **kamatforrása** az osztalékforrással azonosan kizárólag csak a társaság "adózott eredménye" **lehet**.

e. -7: Egész nap izgultam, de a **legidegesebb** a most következő fizika óra előtt **voltam**.

f. -6: Mivel egyetemre szeretnék menni, nagy valószínűséggel igazi **felnőtt** csak 25 éves

DOI: 10.15774/PPKE.BTK.2025.020

korom után **lehetek**.

g. -6: Fáradt voltam halálosan, olyan fáradt, **amilyen** az ember csak húszéves korában tud **lenni**.

h. -6: Azonban helyreállító **ügynök** csak megfelelő bizonyítvány (certificate) birtokában **lehet** valaki.

i. 6: Ezért pl. az egyhangúságot kívánó kérdések is a jelenlévők szavazatainak egyhangúságát jelentik akkor is, ha csak a tagok kisebbsége **van** az egyébként határozatképesnek számító taggyűlésen **jelen**.

A példákból az látszik, hogy a predikatív névszó és a kopula távolsága -8 és 6 szó (mondatrész) között változik. A kopulát megelőző predikatív névszók eltávolodását általában a kontrasztív topikba helyezés okozza.

A távolságok vizsgálata tehát alátámasztja azt az elméletet, mely szerint a predikatív névszó a kopula mellett igemódosító pozíciót tölt be. A tagadó mondatok szórendjének vizsgálata a fókuszálás eseteiről adhat további információt. A tagadó mondatok keresésénél a szűrés feltételei a következők voltak:

- legyen a mondatban predikatív névszó
- a predikatív névszó feje testes elem legyen (ne zéró fej)
- legyen a mondatban tagadószó
- a tagadószó feje a predikatív névszó vagy a kopula legyen
- ne legyen más elem (például az alany) fókuszálva a mondatban (ezeket utólag, a kézi átnézés során szűrtem ki)

A találatok között 42 fókuszos és 588 semleges szórendű mondat szerepelt, tehát az igemódosító pozíció fölénye feltűnő. Ez az adatmennyiség, bár messzemenő következtetések levonására nem alkalmas, lehetővé teszi a szórend kérdésének részletesebb vizsgálatát. A különböző szórendű mondatokat áttekintve azt vizsgáltam, hogy ezek milyen gyakran predikatívak vagy azonosítók, illetve milyen névszók (vagy névszói kifejezések) fordulnak elő az egyes pozíciókban. Az osztályozásnál Kádár (2011b) nyomán predikatívnak tekintettem azokat az eseteket, ahol a predikátum

DOI: 10.15774/PPKE.BTK.2025.020

NP, NumP, QP, AdjP vagy PossP kategóriájú. (94) a, c, e, és g, példái a gyűjtésből származnak, b, d, f, és h, pedig ezek más szórendű változatai.

(94) a. Mindent úgy értek, ahogy mondom, **nem vagyok szimbolista költő**.

b. Mindent úgy értek, ahogy mondom, **nem szimbolista költő vagyok.**

c. Nem mondom, én **sem vagyok egy "szent"**, de tudom, hogy van az ésszerű határ.

d. Nem mondom, én **sem egy "szent" vagyok**, de tudom, hogy van az ésszerű határ.

e. Tudtam, ez a nap **nem olyan lesz**, mint a többi.

f. Tudtam, ez a nap **nem lesz olyan**, mint a többi.

g. Azért szeretnék sofőr vagy műszerész vagy más lenni, mert akkor **nem vagyok más lekötelezettje**, hanem magamnak csinálom amit csinálok, és annyit keresek amennyit dolgozom.

h. Azért szeretnék sofőr vagy műszerész vagy más lenni, mert akkor **nem más lekötelezettje vagyok**, hanem magamnak csinálom amit csinálok, és annyit keresek amennyit dolgozom.

Az azonosító mondatok esetén a predikatív névszói csoport DP. (95) példái szintén tartalmazzák az eredeti korpuszbeli és a módosított szórendi változatokat is.

(95) a. Ahogy bementem a terembe, már éreztem, hogy ez a nap **nem az én napom lesz**.

b. Ahogy bementem a terembe, már éreztem, hogy ez a nap **nem lesz az én napom**.

c. Szerencsére már kinőttem abból a korból és furcsa módon észlelem magamon, hogy már **nem az a kislány vagyok**, mint régen.

d. Szerencsére már kinőttem abból a korból és furcsa módon észlelem magamon, hogy már **nem vagyok az a kislány**, mint régen.

Mind a predikatív, mind az azonosító példákban azt látjuk, hogy a szórendi módosítás általában nem változtat a mondat jelentésén vagy elfogadhatóságán. Vannak azonban kivételes példák: mind a predikatív és azonosító, illetve a fókuszos és semleges mondatok között előfordul olyan eset, amikor a mondat furcsán hat, vagy nem értelmezhető semleges jelentésűként más szórenddel (96).

DOI: 10.15774/PPKE.BTK.2025.020

(96) a. Mondhatom, **nem az én hibám lesz**, ha nem az öreg Győzelem-tömb zászlódísze lesz a leggazdagabb az egész utcában.

b. #Mondhatom, **nem lesz az én hibám**, ha nem az öreg Győzelem-tömb zászlódísze lesz a leggazdagabb az egész utcában.

c. Ide tartozik az is, hogy a szakmai irány változása **nem egy egyszeri döntés volt**, hanem a KFKI "szövetségi részvénytársaság" működési elvéből ered.

d. #Ide tartozik az is, hogy a szakmai irány változása **nem volt egy egyszeri döntés**, hanem a KFKI "szövetségi részvénytársaság" működési elvéből ered.

e. Szerencsére **nem volt túl nagy** a forgalom, így elég hamar leértünk.

f. #Szerencsére **nem túl nagy volt** a forgalom, így elég hamar leértünk.

g. Nekem senkivel szemben nincsenek előítéleteim, legfeljebb az illető **nem lesz a barátom**.

h. *Nekem senkivel szemben nincsenek előítéleteim, legfeljebb az illető **nem a barátom lesz**.

(96b)-ben a semleges szórend furcsán hat, valószínűleg azért, mert a *nem az én hibám* kifejezést általában kontrasztív módon értelmezzük. (96d) esetén a *hanem*-mel bevezetett tagmondat egyértelművé teszi, hogy itt is a kontrasztív olvasat okozza a semleges szórend furcsaságát. (96f és h) esetén azonban már nem nyilvánvaló, miért tűnik a fókuszos változat nehezen vagy egyáltalán nem elfogadhatónak.

A *nem... hanem...* típusú mondatok esetén azt várjuk, hogy a predikatív névszó(i kifejezés) megelőzi az igét. Ez a *nem... hanem...* szerkezet mellérendelő voltával, és emiatt a hangsúlytörlés hiányával magyarázható: (97a)-ban a *nap* valójában nem fókusz, hanem igemódosító, amely a hangsúlytörlés hiányának köszönhetően meg tudott maradni a tagadószó és az ige között. (98a) É. Kiss et al. (2003) példája (É. Kiss 2003:67).

(97) a. De egyszer volt egy, de ez tulajdonképpen **nem is nap volt**, hanem egy hét.

b. *De egyszer volt egy, de ez tulajdonképpen **nem is volt nap**, hanem egy hét.

DOI: 10.15774/PPKE.BTK.2025.020

(98) a. János nem **elment**, hanem elfutott.

b. *János nem **ment el**, hanem elfutott.

Ezt a korpusz adatai igazolták, a semleges szórendű mondatok között csak 4 *nem... hanem...* szerkezetű mondat szerepelt (pl. 94g), míg a fókuszos szórendűek között 17 (ami az összes ilyen szórendű mondat 40,48%-a).

A predikatív és azonosító mondatok szórendi változatainak eloszlását a 13. ábra mutatja. A korpusz adataiból látható, hogy a predikatív mondatoknál egyértelműen preferált (95,33%) a semleges szórend. Az azonosító mondatok esetén azonban a fókuszos szerkesztésmód aránya láthatóan jelentősebb (46,67%). A szórendi változatokon vizsgálva a mondattípusok eloszlását azt tapasztaljuk, hogy a semleges szórendű mondatok 97,27%-a predikatív, és csak 2,24%-a azonosító mondat, a fókuszos mondatoknál azonban ez az arány 66,67%, illetve 33,33%. Kádár (2011b) szerint "az azonosító mondatoknak a magyarban tulajdonképpen nem létezik semleges változatuk", ugyanis vagy az alanyi, vagy a predikatív DP fókusz (Kádár 2011: 75). Az adatok ennek látszólag ellentmondanak, meg kell azonban jegyezni, hogy a predikatív–azonosító megkülönböztetésnél csak formális kritériumokat használtam: azonosítónak címkéztem minden olyan mondatot, ahol a predikátum DP volt. Ezek azonban jelentésükben nem feltétlenül nevezhetők azonosítónak. (100a) esetén például *a NATO tagja* predikátum DP ugyan, de nem azonosítást fejez ki, hanem tulajdonságot. Hasonlóképpen (100b) *az a férfi* predikátuma is inkább *olyan férfi*ként értelmezhető, azaz predikatívnak nevezhető. A formálisan azonosítónak címkézett semleges mondatokat áttekintve a legtöbb esetben hasonlókat tapasztaltam. A korpusz példái közül négy semleges szórendű mondatot nevezhetünk valóban azonosítónak (101), ezek közül három a *Piszkos Fred, a kapitány* című regényből származik, amelynek nyelvezete nem feltétlenül tükrözi a sztenderd nyelvhasználatot. A továbbiakban ezért csak a predikatív mondatok szórendi változataival foglalkozom.

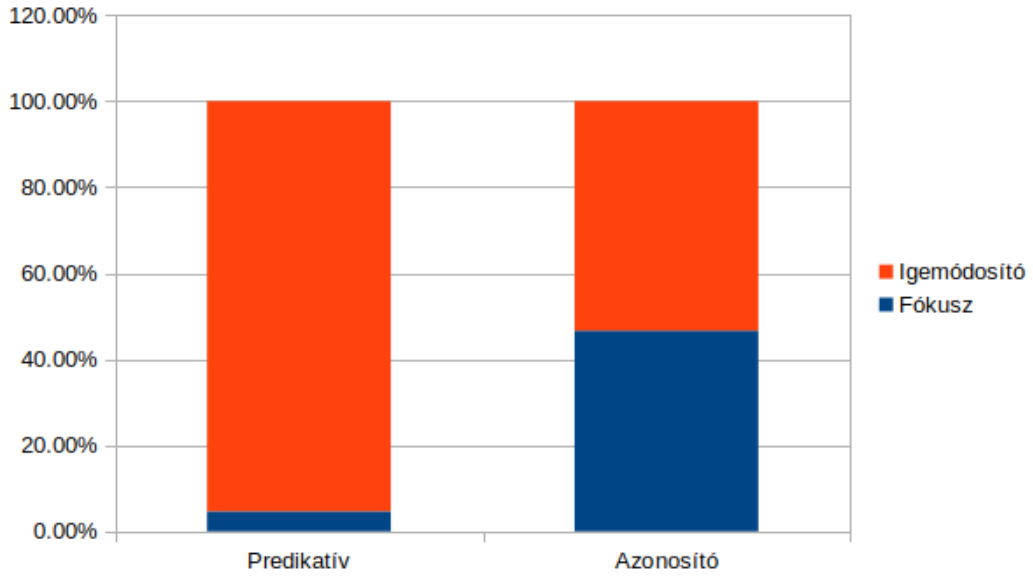
(99) a. Nem vagyok (egy) beteges típus, úgyhogy orvoshoz is csak nagyon ritkán kerültem.

b. Hát az bizony nem volt (egy) túlságosan boldogító érzés.

(100) a. Lantos véletlenszerű egybeesésnek nevezte, hogy hazánk még két hete sem volt **a NATO tagja**, amikor megindultak a légicsapások.

b. Persze ez szimplifikáció, az ember sosem ilyen megfogalmazható, de annyi biztos,

DOI: 10.15774/PPKE.BTK.2025.020



13. ábra. A Szeged Treebank tagadó predikatív és specifikáló mondatainak szórendi eloszlása

hogy rendkívül ügyetlen és inkompetens vagyok a világ dolgaiban, egyáltalában nem vagyok **az a férfi**, akinek nyugodt felsőségében megbízhat egy asszony.

(101) a. Nem vagyok Fülig Jimmy, hanem José Pombio.

b. Te nem vagy Tamás.

c. Maga jól tudja, hogy én nem vagyok a király, mert gyerekkorom óta jól ismer, és homlokán heged a régi sebes kardos nyom.

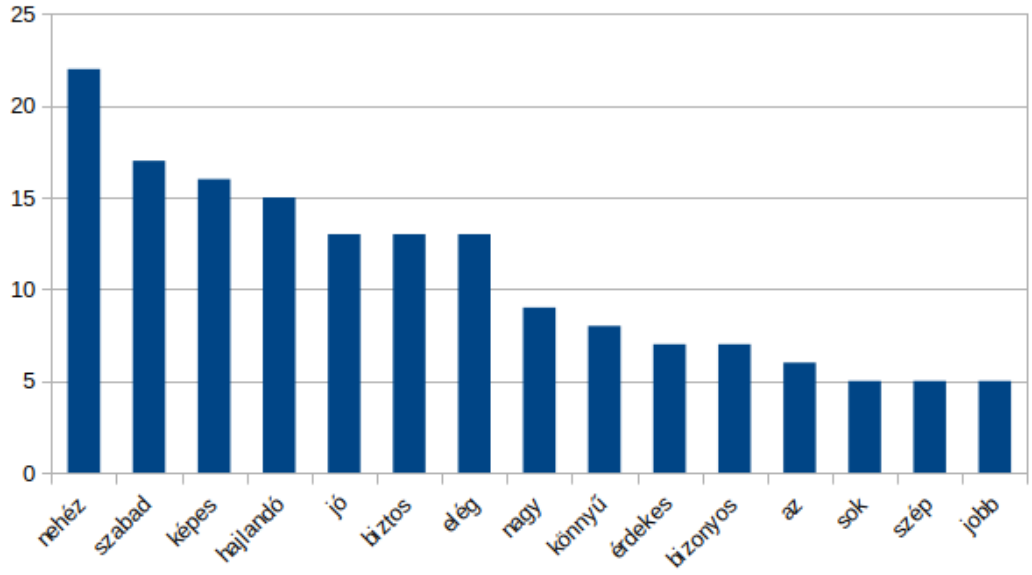
d. Mer ha én nem vagyok a király és idáig hagyott engem fajulni, akkor ő a palástbűnös és ezért mellém ítélik a szomszéd kötélre bíróilag.

A gyakori predikatív névszókat és névszói kifejezéseket a fókuszos esetek kis száma miatt csak a semleges szórendű mondatokban vizsgáltam. A 14. ábra a 15 leggyakoribb igemódosítóként viselkedő névszói állítmányt tartalmazza.

5.2.2. Elméleti keret, hipotézisek

A nominális mondatok szerkezeti elemzésével Kádár (2011b) foglalkozott részletesen. Az általa javasolt predikatív mondatszerkezeti elemzés megengedi az előzőekben tárgyalt fókusz–igemódosító

DOI: 10.15774/PPKE.BTK.2025.020



14. ábra. A leggyakoribb igemódosítóként viselkedő névszói állítmányok a Szeged Treebankben

kettősséget, szerinte ugyanis a névszói predikátum a [Spec TP]-be mozog, ahonnan egy újabb, [Spec FP]-be történő mozgatással maga is fókusz lehet (Kádár 2011b: 81). Arra azonban nem tér ki, mitől függ, hogy egy nominális predikátum – jellemzően - fókuszként funkcionál-e? Kutatásom elméleti kerete (Higgins, 1973) kopula-tipológiájából indul ki, amely négy különböző kopulás mondattípust különböztet meg. Higgins (1973) szerint a névszói állítmány predikatív, ha az alany referenciális NP vagy QP, és a predikatív komplementum a referensről tesz valamilyen állítást. Ezzel szemben a maradék három típus nem predikatív. Az identifikáló mondat alanya valamilyen demonstratív, deiktikus referenssel bíró elem, és a mondat funkciója ennek megnevezése. Az ekvatív típus két referens azonosságát fejezi ki. A negyedik, specifikáló típusnál az alany nem referenciális, hanem egy változót, vagy más szóval halmazt határoz meg. A predikátum ezt a halmazt specifikálja az elemeinek kimerítő felsorolásával. Például:

(102) a. A programozóverseny győztesei Lili és Balázs voltak.

b. A programozóverseny győztese nem Abigél volt.

(102a)-ban az alany által kijelölt halmaz a programozóverseny egy vagy több győztese. Ez a halmaz egzisztenciálisan preszupponált, ami abból is látszik, hogy (102b)-ben a tagadás után is megmarad az az előfeltevés, hogy a versenynek volt győztese, azaz az alany által meghatározott

DOI: 10.15774/PPKE.BTK.2025.020

halmaz létezik, és nem üres. (102a) felsorolással specifikálja a programozóverseny győzteseinek halmazát, míg (102b) azt állítja, hogy a fókuszban álló individuum (Abigél) ennek a halmaznak nem eleme.

Higgins (1973) elmélete alapján É. Kiss (2017) amellett érvel, hogy a magyar fókuszemelés célja valójában egy specifikáló predikatív szerkezet létrehozása, ahol a fókusz a predikátum, a háttér pedig annak alanya. É. Kiss (2017) szerint a specifikáló fókusz háttere mindig preszupponált halmaz, az az értelmezési tartomány, ahol a fókuszált összetevő által meghatározott értéket értelmezzük. Eszerint (90b) a következőképpen értelmezhető:

(103) a. Petőfi nem KÉKSZEMŰ volt. (predikatív)

b. Petőfi szemszíne nem A KÉK volt. (specifikáló)

Azaz a (103a) azt kódolja, hogy Petőfinek volt egy feltételezhető szemszíne, amit kékként azonosítunk. Ezzel szemben (91b) nem igen alakítható specifikáló mondattá:

(104) a. A bíró nem KOPASZ volt. (predikatív)

b. #A bíró <*valamilyen tulajdonsága*> nem A KOPASZ volt. (specifikáló)

(104a) esetében tehát azért tűnik furcsának a nem semleges szórend, mert a kopaszságról mint tulajdonságról nincs a szemszínhez hasonló előfeltevésünk. Eszerint a kopulás mondatok névszói predikátuma akkor fókuszálható, ha az az alany valamilyen inherens tulajdonságára utal (É. Kiss 2017: 10). Inherens tulajdonság alatt azokat értjük, amelyekkel (feltételezésünk szerint) minden individuum rendelkezik, és amelyek lehetséges értékei egy véges halmazba sorokhatók.

A fejezetben ismertetett kutatás kiinduló alternatív elmélete szerint azonban az állítmány fókuszálhatóságát nem az határozza meg, hogy az általa megfogalmazott tulajdonság előfeltételezett-e az alany esetén, hanem az, hogy vannak-e releváns alternatív értékei, amelyekkel a fókusszal jelölt összetevő értéke szembeállítható. Ez az elképzelés a Rooth (1985) által bevezetett fókuszelmélettel rokon, mely szerint a fókusz alternatívahalmazt tételez fel, amelynek egyik eleme a fókuszált összetevő. Ezt az elméletet követi Krifka (2008) fókuszdefiníciója is, és a kutatás hipotézisei is ez alapján magyarázhatók a legkönnyebben:

„Focus indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions." (Krifka 2008: 6)

DOI: 10.15774/PPKE.BTK.2025.020

A releváns alternatívák jelenléte szerint a predikatív szerepet betöltő NP/AP-k két nagy szemantikai csoportba oszthatók: binárisokra és többértékűekre. Bináris jelentésű NP/AP alatt azokat értem, amelyekhez (kontextus nélkül) nem tartoznak alternatívák a kifejezés ellentétén kívül. Ezek tehát olyan értelemben binárisak, hogy az általuk jelölt tulajdonsághoz valójában csak igen/nem értékek tartoznak aszerint, hogy az alany rendelkezik-e az adott tulajdonsággal vagy nem. A Szeged Treebank gyakori igemódosítóként viselkedő névszói predikátumai (14. ábra) nagyrészt ilyenek: *nehéz, szabad, képes, hajlandó, jó, biztos, elég, nagy, könnyű, érdekes, bizonyos, sok, szép*. Ezzel szemben a többértékű NP/AP önmagában, kontextus nélkül is képes alternatív jelentésű predikátumokat aktiválni. (90) esetén a *kékszemű* tehát többértékű predikátum, hiszen az alternatívák között megtalálhatók az egyéb szemszínek. (91)-ban, a *kopasz*hoz azonban nem tudunk alternatív értékeket rendelni azon kívül, hogy nem kopasz. Ez Rooth (1985) és Krifka (2008) elméletei alapján megmagyarázza a (90b) és (91b) elfogadhatósága közti különbséget.

Összegezve, a kutatás hipotézise a következő: a kopulás mondatok névszói predikátuma akkor kerülhet fókuszpozícióba, ha a jelentése többértékűként határozható meg. A bináris jelentésű melléknevek tartalmazó állítmányt jellemzően nem fókuszáljuk. A hipotézist két empirikus módszerű kutatással, az anyanyelvi intuíciót felmérő kérdőíves vizsgálattal és korpuszadatok gyakorisági elemzésével ellenőriztem.

5.2.3. Kérdőíves vizsgálat

A kérdőíves kutatás az anyanyelvi intuícióra hagyatkozva méri fel, milyen arányban választják a résztvevők a fókuszos vagy a semleges szórendet a testes kopulát tartalmazó tagadó mondatokban. A kérdőív 10 egyszerű állító mondatot tartalmazott, és azt kérte a kitöltőktől, hogy adják meg ezek tagadó változatait. A mondatokhoz nem tartozott kontextus, és az instrukció is nagyon egyszerű volt. A kérdőív annyi kikötést tett, hogy a kitöltők magyar anyanyelvűek legyenek, és minden mondathoz csak egy választ adjanak. (Ez utóbbit nem sikerült minden adatközlőnek betartani.) A példamondatok egyik felében tipikusan bináris jelentésű (*magas, nős, jó szakács, tehetetlen, érthető*), a másikban tipikusan többértékű (*orvos, angol, rózsaszín, huszonöt éves, fradista*) névszói predikátumok szerepeltek. Két példa:

(105) a. Tehetetlenek vagyunk.

DOI: 10.15774/PPKE.BTK.2025.020

b. A szobám fala rózsaszín volt.

A felmérés során összesen 261 értékelhető kitöltés érkezett. Az eredményeket a 17. és 18. táblázat tartalmazza.

	Predikátum	**Fókusz**	**Igemódosító**	**Egyéb**
Bináris	tehetetlen	0 (0,0%)	248 (96,1%)	10 (3,9%)
	érthető	0 (0,0%)	248 (96,1%)	10 (3,9%)
	magas	2 (0,8%)	240 (93,4%)	15 (5,8%)
	jó szakács	2 (0,8%)	239 (92,6%)	17 (6,6%)
	nős	1 (0,4%)	224 (86,5%)	34 (13,1%)
Többértékű	orvos	96 (36,8%)	159 (60,9%)	6 (2,3%)
	angol	167 (64,2%)	90 (34,6%)	3 (1,2%)
	rózsaszín	129 (49,4%)	120 (46,0%)	12 (4,6%)
	huszonöt éves	113 (43,3%)	143 (54,8%)	5 (1,9%)
	fradista	80 (30,7%)	169 (64,7%)	12 (4,6%)

17. táblázat. A kérdőíves vizsgálat eredményei szavanként

A 17. táblázat eredményei egyértelműen mutatják, hogy a bináris jelentésű predikátumok esetén a kitöltők szinte soha nem választották a fókuszos szórendet, míg a többértékűeknél a kétféle szórend aránya egyenletesebben oszlik el. Azt is észre lehet venni, hogy a bináris esetekben több az „egyéb” kategóriába sorolt válasz. Ezek nagy részében a kitöltők nem tagadó mondatokat írtak, hanem egy másik állítást az ellentétes jelentésű melléknévvel, például:

(106) Kérdés: József nős volt.

Válasz: József nőtlen volt.

Ez megerősíti a kérdőívhez választott predikátumok bináris jellegét.

A 18. táblázatból az is látszik, hogy a semleges szórend általánosságban nézve is preferáltabb (ami egyébként nem meglepő, hiszen ez a jelöletlenebb), de a többértékű predikátumok esetén ez a különbség nem tűnik annyira jelentősnek.

Összességében tehát elmondható, hogy a kérdőívek eredményei alapján úgy tűnik, valóban van összefüggés a predikátum bináris vagy többértékű jellege és a pozíciója között.

DOI: 10.15774/PPKE.BTK.2025.020

	Fókusz	**Igemódosító**	**Összesen**
Bináris	5	1199	1204
Többértékű	585	681	1266
Összesen	600	1880	2470

18. táblázat. A kérdőíves vizsgálat eredményei összesítve

5.2.4. Korpuszalapú vizsgálat

A kutatás másik része egy korpuszalapú vizsgálat volt, azaz nagy mennyiségű szöveges adat elemzésén alapult. Az adatok forrása az MNSz2 volt. A korpuszban négyféle predikatív kopulás mondatszerkezetre kerestem a következő keresőkifejezésekkel:

1. Melléknév (vagy melléknévi igenév) fókuszpozícióban

   ```
   [msd="FN.(PS[et]\di?.)?(PL.)?NOM|FN_NM.([et]\d.)?NOM"][word="[ns]em"]

   [msd="MN.(PL.)?NOM|(IK.)?IGE._OKEP.NOM|(IK)?.IGE._MIB.NOM"]

   [lemma="van"]
   ```

   Géczy nem dózsás volt.
   Ő nem bekéredzkedett volt.
   Sikere nem politikai volt.

2. NP fókuszpozícióban

   ```
   [msd="FN.(PS[et]\di?.)?(PL.)?NOM|FN_NM.([et]\d.)?NOM"][word="[ns]em"]

   [msd="MN.(PL.)?NOM|(IK.)?IGE._OKEP.NOM|(IK)?.IGE._MIB.NOM"]*

   [msd="FN.(PS[et]\di?.)?(PL.)?NOM"][lemma="van"]
   ```

   Sándor nem kártékony uszító volt.
   Ez nem vadászgép volt.
   Én nem hordár vagyok.

3. Melléknév (vagy melléknévi igenév) igemódosító pozícióban

   ```
   [msd="FN.(PS[et]\di?.)?(PL.)?NOM|FN_NM.([et]\d.)?NOM"][word="[ns]em"]
   ```

DOI: 10.15774/PPKE.BTK.2025.020

```
[lemma="van"][msd="MN.(PL.)?NOM|(IK.)?IGE._OKEP.NOM|
(IK)?.IGE._MIB.NOM"]
```

A testület nem volt döntőképes.

Senki nem volt ideges.

A kép nem volt megnyerő.

4. NP igemódosító pozícióban

```
[msd="FN.(PS[et]\di?.)?(PL.)?NOM|FN_NM.([et]\d.)?NOM"][word="[ns]em"]
[lemma="van"][msd="MN.(PL.)?NOM|(IK.)?IGE_OKEP.NOM|(IK)?.IGE._MIB.NOM"]*
[msd="FN.(PS[et]\di?.)?(PL.)?NOM"]
```

Béla nem volt rossz mérnök.

Találmányai nem voltak világhírű szabadalmak.

Mi nem vagyunk barátnők.

Az 1-2. (fókuszos) típusokra összesen 1665, a 3-4. (igemódosítós) típusokra pedig 34723 találatot kaptam. Ez már önmagában látványos különbség, jól látható tehát, hogy a Szeged Treebankhez hasonlóan az MNSz2-ben is jóval gyakoribb a semleges szórend.

Az adatok feldolgozása nagyrészt automatikusan, egy Python script segítségével történt. A script a mondatokban tipikusan bináris jelentésű és tipikusan többértékű mellékneveket (és azokat tartalmazó NP-ket) keres, és ezek gyakoriságát számolja a különböző mondattípusokban. Tipikusan binárisak és gépi eszközzel jól kereshetők a ható- és fosztóképzős, valamint a -szerű, -képes, -mentes, -ellenes utótagú melléknevek. Tipikusan többértékűek a számneves kifejezések, a színnevek, a nemzetiségnevek és a foglalkozások. Listák segítségével ezek is könnyen kereshetők számítógépes programmal. A listákat vektoros szómodellezés segítségével állítottam össze. A szómodellezéshez a Facebook Artificial Intelligence által fejlesztett FastTextet[11] választottam, a nyílt hozzáférése és egyszerű használata miatt. A FastText nyelvmodellező szolgáltatásai Python modulként használhatók szóbeágyazási modellek (Joulin et al., 2016) készítésére és szövegosztályozásra (Bojanowski et al., 2016). A programcsomag mellett előre betanított szóvektor-modellek is

[11] https://fasttext.cc/

DOI: 10.15774/PPKE.BTK.2025.020

letölthetők, amelyek webes szövegek alapján készültek, és jelenleg 157 nyelven (köztük a magyarul) érhetők el. (Grave et al., 2018) A szólisták elkészítéséhez ezt az előre tanított magyar nyelvű modellt használtam. Egy nyelvtanulóknak szánt online szókincsgyűjteményből[12] válogattam össze a kiinduló listát, majd minden szóhoz lekérdeztem a 10 leghasonlóbb szót a FastText előre tanított magyar nyelvű modellje alapján. Ezt a műveletet megismételtem a bővített listára is. A kapott eredményeket kézzel is át kellett nézni, sok esetben ugyanis ragozott alakok és hibás vagy nem odaillő szavak is szerepeltek a találatok között. Ezzel a módszerrel végül 186 színnevet, 295 nemzetiséget (vagy származásra utaló kifejezést, pl. *nyugat-európai*) és 1106 foglalkozást sikerült listázni.

A szórendi változatok gyakoriságára mért eredményeket a 19-21. táblázatok tartalmazzák.

	Fókusz	**Igemódosító**
Hatóképzős	1 (0,1%)	855 (99,9%)
Fosztóképzős	45 (3,0%)	1460 (97,0%)
-szerű	35 (5,0%)	664 (95%)
-mentes	0 (0,0%)	451 (100,0%)
-ellenes	7 (10,3%)	61 (89,7%)
-képes	1 (0,1%)	1845 (99,9%)

19. táblázat. Bináris melléknevek a korpuszban

A 19. táblázatból látszik, hogy a bináris jellegű predikátumok csak viszonylag ritkán fordulnak elő fókuszpozícióban, semleges szórendű mondatokban viszont elég gyakoriak. A legkisebb különbség az *-ellenes* utótagnál látszik, ebből a típusból azonban meglehetősen kevés az adat, így a fókuszált *-ellenes* utótagú melléknevek arányaiban nagyobb gyakoriságából nem lehet következtetést levonni.

A többértékű esetekre (20. táblázat) már nem volt ennyi találat, és a két pozíció gyakorisága között is kevésbé drasztikus az eltérés, mint a binárisoknál, bár az igemódosító itt is jelentős fölénnyel bír.

A script arról is készített listát, hogy melyek azok a melléknevek, amelyek csak fókuszpozícióban fordulnak elő a korpuszban. Az összes (token szerinti) előfordulást figyelembe véve 89

[12] www.e-spanyol.hu

DOI: 10.15774/PPKE.BTK.2025.020

	Fókusz	**Igemódosító**
Színnév	15 (40,5%)	22 (59,5%)
Számneves kifejezés	27 (14,4%)	160 (85,6%)
Nemzetiségnév	76 (20,2%)	301 (79,8%)
Foglalkozás	65 (8,2%)	725 (91,8%)

20. táblázat. Többértékű melléknevek a korpuszban

	Fókusz	**Igemódosító**	**Összesen**
Bináris	48	5072	5120
Többértékű	183	1208	1391
Összesen	231	6280	6013

21. táblázat. A korpuszvizsgálat eredményei összesítve

ilyen van. Ezeket áttekintve azt találtam, hogy a leggyakoribb (32 darab) csak fókuszban álló jelző származásra utal, például: *helybéli, tatár, szegedi*. Ez a 20. táblázat eredményei után nem meglepő, hiszen ott is a nemzetiségnevek fordultak elő az egyik legnagyobb arányban fókuszpozícióban. A gyakoriság szerint következő szemantikai csoport 7 előfordulással a foglalkozásoké, például: *utcaseprő, gépíró, zöldséges*. Említésre méltó gyakorisággal (5-5 darab) fordultak még elő a valamilyen közösséghez tartozást (például párt tagja, sportcsapat szurkolója, szervezet dolgozója) jelölő melléknevek (például: *komcsi, dózsás, MÁV-os*) és a számneves kifejezések (például: *56-os, ötperces*).

Összességében tehát az mondható el a korpuszadatokról, hogy az igemódosító pozíció mindkét predikátumtípus esetén lényegesen gyakoribb, de az arányok közötti különbség a binárisok esetén jóval nagyobb. A többértékű predikátumok között, ha kisebb arányban is, de jelentős számban előfordul a fókuszálás, ezzel szemben a binárisoknál a fókuszpozíció előfordulását kifejezetten ritkának mondhatjuk.

DOI: 10.15774/PPKE.BTK.2025.020

5.2.5. Következtetések, összegzés

Mindkét vizsgálat világosan kimutatta, hogy van összefüggés a névszói predikátum bináris vagy többértékű tulajdonsága és a pozíciója között. A semleges mondat általánosságban véve preferáltabb, ez főleg a korpuszadatokon látszik. A többértékű predikátumok esetén a fókuszmozgatás opcionálisnak tűnik, vagy legalábbis az elvégzett vizsgálatok nem adnak információt annak pontos okáról. Úgy tűnik azonban, hogy bizonyos szemantikai kategóriák tagjai, például a származást, közösséghez tartozást vagy foglalkozást jelölő névszói predikátumok hajlamosabbak a fókuszmozgatásra. A bináris jelentésű melléknevek (és az azokat tartalmazó NP-k) ezzel szemben nagyon ritkán kerülnek fókuszpozícióba.

Mindebből arra lehet következtetni, hogy a fókusz valóban a releváns alternatívák jelenlétét jelöli. Ha a predikátum bináris jelentésű, akkor az általa jellemzett tulajdonsághoz nem tartozik alternatíva halmaz, ezért ezek nem lehetnek fókuszban.

A kutatás tehát részben választ adott a fejezet bevezetőjében feltett kérdésekre. Mind a kérdőív, mind a korpusz adatai alátámasztják azt az elképzelést, hogy ha a predikátummal jellemzett tulajdonsághoz nem tartoznak alternatív értékek a tulajdonság hiányán vagy annak ellentétén kívül, akkor a predikátum nem lehet fókusz. Rooth (1985) és Krifka (2008) fókuszelméleteit figyelembe véve nem meglepő, hogy a tipikusan bináris jelentésű predikátumok szinte kizárólag igemódosító pozícióban fordultak elő.

DOI: 10.15774/PPKE.BTK.2025.020

6. A predikatív névszó (segéd)igei tulajdonságai

Az 5.1.4. és 5.1.5. fejezetekben azt láttuk, hogy a kopula nélküli mondatok szórendje bizonyos esetekben megváltozik a kopulásokhoz képest, és a predikatív névszó igei viselkedést mutat a szórend tekintetében. Ez a rövid zárófejezet ezekre az esetekre és a predikatív névszók egyéb igei tulajdonságaira tér ki.

6.1. Igére jellemző szórend

6.1.1. Igéből képzett névszók és igekötő-leválás

Az igére jellemző szórend az igekötős igéből képzett névszói predikátumok esetén figyelhető meg a legegyértelműbben. Nem semleges mondatokban az igékről leválik az igekötő, és fordított (107a) vagy megszakított (107b) szórendről beszélhetünk. Ez az alfejezet azt vizsgálja, hogy mely esetekben jellemzőek ezek a szórendi változatok az igéből képzett névszókra.

(107) a. Még nem fejeztem be a levelet.

b. El sem olvasta a levelemet.

Ehhez először kigyűjtöttem a Szeged treebankből azokat a mondatokat, ahol egy névszónak van igekötője (`PREVERB` relációjú függősége), a keresést egyelőre nem szűkítve le a predikatív névszókra. A találatokat kézzel áttekintve a következő esetekben találkoztam a fordított szórenddel:

- **Tagadó mondatokban**
  - Hatóképzős névszóknál (108a)
  - -Andó kézpős névszóknál (108b)
- **Kérdő mondatokban**
  - Hatóképzős névszóknál (108c)
- **Egyéb fókuszos mondatokban**
  - Hatóképzős névszóknál (108d)

DOI: 10.15774/PPKE.BTK.2025.020

  - -Andó képzős névszóknál (108e)

(108) a. Rendszerkötet nem **terjeszthető ki**.

b. A biztonsági mentés nem **tévesztendő össze** az archiválással

c. De hogyan **határozható meg**, hogy a cím melyik része a hálózat?

d. A természet csak jelek segítségével **fordítható le** a piktúra nyelvére.

e. Szintén a szerződésekre irányadó általános szabályok köréből **emelendő ki** a Ptk. 217. § (1) bekezdése

A találatok túlnyomó többsége hatóképzős volt, a tagadó mondatoknál 3, az egyéb fókuszosoknál 1 példa volt az *-Andó* képzőre.

Más igéből képzett névszók esetén nem fordult elő a fordított szórend ((109) példái közül (109a) szerepelt a korpuszban).

(109) a. Iskolaidőben ez nem meglepő – *Iskolaidőben ez nem lepő meg.

b. Iskolaidőben ez nem meglepetés – *Iskolaidőben ez nem lepetés meg.

Míg a fordított szórend a keresés alapján csak a predikatív névszókra jellemző, addig a megszakított szórend előfordult nem predikatív névszók esetén is. A kapott megszakított szórendű találatok az alábbi típusokba oszthatók:

- Is/Sem
  - Hatóképzős predikatív névszóknál (110a)
- Tagadás
  - Nem predikatív főneveknél (110b)
  - Nem predikatív folyamatos melléknévi igeneknél (110c)
  - Nem predikatív hatóképzős névszóknál (110d)
  - Nem predikatív -Andó képzős névszóknál (110e)
  - Nem predikatív befejezett melléknévi igeneveknél (110f)

DOI: 10.15774/PPKE.BTK.2025.020

(110) a. Eszedbe jutott már, hogy ez a jelmondat **meg is fordítható**?

b. Legfelsőbb Bírósági döntés azonban nem született ebben a vitás kérdésben, így maradt a gyakorlatban az **egyet nem értés** és a bizonytalanság.

c. Gyufája nem volt, de a soha **meg nem szólaló** őr, aki az ételt hozta, hajlandó volt tüzet adni.

d. Pillanatok alatt minden bútort odarángatott a **be nem zárható** ajtó elé

e. Az összeg fele **vissza nem térítendő** támogatás.

f. Ez minimálisan a **ki nem fizetett** szerzői díjat jelenti.

A talált elváló példák közül a tagadó alakok a legjobban kereshetők, ezért ezekkel végeztem el az összehasonlítást a nem elváló megfelelőikkel. Ehhez azokat a mondatokat gyűjtöttem ki, ahol egy névszóhoz tagadószó tartozik, majd elvégeztem a névszók részletes morfológiai elemzését az emMorph (Novák, 2014) elemző segítségével. A találatok közül azokat tartottam meg, ahol a névszó elemzése tartalmazott igekötőt. Mivel a kutatás célja az igékre jellemző szórend általános vizsgálata, a továbbiakban nem különítettem el a fordított és megszakított szórendű példákat, ezekre együttesen "igekötő-leválás"-ként fogok hivatkozni.

A 22. táblázatban látható az igekötő-leválás gyakorisága a különböző típusú melléknévi igenevek esetén. (A "folyamatos" csoportba azok a folyamatos melléknévi igenevek kerültek, amelyek nem hatóképzősek.) Ez alapján az igekötő leválása a predikatív hatóképzős (97,97%) és a nem predikatív befejezett (93,90%) melléknévi igenevekre a legjellemzőbb.

	Predikatív		**Nem predikatív**	
	elváló	nem elváló	elváló	nem elváló
Hatóképzős	145	3	5	12
-Andó	3	1	7	0
Folyamatos	0	7	16	23
Befejezett	0	2	77	5

22. táblázat. Az igekötő-leválás gyakorisága melléknévi igeneveknél

Látható, hogy az elválás általánosan is preferált: az összes példából 253 volt elváló, és mind-

DOI: 10.15774/PPKE.BTK.2025.020

össze 53 nem elváló. A predikatív és nem predikatív esetek aránya nagyjából kiegyenlített (161, illetve 145 példa), a szófaji típusok gyakoriságában viszont jelentős eltérések vannak, így első ránézésre nem egyértelmű, hogy a predikatív tulajdonság vagy a szófaji besorolás-e a meghatározóbb az igekötő-leválásban. Ha csak a predikatív tulajdonság mentén vizsgáljuk az igekötő-leválás arányát (23. táblázat), azt látjuk, hogy az elváló igekötős eseteknek kicsit több, mint a fele predikatív, míg a nem elváló esetek háromnegyedét a nem predikatív példák teszik ki. Ez arra utal, hogy a predikatív igeneveknél gyakoribb az igékre jellemző szórend. Nem szabad azonban figyelmen kívül hagyni, hogy a predikatív elváló igekötős esetek szinte mindegyike hatóképzős, így egyáltalán nem biztos, hogy valóban (csak) a predikatív tulajdonság megléte vagy hiánya a döntő szempont az igekötő-leválásban.

	Predikatív	**Nem predikatív**
Elváló	58.50%	41.50%
Nem elváló	24.53%	75.47%

23. táblázat. Az igekötő-leválás aránya predikatív és nem predikatív melléknévi igeneveknél

Az igenevek morfológiai típusa szerint nézve az arányt (24. táblázat) szinte minden esetben 90% fölötti az igekötő-leválás aránya. Ez alól kivételt képeznek a nem hatóképzős folyamatos melléknévi igenevek, ezek gyakrabban maradnak együtt az igekötőjükkel.

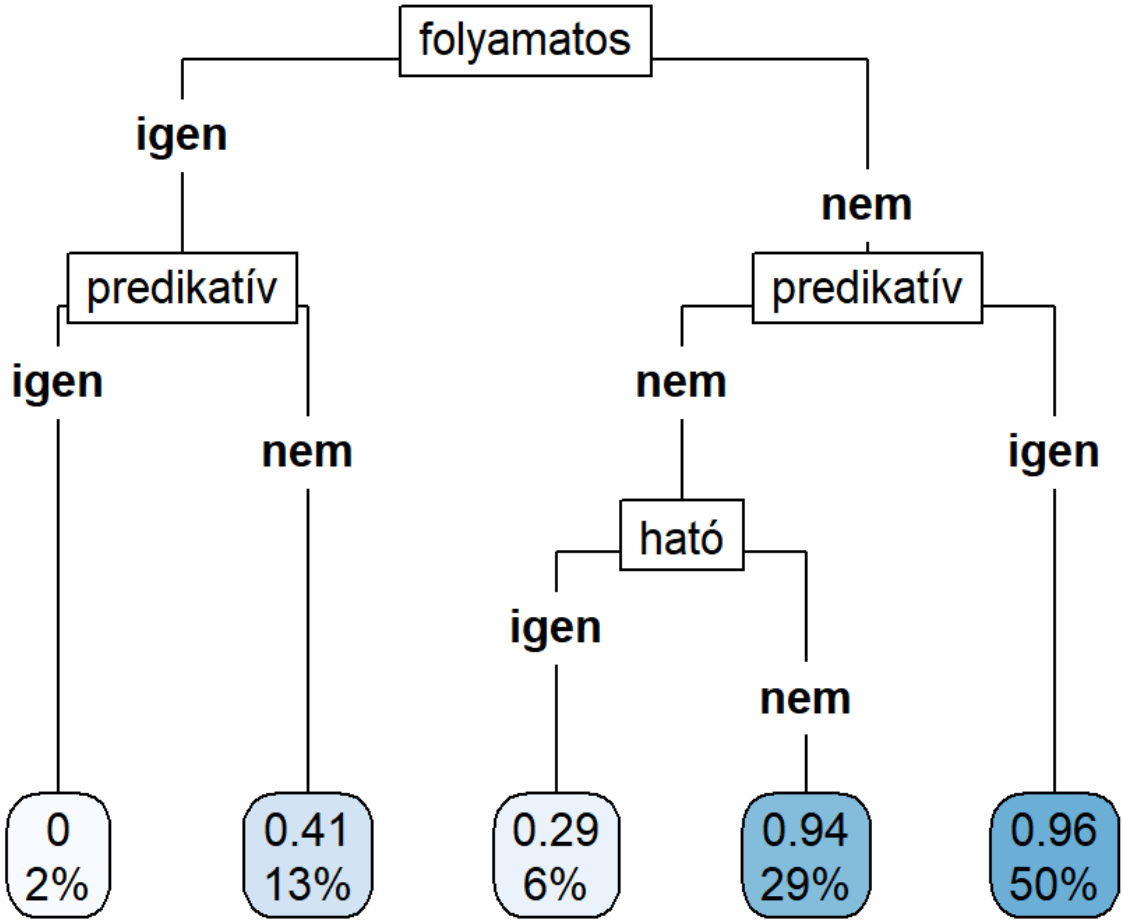
	Elváló	**Nem elváló**
Hatóképzős	90.91%	9.09%
-Andó	90.91%	9.09%
Folyamatos	34.78%	65.22%
Befejezett	91.67%	8.33%

24. táblázat. Az igekötő-leválás aránya melléknévi igeneveknél

A két szempont (a predikatív tulajdonság és a morfológiai típus) igekötő-leválásra gyakorolt hatását egy döntési fában (Breiman et al., 1984) foglaltam össze (15. ábra)[13]. A döntési fa egy grafikus modell, amely azt ábrázolja, hogy az egyes bizonytalan kimenetelű döntések milyen hatással vannak a várható kimenetelre (esetünkben az igekötő-leválásra). A kimeneten (a fa levelein) a felső

[13] A döntési fát Szappanos Balázs készítette R programnyelv segítségével.

DOI: 10.15774/PPKE.BTK.2025.020



15. ábra. Döntési fa az igekötő-elválás eseteiről

sorban látható számok az adott feltételeknek megfelelő példák arányát mutatják, például a nem predikatív folyamatos melléknévi igeneveknek 41%-a elváló. Az alsó sorban látható százalékok pedig egy összesített, becsült valószínűséget adnak, eszerint egy nem predikatív folyamatos melléknévi igenév az adatainkból becsülve 13%-os valószínűséggel fog elválni az igekötőjétől.

A 15. ábrán látható döntési fánk szerint a legnagyobb súllyal bíró döntés az, hogy a kérdéses igenév folyamatos-e. A folyamatos melléknévi igenekvnél az igekötő-leválás valószínűsége 15%, míg a nem folyamatosoknál 85%. A második elágazás a fában a predikatív tulajdonság, ez a folyamatosoknál csökkenti, a nem folyamatosoknál növeli az elválás esélyét. A nem predikatív nem folyamatosoknál pedig az a lényeges kérdés, hogy az igenév hatóképzős-e. A nem predikatív hatóképzős igenevek ugyanis ritkán válnak el az igekötőjüktől, a nem hatóképzősök (tehát kizárásos

DOI: 10.15774/PPKE.BTK.2025.020

alapon a befejezettek és az -Andó képzősök) jóval gyakrabban.

Összességében elmondható, hogy a szintaktikai szerepnek (predikatív tulajdonságnak) jelentős, de nem kizárólagos a szerepe az igére jellemző szórend gyakoriságában a vizsgált mondattípusokban. Az eredmények alapján azt gyanítom, hogy a másik szempont valójában nem a képzők típusa, hanem a melléknévként való lexikalizálódottság mértéke lehet. Ez magyarázhatja a folyamatos és a hatóképzős igenevek kirívó viselkedéseit. A folyamatos melléknévi igenevek ugyanis már gyakran lexikalizáltak melléknévként (pl. *meglepő, megfelelő*), a hatóképzősökre viszont valóban mondhatjuk, hogy ezek a "legigeibbek", hiszen az igei tő mellett a képzőjükben is őriznek igei tulajdonságot (a modalitást).

Mindez nem mond ellent az 5.1.4. fejezet azon állításának, miszerint az elváló igekötőjű hatóképzős igenevet tartalmazó tagadó mondatok esetén szabály alapon azonosítható a zéró kopula, "más esetben ugyanis az igekötő nem válhat le a hatóképzős névszóról", hiszen ott csak a fordított szórendről volt szó, itt pedig a megszakított szórendet is vizsgáltuk. A nem predikatív esetekre csak ez utóbbi szórend jellemző.

6.1.2. Az *-e* kérdő partikula helye

A másik predikatív szereppel (és igeiséggel) összefüggésbe hozható szórendi kérdés az *-e* kérdő partikula helye. A SMny. szerint az *-e* partikula az eldöntendő kérdések kérdő operátorának szerepét tölti be, melynek hatóköre a VP, ezért a VP-hez csatolt pozícióban jelenik meg (a 111. példa a SMNy-ból származik). Nominális mondatok esetén az *-e* kérdő partikula kanonikus helye a kopulán (112a), annak hiányában a predikatív névszón (112b) van. Néha azonban a kopula megléte esetén is a predikatív névszóra kerül a kérdő operátor (113), itt tehát egyfajta versengés figyelhető meg a szintaktikai szerep (melyik szó a predikátum) és a szófaj között (melyik szó az ige). A következőkben ennek a jelenségnek a korpuszalapú vizsgálatáról lesz szó.

(111) Találkoztál-e Imrével?

(112) a. Csakhogy én nem tudom, hogy ez tényleg igaz **volt-e**?

b. Csakhogy én nem tudom, hogy ez tényleg **igaz-e**?

DOI: 10.15774/PPKE.BTK.2025.020

(113) Jó-e volt a műsor?

A kutatás fő kérdése az, hogy mely esetekben fordul elő a 113. példában látott szórend, és hogy a jelenség összefügg-e a létige elhagyhatóságával? Azaz: hajlamosabb-e a predikatív névszó (vagy más nem igei elem) "átvenni" a létigétől az *-e* kérdő partikulát, ha a létige elhagyható?

A kutatáshoz az MNSz2-t használtam. A 113. példához hasonló mondatokat a következő keresőkifejezés segítségével kérdeztem le:

```
(114) <s/> containing ([msd="EKSZ"][msd="(IK\.)?IGE\.T?[MPF]?[et][123]"
      & lemma="van|lesz|van_lesz"])
```

A találatokból az MNSz2 keresőfelületének szűrő funkciójával kiszűrtem azokat, ahol a *van* szótő valójában *volna* segédigeként van jelen (*Tehetett-e volna mást?*), a maradék találatot pedig kézzel néztem át, és kiszűrtem a hibás annotáció miatti nem releváns találatokat, illetve azokat az eseteket, ahol az *-e* kérdő operátor több szón is szerepel (115). Így mindössze 38 releváns találat maradt. Ezekben azt vizsgáltam meg, hogy elhagyható-e belőlük a létige az alapértelmezett esetben.

A 25. táblázatban az "igen, mindig" sor azokra a mondattípusokra vonatkozik, ahol az alapértelmezett esetben soha nincs hangzó kopula. Az "igen, opcionálisan" kategória a bizonyos esetekben opcionálisan elhagyható létigéket foglalja magában (a példamondatok nem feltétlenül az elhagyható esetet képviselik), míg a "nem" sor esetén a mondat semmiképpen sem lehet jólformált hangzó kopula nélkül.

(115) Ezt akkor most szabad-e volt-e megfejtenem-e, vagy sem(e)?

Elhagyható-e a létige	**Gyakoriság**	**Példa**
igen, mindig	25	Na most, hogy ez elegendő-e volt?
igen, opcionálisan	6	Már ott-e lesz a Hajduk Split elleni BL-selejtezőn?
nem	7	igazam-e volt a témában

25. táblázat. A létige elhagyhatósága a *jó-e volt* típusú mondatokban

Bár a rendelkezésre álló adat nagyon kevés, az mégis látszik, hogy az *-e* kérdő operátor leginkább a "klasszikus" kopulás esetekben hajlamos a predikatív névszóra kerülni. Meglepő azonban,

DOI: 10.15774/PPKE.BTK.2025.020

hogy az opcionálisan elhagyható létigék esetén nem gyakoribb ez a szórend, mint a minden esetben kötelezően jelen levő létigéknél. Ez utóbbi két kategória példáit érdemes egyesével is megvizsgálni.

(116) a. Másnap Janicsányi szakaszvezető megkérdezte, hogy bélférgileg **rendben-e** van a honvéd.

b. Már **ott-e** lesz a Hajduk Split elleni BL-selejtezőn?

c. **Bent-e** vannak.

d. ...két azonos szóalak, s az egyik rossz, de ugyanabban az alakban benne van, **benne-e** van, benne-e, mindkét változat.

e. **Újév-e** lesz?

f. **Csütörtök-e** van, népek?

(117) a. Vizsgáld meg, hogy nem valamiféle **paranoia-e** van a dologban szerintem.

b. Arra gyűjtsél te, aztán linkeld be ide, hogy elolvashassam és megfontoljam, **igazam-e** volt a témában.

c. **Igy-e** van Jandro koma?

d. Nem tudom, **front-e** van, időjárás, alacsony nyomású napkitörés, telihód, teli-Jupiter, valami aszcendens vagy horoszkóp, de izzik a fejem folyton, bámulok bután magam elé egyre, kibírhatatlan ízű teákat iszom, és nem eszem mást, csak naponta egy fej fokhagymát.

e. És akkor ebből már jön az a kérdésem, hogy **igazuk-e** van azoknak, akik elismerik Makovecznek a nagyságát, de azt mondják, hogy nem annyira városi terekre való az amit ő tervezett.

f. És akkor összepuszilgatott engemet, s mondta, hogy vigyázzak magamra, meg hogy **jól-e** vagyok, hogy érezem magam, hol a Feri?

g. Mi továbbra is folytatjuk a konyhatakarítást Linus-szal, és a gyerekek továbbra is az

DOI: 10.15774/PPKE.BTK.2025.020

ajtónknál nyüzsögnek, kiabálva azt hogy: Hodi, hodi!, ami annyit jelent, hogy **Otthon-e** van valaki.

Az opcionálisan elhagyható létigét tartalmazó példák (116) között három lokatív mondat szerepel, mindegyik tartalmaz valamilyen helyre utaló demonstratív elemet (*ott, bent, benne*). Ezekről a 4.2. fejezetben megállapítottuk, hogy a demonstratív elem is tartalmazza a létezés jelentését, így a létige tartalmas jelentése gyengül és inkább inflexióhordozó segédigévé válik. A *rendben (van)* (116a) és az időre vonatkozó (116e-f) kifejezéseket szintén említettük már a 4. fejezetben. Utóbbiak érdekessége, hogy jelen időben kötelezően megjelenik a létige (*Újév van.; Szerda van.; 6 óra van.*), jövő időben viszont elhagyható (*Holnap újév.; Holnap szerda.; Mindjárt 6 óra.*).

A minden esetben kötelező létigés példák között kétszer fordul elő az *igaza van* kifejezés. Ez formálisan ugyan birtokos mondat, a jelentését tekintve azonban kérdéses, hogy valóban birtoklást fejez-e ki (vö. az angolban pl. kopulás mondattal fejezik ki ugyanezt: *they are right*). Látunk még (116a)-hoz hasonló, határozói bővítményes példákat (117c, g) és lokatív mondatot (117h) is, ezekből azonban nem hagyható el a létige. (117d) a Kádár (2004)-féle környezeti kopulás szerkezet körébe tartozik. A példák közül (117a) az egyetlen, amelyben a létige vitathatatlanul "tartalmas", pusztán valaminek a létezését kifejező ige, és egy "kicsit sem" kopula.

Elmondhatjuk tehát, hogy az *-e* kérdő partikula jellemzően valóban akkor nem kerül a létigére, amikor az kopulaszerű, a kopulaszerűség azonban nem feltétlenül jár együtt az elhagyhatósággal. Ha a létige tartalmas (létezést kigejező) igei jelentése gyengül, akkor az *-e* kérdő partikulát (a zéró kopulás mondatokhoz hasonlóan) a logikai állítmány szerepét betöltő mondatrész veszi át, ami az esetek többségében a predikatív névszó.

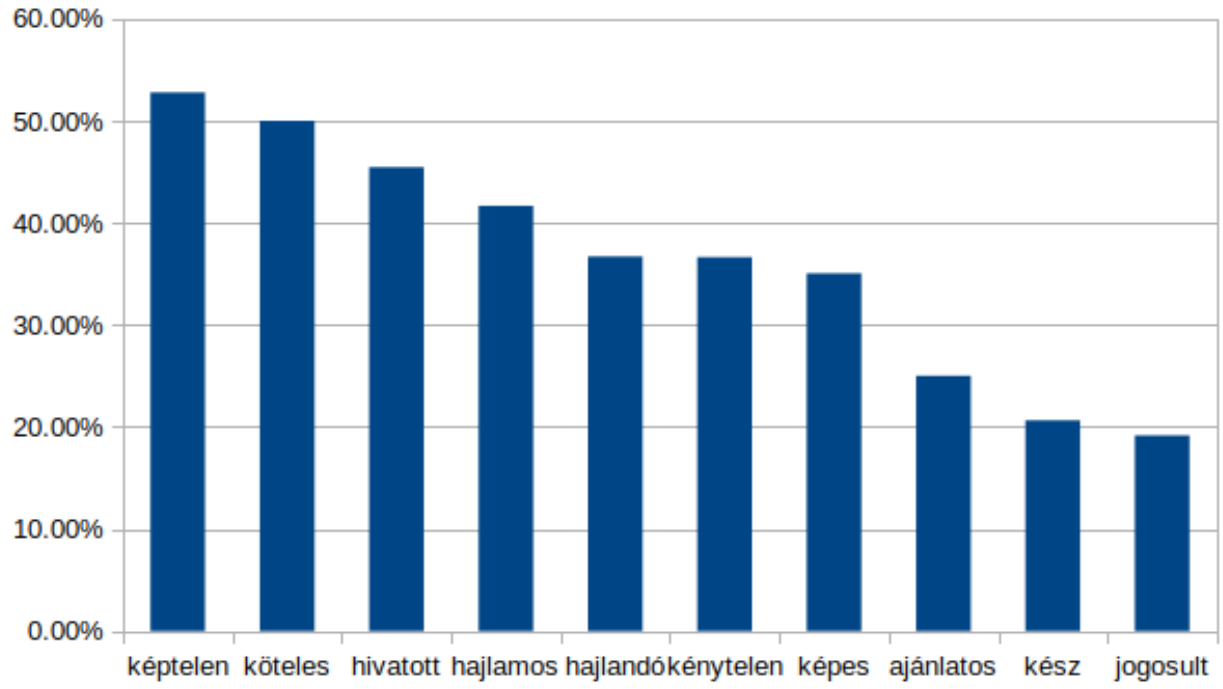
6.2. A predikatív névszók vonzatai

A predikatív névszók a szórend mellett a vonzatok tekintetében is hasonlóak az igékhez. A SMny. régensnek nevezi a vonzattal bíró szavakat, azaz azokat a szótári tételeket, "melyek meghatározott számú nyelvi egység jelenlétét követelik meg az őket befogadó mondatokban, és ezek mindegyikétől megkövetelik egy-egy grammatikai tulajdonságegyüttes kielégítését is". A régens általában ige, de előfordulnak más szófajok is régens szerepben, tipikusan ilyenek például az állítmányi névszók.

Ez a fejezet a predikatív névszók nem alanyi vonzatait veszi számba korpuszvezérelten, azon

DOI: 10.15774/PPKE.BTK.2025.020

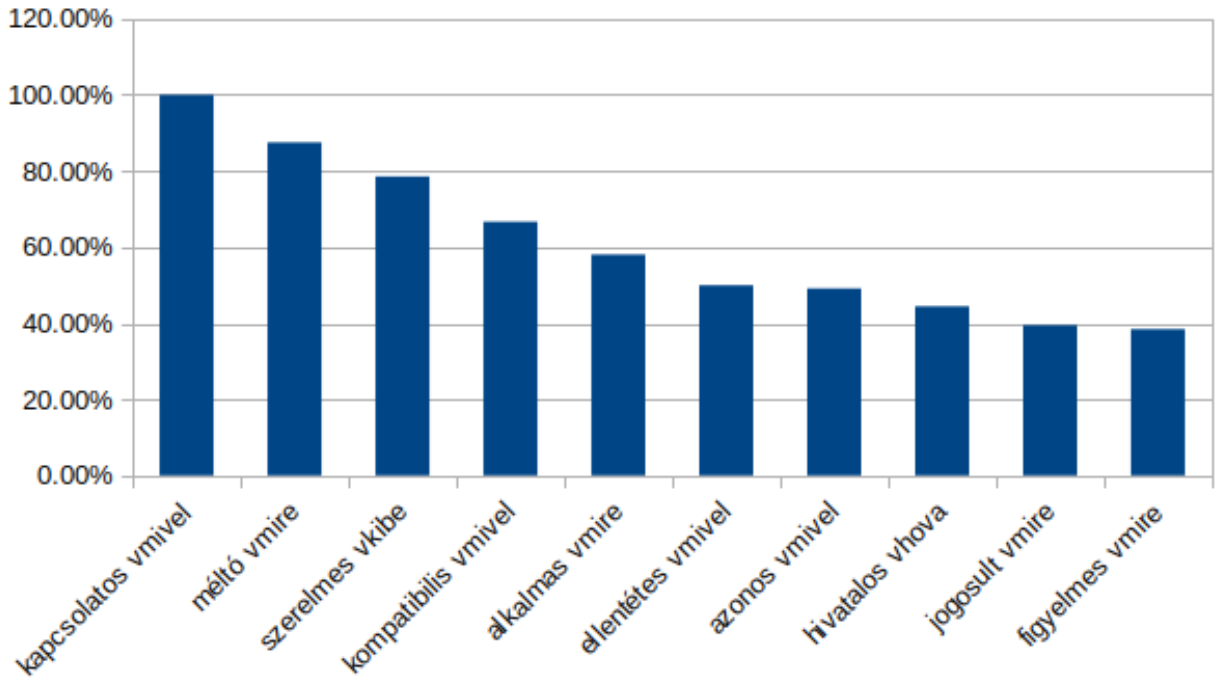
belül is a határozói és az infinitívuszi vonzatokat. Ehhez a Szeged Treebank függőségi annotációját használtam. A korpuszból (automatikusan) kigyűjtöttem az összes `PRED` relációban álló névszót és a hozzájuk tartozó infinitívuszi (`INF`) és határozói (`OBL`) vonzatokat. (Kerestem tárgyi vonzatokat is, de arra nem kaptam jó találatot.) Utóbbiaknál kézzel kiszűrtem azokat a találatokat, ahol a névszó igéből képzett és az ige vonzatait őrizte meg (pl.: *használható valamire*). Így infinitívusszal összesen 43-, határozóval pedig 82-féle vonzatos névszót kaptam. A találatokat relatív gyakoriság szerint rendeztem: az adott névszó vonzatos találatainak számát elosztottam a névszó összes predikatív szerepben való előfordulásának számával. A leggyakoribb infinitívuszi és határozói vonzatos predikatív névszókat a 16. és 17. ábrák mutatják. (Nem vettem figyelembe azokat a névszókat, amelyeknek az összes predikatív előfordulása 5-nél kevesebb volt.) A gyakorisági listák végén (1% alatti relatív gyakorisággal) az infinitívuszi vonzatoknál az *elegendő, könnyebb, jobb, jó, fontos*, a határozósoknál pedig a *jó valamire, nagy valakire, igaz valamire/valakire* áll.



16. ábra. A leggyakoribb infinitívuszi vonzatos predikatív névszók a Szeged Treebankben

Mint látható, az infinitívuszos és határozói vonzatos találatok átfednek. A 18. ábra mutatja a leggyakoribb infinitívusszal és határozói vonzattal is előforduló névszókat. Ez alapján a *hajlamos*, a *képtelen*, a *jogosult*, a *képes* és a *köteles* esetén gyakoribb a valamilyen vonzatos előfordulás a

DOI: 10.15774/PPKE.BTK.2025.020



17. ábra. A leggyakoribb határozói vonzatos predikatív névszók a Szeged Treebankben

vonzat nélkülinél.

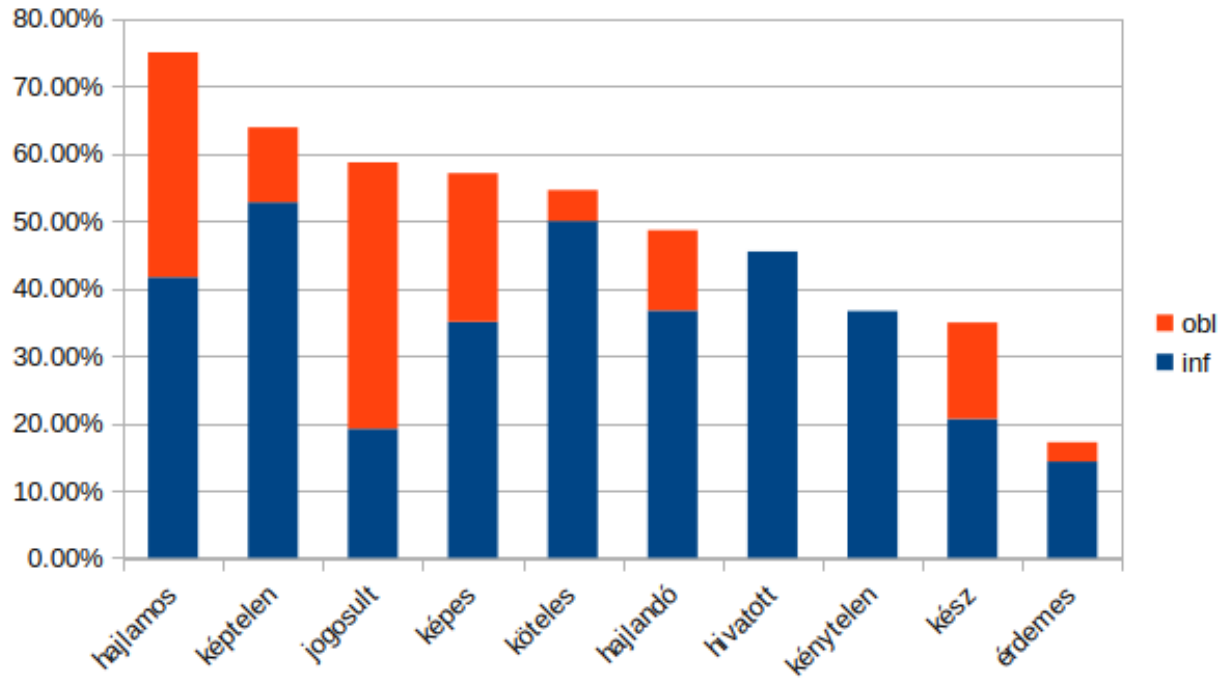
Mindehhez meg kell jegyezni, hogy a Szeged Treebank nem nagy, ezért több határozói vonzatos névszó azért nem került be a fenti ábrákba, mert csak egyetlen előfordulása volt (és így 100% relatív gyakorisága). Ilyenek voltak a következők: *vevő valamire, egyenértékű valamivel, hírhedt valamiről, mámoros valamitől, takarékos valamivel*. A ritka találatok közül érdemes még külön megemlíteni a *fenékig tejfel* és a *talpig úriember* szerkezeteket is. Ezeknél a vonzat lexikalizálódott, nem beszélhetünk *valameddig* tejfelről és úriemberről.

A méréseket érdekes lehet nagyobb korpuszon megismételni, egyelőre azonban nem áll rendelkezésre megbízható függőségi elemző eszköz vagy nagyobb méretű függőségi elemzett korpusz.

6.2.1. Infinitívuszi vonzatok

Az infinitívuszi vonzattal bíró predikatív névszókkal érdemes a segédigeiség szempontjai szerint is foglalkozni, a segédigékkel foglalkozó szakirodalom ugyanis gyakran említ infinitívuszi vonzattal bíró predikatív névszó + kopula szerkezetek is a lehetséges segédigék között.

DOI: 10.15774/PPKE.BTK.2025.020



18. ábra. Gyakoribb infinitívuszi és határozói vonzatos predikatív névszók a Szeged Treebankben

Kálmán C. et al. (1989) a magyar segédigék körét két fő ismérv, a csüggés és a beférkőzés alapján próbálja meghatározni. Csüggésről beszélünk, ha az infinitívusz melleti ragozott ige semleges körülmények között hangsúlytalan (tehát az infinitívusz van fókuszhelyzetben). A beférkőzés pedig az ige beférkőzését jelenti az infinitívusz és annak igekötője közé.

A "ragozott ige + *-ni* toldalékos alakok" segédigeiségének vizsgálatánál Kálmán C. et al. (1989) az ige fogalmába beleérti a névszó + (akár zéró) kopula felépítésű "összetett igéket" is. Ezek közül az alábbiak kerültek be a segédige-jelöltek közé: *hajlandó, hülye (lesz), ideje, jó, kár, kénytelen, képes, kész, köteles, muszáj, szabad, szokás, szükséges, tilos*. A két vizsgált ismérv alapján ezek közül a *szabad* és a *szokás* mutatott segédigei tulajdonságot. Mint Kálmán C. et al. (1989) is észreveszi, éppen ezek azok a gyakran infinitívusszal együtt álló predikatív névszók, amelyeknek vannak igei alakjai is: *szokott*, illetve *szabadott, szabadna, szabadjon*.

Kalivoda és Prószéky (2023) korpuszadatok alapján azt állapította meg, hogy az infinitívusszal együtt járó "segédigeszerű elemek" 11,2%-a névszói, és 88,8%-a igei tövű. A kettő között találhatók a Kálmán László[14] által "féligék"-nek vagy "alig-igék"-nek nevezett elemek, mint a *kell*, a

[14] https://m.nyest.hu/hirek/feligek-es-alig-igek

DOI: 10.15774/PPKE.BTK.2025.020

szabad és a *muszáj*. A *kell* alig-ige státusza mellett szól (igen korlátozott pradigmája mellett) az északkeleti nyelvjárásra jellemző *kellesz* alak, amely a névszói tövű segédigeszerű elemek mintájára a *kell* + *lesz* kopulás szerkezetből jött létre. A *szabad* és a *muszáj* pedig, bár alapvetően a névszók közé soroljuk őket, előfordulnak igei ragokkal is: *muszájna, muszájjon, szabadott, szabadna, szabadjon*. Míg a *muszáj* esetén ezek nem tekinthetők sztenderd alakoknak, a *szabad* bizonyos igei alakjai (*szabadna, szabadjon*) már az irodalmi nyelvben is elfogadottnak számítanak, itt tehát úgy tűnik, valódi átmeneti szófajról van szó. Kenesei (2001) segédigéket leíró munkájában az igévé válás útján levő szóként hivatkozik rá ("on the way of becoming a defective or irregular verb").

Megvizsgáltam a kérdéses fél-igék gyakoriságát az MNSz2-ben. A *muszájna* alakra 14 (118a), a *muszájjon*-ra 2 (118b) találatot kaptam, ezek tehát valóban marginális jelenségnek tűnnek. A *szabad* igei alakjainak gyakoriságai azonban már nem elhanyagolhatók, sőt, ezeknek az annotációjában is megjelenik az igei szófajcímke. (A *muszáj* esetén az előbb említett alakok `UNKNOWN` címkét kaptak.) A *szabadott* alak 812-szer (119a), a *szabadna* 4632-szer (119b), a *szabadjon* 492-szer (119c) fordult elő a korpuszban. A következő alfejezetben ezért ez utóbbiakat, a *szabad* igei ragos alakjait vetem össze a kopulás alakokkal.

(118) a. Hát azért rögtön kitenni őket nem **muszájna**.

b. Rajtam keresztül **muszájjon** csak nektek egymás fejét beverni?

(119) a. A civileknek látni sem **szabadott** őket, még csak sejteni sem.

b. Ilyet mégsem **szabadna** mondani.

c. Ezzel kapcsolatban **szabadjon** még egy dolgot elmondani

6.2.2. Az igévé válás útján: *szabad*

Mint említettem a *szabad* melléknévként és igeként annotálva is előfordul az MNSz2-ben, az annotálás pontos elveiről azonban nem érhető el dokumentáció, így elsőként a szófaji besorolás szempontjait kell feltérképezni.

A *szabad* szótőre keresve összesen 273700 találatot kapunk, a címkegyakoriságot lekérdezve pedig kiderül, hogy ebből 97606 alak, azaz a találatok 35%-a kapott igei címkét. Ezek eloszlását

DOI: 10.15774/PPKE.BTK.2025.020

Címke	Szóalak	Gyakoriság (db)	Gyakoriság (%)
IGE.e3	*szabad*	91671	93,92%
IGE.Fe3	*szabadna*	4630	4,74%
IGE.Me3	*szabadott*	812	0,83%
IGE.Pe3	*szabadjon*	492	0,50%
IGE.e3.KSZ	*szabad-e*	1	0,00%

26. táblázat. A *szabad* szótő igei annotációi az MNSz2-ben

a 26. táblázat mutatja. A Gyakoriság (%) oszlop az összes igei címkéhez viszonyított gyakoriságot jelenti. Az utolsó, egyetlen találatos sor pontos példája: *SZABAD-E SZÓLNI ÖNMAGUNK BAJAIRÓL?* Az IGE.e3.KSZ címke ebben az esetben hibás tokenizálás miatt keletkezett.

Az igeragos alakok néhány kivételtől (annotálási hibától) eltekintve minden esetben igei címkét kaptak. (A *szabadna* 2 esetben a SKIP címkével van annotálva.) Kérdés azonban, hogy a ragtalan *szabad* alakot mikor tekinti a korpusz igének, és ebben milyen szerepet játszik a kopula jelenléte vagy hiánya.

Elsőként az infinitívuszi vonzat és az igei annotáció összefüggését vizsgáltam meg. Ehhez először végeztem egy lekérdezést a *szabad* szóalakra (kis- és nagybetűvel is), és a találatokat szűrtem azokra, amelyekben van infinitívusz a *szabad*-tól legfeljebb 3 token távolságra. (kipróbáltam 5 tokenes keresési ablakkal is, de ez nem változtatott az eredményeken.) Ugyanezt megismételtem negatív szűréssel. Az infinitívuszos találatoknál újabb szűréseket csináltam a kopulára (*van*, *lesz* és *van__lesz* szótövek +/- 5 tokenes ablakban). Az igei címkék száma és aránya az így kapott találatokban a 27. táblázatban látható. Ez alapján azt lehet elmondani, hogy a segédigeszerű *szabad* esetén jellemzőbb az igei annotáció, de egyéb (nem infinitívuszos) esetekben sem kimondottan ritka. A kopula jelenléte valamelyest csökkenti az igei címkék arányát, ebben az esetben az igei vagy melléknévi annotálás randomnak tűnik.

Ha tehát abból indulunk ki, hogy a segédigeszerű *szabad*-nak kellene igei címkét kapnia, akkor azt tapasztaljuk, hogy az MNSz2-ben ehhez képest nagyon sok a fals pozitív találat. A 28. táblázatban az igei címkék pontosság- és fedésértékei láthatók, külön sorokban szerepeltetve azt az esetet, amikor a kopulás alakokat is igeinek tekintjük, illetve, amikor nem. Mindkét esetben nagyon alacsony pontosságot és közepes fedést kapunk. A segédigeszerű *szabad* lekérdezésénél tehát

DOI: 10.15774/PPKE.BTK.2025.020

Találatok típusa	Igei címkék száma	Igei címkék aránya
Infinitívuszos összes	13662	69,65%
Infinitívuszos kopulás	2240	52,67%
Infinitívuszos kopula nélkül	11197	74,35%
Nem infinitívuszos	77769	34,17%

27. táblázat. A *szabad* szóalak igei címkéinek aránya az infinitívusz jelenlétének függvényében

érdemesebb az infinitívusz jelenlétére hagyatkozni a szófajcímkék helyett.

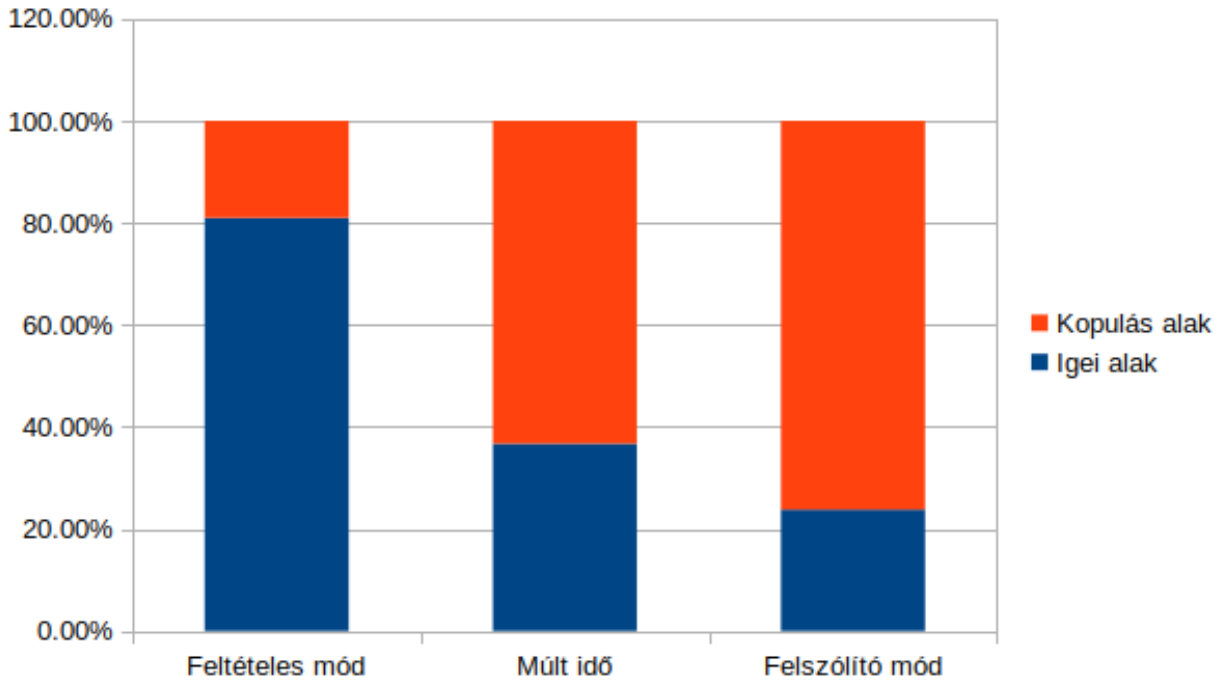
	Pontosság	**Fedés**
Kopulás alakokkal együtt	14,94%	69,65%
Kopulás alakok nélkül	12,81%	74,35%

28. táblázat. A *szabad* igei annotációinak pontosság- és fedésértékei az MNSz2-ben

A segédigeszerű *szabad* igévé válása leginkább abban mérhető, hogy milyen arányban fordul elő a korpuszban az igeragos és a kopulás változat. Az igeragos alakok gyakoriságához elegendő magukat a kérdéses szóalakokat (*szabadna, szabadott, szabadjon*) lekérdezni. A kopulás alakokhoz a szófajcímkék vizsgálatának tanulságai alapján a 27. táblázathoz használt infinitívuszos kopulás lekérdezés eredményét vettem figyelembe. Ebből automatikusan (Python script segítségével) állítottam elő egy gyakorisági listát a különböző kopulás alakokról. A részletes eredményeket a 29. táblázat, az eredmények összefoglalását pedig a 19. ábra mutatja. Utóbbi az igeragozott alakok gyakoriságát mutatja a hasonló jelentésű kopulás alakokhoz képest.

Amint látható, az igévé válás a feltételes módban a leglátványosabb: a *szabadna* jelentős többségben van a feltételes módú kopulás alakokhoz képest. A múlt idejű igealak (*szabadott*) az összes múlt idős eset valamivel több, mint egyharmadára jellemző. A legkevésbé preferált igei inflexió a felszólító módé: a *szabadjon* aránya nem éri el a 25%-ot. A feltételes mód fölényét magyarázhatja az is, hogy az összes nem kijelentő mód jelen idejű segédigeszerű *szabad* közül - az igeragozott és kopulás alakokat is figyelembe véve - általában is a feltételes módúak voltak a leggyakoribbak, sőt, amint a 29. táblázatban látható, múlt időben is gyakoribb a feltételes mód a kijelentőnél, mind az igei, mind a kopulás alakok esetén. Így nem meglepő, hogy az igévé válás folyamata is a feltételes módban mutatkozik meg leginkább.

DOI: 10.15774/PPKE.BTK.2025.020



19. ábra. A *szabad* igei és kopulás alakjainak aránya az MNSz2-ben

A 29. táblázatban észre lehet még venni, hogy a kopulás alakoknál feltűnően gyakori a fókuszos szórend. Bár nem vizsgáltam részletesen a jelenséget, de valószínűnek tartom, hogy ez a tagadó és tiltó mondatok gyakorisága miatt lehet.

6.3. Összefoglalás

A fejezetben ismertetett korpuszadatok két szórendi jelenségben, az igekötő-leválásban és az *-e* kérdő partikula helyében is kimutatták a predikatív névszók igei viselkedését, utóbbi olyan esetekben is, ahol a predikatív névszó nem igéből képzett. Az igekötős igéből képzett névszókra nem csak a predikatív esetben lehet jellemző az igekötő leválása, a nem predikatív eseteknél azonban csak a megszakított szórendet figyeltük meg, a fordított szórend kizárólag a predikatív névszókra jellemző. Az *-e* kérdő partikula esetén azt láttuk, hogy az leginkább a kopulákról és a kopulaszerű létigékről hajlamos a névszókra (vagy határozószókra) átkerülni.

A vonzatok vizsgálatánál is azt tapasztaltuk, hogy az igékre jellemző vonzatok számos nem igéből képzett predikatív névszónál is előfordulnak. Néhány infinitívuszi vonzatos névszó a segédigékkel foglalkozó irodalomban is feltűnik mint segédigeszerű elem. Ezek közül néhány (*kell,*

DOI: 10.15774/PPKE.BTK.2025.020

Feltételes mód		**Múlt idő**		**Feltételes mód**	
szabadna	4630	*szabadott volna*	435	*szabadjon*	492
lenne szabad	684	*szabadott*	377	*legyen szabad*	1376
volna szabad	377	*lett volna szabad*	829	*szabad legyen*	197
szabad volna	14	*volt szabad*	492	*lett légyen szabad*	1
szabad lenne	13	*szabad volt*	61		
		szabad lett volna	18		

29. táblázat. A *szabad* igeragos és kopulás alakjai az MNSz2-ben

muszáj, szabad) már el is indult a segédigévé válás útján, sőt a *szabad* esetén már van olyan igeragozott alak (*szabadna*), ami preferáltabb a kopulás megfelelőjénél (*szabad lenne/volna*). Ebben - ha még csak kezdeti stádiumban is - tetten érhető a kopula igei inflexióként való grammatikalizálódása.

DOI: 10.15774/PPKE.BTK.2025.020

7. Összegzés

Összegezve, a dolgozat fő eredményei a következők:

1. Egy angol-magyar párhuzamos korpusz segítségével létrehoztam egy olyan magyar nyelvű szövegkorpuszt, amelyben speciális címkék jelzik a zéró kopulák helyét. Ennek segítségével munkatársaimmal betanítottunk egy gépi fordítót arra, hogy tetszőleges mondatba beillessze a(z esetleges) zéró kopulát. Egy ilyen eszköz hasznos lehet a számítógépes mondatelemzés számára, a függőségi elemzési modellek ugyanis sok esetben (pl. a Szeged Treebank esetén is) alkalmaznak üres igei fejeket a kopula nélküli nominális mondatok elemzésére. Az általunk fejlesztett eszköz a tanítókorpusszal megegyező forrásból származó szövegeken közel 90%-os pontosságot ért el. Más szövegtípusokon tesztelve azonban ennél jóval alacsonyabb eredményt kaptunk. Ennek az az oka, hogy a tanítókorpusz szövegei, a speciális módszertan miatt, csak egyetlen műfajra (filmfeliratok) korlátozódnak. A zérókopula-beillesztő eszköz általánosan alkalmazhatóvá tételéhez szükség van a tanítóanyag más szövegtípusokkal való kibővítésére.

2. Korpuszmérésekkel kimutattam, hogy a zéró létige a magyarban nem csak a hagyományos értelemben vett névszói állítmányt érintő jelenség, hanem különböző feltételek mellett más típusú létigék is opcionálisan elhagyhatók jelen idő harmadik személyben. Ezek jelentős része állandósult kifejezés (pl. *rendben, vége, tele*, stb.) vagy műfajspecifikus jelenség (pl. irodalmi leírások). Néhány opcionálisan elhagyható létigét tartalmazó mondattípus azonban produktívnak mondható, ilyenek a helyre utaló demonstratív elemet (*itt/ott/hol*) vagy tagadó kifejezést (*semmi, sehol*) tartalmazó mondatok.

3. Az MNSz2-ből mért adataim alátámasztják azt a hipotézist, miszerint az igemódosító *itt/ott* gyengíti a módosított ige deskriptív tartalmát, így az valójában a létige (valamivel specifikusabb) szinonímájává válik. Ha a módosított ige maga a létige (*van*), akkor az emiatt szemantikailag redundánssá válik, tartalmas jelentése gyengül, a funkciója inkább kopulaszerű inflexióhordozóhoz lesz hasonló. Ezért ha a mondat jelen idő, egyes szám 3. személyben áll (azaz nincs jelölt inflexió), a létige zéró hangalakú is lehet. Ez elméleti magyarázattal szolgál arra a jelenségre, miszerint opcionálisan a lokatív és egzisztenciális létigék is elhagyhatók

DOI: 10.15774/PPKE.BTK.2025.020

a mondatokból, ha valamilyen, lehetőleg minél általánosabb jelentésű helyre utaló névmás (*itt/ott/hol*) előzi meg őket.

4. A tagadó formájú (*nincs, sincs*) létigék csak abban az esetben hagyhatók el a mondatból, ha az tartalmaz valamilyen egyéb tagadó értelmű kifejezést (*semmi, sehol, se*). Korpuszadatok alapján kimutattam, hogy bizonyos *semmi* + főnév kifejezések esetén különösen gyakori az ige nélküli szerkesztésmód, például: *semmi baj, semmi gond, semmi szükség (valamire), semmi ok (valamire), semmi kétség (valami felől), semmi köze (valamihez), semmi értelme (valaminek), semmi baja (valaminek/valakinek), semmi akadálya (valaminek), semmi kedve (valamihez)*.

5. Bizonyos speciális esetekben nem csak a létige, hanem annak hatóképzős alakjával azonos alakú segédige (*lehet*) is elhagyható a mondatokból. Korpuszgyakorisági vizsgálatok alapján megállapítottam, hogy a modalitást magába sűrítő, segédige nélküli infinitívuszos szerkesztésmód az experiens-alanyú, többargumentumú (elsősorban tranzitív) igékre jellemző leginkább, közülük is leggyakrabban az észlelést-érzékelést kifejező igékre. Bár a segédige nélküli infinitívuszos mondatok jelentése hasonló a hatóképzős névszói állítmányos mondatokéhoz, a két mondattípus között az igető-gyakoriságra koncentráló kutatásaim nem mutattak ki egyértelmű rokonságot. Megfigyelhető azonban, hogy az infinitívuszos és hatóképzős névszói állítmányos szerkesztésmód között stiláris különbség lehet.

6. Korpuszadatok alapján megmutattam, hogy a hagyományos "lexikális ige" – "inflexióhordozó funkcióelem" megkülönböztetés nem minden esetben magyarázza egyértelműen a létige elhagyhatóságát. Bizonyos esetekben ugyanaz a lexikális elem előfordul testes létigével és anélkül is. Ennek oka lehet a szófaji többértelműség (*kész, tele*), illetve a mennyiségre (*sok, kevés*) vagy állapotra (*rendben*) utalás. A "környezeti helyzetre" utaló névszók (*hideg, meleg, sötét*) esetén akkor szükséges a létige jelenléte, ha a mondatban nincs (explicit) alany, amire a jelzett tulajdonság vonatkozik. A *késő* esetén jelentésváltozás figyelhető meg a létigés és az ige nélküli szerkesztésmód között: előbbi általában napszakra, utóbbi pedig elkésett szándékra vonatkozik.

7. A kopulás mondatok szerkezetével kapcsolatban kérdőíves és korpuszalapú módszerrel iga-

DOI: 10.15774/PPKE.BTK.2025.020

zoltam, hogy az általában igemódosító pozícióban álló névszói predikátumok akkor fókuszálhatók, ha a jelentésük többértékűként határozható meg, azaz önmagukban is képesek alternatív jelentésű predikátumokat aktiválni. Jellemzően ilyenek a színnevek, számneves kifejezések, illetve a származást,közösséghez tartozást vagy foglalkozást jelölő névszói predikátumok. Ezzel szemben a bináris jelentésű melléknevek (amelyekhez nem tartoznak alternatív értékek az ellentétükön kívül) szinte soha nem kerülnek fókuszpozícióba. Ez megerősíti Rooth (1985) és Krifka (2008) elméletét, miszerint a fókuszálás a releváns alternatívák jelenlétét jelöli.

8. A *szabad* igeragozott alakjainak korpuszvezérelt vizsgálatával demonstráltam a kopula igei inflexióként való grammatikalizálódásának (kezdeti) folyamatát. A *szabad* három különböző igeragozott formában található meg a sztenderd köznyelvben (*szabadna, szabadott, szabadjon*), melyek közül a feltételes mód esetén a korpuszadatok alapján már jóval preferáltabbnak bizonyult az igeragos alak a kopulásnál.

DOI: 10.15774/PPKE.BTK.2025.020

Hivatkozások

Alberti Gábor és Medve Anna. *Generatív grammatikai gyakorlókönyv I-II*. Janus-Books, Budapest, 2002.

Balogh Péter. A főnévi igenév mint állítmány. *Magyar Nyelvőr*, 1890, 109–112.

Bohnet Bernd. Top Accuracy and Fast Dependency Parsing is not a Contradiction. In: *Proceedings of Coling 2010*, 2010, 89–97.

Bojanowski Piotr, Grave Edouard, Joulin Armand és Mikolov Tomas. Enriching Word Vectors with Subword Information. *CoRR*, 2016.

Breiman Leo, Friedman Jerome, Stone Charles J. és Olshen R. A. *Classification and Regression Trees*. CRC Press, 1984.

Clark E. V. Locationals: Existential, Locative and Possessive Constructions. In: Greenberg J. H. szerk. *Universals of Human Language IV. Syntax*. Stanford Univerity Press, Stanford, 1978, 85–126.

Deme László. A nyomatéktalan mondat egy fajtájáról. Az ott határozószó igekötőszerű használata. *Magyar Nyelv*, 1959, 55(2):185–198.

Dömötör Andrea, Yang Zijian Győző és Novák Attila. Much Ado About Nothing – Identification of Zero Copulas in Hungarian Using an NMT Model. Marseille, France. European Language Resources Association, 2020a, 4802–4810.

Dömötör Andrea, Yang Zijian Győző és Novák Attila. Nesze semmi, fogd meg jól! Zéró kopulák automatikus felismerése neurális gépi fordítással. In: Berend Gábor, Gosztolya Gábor és Vincze Veronika szerk. *XVI. Magyar Számítógépes Nyelvészeti Konferencia*, Szeged. Szegedi Tudományegyetem TTIK, Informatikai Intézet, 2020b, 385–398.

Dyer Chris, Chahuneau Victor és Smith Noah A. A Simple, Fast, and Effective Reparameterization of IBM Model 2. In: *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, 2013, 644–648.

DOI: 10.15774/PPKE.BTK.2025.020

É. Kiss Katalin. *The syntax of Hungarian*. Cambridge University Press, Cambridge, 2002.

É. Kiss Katalin. Egy igekötőelmélet vázlata. *Magyar nyelv*, 2004, 100(1):15–43.

É. Kiss Katalin. Deriving the Properties of Structural Focus. In: Lee C., Kiefer F. és Krifka M. szerk. *Contrastiveness in Information Structure, Alternatives and Scalar Implicatures. Studies in Natural Language and Linguistic Theory*, 91. kötet. Springer, 2017.

É. Kiss Katalin, Kiefer Ferenc és Siptár Péter. *Új magyar nyelvtan*. Osiris Kiadó, Budapest, 2003.

Grave Edouard, Bojanowski Piotr, Gupta Prakhar, Joulin Armand és Mikolov Tomas. Learning Word Vectors for 157 Languages. In: *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.

Hajič Jan, Ciaramita Massimiliano, Johansson Richard, Kawahara Daisuke, Martí Maria Antònia, Màrquez Lluís, Meyers Adam, Nivre Joakim, Padó Sebastian, Štěpánek Jan, Straňák Pavel, Surdeanu Mihai, Xue Nianwen és Zhang Yi. The CoNLL-2009 Shared Task: Syntactic and Semantic Dependencies in Multiple Languages. In: Hajič Jan szerk. *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL 2009): Shared Task*, Boulder, Colorado. Association for Computational Linguistics, June, 2009, 1–18.

Heggie Lorie. *The Syntax of Copular Structures*. Doktori disszertáció, University of Southern California, 1988.

Higgins Roger. *The Pseudo-Cleft Construction in English*. Garland Press, New York, 1973.

Indig Balázs, Sass Bálint, Simon Eszter, Mittelholcz Iván, Kundráth Péter és Vadász Noémi. emtsv – Egy formátum mind felett. In: Berend Gábor, Gosztolya Gábor és Vincze Veronika szerk. *XV. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2019)*, Szeged. Szegedi Tudományegyetem Informatikai Tanszékcsoport, 2019, 235–247.

Joulin Armand, Grave Edouard, Bojanowski Piotr és Mikolov Tomas. Bag of Tricks for Efficient Text Classification. *CoRR*, 2016.

Junczys-Dowmunt Marcin, Grundkiewicz Roman, Dwojak Tomasz, Hoang Hieu, Heafield Kenneth, Neckermann Tom, Seide Frank, Germann Ulrich, Aji Alham Fikri, Bogoychev Nikolay,

DOI: 10.15774/PPKE.BTK.2025.020

Martins André F. T. és Birch Alexandra. Marian: Fast Neural Machine Translation in C++. In: *Proceedings of ACL 2018, System Demonstrations*, Melbourne, Australia. Association for Computational Linguistics, July, 2018, 116–121.

Kádár Edit. Hány van van? In: Büky László szerk. *A mai magyar nyelv leírásának újabb módszerei VI.* Szeged, 2004, 53–62.

Kádár Edit. Environmental Copula Constructions In Hungarian. *Acta Linguistica Hungarica*, 2011a, 58(4):417–447.

Kádár Edit. *A kopula és a nominális mondatok a magyarban*. Akadémiai Kiadó, Budapest, 2011b.

Kalivoda Ágnes és Prószéky Gábor. Infinitívuszi szerkezetek a magyarban: adatvezérelt megközelítés. *Nyelvtudományi Közlemények*, 2023, 119:223–245.

Kálmán C. György, Kálmán László, Nádasdy Ádám és Prószéky Gábor. A magyar segédigék rendszere. *Általános nyelvészeti tanulmányok*, 1989, 17:49–103.

Kálmán László. *Magyar leíró nyelvtan. Mondattan I.* Tinta Könyvkiadó, Budapest, 2001.

Kenesei István. Criteria for auxiliaries in Hungarian. In: Kenesei István szerk. *Argument structure in Hungarian*. Akadémiai Kiadó, Budapest, 2001, 73–106.

Kent Allen, Berry Madeline M., Luehrs Fred U., Jr. és Perry J.W. Machine literature searching VIII. Operational criteria for designing information retrieval system. *American Documentation*, 1955, 6(2):93–101.

Kiefer Ferenc szerk. *Strukturális magyar nyelvtan 1., Mondattan*. Akadémiai Kiadó, Budapest, 1992.

Kiefer Ferenc. A kétféle igemódosítóról. *Nyelvtudományi Közlemények*, 2003, (100):177–186.

Kiefer Ferenc és Ladányi Mária. Az igekötők. In: Kiefer Ferenc szerk. *Strukturális magyar nyelvtan 3: Morfológia*. Akadémiai Kiadó, Budapest, 2000, 475–480.

Kocsány Piroska. Az ott szokatlan szerepei. Igekötő, diskurzuspartikula. *Jelentés és Nyelvhasználat*, 2021, 8(1):47–80.

DOI: 10.15774/PPKE.BTK.2025.020

Komlósy András. Régensek és vonzatok. In: Kiefer Ferenc szerk. *Strukturális magyar nyelvtan 1., Mondattan*, Budapest. Akadémiai Kiadó, 1992, 299–527.

Kratzer Angelika. Stage-level and individual-level predicates. In: Carlson G. N. és Pelletier F. J. szerk. *The Generic Book*. University of Chicago Press, 1995, 125–175.

Krifka Manfred. Basic notions of information structure. *Acta Linguistica Hungarica*, 2008, 55 (3–4):243–276.

Kudo Taku és Richardson John. SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Brussels, Belgium. Association for Computational Linguistics, November, 2018, 66–71.

Lengyel Klára. A mondatrészek. In: Keszler Borbála szerk. *Magyar grammatika*. Nemzeti Tankönyvkiadó, Budapest, 2000a, 399–465.

Lengyel Klára. A segédigék és származékaik. In: Keszler Borbála szerk. *Magyar grammatika*. Nemzeti Tankönyvkiadó, Budapest, 2000b, 252–258.

Lengyel Klára. Az állítmány. In: Keszler Borbála szerk. *Magyar grammatika*. Nemzeti Tankönyvkiadó, Budapest, 2002, 394–404.

Lison Pierre és Tiedemann Jörg. OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles. In: Chair) Nicoletta Calzolari (Conference, Choukri Khalid, Declerck Thierry, Goggi Sara, Grobelnik Marko, Maegaard Bente, Mariani Joseph, Mazo Helene, Moreno Asuncion, Odijk Jan és Piperidis Stelios szerk. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France. European Language Resources Association (ELRA), may, 2016.

Mejías-Bikandi Errapel. The nature of the stage/individual-level distinction and its syntactic reflex: evidence from Spanish. In: *Proceedings of the 11th West Coast Conference on Formal Linguistics*, 1993, 326–340.

Mel'čuk I. A. *Dependency Syntax: theory and practice*. State University of New York Press, Albany, NY, 1988.

DOI: 10.15774/PPKE.BTK.2025.020

Mel'čuk I. A. Levels of Dependency in Linguistic Description: Concepts and Problems. In: Agel V., Eichinnger L., Eroms H.-W., Hellwig P., Herringer H. J. és Lobin H. szerk. *Dependency and Valency. An International Handbook of Contemporary Research*, 1. kötet. W. de Gruyter, Berlin-New York, 2003, 188–229.

Milsark Gary. Towards an explanation of certain peculiarities of the existential construction in English. *Linguistic Analysis*, 1977, (3):1–29.

Minnen Guido, Carroll John A. és Pearce Darren. Applied morphological processing of English. *Natural Language Engineering*, 2001, 7(3):207–223.

Mitchell Tom. *Machine Learning*. McGraw Hill, New York, 1997.

Nádas Tímea. Hányféleképpen mondhatni? *Magyar Nyelvőr*, 2002, 3:329–351.

Novák Attila. A New Form of Humor – Mapping Constraint-Based Computational Morphologies to a Finite-State Representation. In: Calzolari Nicoletta, Choukri Khalid, Declerck Thierry, Loftsson Hrafn, Maegaard Bente, Mariani Joseph, Moreno Asuncion, Odijk Jan és Piperidis Stelios szerk. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland. European Language Resources Association (ELRA), May, 2014, 1068–1073. ACL Anthology Identifier: L14-1207.

Novák Attila, Laki László János és Novák Borbála. Mit hozott édesapám? Döntést – Idiomatikus és félig kompozicionális magyar igei szerkezetek azonosítása párhuzamos korpuszból. In: *XV. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2019)*, Szeged. Szeged University, 2019, 63–71.

Oravecz Csaba, Váradi Tamás és Sass Bálint. The Hungarian Gigaword Corpus. In: Calzolari Nicoletta és et al. szerk. *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*, Reykjavik, Izland. European Language Resources Association (ELRA), 2014, 1719–1723.

Orosz György és Novák Attila. PurePos 2.0: a hybrid tool for morphological disambiguation. In: *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2013)*, Hissar, Bulgaria. Incoma Ltd. Shoumen, Bulgaria, 2013, 539–545.

DOI: 10.15774/PPKE.BTK.2025.020

Pete István. Zérók és nem zérók, kis pro-k és nagy PRO-k a magyarban. *Magyar Nyelvőr*, 2004, 128(3):326–339.

Rooth Mats. *Association with focus*. Doktori disszertáció, University of Massachusetts, Amherst, 1985.

Schuster Sebastian és Manning Christopher D. Enhanced English Universal Dependencies: An Improved Representation for Natural Language Understanding Tasks. In: Calzolari Nicoletta, Choukri Khalid, Declerck Thierry, Goggi Sara, Grobelnik Marko, Maegaard Bente, Mariani Joseph, Mazo Helene, Moreno Asuncion, Odijk Jan és Piperidis Stelios szerk. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, Portorož, Slovenia. European Language Resources Association (ELRA), May, 2016, 2371–2378.

Seeker Wolfgang, Farkas Richárd, Bohnet Bernd, Schmid Helmut és Kuhn Jonas. Data-driven Dependency Parsing With Empty Heads. In: *Proceedings of COLING 2012: Posters*, Mumbai, India. The COLING 2012 Organizing Committee, December, 2012, 1081–1090.

Simkó Katalin Ilona és Vincze Veronika. Hungarian Copula Constructions in Dependency Syntax and Parsing. In: *Proceedings of the Fourth International Conference on Dependency Linguistics (Depling 2017)*, Pisa,Italy. Linköping University Electronic Press, September, 2017, 240–247.

Stassen Leon. Typology Versus Mythology: The Case of the Zero-Copula. *Nordic Journal of Linguistics*, 1994, 17(2):105–126.

Surányi Balázs. Ott 'there' in Hungarian and the theory of expletives. Kézirat, 2005.

Szabolcsi Anna. From the definiteness effect to lexical integrity. In: Werner Abraham és de Meij Sjaak szerk. *Topic, Focus and Configurationality*. John Benjamins, Amsterdam/Philadelphia, 1986, 321–348.

Szabolcsi Anna. *A birtokos szerkezet és az egzisztenciális mondat*. Akadémiai Kiadó, Budapest, 1992.

Szécsényi Tibor. Esetegyeztetés a predikatív főnévi csoportban, 2000.

DOI: 10.15774/PPKE.BTK.2025.020

Temesi M., Rónai B. és Vargha K. *Anyanyelvünk*. Művelt Nép Tudományos és Ismeretterjesztő Kiadó, Budapest, 1955.

Tesnière Lucien. *Eléments de Syntaxe Structurale*. Klincksieck, Paris, 1959.

Teubert Wolfgang. My version of corpus linguistics. *International Journal of Corpus Linguistics*, 2005, 10(1):1–13.

Tognini-Bonelli Elena. *Corpus Linguistics at Work*. John Benjamins, 2001.

Toutanova Kristina, Klein Dan, Manning Christopher D. és Singer Yoram. Feature-rich Part-of-speech Tagging with a Cyclic Dependency Network. In: *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*, NAACL '03, Stroudsburg, PA, USA. Association for Computational Linguistics, 2003, 173–180.

Vadász Noémi. KorKorpusz: kézzel annotált, többrétegű pilotkorpusz építése. In: Berend Gábor, Gosztolya Gábor és Vincze Veronika szerk. *XVI. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2020)*, Szeged. Szegedi Tudományegyetem, TTIK, Informatikai Intézet, 2020, 141–154.

Váradi Tamás, Simon Eszter, Sass Bálint, Mittelholcz Iván, Novák Attila, Indig Balázs, Farkas Richárd és Vincze Veronika. E-magyar – A Digital Language Processing System. In: chair) Nicoletta Calzolari (Conference, Choukri Khalid, Cieri Christopher, Declerck Thierry, Goggi Sara, Hasida Koiti, Isahara Hitoshi, Maegaard Bente, Mariani Joseph, Mazo Hélène, Moreno Asuncion, Odijk Jan, Piperidis Stelios és Tokunaga Takenobu szerk. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA), May 7-12, 2018, 2018.

Vaswani Ashish, Shazeer Noam, Parmar Niki, Uszkoreit Jakob, Jones Llion, Gomez Aidan N, Kaiser Ł ukasz és Polosukhin Illia. Attention is All you Need. In: Guyon I., Luxburg U. V., Bengio S., Wallach H., Fergus R., Vishwanathan S. és Garnett R. szerk. *Advances in Neural Information Processing Systems 30*. Curran Associates, Inc., 2017, 5998–6008.

DOI: 10.15774/PPKE.BTK.2025.020

Vincze Veronika, Szauter Dóra, Almási Attila, Móra György, Alexin Zoltán és Csirik János. Hungarian Dependency Treebank. In: *Proceedings of LREC 2010*, Valletta, Malta. ELRA, May, 2010.