

KALIVODA ÁGNES

IGEKÖTŐS SZERKEZETEK A MAGYARBAN

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, 2021

Tartalomjegyzék

Köszönetnyilvánítás	5
A fonológiai jelölések jegyzéke	7
1. Bevezetés	8
2. Módszertan	10
2.1. Korpuszvezéreltség	10
2.2. A kutatás alapjául szolgáló erőforrások	11
2.2.1. Magyar Nemzeti Szövegtár	11
2.2.2. A PREVLEX táblázat	14
2.3. Összegzés	17
3. Az igekötő-állomány meghatározása	18
3.1. A meghatározás problémái	18
3.2. Az igekötők prototípus-elméleti keretben	20
3.3. A vizsgált jegyek	21
3.3.1. Morfológiai produktivitás	22
3.3.2. Reduplikáció	25
3.3.3. A <i>-fele/felé</i> morféma csatolhatósága	28
3.3.4. Fokozhatóság	30
3.3.5. Személyragozhatóság	31
3.3.6. Szerkesztettség, szótagszám és gyakoriság	32
3.4. Igekötő × Jegy mátrixok	34
3.5. Jegyek közötti korrelációk	36
3.6. A prototípus és a prototipikus jegyek	40
3.7. Az igekötők jegyalapú osztályozása	41
3.7.1. Smith et al. (1988) módszere	41
3.7.2. Klaszteranalízis	43

3.7.3.	A módszerek közötti választás	45
3.8.	Összegzés	47
4.	Az igekötős szerkezetek szórendi mintázatai	50
4.1.	Szinkrón vizsgálat: a mai magyar nyelvállapot	51
4.1.1.	A kutatás módszere	51
4.1.2.	Finit ige	53
4.1.3.	Infinitívusz	61
4.1.4.	Határozói igenév	64
4.1.5.	Melléknévi igenév	66
4.1.6.	A beférkőző <i>nem, sem, se, is</i> az igei származékokban	72
4.1.7.	<i>El kell, hogy menjek – El kell menjek</i>	75
4.1.8.	Igekötős igék topikalizációja	79
4.2.	Diakrón vizsgálat: az ómagyar kortól napjainkig	81
4.2.1.	A kutatás módszere	82
4.2.2.	A prototipikus igekötők pozíció szerinti eloszlása	84
4.2.3.	A kétféle tagadó szórend arányának változása	87
4.2.4.	A fókuszos szerkezetek arányának növekedése	89
4.3.	Összegzés	89
5.	Az igekötők produktív kapcsolódási mintái	93
5.1.	A kutatás módszere	94
5.2.	Az igealkotás produktív esetei	96
5.2.1.	Névszó + igeképző	97
5.2.2.	Ige + igeképző	100
5.2.3.	Hangzásséma	101
5.2.4.	A vizsgált igealkotási módok arányai az igekötős szerkezetekben	104
5.3.	PREVCONS: Igekötős igei hapaxok adatbázisa	106
5.4.	Az igekötők ontológiája	110
5.4.1.	A poliszémiából adódó nehézségek	112
5.4.2.	Ontológiai entitások: igekötők és jelentések	114

5.4.3. Szemantikai relációk	117
5.5. Összegzés	120
6. Visszatekintés, újragondolás	122
6.1. Visszatekintés az igekötők állományára	122
6.1.1. Disztribúciós és szemantikai jegyek	123
6.1.2. A prototípus-elméleti keret revíziója	124
6.2. A szerkezetek felőli megközelítés alapjai	124
6.2.1. Az elképzelés	124
6.2.2. A kiindulás: a PREVDISTRO adatbázis	126
6.2.3. Adatelemzés a PREVDISTRO felhasználásával	128
6.3. Hogyan tovább?	131
6.4. Összegzés	132
7. Összefoglalás, új tudományos eredmények	134
Hivatkozások	141
Melléletek	155
I. számú melléklet	155
II. számú melléklet	156
III. számú melléklet	157
IV. számú melléklet	158
Magyar nyelvű összefoglaló	159
Summary in English	161

Köszönetnyilvánítás

Sokaknak tartozom hálával azért, hogy ez a disszertáció elkészülhetett. Mindenekelőtt a családomnak, amiért kezdettől fogva támogattak, biztattak és elviseltek. Köszönöm Apának, hogy lektorálta a dolgozatot, Anyának az első és az utolsó mottót, férjemnek, Tominak a spontán igekötős szerkezeteit, amelyekkel több alfejezethez ötletet adott.

Köszönöm a témavezetőmnek, Prószték Gábornak a belém vetett töretlen bizalmát és az inspiráló beszélgetéseket. Köszönöm, hogy azt mondta: „Azért gyere konzultálni, mert nekem jó!” Hálás vagyok azért, hogy támogatta az ötleteimet, többek között azt az – egyesek szerint merész – gondolatot, hogy É. Kiss Katalint és Kálmán Lászlót kérjük fel opponensnek.

Nagyon köszönöm nekik a disszertáció első verziójára adott alapos bírálatokat, melyeknek hatására, úgy érzem, sokat javult a végső változat. É. Kiss Katalinnak hálás vagyok azért is, hogy évek óta támogatta a munkámat tanáromként és projektvezetőmként, és amiért megtanított kutatni, cikket írni, előadni. Kálmán Lászlónak pedig azért tartozom külön köszönettel, mert a bírálata hatására újragondoltam sok mindent, és maradt gondolkodnivalóm a következő negyven évre is.

Köszönöm a munkahelyi vitám további bizottsági tagjainak, Surányi Baláznak, Ladányi Máriának és Reményi Andreának, valamint a vita iránt érdeklődő résztvevőknek az értékes megjegyzéseiket, és azt, hogy utána többen e-mailben is tettek még javaslatokat.

Köszönet illeti a Pázmány Péter Katolikus Egyetem Nyelvtudományi Doktori Iskoláját, ahonnan ösztöndíjban részesültem, valamint a Doktori Iroda munkatársait, amiért végtelen türelemmel viseltettek a lassúságom iránt. Külön köszönet Mocsári Andreának a kedvességéért és a segítőkészségéért, amire a leadás előtti napokban nagy szükségem volt. Köszönöm a Doktori Iskola vezetőjének, Surányi Baláznak, hogy egy elmaradt EDHT ülés kapcsán személyesen felügyelte a disszertációm sorsát.

Köszönöm a Nyelvtudományi Kutatóközpontban a legközelebbi kollégáimnak, Simon Lacinak, Lipp Verának és Sass Bálintnak, hogy nagy szeretettel támogattak, és időről időre emlékeztettek, hogy most még nem a nagydoktorit kell megírni. Bálintnak külön köszönöm, hogy „terjesztette az ígét”, és a munkássága rám is nagy hatással volt.

Óriási hálával tartozom a volt és jelenlegi kollégáimnak, a szolid italozások résztvevőinek, amiért éveken át támogattak, és a leadás utolsó pillanataiban is tartották bennem a lelket. Köszönet

ezért Simon Eszternek, Mittelholcz Ivánnak, Vadász Noéminek – aki a legjobb szobatárs –, Makrai Marcinak, Indig Baláznak, Arató Verának, Halász Dávidnak, Nyéki Bencének és Lévai Daninak. Külön köszönet illeti Esztert, amiért megvédett engem a világ minden gonoszságától, aztán megtanított rá, hogyan védjem meg saját magam. És külön köszönet Baláznak a sok szakmai és lelki támogatásért, de legfőképp a barátságáért, ami nagyon sokat jelent nekem.

Köszönöm a közös munkát Laki Lacinak, Yang Zijian Győzőnek és Ligeti-Nagy Noéminek. Köszönet illeti Pintér Lillát, aki mindig segítőkészen válaszolt a doktori képzéssel kapcsolatos kérdéseimre. Köszönet a csoporttársaimnak és barátaimnak, Dömötör Andinak és Langó-Tóth Áginak az együtt töltött évekért, azért, hogy néha együtt pánikoltunk és sokszor együtt ünnepeltünk. Köszönet a legjobb barátnőmnek, Nguyen-Dang Nórinak, mert nélküle megbuktam volna szintaxisból.

Köszönet minden egykori tanáromnak, különösen azoknak, akik nagy hatással voltak rám, bár erről talán nem mindegyikük tudott: Pécsváradiné Tóth Anikónak, Annus Ildikónak, Dési Editnek, Janurik Szabolicsnak, Ričardas Petkevičiusnak, Kohlmann Ágnesnek, Cser Andrásnak. Köszönöm Mádl Péternek, hogy programozni – és ami még fontosabb – gondolkodni tanított.

I would like to thank Farrell Ackerman and Rob Malouf for showing me new ways of linguistic thinking, and for our Saturday lab which was a huge help in reshaping my thesis. I am grateful for each time when *rájuk jön a hozzászólhatnék*.

És végül szívből köszönöm Olsvay Csabának mindazt, amit értem tett: a sok vigasztalást és biztatást, a dolgozat minden verziójának a gondos átolvasását és javítását. Az ő segítségével ez a disszertáció nem készülhetett volna el.

A fonológiai jelölések jegyzéke

C	mássalhangzó
C:	hosszú mássalhangzó
V	magánhangzó
V ₁ ... V ₁	azonos magánhangzók
A	<i>a</i> vagy <i>e</i>
O	<i>o</i> , <i>ö</i> vagy <i>e</i>
U	<i>u</i> vagy <i>ü</i>
Ó	<i>ó</i> vagy <i>ő</i>
/.../	fonéma (pl. /j/ → <i>j</i> vagy <i>ly</i>)
(...)	nem mindig megjelenő elem

1. Bevezetés

„A tárgy [...] rendkívül aktuális, újszerű és hatalmas szakirodalma van, ugyanakkor még senki nem foglalkozott vele.”

(Kálmán C. György: Módszertani útmutató opponensi véleményhez)

A mottóként idézett önellentmondás viccből hangzott el, mégis nagyon illik ehhez a dolgozathoz. A magyar nyelv igekötős szerkezeteivel sokan és sokféleképpen foglalkoztak már: elsőként Sylvester Jánost említhetjük, aki az 1539-es *Grammatica Hungarolatina* című művében *előlvető* névvel tárgyalta az igekötőt a kor szellemének megfelelően, a latin prepozíciók mintájára (Jakab 1982: 11-12). A későbbi nyelvtanírók fokozatosan távolodtak a latin nyelvtani szemlélettől, és felismerték az igekötők olyan – főként mondattani – tulajdonságait, amelyek alapján nem rokoníthatók a latin prepozíciókkal. Ezeket a felismeréseket tükrözik a jelenség későbbi megnevezései is: például a *Debreceni Grammatika* (1795) *elegyes előlutóljáró*, Virág Benedek (1810) *elő- s utol-vető* műszóval hivatkozik az igekötőre (Jakab 1982: 68-69).

Napjaink nyelvészeti szakirodalmában is élénk érdeklődés mutatkozik az igekötők iránt. Kerekes (2011: 110) a következőt írja erről: „A szakirodalom egységes abban, hogy egy nagyon egyedi, mégis összetett, mi több, ellentmondásos kategóriával van dolgunk.” Ezzel lényegében össze is foglalta azt, amiben a szakirodalom az igekötőket tekintve egységes. Máskülönben nagyon eltérő nézetekkel találkozunk, például azokban a kérdésekben, hogy az igekötő önálló szófaj-e vagy sem, önálló szó-e vagy képzőszerű elem, és egyáltalán melyik szavakat soroljuk az igekötők közé.

A mottó második fele azért állja meg itt a helyét, mert az igekötők állományának és az igekötős szerkezeteknek korábban nem készült korpuszvezérelt leírása. A dolgozatomban ennek a megvalósítására törekszem. A korpuszvezéreltség egy hangsúlyozottan adatközpontú szemlélet, amelyben a kutatás nagy mennyiségű szöveganyag automatikus elemzéséből indul ki, és elsődleges célja az introspekcióval nem feltárható jelenségek kimutatása. A kutatásom során többször a kognitív nyelvészethez fordulok, de nem köteleződök el egyetlen nyelvészeti irányzat mellett sem. Ezzel kapcsolatban Ladányi (2007: 13) gondolatát szeretném idézni – hozzá hasonlóan én is vállalom inkább „az eklekticizmus vádját, mintsem egyetlen választott elméleti keret (esetleg a szükségesnél szűkebbre szabott) korlátait.”

A dolgozat következő fejezetében a kutatás módszertani háttérét ismertetem. Ezután egy-egy

nagy fejezetben foglalkozom a főbb kutatási kérdéseimmel, amelyek a következők: (1) Mely lexikai elemeket soroljuk az igekötők közé, és mi alapján végezzük a besorolást? (2) Milyen mintázattípusokat mutatnak az igekötős szerkezetek, mikor és mennyire távolodhat el az igekötő az igétől, igenévtől, igei származéktól? (3) Hogyan írhatjuk le az igekötők produktív kapcsolódási mintáit, és ez alapján milyen megállapítások tehetők az igekötők szemantikájáról?

Ezeket egy rövid, de a mondanivalóját tekintve nagyon lényeges fejezet követi: először visszatérek a dolgozat elején bevezetett megközelítés kiértékelésére, majd ennek a problémáit látva felvázolok egy másik megközelítést, amelyben az igekötős szerkezetek kutatása nem az igekötőkből, hanem a szerkezetekből indul ki. A dolgozatot a kutatás főbb eredményeinek összefoglalásával és a téziseim megfogalmazásával zárom.

2. Módszertan

„Adatod lehet információ nélkül,
de információd nem lehet adat nélkül.”

(Daniel Keys Moran)

Elsőként röviden ismertetem azt az erősen adatközpontú szemléletet, amely a dolgozat egészét jellemzi. Ezután bemutatom azokat az erőforrásokat, amelyeket a teljes kutatás során felhasználtam. Az egyes kutatási kérdések megválaszolásához használt eszközöket és módszereket mindig a releváns fejezeteknél tárgyalom.

2.1. Korpuszvezéreltség

Néhány évtizede még élénk viták folytak arról, hogy a korpuszok – különféle szempontok szerint összeállított szövegtárak – szükségesek-e a nyelvi jelenségek vizsgálatához, mára viszont a nyelvészet legtöbb területén magától értetődővé vált ezek használata. A kompetenciát, vagyis az anyanyelvi beszélő nyelvi képességét kutató nyelvészek körében is egyre jellemzőbb, hogy nem kizárólag introspekcióval dolgoznak, hanem korpuszadaton is ellenőrzik a hipotéziseiket.

Ugyanakkor fontos hangsúlyozni, hogy a korpuszok használatának nem az az ideális módja, amikor a kutató célzottan az elméletét alátámasztó példákat gyűjt a korpuszból. A korpusznyelvészet túl akar lépni ezen: törekszik olyan jelenségek leírására is, amelyekre – egyelőre – nincs magyarázat (Teubert 2005: 125-126). Fontos szerepet kapnak benne a performancia-hibák, valamint minden olyan jelenség, amely a kompetencia felől nézve anomáliák sorának tűnhet. Ilyenekre bőven lesz példa ebben a dolgozatban.

Francis (1993) óta kétféle módszert szokás megkülönböztetni a korpuszvizsgálatokban. Az egyik a korpuszalapú (corpus-based), amellyel egy kialakult elméletet tesztelünk, és az adatok alátámasztják vagy cáfolják az elméletet. A másik a korpuszvezérelt (corpus-driven), amely kifejezetten azt hangsúlyozza, hogy előzetes elmélet nélkül kezdjük a kutatást, introspekcióval nem feltárható jelenségek kimutatása a cél, és tisztán az adatból kiindulva, indukción keresztül jutunk eredményre. A két módszer alapos leírását adja Sass (2011: 15-17).

Stubbs (2013) úgy véli, hogy a korpuszvezérelt módszert jellemzően objektívebbnek tartjuk a

korpuszalapúnál, de gyakran túlhangsúlyozzuk az indukció szigorúságát. Tanulmányában kifejti, hogy a tiszta indukciót már több mint négyszáz éve megkérdőjelezzük, az elmélettől mentes indukció lehetőségét pedig – legalábbis a társadalomtudományokban – általánosan elutasítják. Arra szeretné felhívni a figyelmet, hogy egy jelenség vizsgálatánál tulajdonképpen sosem a semmiből indulunk ki. Ha nincs legalább minimális elképzelésünk arról, hogy mit keresünk, akkor kevés esélyünk van arra, hogy megtaláljuk.

Az általam alkalmazott módszer alapvetően korpuszvezérelt, de nem mentes bizonyos előfeltevésektől. Feltételezem például, hogy az igekötő bármi olyan szótól elválhat, amely igekötős igéből képzett. Természetesen ez ilyen formában nem lesz igaz, de jó kiindulópontja a korpuszméréseknek, és segíti az olyan szerkezetek feltárását, amelyekkel a szakirodalom ritkán vagy esetleg egyáltalán nem foglalkozik. Emellett több esetben elméleti forrás adja az ötletet egy korpuszvizsgálathoz (ilyen például a növekvő összetevők törvényével kapcsolatos korpuszmérés, bővebben ld. a 4.1.2. fejezetben). Úgy vélem, hogy ennek a kutatásnak előnyére válik az induktív és deduktív módszerek keveredése.

2.2. A kutatás alapjául szolgáló erőforrások

A továbbiakban arról a két erőforrásról lesz szó, amelyet a disszertációhoz kapcsolódó összes kutatásomban felhasználtam. Elsőként bemutatom a Magyar Nemzeti Szövegtár 2.0.4 verzióját (Oravecz et al. 2014), amelyen a korpuszméréseim alapulnak. Ezután egy általam létrehozott lexikai erőforrásról lesz szó, a PREVLEX táblázatról (Kalivoda 2019). A PREVLEX 53 535 igekötős igét tartalmaz a gyakorisági adataikkal együtt. Manuálisan ellenőriztem, így a minősége és a mérete is alkalmassá teszik arra, hogy a különböző korpuszmérések során a hibás találatok jelentős részét ki lehessen szűrni a segítségével.

2.2.1. Magyar Nemzeti Szövegtár

A 20–21. századi szövegeket tartalmazó Magyar Nemzeti Szövegtár 2.0.4 verziója – a továbbiakban MNSZ2 – összesen 1,348 milliárd tokenből áll, amely írásjelek nélkül számolva 1,04 milliárd szövegszót jelent.¹ A korpusz hat stílusrétegből tartalmaz szövegeket: sajtó, szépirodalom, tudó-

¹ Az MNSZ korpuszának 2018. június 8-án megjelent egy újabb verziója, a 2.0.5-ös. Ez az összetételét tekintve nem változott az általam használt 2.0.4-es verzióhoz képest, viszont a meglévő szöveganyagban történtek javítások. Erről

mányos, hivatalos, személyes, beszélnyelvi. Ezen belül öt regionális nyelvváltozatra tagolódik: magyarországi, szlovákiai, kárpátaljai, erdélyi, vajdasági.²

A teljes szöveganyag mondatokra bontott, lemmatizált – tehát a szóalakok vissza vannak vezetve szótári alakokra –, valamint morfológiailag és szófajilag elemzett. Ennél magasabb szintű annotációt – pl. szintaktikai elemzést – a korpusz nem tartalmaz.

Két okból döntöttem az MNSZ2 használata mellett. Az első az volt, hogy a szövegek el vannak látva metaadatokkal is, amelyek közül többet – főleg a stílusréteggel kapcsolatos adatokat – külön is vizsgáltam, amikor ez a korpuszmérések eredménye alapján indokoltnak tűnt. A másik ok teljesen gyakorlati jellegű: lehetőséget kaptam arra, hogy közvetlenül az MNSZ2 forrásfájllal dolgozzak, a webes lekérdező felület használata nélkül. Ez egyrészt meggyorsította a különféle kereséseket, másrészt lehetővé tette azt, hogy a korpusz annotációját a kutatás igényeihez igazodva módosítsam. Erre a következő esetekben volt szükség:

1. Minden igekötős finit igénél és igenévnél, ahol kiderül ugyan az annotációból, hogy igekötős a szó (az elemzés IK.-val kezdődik), de sem a szóalakban, sem a lemmában nem látszik, hogy hol van az igekötő és az igei rész közötti határ. Így például automatikus adatfeldolgozás szempontjából a *felültettek* szó lehet *felül+tettek* és *fel+ültettek* is.
2. Olyan igekötős igéknél, igeneveknél és egyéb igei származékoknál, amelyek UNKNOWN elemzést kapnak, pedig egyébként létező szavak (pl. *bedepizni*, *összelájkolnotok*). Erről bővebben lesz szó következő fejezetben.
3. Igei származékoknál – amelyek nem finit igék vagy igenevek – nem derül ki az MNSZ2 jelenlegi annotációjából, hogy a képzés bemenetül szolgáló szónak van-e igekötője vagy sem. Tehát például a *megértettség* szó elemzése FN.NOM (alanyesetű főnév), az *elvárhatóan* elemzése HA (határozószó). Ezek mind helyes annotációk, de kevésnek bizonyulnak akkor, ha az igekötő szórendi helyét akarjuk vizsgálni – és a példaként hozott két szerkezet állhat elvált igekötővel is: *meg nem értettség*, *el is várhatóan*.
4. A -hAtÓ végződésű igeneveknél, amelyek mindig melléknévként annotáltak, pedig állítmányi helyzetben nagyon hasonlít a viselkedésük a finit igékéhez: az igekötőjük elválhat, lehet

bővebben itt lehet tájékozódni: <http://clara.nytud.hu/mnsz2-dev/hirek.html>

² Az egyes alkorpuszok méretére vonatkozó adatok ezen a linken érhetők el: <http://clara.nytud.hu/mnsz2-dev/stat.html>

hátravetett, és ekkor messzire is kerülhet az igenevektől (erről bővebben ld. a 4.1.5. fejezetet).

Az említett módosításokon túl három szűrést végeztem a forrásfájlon annak érdekében, hogy a lehető legjobb minőségű szöveganyagot kapjam. Egyrészt kiszűrtem a verseket, mivel sokuk nem természetes nyelvhasználatot tükröz. Másrészt – amennyire csak lehetett – eltávolítottam az idegen nyelvű, valamint a magyar, de ékezet nélkül írt mondatokat, mert torzíthatták volna a keresések eredményét. Például az ékezetet eleve nem tartalmazó igekötős igék sokkal gyakoribbnak tűntek volna, mint az ékezetesek. Ehhez azt a heurisztikát alkalmaztam, hogy töröltem minden olyan mondatot, amelyben a tokenek 80%-a UNKNOWN vagy SKIP elemzést kapott. Ez a módszer inkább a pontosságnak, mintsem a fedésnek kedvezett. Végül igyekeztem kiszűrni a korpuszban található duplumokat. Itt is a pontosságot tartottam szem előtt. Csak a nyolc tokennél hosszabb mondatokat vettem figyelembe a szűrésnél, feltételezve, hogy ennél rövidebb mondatoknál – pl. köszönéseknél – természetes lehet a többszörös jelenlét. Még ezzel az óvatos módszerrel is rendkívül magasnak bizonyult a duplumok aránya (20,12%), a személyes alkorpuszon belül akadt olyan – meglehetősen hosszú – mondat, amely száznál többször ismétlődött. A szűrések eredményét az 1. táblázat foglalja össze. Amikor az MNSZ2-re hivatkozok a dolgozat további részeiben, mindig az itt bemutatott, módosított verziót értem ezalatt.

korpusz	token	százalék
eredeti MNSZ2	1 348 000 000	100
versek	5 661 000	0,42
UNKNOWN / SKIP	26 825 200	1,99
duplumok	271 217 600	20,12
módosított MNSZ2	1 044 296 200	77,47

1. táblázat. Az MNSZ 2.0.4 mérete a versek, értelmes elemzés nélküli mondatok, valamint a duplumok szűrése előtt és után. A tokenszám írásjelekkel együtt értendő.

Úgy ítélt meg, hogy az MNSZ2 jelenleg az egyik legjobb minőségű, mai magyar nyelvállapotot reprezentáló korpusz, és az itt bemutatott módosítások után alkalmasnak találtam a kutatás céljaira.

2.2.2. A PREVLEX táblázat

Az MNSZ2 előző fejezetben bemutatott, módosított verziójából hoztam létre a PREVLEX-et,³ amely az igekötős igék jelenleg legbővebb (53 535 szavas), manuálisan ellenőrzött táblázata. Részét képezik a hapaxok – *hapax legomena*, azaz egyszer előforduló szavak –, valamint azok a szavak is, amelyek a korpuszban UNKNOWN címkével jelennek meg. Először arról lesz szó, hogy milyen munkafolyamattal, milyen megfontolások mentén készült a PREVLEX. Ezután az erőforrás szerkezetét ismertetem.

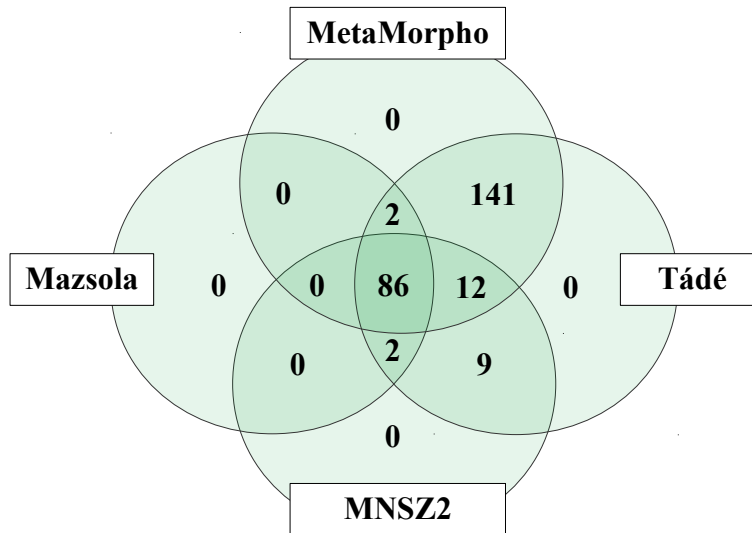
A PREVLEX előállításához először is szükség volt a lehetséges igekötők listájára, amelyet a Manócska⁴ integrált igei vonzatkeret adatbázisból nyertem ki (Kalivoda et al. 2018). Ebben minden olyan lexikai elem szerepel, amelyet a magyar igei vonzatkerettárak közül legalább egy igekötőnek jelöl. A Manócska készítése során a szerzőtársaimmal, Vadász Noémivel és Indig Balázssal együtt átnéztük ezeket a szavakat, és javítottuk az egyértelműnek tűnő hibákat (pl. a *vissz*, *nyug* igekötőnek jelölését a *visszhangoz*, *nyugdíjaz* szavak esetében). Ennek ellenére a végső lista hosszú – összesen 252 tagot számlál –, és több mint kétharmada esetében (pl. *szénné*, *pofon*, *zsebre*) az igekötői státusz vitatható. Ez jól tükrözi azt is, mennyire nincs egyetértés abban, hogy mely szavakat tekintjük igekötőnek (a problémakörrel részletesen foglalkozom a 3. fejezetben). Az 1. ábra azt szemlélteti, hogy melyik erőforrás hány igekötőt nevez meg. A MetaMorpho (Prószték et al. 2004) és a Tádé (Kornai et al. 2016) kezelik a legtágabban ezt a kategóriát. Az MNSZ2 alapján készült listák (Kalivoda 2016) feleannyi jelöltet sem tartalmaznak, a legszigorúbb pedig a Mazsola (Sass et al. 2010; Sass 2015), amelyben 90 igekötő szerepel.

A kiinduló, 252 szavas listában további 17 hibát találtam (ilyen pl. a *vízi*, amely egyszerűen bizonyos szóösszetételek első tagja). Így végül 235 lexikai elem maradt, amely az ismertett források valamelyike szerint igekötő, és a kutatásnak ebben a szakaszában én is fenntartottam ennek a lehetőségét.

A következő lépésben lekértem az MNSZ2-ből minden olyan finit igeként vagy UNKNOWN-ként annotált szót, amely egy adott igekötő-jelölttel kezdődik. Ennek a döntésnek két része is magyarázatra szorul. Először is az, hogy miért csak a finit igéket vettem figyelembe, amikor az igekötők például igenevekhez is kapcsolódhatnak. A korpuszvizsgálat során azt tapasztaltam, hogy

³ <https://github.com/kagnes/prevlex>

⁴ <https://github.com/ppke-nlpg/manocska>



1. ábra. A Manócskában szereplő erőforrások összesen 252 szót minősítenek igekötőnek. A halmazok azt mutatják, hogy az egyes erőforrások hány másikkal és hány darab szót illetően értenek egyet.

az igenevek esetében erős a tendencia az igemódosító és az igenév egybeírására (pl. *jóltáplált vendég, földreszállt angyal*), míg ugyanezeket az igemódosítókat a finit igével már kevésbé írják egybe. Valószínű, hogy az igenevek figyelembevétele nem változtatott volna jelentősen a PREVLEX összetételén, viszont sokkal több ellenőrizendő adathoz vezetett volna. Másodsor, az UNKNOWN szavakra azért volt szükség, mert sok jó találat csak így jelenik meg (pl. *visszacuccol, felstócol, benyammog*). Ugyanakkor az UNKNOWN szavak legnagyobb része hibás találat (elírt vagy idegen nyelvű szó), és a finit igék között is akadnak álpozitív találatok (pl. a *túlélősködik* mint igekötős ige). Emiatt az eredményül kapott, közel 178 000 szavas listát át kellett nézni.

Először eltávolítottam a lehetséges igekötőket a szavak elejéről, és a megmaradó szórészeket néztem át aszerint, hogy egyáltalán igék-e vagy sem. Ezután a már jóval rövidebb listát átnéztem úgy, hogy az ige az adott igekötő-jelölttel is megfelelő-e (ezen a szinten szűrtem ki pl. a *túlélősködik* és *feltűnősködik* igéket). Néhány olyan esetben, ahol az igekötő+ige kombináció nem volt értelmetlen, viszont nagyon valószínűtlennek tűnt, csak a konkrét szövegbeli előfordulások segítségével tudtam dönteni (pl. a *túltejesít*-ről így derült ki, hogy mindig a *túteljesít* hibásan írt változata). Ezután lokálisan újraelemeztem a forrásfájlt a javított adatokkal (pl. a korábban UNKNOWN *hype-olok, hype-ol* szavakat összevontam egy lemmává). A javított korpuszból állt elő a PREVLEX végső változata.

Az adatok átnézése során többször felmerült a kérdés, hogy bizonyos szóalakokat nem kellene-e

valahogyan normalizálni. Három esetben az ige okozott bizonytalanságot, mert (1) teljesen azonos jelentésű igeik történetileg eltérő tőváltozattal rendelkeznek (pl. *verekedik* – *verekszik*), (2) két igenek minimálisan eltérő töve van (pl. *gyömöszköl* – *gyömöcköl*), (3) egy-egy neologizmus többféle írásváltozatban létezik (pl. *dizájnol* – *design-ol* – *designol*). Egyedül az utóbbi csoport kapcsán voltam biztos abban, hogy a különbség csak ortográfiai jellegű. Ezeket a szavakat normalizáltam – rendszerint a magyar kiejtés szerint írt változatra –, mindenhol megőrizve az eredeti szóalakot is.

Elkülöníthető továbbá három olyan probléma, amely a képzőt érinti: (1) két vagy több ige képzőjében csak a kötőhang tér el (pl. *feccel* – *feccöl* – *feccol*), (2) opcionálisan *-ik*-es végződésű az ige (pl. *szörföz* – *szörfözik*), (3) ugyanaz az ige *-Vz(ik)* és *-Vl* képzővel is előfordul (pl. *offtopicol* – *offtopicoz*). Bár itt is szólhatnak érvek a normalizálás mellett, annyi biztos, hogy nem egyszerű ortográfiai különbségekről van szó. A (3)-asban látható példák egyelőre még ugyanazt jelentik, de elképzelhető, hogy idővel kis jelentésbeli eltérés kapcsolódik hozzájuk (ahogy azt pl. a *házal* – *házaz* párnál látjuk). A normalizálást ezekben az esetekben nem végeztem el.

Az erőforrás egy TSV fájlként érhető el, amely öt oszlopból áll. Az első oszlopban szerepel az ige (igekötő+igelemma formában). Ezt követi az MNSZ2-ben mért tokengyakoriság. A harmadik oszlopban kétféle érték szerepelhet attól függően, hogy az ige kapott-e megfelelő annotációt az MNSZ2-ben (FIN, ha igen és UNKNOWN, ha nem). A negyedik oszlop azt jelzi, hogy az ige hány dokumentumban fordult elő. Ez fontos információ lehet akkor, ha a tokengyakoriság és a tartalmazó dokumentumok száma nincs arányban (pl. az ige százszor fordul elő, de mindössze egy dokumentumban). Utolsóként szerepel a normalizált alak, amely csak a neologizmusok esetében térhet el az első oszlop tartalmától.

Bár az igekötős ige lista manuálisan ellenőrzött, a gyakorisági adatok fenntartással kezelendők. Néhány igealak ugyanis több lehetséges elemzéssel rendelkezik (pl. a *leszel* egyik lehetséges elemzése a *lenni* E/2. alakja, a másik a *leszel* igekötős ige). Ezek az elemzések sokszor eleve rosszak a forrásfájlban, így kissé torzíthatják a gyűjtött statisztikát.

A 2. táblázat számszerű áttekintést ad a PREVLEX-ről. A várakozásnak megfelelően az igekötős ige Zipf-eloszlást (Zipf 1932) mutatnak: néhány ige rendkívül nagy tokengyakorisággal bír, míg a hapaxok ritkák, de sokfélék.

kategória	típus	token
összes igekötős ige	53 535	12 145 914
hapaxok	21 116	21 116
UNKNOWN szavak	5 112	25 234
UNKNOWN hapaxok	3 543	3 543

2. táblázat. A PREVLEX számokban. Az értékek az eredeti igealakokra vonatkoznak, nem a normalizáltakra.

2.3. Összegzés

Ebben a fejezetben bemutattam az igekötős szerkezetekkel kapcsolatos kutatásaim módszertani hátterét. Röviden ismertettem a korpuszvezérelt megközelítést, és jeleztem, hogy bár alapvetően ez a szemlélet jellemzi a munkámat, a megállapításaim jelentős részéhez nem tisztán induktív általánosítással jutottam el, hanem különböző elméletek is inspiráltak. Ezután bemutattam a Magyar Nemzeti Szövegtár 2.0.4-es verzióját, amelyen a korpuszméréseim alapulnak, majd felsoroltam a módosításokat, amelyeket el kellett végezni a korpuszon ahhoz, hogy a kutatás céljaira alkalmas legyen. Végül beszámoltam az 53 535 szavas PREVLEX táblázatról, amelyet az MNSZ2 felhasználásával hoztam létre, és jelenleg az igekötős igék legbővebb gyűjteményének tekinthető.

3. Az igekötő-állomány meghatározása

„Ha lenne két nyelvész,
aki mindenben egyetértene,
az egyik fölösleges lenne.”

(Mádl Péter egy svéd nyelvórán)

Ebben a fejezetben arra teszek kísérletet, hogy körülhatároljam a vizsgálat körét, azaz meghatározom, hogy mit értek ‘igekötő’ alatt. Ahogy korábban is láttuk, ennek nincs egységesen elfogadott definíciója: a kapcsolódó szakirodalmat áttekintve alig találni arra példát, hogy két szerző azonos szempontok mentén határozná meg az igekötőket, és ugyanazokat a szavakat címkézné fel ezzel a terminussal. Így én is felteszem a kérdést, amelyet előttem már sokan feltettek: Mely lexikai elemeket soroljuk az igekötők közé, és mi alapján végezzük a besorolást? Ezzel kapcsolatban az nem reális célkitűzés, hogy a válasz egyezményesen elfogadható legyen, az viszont igen, hogy az oda vezető út konzisztens és adatvezérelt legyen, és az eredmény minél nagyobb átfedést mutasson a terminus korábbi használataival.

Elsőként nagy vonalakban áttekintem az igekötők meghatározásának elméleti és gyakorlati problémáit. Ezután – immár az általam javasolt megoldáshoz közelítve – ismertetem a prototípus-elméletet, amely lehetővé teszi az elmosódott kategóriahatárok rugalmas kezelését. A továbbiakban azzal foglalkozom, hogy a prototípus-elmélet hogyan ültethető át a gyakorlatba.

Ennek keretében néhány olyan automatikusan mérhető jegyet veszek sorra, amelyek segíthetnek körülhatárolni az igekötők csoportját. A fejezet legfontosabb szakasza az egyes jegyek részletes leírása, valamint ezek együttállásainak, korrelációinak megállapítása. Ezt követően arról lesz szó, hogy ha egyszerű gyakorlati okokból meg akarjuk húzni a határvonalat az igekötők és más igemódosítók között, arra milyen lehetőségeink vannak – az erre irányuló módszereket jegyalapú osztályozásnak nevezem. A fejezetet az eredmények összefoglalása zárja.

3.1. A meghatározás problémái

A definíciós problémák azzal a kérdéssel kezdődnek, hogy az igekötő önálló szófaj-e vagy sem. A hagyományos nyelvtanok az előbbi állásponton vannak, míg például a generatív szakirodalomban az a jellemző, hogy az ‘igekötő’ terminus a határozószóknak egy alosztályát jelenti (mint ahogyan

a tulajdonnevek a főneveknek egy alosztályát képviselik). Eszerint igekötők azok a határozószók, amelyek semleges mondatban igemódosító pozícióban – közvetlenül a finit ige előtt – állnak.

Akár önálló szófajnak, akár a határozószók egyik típusának tekintjük az igekötőt, azzal szembeülünk, hogy nincs éles kategória-határ. Az első esetben az a kérdés, hogy az igekötőket hogyan lehet elhatárolni más szófajú szavaktól, amelyek hozzájuk hasonló funkciót töltenek be, a második esetben pedig az, hogy az igekötőszerűvé vált, pusztán névszói igemódosítók szófaja névszó vagy határozószó. Ezért nem részletezem tovább a kérdést – az általam javasolt megközelítésnek akkor is megvan a létjogosultsága, ha az igekötőt önálló szófajnak, és akkor is, ha a határozószók egyik típusának tekintjük.

Az igekötő-állomány meghatározását az teszi problémássá, hogy az igekötők jelentéstani és mondattani viselkedésük alapján is hasonlítanak más lexikai elemekhez, amelyeket a továbbiakban pusztán névszói igemódosítónak nevezek (Kiefer 2007: 226-227). A két kategória szintaktikai hasonlóságát a 3. táblázat szemlélteti. Ez a hasonlóság még sok másra is kiterjed a mondatbeli pozíción túl, például arra, hogy jellemzően milyen szavak férkőzhetnek be az ige és az igemódosítója közé. Erről részletesen lesz szó a 4. fejezetben.

Igekötő	Pusztán névszói igemódosító
Kifestem a kerítést.	Kékre festem a kerítést.
Ki is festem a kerítést.	Kékre is festem a kerítést.
Nem festem ki a kerítést.	Nem festem kékre a kerítést.

3. táblázat. Az igekötő és a pusztán névszói igemódosító szintaktikai hasonlósága.

A 3. táblázatban szereplő *kékre* nem szokott felmerülni kérdéses esetként, de mi a helyzet az olyan szavakkal, mint például az *észre* (*észre+vesz*) vagy a *pofon* (*pofon+vág*)? Jakab (1976: 9-10) hét forrásművet vizsgál abból a szempontból, hogy melyik milyen szavakat sorol az igekötők közé. Arra jut, hogy a forrásművek összesen 28 elemet tekintve értenek egyet, ez a vizsgált 75 elemű halmaznak mindössze 37,3%-a.⁵ Az egyetértés hiánya lényegében változatlanul fennáll Jakab 1976-os tanulmánya óta is, erről bővebben ld. Komlósy (1992) és Kiefer (2007) munkáját.

A bizonytalanság nemcsak elméleti szinten jelentkezik, hanem a gyakorlatban is megnyilvánul.

⁵ A 28 elem a következő: *abba, agyon, alá, be, bele, el, fel, félbe, félre, felül, fenn, hátra, helyre, hozzá, ide, keresztül, ki, körül, közbe, le, meg, neki, oda, össze, rá, rajta, túl, vissza*. Ha ezt a halmazt további szakirodalmakkal vetnénk össze, még inkább csökkenne az egységesen igekötőnek minősített szavak aránya. Kiefer (2003b) például kizárná a *fenn*-t.

Ezzel kapcsolatban visszautalnék a 2. fejezetben bemutatott az 1. ábrára, ahol látható, hogy különféle lexikai erőforrásokban, korpuszokban mekkora eltérések tapasztalhatók az igekötőnek címkézett szavak mennyiségében. Itt említendő még az egybe- és különírás problémája is. Ha például az MNSZ2-ben bizonyos igemódosítók egybe vannak írva az igével – pl. *pórujár, békénhagy* – akkor a teljes szó UNKNOWN elemzést kap, míg a Tádében ezek igekötős igeeként szerepelnek.

Az igekötők és a puszta névszói igemódosítók elhatárolásának nehézsége nem elszigetelt eset. A legtöbb nyelvi jelenségnél megfigyelhető, hogy a kategóriahatárok elmosódottak (Langacker 1987: 14), és bár ilyen a jelenségek természete, az azokat leírni próbáló modellek számára nem praktikus, hogy elmosódottnak tekintsék a kategóriahatárt. Jellemzőbb az, hogy szükséges és elégséges feltételek mentén elhatárolt kategóriákkal dolgoznak. Ennek a módszertani mellékhatásaként szemmet kell hunyniuk az említett tapasztalati tények felett. A következő fejezetben egy olyan modellről lesz szó, amely megpróbálja elkerülni ezt a mellékhatást, és az elmosódó kategóriahatárok rugalmas kezelésére törekszik: ez a prototípus-elmélet.

3.2. Az igekötők prototípus-elméleti keretben

A prototípus-elmélet a fogalmi tudás kategóriákba rendezésének fontos eszköze a kognitív tudományokban. Az elmélet részletes kidolgozása Eleanor Rosch nevéhez köthető (Rosch 1973), aki elősorban kognitív és antropológiai szempontból foglalkozott az emberek kategorizálási képességével és a prototipikus egyed fogalmával. A nyelvészetben csak később jelent meg a prototípus-elmélet, először a holisztikus kognitív szemantika eszköztárában. Az egyik legalaposabb, magyar nyelvű ismertetését Kiefer (2007: 84-114) adja.

Az elmélet szerint azért tekintünk valamit egy adott kategória tagjának, mert a tulajdonságai elég közel állnak annak a példánynak a tulajdonságaihoz, amelyről úgy véljük, hogy a kategóriát legjobban képviselő példány (ez a prototípus). A kategóriához tartozás fokozat kérdése: egyes példányok jobban hasonlítanak a prototípusra, mások kevésbé. Ha egy magyar kultúrkörben nevelkedett beszélőnek fel kellene sorolnia az első három madarat, ami eszébe jut, valószínűleg legalább egyet említene ezek közül: veréb, fecske, galamb. Ezek a madár-kategória prototipikus példányai, mert rendelkeznek a „madárság” tipikus jegyeivel (ún. tipikalitási feltételeivel): tudnak repülni, van szárnyuk, van tollazatuk, van csőrük, stb. A kategóriának kevésbé tipikus példánya a strucc vagy a pingvin, mert repülni tudás híján mindjárt az első jeggyel nem rendelkeznek. Ugyanakkor

nem minden jegy egyformán fontos abból a szempontból, hogy egy adott példányt mennyire érzünk a kategória tagjának.

Az a hipotézis, miszerint az ‘igekötő’ fogalom a természetes kategóriák fogalmaihoz hasonlóan megragadható a prototípus-elmélettel, nem ismeretlen a szakirodalomban. A magyar igekötő-állomány meghatározása kapcsán elsőként Forgács (2005) tanulmányában jelenik meg a prototípus-elmélethez közeli gondolat, ő a prágai iskola centrum – periféria modelljét használja az igekötők kategorizálásához. Az igekötőség jegyeit elsősorban szemantikai szempontból, valamint a morfológiai szerkesztettség szempontjából közelíti meg. Kerekes (2011) már a prototípus-elméletre hivatkozva vizsgálja a magyar igekötő-állományt. Vizsgál szemantikai és morfológiai jegyeket – ezekhez Forgács (2005) és Jakab (1976) elemzését veszi alapul –, és bevezet újabb, formai jegyeket is, amelyek az egyes igekötőjelöltek különböző szerkezetekben való előfordulását jellemzik. Ezeknek a jegyeknek az értékeit kísérletes módszerrel, anyanyelvi beszélők ítéletei alapján állapítja meg. Én Kerekes (2011) módszerét követem annyiban, hogy empirikusan vizsgálom bizonyos jegyek meglétét az egyes igekötőjelöltek esetében. Abban viszont eltérek tőle, hogy ezt nem anyanyelvi beszélők tesztelésével, hanem korpuszméréssel valósítom meg. Az általam vizsgált jegyek elsősorban szóalaktani jellegűek, de a disszertáció végén kitérek a disztribúciós és szemantikai jegyek számításba vételére is (ld. a 6.1.1. fejezetet).

Úgy döntöttem, hogy bevonok a korpuszvizsgálatba minden szót, amely a PREVLLEX erőforrásban – ld. 2.2.2. fejezet – igekötőként szerepel, így a kiinduló adathalmaz összesen 235 tagot számlált.⁶ Minden szó esetében kigyűjtöttem az MNSZ2-ből a vizsgált jegyek előfordulásait.⁷ A továbbiakban ezeknek a korpuszméréseknek az eredményét mutatom be.

3.3. A vizsgált jegyek

Ebben a fejezetben az ‘igekötő’ fogalom olyan lehetséges jegyeit veszem sorra, amelyek (1) megfigyelhetők legalább egy lexikai elem esetében abból a 28 elemű halmazból, amelyet a Jakab által vizsgált összes forrásmű igekötőnek tekint – ld. a 3.1. fejezetet –, és (2) automatikus módszerrel

⁶ A továbbiakban mind a 235 lexikai elemet igekötőnek hívom az egyszerűség végett, amíg nem sikerül árnyaltabb felosztást találni.

⁷ Az egyes jegyek előfordulási gyakoriságát az olyan tagmondatokban vizsgáltam, ahol az igekötő finit igével állt.

viszonylag könnyen mérhetők. Ezek a morfológiai produktivitás⁸ (azaz hogy egy affixum mennyire vett és vesz részt új szavak alkotásában), a reduplikáció (pl. *meg-megállt*), a *-fele/felé* morféma csatolhatósága (pl. *szontyolodik elfele*), a fokozhatóság (pl. *összébbszűz*) és a személyragozhatóság (pl. *ráműr*).⁹ Végűl három általános jegyről lesz szó: a szerkesztettségűr, a szótagszáműr és a gyakoriságűr. Ezek minden szót jellemeznek, mégis érdemes vizsgálni űket, mert az igekötűk esetében – ahogy látni fogjuk – jellegzetes értékeik vannak.

3.3.1. Morfológiai produktivitás

Egy szóalkotási mintát akkor tekinthetűnk produktívűnak, ha a minta alapján tetszőleges számű, szemantikailag transzparens szó hozható létre egy adott szemantikai tartományban (Kiefer és Ladányi 2000c: 149). A morfológiai produktivitás esetében is – mint minden nyelvi jelenségűnél – célszerű rávilágítani a kompetencia és a performancia közötti különbségre. A nyelvi rendszer szintűjén a produktivitás egy lehetőség, a minden további nélkül létrehozható szóalakok nem biztos, hogy ténylegesen létrejönnek, és még kevésbé valószínű, hogy benne lesznek egy korpuszban. Emiatt több elméleti morfológus (pl. Dressler 2003) nem tartja helyesnek a kompetenciaszintű lehetőség performanciaszintű valószínűségeű alapuló vizsgálatát. Én mégis az utóbbit célzom meg – automatikus eszközűkkel jelenleg nincs is igazán más választás –, abban bízva, hogy az eredmény tendenciák szintűjén érdekes lehet a kompetenciát kutatóknak is.

A morfológiai produktivitás kvantitatív meghatározása Baayen (1989, 2009) nevéhez köthető. Három típusat különböztet meg: a megvalósult (*realized*), a terjeszkedű (*expanding*) és a lehetséges (*potential*) produktivitást. A következűkben arról lesz szó, hogy pontosan mik ezek a típusok, és hogyan jellemezhetnek egy-egy igekötűt.

A **megvalósult produktivitás** (P_m) annak a mértéke, hogy egy adott affixum a mérés időpontjáig mennyire vett részt a szóalkotásban, tehát a múltbeli és a jelenlegi szerepe jellemezhető ezáltal. Úgy kapjuk meg, hogy az affixumot (itt: igekötűt) tartalmazó lemmák darabszámát elosztjuk a korpuszban (itt: a PREVLEX-ben) található összes lemma darabszámával.

⁸ Erről külön publikáció született – Kalivoda (2019) –, ahol kissé eltérű számadatok szerepelnek, mint ebben a dolgozatban. Ennek az oka az, hogy időközben készűlt a PREVLEX-nek egy újabb, javított verziója, és ezen megismételtem a méréseket.

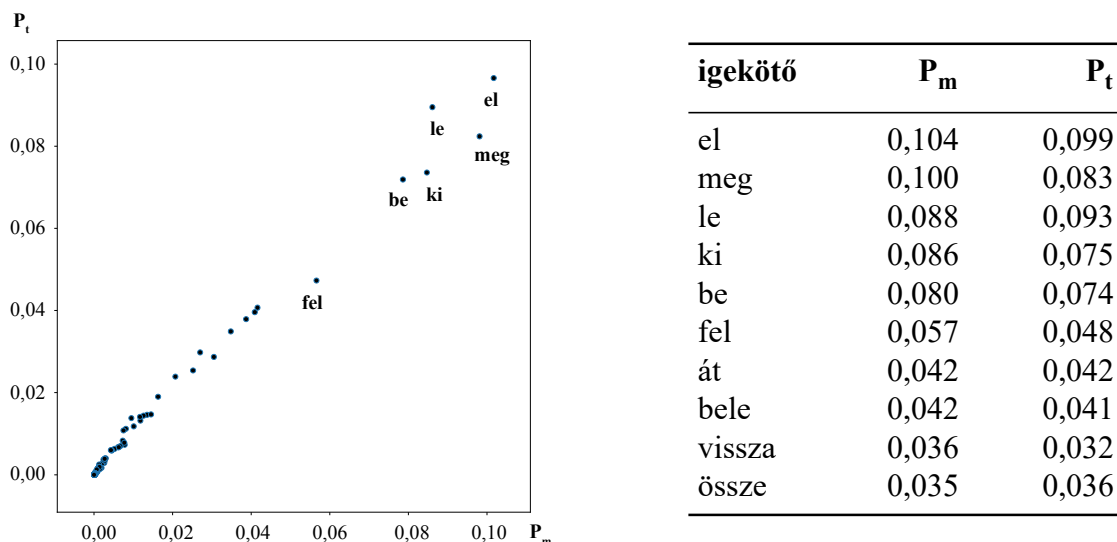
⁹ Azért írom ezeket egybe az igével, mert korpuszpéldát idézek. Helyesírási kérdésekre nem térek ki a dolgozatban.

$$P_m = \frac{\text{az affixumot tartalmazó lemmák (típus)}}{\text{a korpusz összes lemmája (típus)}}$$

A **terjeszkedő produktivás** (P_t) arról ad jóslatot, hogy az affixumnak várhatóan mekkora szerepe lesz a szóalkotásban a közeljövőben. Ehhez az affixumot tartalmazó hapaxok darabszámát osztjuk el a korpuszban található összes hapax darabszámával. A hapaxok jelentése jellemzően kompozicionális, ezért kevesebb „hamis produktív” találat adódik hozzá az eredményhez, mint a megvalósult produktivás esetében.

$$P_t = \frac{\text{az affixumot tartalmazó hapaxok}}{\text{a korpusz összes hapaxa}}$$

A PREVLEX igekötőinek megvalósult és terjeszkedő produktivásáról a 2. ábra ad áttekintést. Bal oldalt a két mérték összefüggése síkban ábrázolva látható, az X-tengelyen a P_m , az Y-tengelyen a P_t értékeivel. Jobb oldalt a tíz legmagasabb értékű igekötő szerepel.¹⁰



2. ábra. Az igekötők megvalósult (P_m) és terjeszkedő (P_t) produktivitása.

Az ősi igekötők (*meg*, *el*, *le*, *ki*, *be*, *fel*) mindkét mérték szerint kiugróan produktívak, bár a *fel* elmarad a többitől. Látható az is, hogy a kétféle produktivás nagyjából egyenesen arányos. A

¹⁰ Bizonyos igekötők (pl. *fel* – *föl*) egymás alakvariánsaiként is kezelhetők, és a mérés előtt összevonhatók. A kutatás során minden jegy értékét kimértem úgy, hogy ezek az alakvariánsok önálló tételként szerepeltek, és úgy is, hogy összevontam őket (erről bővebben a 3.4. fejezetben lesz szó). Itt csak az összevont verzió eredményeit mutatom be részletesen.

tendenciától csak a *le* és a *meg* térnek el. A *le* azért is figyelemre méltó, mert a terjeszkedő produktivitása nagyobb, mint a *meg*-é. Ez azt jelenti, hogy várhatóan az igekötős igék alkotásában is egyre nagyobb szerepe lesz.

A Baayen által definiált produktivás-típusok harmadik tagja a **lehetséges produktivás** (P_1). Ez az egészen távoli jövőről ad jóslatot: mik azok a most még viszonylag ritkán előforduló affixumok, amelyeknek jó esélye van arra, hogy később sok szó képzésében vegyenek részt? A terjeszkedő produktiváshoz hasonlóan ez is hapax-alapú mérték, de az eddig látottaktól lényegesen eltérő eredményt hoz. Úgy kapjuk meg, hogy egy adott affixumhoz tartozó hapaxok tokengyakoriságát elosztjuk az affixumhoz tartozó összes szó tokengyakoriságával.

$$P_1 = \frac{\text{az affixumot tartalmazó hapaxok}}{\text{az affixumot tartalmazó összes szó (token)}}$$

A méréshez célszerű gyakorisági küszöböt választani. Minél kisebb tokengyakorisággal osztunk, annál magasabb – és annál kevésbé informatív – lesz a lehetséges produktivás. A mérést 5-ös és 5000-es küszöbvel (ld. 4. táblázat) végeztem el.

igekötő	token	hapax	P_1	igekötő	token	hapax	P_1
szénné	5	5	1,000	agyon	7 588	264	0,035
égbe	5	3	0,600	körbe	11 285	287	0,025
torkon	9	5	0,556	ide	15 423	296	0,019
oldalba	18	9	0,500	körül	12 590	226	0,018
szarrá	19	9	0,474	hátra	8 256	137	0,017
tarkón	12	5	0,417	utána	8 415	126	0,015
fejen	16	6	0,360	végig	42 060	619	0,015
porba	8	3	0,380	előre	12 366	165	0,013
földre	27	9	0,333	újra	24 230	312	0,013
seggre	9	3	0,333	egybe	9 831	123	0,013

4. táblázat. Lehetséges produktivás (P_1), a 10 legmagasabb értékű szó 5-ös küszöb (bal oldalon) és 5000-es küszöb (jobb oldalon) mellett.

A 4. táblázat bal oldalán szereplő lexikai elemeket egyetlen nyelvészeti szakirodalom sem sorolná az igekötők közé. Tény viszont, hogy igekötőszerűbben viselkednek sok más igemódosítónál, és emiatt nem ritka, hogy az anyanyelvi beszélők egybe is írják ezeket az igével. Az előfordulásaik több dokumentumra oszlanak el, tehát nem csak egy ember szóhasználatát látjuk. A többségük egy

jól meghatározható szemantikai tartományban mutatja a produktivitás jeleit, például az *oldalba*, *torkon*, *tarkón* támadást kifejező igékkel állnak (pl. *szúr*, *vág*, *rúg*). A *szarrá* (*ázik*, *fagy*, *bombáz*) már absztraktnan értendő – és tulajdonképp a *szét* stilisztikailag jelölt változata.

A 4. táblázat jobb oldalán rangsorolt szavak státusza kevésbé megosztó: a szakirodalomban is felbukkannak igekötőként, bár az egyetértés ezeket illetően sem általános. Könnyen elképzelhető, hogy idővel lényegesen több új szó alkotásában vesznek majd részt. Az itt látható lexikai elemek mindegyike kapcsolódhat névszóból képzett igéhez (pl. *végigszambáz*, *telekommentel*, *agyonpár-náz*, *körbekordonoz*, *idekontárkodik*), ami táptalajt nyújt a kreatív szóalkotásnak.

3.3.2. Reduplikáció

A reduplikáció az a morfológiai művelet, amikor a tő egy részének az ismétlése valamilyen nyelvtani vagy szóképzési-szemantikai kategóriát fejez ki (Cser 2017: 12). Az igekötők esetében a reduplikáció teljes, vagyis a tő egészét ismételjük, például *fel-felnézett*, *vissza-visszatért*.¹¹

Az egyik legalaposabb munka, amely kizárólag ezzel a jelenséggel foglalkozik, Kiefer (1995) tanulmánya. A továbbiakban ennek a főbb pontjait foglalom össze, és ahol csak lehet, korpuszmé-
résekkel támasztom alá vagy kérdőjelezem meg a benne foglaltakat.

Kiefer a bevezetőjében leírja, hogy az igekötő kettőzésének iteratív funkciója van, tehát a cselekvés, történés ismétlődő jellegét fejezi ki. Az iteratív funkció képzése az akcióminőség-képzéshez tartozik, ezért az alapige jelentését nem módosítja. A tanulmány kutatási kérdései közül hármat emelnék ki, amelyek a jelen fejezet szempontjából relevánsak: (1) Mindegyik igekötő reduplikálható? Ha nem, mik ennek a korlátai? (2) A magyar nyelvnek van iteratív funkciójú szuffixuma is – ez az -(O)gAt –, felmerül tehát a kérdés, hogy milyen szemantikai különbségek vannak a reduplikált igekötős és a szuffixált igealakok között, már amennyiben vannak különbségek? (3) Eltérnek-e a reduplikált igekötős igék mondattani tulajdonságai a nem reduplikált igekötős igékétől? Ha igen, mivel magyarázható ez?

Az első kérdésre – reduplikálható-e minden igekötő? – Kiefer nemleges választ ad. Ezzel egyúttal cáfolja J. Soltész (1959: 176) állítását, miszerint „Önmagával ikerítődve minden igekötő alkothat megismételt igekötőt, akár van még konkrét jelentéstartalma, akár pusztán aspektus-kifejező

¹¹ Az igekötők reduplikációjára használt egyéb terminusok: megismételt igekötők (J. Soltész 1959), megkettőzött igekötők (Kiefer és Németh 2012), duplikált igekötők (Kerekes 2011). Az igekötők reduplikációjával kapcsolatban ld. még Piñón (1991), Ackerman (2003).

vagy képzőszerű funkciója van.” Kiefer négy megszorítást fogalmaz meg, amelyek a reduplikáció megjelenését korlátozzák. Az első megszorítás a vizsgált elemek fonológiai súlyára vonatkozik: eszerint két szótagonál hosszabb igekötők nem reduplikálhatók. Ezt az állítást később minden igemódosítóra kiterjeszti. Ez alól egy kivételt találtam, az *előre* elemet. Kiefer ezt nem tekinti igekötőnek, de az igemódosítók közül nem zárja ki. Besorolástól függetlenül tény az, hogy a vizsgált anyagban ez többször is megjelenik reduplikált formában, ld. az (1) példasort. A metaadatok tanúsága szerint különböző szerzők használják így. Ettől a kivételtől eltekintve a fonológiai súlyra vonatkozó megszorítás érvényes. A korpuszadatokon az is megfigyelhető, hogy a reduplikáció egy szótagú igekötők körében gyakoribb jelenség, mint két szótagúaknál.

- (1) a. [...] lovasát a lejtőn **előre-előrerúgta**.
- b. [...] s édesanya is **előre-előrejött** hozzánk, [...]
- c. Kesserű mama **előre-előreszól** a fiatalokhoz, [...]
- d. [...] és minden tanévben egyre-egyre jobban **előre-előrelépnek** ebben is.

Kiefer további megszorításai már nemcsak az igekötővel kapcsolatosak, hanem a konkrét igétől is függenek. Ilyen a konceptuális megszorítás, amely szerint az intenzitást, túlzásba vitelt kifejező igekötős igék esetében nem jelenhet meg reduplikáció. Tehát nem működhet például a **túl-túlkompenzál* vagy az **agyon-agyondicsér*. Az akcióminőséggel, szemantikával kapcsolatos megfigyeléseket nem igazán lehet az MNSZ2-ben ellenőrizni az ehhez szükséges annotáció hiányában. Annyi viszont biztos, hogy a *túl* igekötő egyáltalán nem jelenik meg reduplikált formában, az *agyon* pedig csak úgy, hogy az igével együtt szó szerinti értelemben (‘megöl’) szerepel, ld. a (2) példát.

- (2) [...] azokat **agyon-agyonverik** bevapumpával, akár rájuk igazolódik valami, akár nem.

Végül Kiefer megfogalmaz két szemantikai megszorítást is. Az egyik azt mondja ki, hogy statikus igék esetében nem lehetséges az igekötő reduplikációja, tehát nem fordulnak elő például a következők: **el-eláll (az étel)*, **el-elfér (valami valahol)*. Kiefer szerint ezek a statikus igék nem tűrik meg az iteratív értelmű -(O)gAt szuffixumot sem. A második szemantikai megszorítás szerint az olyan állapotigék, amelyek visszafordíthatatlan változást jelentenek, szintén nem engedik az igekötő reduplikációját. Nem működik tehát az, hogy **a szomszéd meg-megöregedett*, mivel

ez a történés nem megismételhető egy konkrét alanyra nézve. Ez ugyanakkor nem az igekötős ige állandó tulajdonsága, hanem a szöveggörnyezettől is függ.

Kiefer a második kutatási kérdése kapcsán – van-e szemantikai különbség a reduplikált igekötős és az -(O)gAt képzős igealakok között – tisztázza, hogy az -(O)gAt képzőnek két funkciója van, az egyik a cselekvés vagy történés ismétlődő jellegének a kifejezése, a másik a csökkentett intenzitás kifejezése. Ha az igealakhoz reduplikált igekötő járul, és az -(O)gAt képző is megjelenik rajta, akkor a képző inkább a második funkciójával járul hozzá az ige jelentéséhez, azt érzékeltetve, hogy az esemény „felületesen, csak egy kicsit” történt (pl. *meg-megérintget, fel-felütöget*).

A kiinduló kérdésre azt a választ adja, hogy a két jelenség között van szemantikai különbség, és ez a környezetükben megjelenő időhatározók segítségével érhető tetten. Megfigyelése szerint a reduplikált igekötős formák nem tűnnek meg olyan időhatározót, amely az esemény rendszeres voltát fejezi ki, állhatnak viszont olyanokkal, amelyek rendszertelen időközökre utalnak. Ezt az állítást megvizsgáltam úgy, hogy a korpuszban a következő mintát kerestem: a *rendszerint, rendszeresen, rendszerességgel, folyamatosan, állandóan, hetente, havonta, naponta* szavak egyike, utána reduplikált igekötős finit igealak. Összesen 45 illeszkedő mondatot találtam – néhányhoz ld. a (3) példasor –, ez a szám tehát azt mutatja, hogy Kiefer állítása finomításra szorul. Nem jellemző, hogy a reduplikált igekötős alakok mellett rendszerességre utaló időhatározó áll, de előfordulhat.

- (3) a. [...] az emigrációból is **állandóan bele-beleszólt** az argentin politikába, [...]
- b. Németországban **rendszeresen fel-felröppennek** iparikém-ügyek.
- c. Volt néhány törzshelyem, ahova **rendszeresen vissza-visszatértem** [...]
- d. [...] **folyamatosan fel-felmerül** az ország területi egységének fenntarthatósága.
- e. [...], **naponta bele-belekandikáljon** a sűrűn egymás alá rótt naplósorokba?...

Kiefer utolsó kutatási kérdése a következő: Van-e különbség a reduplikált és a nem reduplikált igekötők mondatbeli eloszlásában? A kérdésre határozott igen a válasz, a két típus szintaktikai viselkedése jelentősen eltér. Preverbálisan a reduplikált igekötő is képes elválni az igétől, de – Kiefer megfigyelése szerint – ekkor csak az *is* szócska ékelődhet az igekötő és az ige közé. A korpuszvizsgálatból viszont az derül ki, hogy bár az elválás mértéke korlátozottabb, a preverbális reduplikált igekötős szerkezetek esetében is megfigyelhetők azok a mintázattípusok, amelyekről

bővebben a 4. fejezetben lesz szó, ld. a (4) példasort.

- (4) a. [...] mert nem tudnám megállni, hogy vadul **fel-fel** ne **nyihogjak**, [...]
- b. [...] a magam kurtácska végtelenségét, amely pillanatokra még **el-el** tud **bolondítani**.
- c. [...] a biztonságát olykor **meg-meg** kellett, hogy **rendítse** az a fölényes és boldog nyugalom, [...]

A reduplikált igekötő Kiefer szerint nem lehet hátravetett helyzetben. Ezt a korpuszmérések tükrében úgy szükséges módosítanunk, hogy nagyon ritkán fordul elő, az eseteknek mindössze 0,1%-ában. Ez élesen szembenáll azzal, hogy a nem reduplikált igekötők az esetek több mint egyharmadában hátravetettek (erről bővebben ld. a 4.1.2. fejezetet). A posztverbális helyzetet az (5) példasor szemlélteti.

- (5) a. Fokozódó élvezettel **nézegetek bele-bele** a tévén a közvetített bugyuta színdarabba.
- b. [...] vajon mifajta szomjúságból **tér vissza-vissza** a „szavalathoz”, vagyis Petőfihez.
- c. [...] aztán Donáth és Márkus, néha Zimányi Tibor **szólt közbe-közbe**.

Egy későbbi munkájában Kiefer (2003b) a reduplikáció lehetőségét fontos szempontnak tartja az igekötők és a puszta névszói igemódosítók elkülönítésében. Kerekes (2011: 114-116) szintén fontos, kritérium-értékű jegynek tartja a reduplikációt. Bár megjelenhet olyan esetekben is, mint pl. *lassan-lassan végez, egyre-egyre közeledett*, ezek az alakulatok elég távol esnek az igemódosítói kategóriától, attól könnyen elkülöníthetők.

Végül érdekes adat, hogy a korpuszvizsgálat tanúsága szerint mindössze 39 igekötő esetében találunk példát a reduplikációra. Ennek az oka nem annyira az adathiány, mint inkább a már tárgyalt konceptuális és szemantikai megszorítások lehetnek.

3.3.3. A *-fele/felé* morféma csatolhatósága

Olyan mondatokban, amelyek felszólítók vagy progresszív olvasatúak – azaz épp zajló, folyamatban lévő cselekvést vagy történést fejeznek ki –, az igekötőkhöz a *-fele/felé* morféma kapcsolható (Kerekes 2011: 115), pl. *mennek felfele, gyűrődik összefele*. A *-fele/felé* jelentése ‘valamilyen irányba’, így elsősorban irányjelentésű igekötőkhöz kapcsolódik.

Erről a jelenségről nagyon kevés szó esik a szakirodalomban. Kerekes (2011) említi, mint lehetséges morfológiai jegyet, de nem tulajdonít neki különösebb jelentőséget. Nádasdy (2005) arról ír, hogy a folyamatos szemlélet kifejezésére már önmagában az igekötő hátravetése is alkalmas, a *-fele/felé* kapcsolása az igekötőhöz csak kihangsúlyozza a folyamatosságot. Blogbejegyzések és kommentek formájában viszont több anonim szerzőnél is felbukkan ez a téma.¹² Ezekből az derül ki, hogy a *-fele/felé* kapcsolása az igekötőhöz főleg népnyelvi jelenség – a nyírségi nyelvjárásban a legjellemzőbb –, de a köznyelvben sem ritka.

Minden igekötőjelölt esetében megvizsgáltam ennek a jegynek a jelenlétét a korpuszadatokban, de csak a leggyakoribb esetet vettem figyelembe, a közvetlenül ige utáni pozíciót. Egyedül a *megfele/megfelé* szóalakok esetében döntöttem úgy, hogy egy külön mérés keretében érdemes lehet megvizsgálni a mondatbeli eloszlásukat a finit igehez viszonyítva. Itt ugyanis nem merül fel az a lehetőség, hogy „tisztán határozói” szerepük lenne, és igekötői nem. A mérés eredményét az 5. táblázat mutatja be.

pozíció	-1	+1	+2	+3	+4
db	3	57	13	2	1
%	3,9	75	17,1	2,7	1,3

5. táblázat. A *megfele/megfelé* szóalakok pozíciói a finit igehez képest.

Látható, hogy a *megfele/megfelé* az esetek 96,1%-ában posztverbális, vagyis az ige után áll. Ennek az az oka, hogy mind a progresszív olvasatú, mind a felszólító mondatok esetében az ‘ige – igekötő’ szórend jellemző. A preverbális, azaz ige előtti helyzet valószínűleg csak azért fordulhat elő, mert a *-fele/felé* szócska már önmagában hordozza a folyamatosságot, ezért a szórend megváltoztatása nem feltétlenül szükséges. A (6) példasor egy-egy mondattal illusztrál minden pozíciót.

- (6) a. Ami vót a lábán, mikor **megfele halt**, azt a cipőt. -1
 b. **Tisztuljál megfele** szépen. +1
 c. [...], hanem **oldja** azokat **megfelé** legalább generációsan. +2
 d. Rohamosan **tisztul** a közélet **megfele**. +3

¹² <https://www.nyest.hu/hirek/a-nap-kommentje>
https://itcafe.hu/tema/magyar_szavak_amiket_nem_mindenki_ismer/hsz_1069-1069.html

- e. **Nézzed** már a dátumot **megfele**, ő írta előbb! +4

A vizsgált adatban a *meg*-en kívül két olyan igekötő van, amelyeknek nincsen irányjelentése, mégis megkapják a *-fele/felé* morfémát – ezek a *tönkre* és a *ketté*, ld. a (7) példáit.

- (7) a. [...], az autóban lévő motor meg nyivákol, mert hát ugye ettől **megy tönkrefele**.
 b. Most **hasadjak kettéfele**?

3.3.4. Fokozhatóság

Bizonyos, főként határozószói eredetű igekötők középfokban is állhatnak. Ez jellemzőbb akkor, ha a szerkezet jelentése kompozicionális (pl. *összébhhúzódik*), de nem kompozicionális igekötős igék esetében is előfordul (pl. *ettől persze mindig csak kijebb és kijebb ábrándultam*). Kálmán (2013) elsősorban az *el* igekötő kapcsán, egy blogbejegyzésben foglalkozik a témával, amelyet ő fokozás helyett inkább összehasonlításnak nevez, „[...] hiszen arról van szó, hogy két vagy több jelenséget összehasonlítunk abból a szempontból, hogy egy tulajdonsággal milyen fokban rendelkeznek”.

Az *el* igekötő hasonlító alakjai az *elebb*, még ritkábban az *eljebb*. Kálmán azt írja, hogy ezek esetleg tájnyelvi alakok is lehetnek, ezt megerősíteni és cáfolni sem tudjuk, mert nincs olyan felmérés, amelyből kiderülhetne. Fenntartja a lehetőségét annak is, hogy egyszeri szóhasználatról – okkazonalizmusról – van szó, mert korpuszok alapján sem adatolható. Ez a cikk írásának idején még bizonyára így lehetett. Az MNSZ2 itt felhasznált verziójában, amely 2016 augusztus végétől érhető el, már találunk példát az *el* hasonlító alakjaira. Ezek különböző dokumentumokból származnak, ami nem garantálja azt, hogy a dokumentumok szerzői is mind különbözőek, de valószínűsíti. Összesen öt példa van az *elebb/eljebb* alakokra, ld. a (8) példasort.

- (8) a. [...], most már ha lesz, akkor se tudok menni -, kicsit **elvadultam**, még **elebb**.
 b. [...], most szakadt el a cérna. Most **szakadt elebb**.
 c. Nem lehet azt mondani, hogy Csucsut „**elebb hanyagoltam**” valamelyest.
 d. [...], jobb, ha Holt Telkek címszó alatt „**elebb temetem**” szegényeket.
 e. [...], sőt a magyar deviza gyengülése számos terméke esetében **eljebb tolt**a az árakat.¹³

¹³ A kontextus alapján elképzelhető, hogy itt egyszerűen elgépelésről van szó, *feljebb* → *eljebb*.

Kálmán azt is megemlíti, hogy az *át* és a *rá* esetében még kevésbé képzelhető el a fokbeli összehasonlítás, mint az *el*-nél. Ezek az én korpuszmérésem adatai közt sem szerepelnek. Az *át* hasonlító alakjára elvétve lehet internetes példákat találni, ld. a (9) példasort.

- (9) a. [...] mely korlátokon **átlépünk**, természetesen, egyre **átabb**, [...]
- b. **Átalakítottam**, és folyamatosan még **átabb alakítom** a terápiás oldalamat.
- c. [...], mert mi még mindig **átvesszük**. Sőt! Még **átabb**.

Ugyancsak egy-két internetes példa akad arra, hogy a középfok jele a *meg* és a *szét* igekötőkhöz járul, ld. a (10) példáit.¹⁴

- (10) a. A lelkem is **megebb nyugodott**, hogy talán tehetek valamit érte.
- b. Ezekre nyilván nekem is **megebb kellett mozdulnom**, mint addig, [...]
- c. És ha már Mona Lisa, ma „**szétebb**” **néztem** a vatera könyvrészlegén.

A hasonlító alakok nemcsak finit ige mellett jelenhetnek meg, hanem például infinitívusszal és egyéb igei származékokkal együtt is, ld. a (11) példákat.

- (11) a. Szóval ez egy nagyon szerteágazó történet, ami még **szertébb fog ágazni**, [...]
- b. [...] amihez hozzá veendő még az erdélyiek konzervatívabb volta (a **zokonabb vevés**).

Elképzelhető, hogy az igekötők hasonlító alakjainak használata főként beszélt nyelvi jelenség. Szóbeli megnyilatkozások során gyakoribbnak tűnik az, hogy rövidsége törekszünk, és ezért mondjuk például a *még inkább összevesztek* helyett azt, hogy *összebbvesztek*. Mindez csak feltételezés szintjén marad addig, amíg nem vonunk be beszélt nyelvi korpuszt is a vizsgálatba.

3.3.5. Személyragozhatóság

A fokozhatósághoz hasonlóan ritka jelenség az, hogy bizonyos igekötők, igekötőszerű elemek személyragozhatók, és sokszor ilyen formájukban is egybeírják őket az igével, különösen akkor, ha az ige rövid, ld. a (12) példasort.

¹⁴ Az egységes módszertan és az adatok összehasonlíthatósága érdekében csak az MNSZ2 adatait vettem bele a mérési eredményekbe, az egyéb forrásból származó példákat csak érdekességképp tüntetem fel.

- (12) a. Ha **hozzámnyúlsz**, megöllek!
 b. [...] ha MOL kútnál tankolsz, akkor utána mindig **nekedmennek** hátulról.
 c. Verjünk tábort, **ránkfér** a pihenés!
 d. **Utánukvágom**, egyék meg, ha az kell nekik!

Az igekötők személyragozhatóságának kérdésköre gyakran tárgyalt téma a szakirodalomban.¹⁵ Élénken vitatott kérdés az, hogy a személyragozott alakok esetében nem történik-e szófajváltás, vagyis igekötő-e még a fenti példákban a *hozzám*, *neked*, *ránk*, *utánuk*, vagy névmási illetve névutói elemeknek tekintendők. Az én megközelítem az, hogy mindazokat a személyragos elemeket is igekötőnek lehet tekinteni, amelyek előfordulnak a tipikusan igekötői szerkezetekben. Emellett szólnak a (13) példásor adatai, amelyek a személyragos alakok reduplikálhatóságát mutatják. Ez a jelenség kétségtelenül ritka – összesen 40 példát lehet rá találni az MNSZ2-ben –, de többféle lexikai elem esetében adathozható finit igei, infinitívuszi és igeneves szerkezetekben egyaránt, ezáltal felvet érdekes, nyelvtipológiai kérdéseket is.

- (13) a. [...] akkor is **belém-belém sajdult** karáncsonyfa díszítés közben, [...]
 b. [...] de a **ránk-ránktörő** helyi rosszarcok is hajlamosak kárt tenni bennünk.
 c. [...] ilyen bohókás komolyan **rám-rámvetve** tekintetét; [...]
 d. A szél **nekem-nekem csap**, s megvacogtat már.

A személyragozható igekötőkkel kapcsolatban részletes korpuszvizsgálatot végeztem, amely ennek az alfejezetnek a kereteit sokszorosán meghaladja. A korpuszmérések eredményeit egy hamarosan megjelenő, Farrell Ackermannal és Rob Malouffal közösen írt tanulmányunk mutatja be olyan megközelítésben, amely konstrukciók hálózatmodelljén alapul (Ackerman et al. 2021).

3.3.6. Szerkesztettség, szótagszám és gyakoriság

Ez a fejezet olyan jegyeket mutat be, amelyek minden szónál megfigyelhetők – emiatt úgy tűnhet, hogy túl általánosak, és nem formálható belőlük olyan tipikalitási feltétel, amely segítené az igekötői

¹⁵Ld. többek között É. Kiss (1998), Kálmán és Trón (2000), Surányi (2006, 2009a,b,c), Rákosi és Laczkó (2011), Hegedűs (2013), Rákosi (2014).

kategória azonosítását. Valójában viszont fontos lehet az, hogy ezek a jegyek milyen értéket vesznek fel a vizsgált szavak esetében.

A **szerkesztettség** azt jelenti, hogy az adott szó morfémákra bontható (Kiefer és Ladányi 2000c: 138). Az ősi igekötőkre – *meg, el, fel, le, ki, be* – ez nem áll, és a hiányát több szerző kiemelten fontos szempontnak tartja az igekötő-állomány meghatározásában. Jakab (1976) megkülönbözteti az alaki szerkesztettséget – szétválasztható-e az adott lexikai elem tő- és ragmorfémára –, valamint a jelentésanit – kompozicionális-e a létrejövő szó jelentése –, és ebből a két szempontból külön-külön elemzi az igekötőjelölteket. Forgács (2005) is hasonlóan jár el: a morfológiai és a szemantikai transzparencia megszűnését külön tárgyalja. Ladányi (2000) ezzel szemben morfoszemantikai jellemzőként írja le a szerkesztettséget, ami azt jelenti, hogy nem áll elő olyan helyzet, amelyben egy lexikai elem formailag elemezhető lenne, de szemantikailag nem. Ez jelentős különbséghez vezet például az *agyon* és *tönkre* elemek esetében. Ladányi mellett érvel, hogy ezeknek a formai egybeesése ragos főnevekkel látszólagos: grammatikalizációs folyamaton mentek keresztül, és az igekötői funkciójukban már nincs közük azokhoz a ragos főnevekhez, amelyekből a grammatikalizáció kiindult.

Az általam használt módszer nem tesz lehetővé ilyen precizitású morfoszemantikai elemzést. Eszerint szerkesztettnek minősül minden olyan lexikai elem, amelyet az emMorph automatikus morfológiai elemző (Novák et al. 2016, 2017) szótőre és suffixumra képes bontani. Ez tehát csak morfológiai szempontú elemzés, és a fent leírtak ismeretében beláthatók a komoly korlátai.

A **szótagszám** a szó fonológiai tulajdonsága, és bár kétségtelenül összefügg a szerkesztettséggel – a morfológiailag komplex szavak legnagyobb része több szótagú –, a két jegy nem ugyanaz. Ezek kapcsolatát a 6. táblázat szemlélteti.

	szerkesztetlen	szerkesztett
egy szótagú	pl. <i>be, meg</i>	pl. <i>jót, részt</i>
több szótagú	pl. <i>össze, vissza</i>	pl. <i>észre, fejbe</i>

6. táblázat. A morfológiai szerkesztettség és a szótagszám kapcsolata.

A **gyakoriság** egy szóalak vagy szótípus előfordulásainak száma egy adott korpuszban. A kognitív és a használatalapú nyelvészeti irányzatok ennek alapvető jelentőséget tulajdonítanak, így fontos szerepet kap a prototípus-elméletben is. Elképzelhető, hogy egy adott kategóriának a leggyakoribb

példányát ítéljük prototipikusnak, és ebből a példányból származtathatóak a tipikalitási feltételek – írja Kiefer (2007: 87) Kálmán László személyes közlését idézve.

3.4. Igekötő × Jegy mátrixok

Az előző fejezet összesen 10 olyan jegyet mutatott be, amelyek értékei korpuszmérésekkel minden igekötő esetében meghatározhatók voltak. Ezek az értékek elrendezhetők egy mátrixban, ahol a sorokat az igekötők, az oszlopokat pedig a vizsgált jegyek alkotják. Az egyes cellák tartalma az adott jegy értéke az adott igekötő esetében, mintaként ld. a 7. táblázatot.

igekötő	redupl.	-fele	fokoz.	szerk.	szótagsz.	gyak.	P_m	...
...
ki	553	2 067	263	0	1	1 354 223	0,086	...
kívül	0	0	1	0	2	87	0,001	...
körbe	88	0	0	1	2	11 285	0,012	...
körül	3	1	0	0	2	12 590	0,008	...
közbe	17	0	0	1	2	6 636	0,002	...
...

7. táblázat. Részlet az Igekötő × Jegy mátrixból, amely a korpuszban mért jegyértékek abszolút gyakoriságát tartalmazza. A fejlécben szereplő rövidítések feloldása: *redupl.* = reduplikáció, *-fele* = a *-fele/felé* szócska csatolhatósága, *fokoz.* = fokozhatóság, *szerk.* = szerkesztettség, *szótagsz.* = szótagszám, *gyak.* = gyakoriság, P_m = megvalósult produktivitás.

A korpuszadatokból többféle mátrix is létrehozható: a sorok alakulását befolyásolja az, hogy az igekötők alakvariánsairól hogyan döntünk, az oszlopok esetében pedig az a kérdés, hogy biztosan az abszolút értékek mondanak-e el a legtöbbet egy-egy jegyről.

Nézzük először a sorokat: Vegyünk minden szóalakot külön, vagy az alakvariánsokat számoljuk inkább egybe? Alakvariánsnak a következők tekinthetők: *bele – belé, be – bé, fel – föl, felül – fölül, fenn – fönn, odább – odébb, rá – reá, tele – teli*. Arra jutottam, hogy hasznos lehet ezeket összevonni, mert a párok tagjai között inkább csak stílusbeli, esetleg nyelvjárásbeli eltérés van.

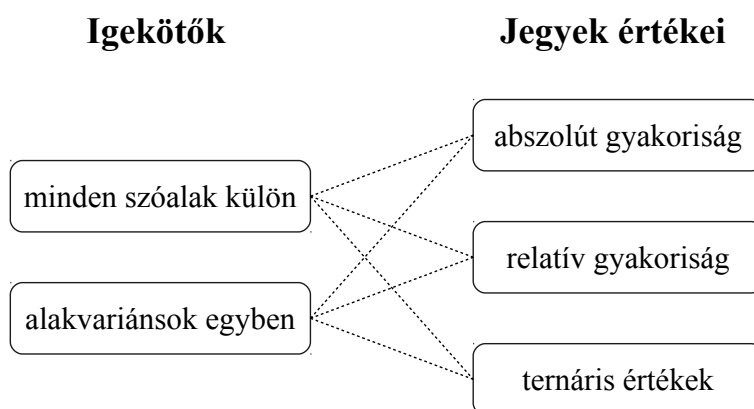
Az oszlopokra vonatkozó kérdés – hogyan érdemes megjeleníteni a jegyek értékét? – már összetettebb. Itt nagyon sok elvi lehetőség adódik, amelyek közül hármat ültettem át a gyakorlatba. Ezek a következők:

1. **Abszolút gyakoriság.** Ez azt mutatja, hogy az adott jegy hány darab szóalagnál valósult

meg az adott igekötő esetében, erre hoztam példát a 7. táblázatban. A jegyek értékei ebben az esetben nagyon széles skálán mozognak (a nullától kezdve a milliós nagyságrendig). Az abszolút gyakoriság a legfontosabb abból a szempontból, hogy a legrészletesebb adatot szolgáltatja, és minden más értékadás ebből vezethető le. Ugyanakkor nehezzé teszi az adat elemzését, mert a gyakoriság mindent ural: a milliós nagyságrendben előforduló igekötőknek minden jegyükben esélyük van nagyobb értékeket felvenni a csak ezres nagyságrendben előfordulóknál.

2. **Relatív gyakoriság.** Ezt úgy kapjuk meg, hogy a jegyek az adott igekötő esetében mért abszolút gyakoriságát elosztjuk a jegy összes előfordulásával. Így minden igekötőhöz egy százalékos értéket kapunk, amely azt mutatja, hogy az mekkora részt vesz ki a vizsgált jegy előfordulásaiból. Ez az adat a könnyebb áttekinthetősége miatt – minden érték 0-tól 100-ig tartó skálán mozog – manuális adatelemzésre is alkalmas. Automatikus módszerek, például a klaszterelemzés esetében – bővebben ld. a 3.7.2. fejezetet – viszont fenntartással kezelendő, mert az igekötők körében ritkább jegyek (pl. a személyragozhatóság) jellemzően nagyobb értékeket kapnak, mint a sűrűbb jegyek, így súlyozottabban vannak jelen. Ugyanez a probléma akkor is fennállna, ha a jegy adott igekötő esetében mért abszolút gyakoriságát az igekötő abszolút gyakoriságával osztanánk le. Ekkor felértékelődnének az olyan esetek, amikor egy nagyon ritka igekötőnél kicsit véletlenszerűen felbukkan egy jegy.
3. **Ternáris értékek.** Ekkor az abszolút gyakoriságot mindössze három értékre egyszerűsítjük le: egy jegy nincs (0), minimálisan van jelen (0,5) vagy jelen van (1) az adott lexikai elem esetében. A 0 érték akkor áll elő, ha a jegy abszolút gyakorisága 0. Ez alól csak maga a *gyakoriság* jegy a kivétel, ahol Sass (2011: 42) javaslata alapján 5-ös küszöböt alkalmaztam: az igekötőnek legalább ötször kellett előfordulnia ahhoz, hogy ne 0-t kapjon a ternáris mátrixban. 1-es értéket a következő gyakorisági küszöböknél adtam: a produktivitási mértékek esetében 0,0005, a *gyakoriság* jegynél 250 – Sass (2011: 78) alapján –, minden egyéb jegynél 5. A ternáris mátrix 0,5-ös értéke értelemszerűen akkor jelenik meg, ha az adott jegy abszolút gyakorisága nem éri el a szükséges küszöbértéket. Bár a háromértékű jegyek elnagyoltak, leegyszerűsítettek az abszolút és a relatív gyakoriságúakhoz képest, mégis jól használhatónak bizonyulnak akár a kézi, akár a gépi adatelemzésben.

Összegezve tehát hat mátrixot hoztam létre a sorok és az oszlopok ábrázolásának kombinációjából, ahogy a 3. ábra mutatja.



3. ábra. Az igekötők, valamint a jegyek értékeinek ábrázolásai. A vizsgált lehetőségekből összesen hat mátrix állítható elő, a kombinációkat a szaggatott vonalak mutatják.

Ezek az Igekötő \times Jegy mátrixok nyílt hozzáférésűek, és további kutatások alapjául szolgálhatnak.¹⁶ A soron következő fejezetekben csak azokról lesz szó bővebben, amelyek a korreláció-vizsgálat és később a jegyalapú osztályozás szempontjából a leghasznosabbnak bizonyultak.

3.5. Jegyek közötti korrelációk

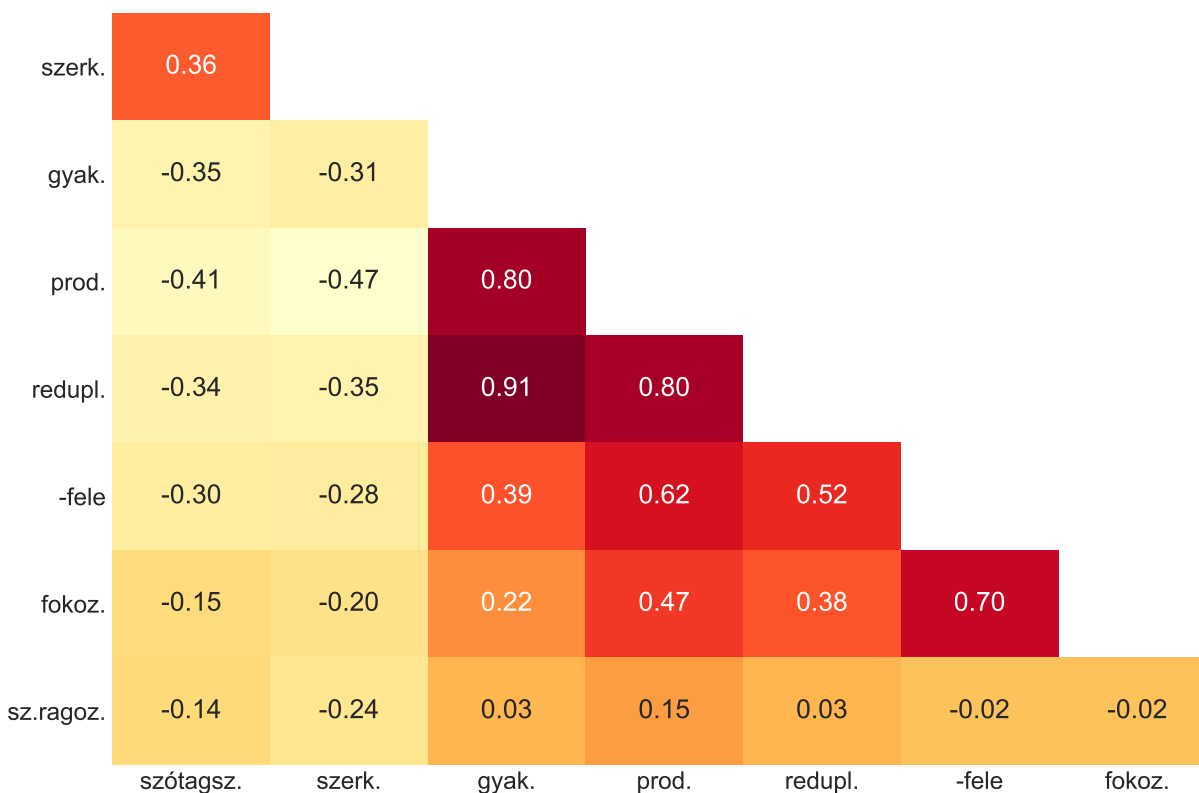
Az előző fejezetben ismertetett mátrixok lehetővé teszik minden jegypár közötti korreláció kiszámolását. Ezáltal kimutatható, hogy van-e kapcsolat az egyes jegyek között, és ha igen, ez mennyire szoros. Az eredményül kapott korrelációs együttható (r) egy 0 és 1 közötti szám, amely az együttállás szorosságát jelzi: a kapcsolat annál gyengébb, minél közelebb van ez a szám a 0-hoz. Az együttható előjele a jegyek közti kapcsolat irányát mutatja: negatív előjel esetén minél magasabb az egyik jegy értéke, annál alacsonyabb a másiké, pozitív esetén pedig az egyik jegy magasabb értékeihez általában a másik jegynek is magasabb értékei párosulnak. Ahogy látni fogjuk, a korrelációszámítással statisztikailag igazolhatóvá válik több olyan állítás, amelyet a szakirodalmak jellemzően tapasztalati tényként kezelnek.

Elsőként az abszolút gyakoriságok mátrixán végeztem el a vizsgálatot,¹⁷ erről a 4. ábra ad áttekintést. Az eredmények magyarázatát az ábrán látható háromszög felső harmadával kezdem: a

¹⁶ <https://github.com/kagnes/prevmatrix>

¹⁷ A relatív gyakoriságokon végzett korreláció-számítás pontosan ugyanezt az eredményt adja, ezért nem tárgyalom külön. Ez ellenőrizhető a <https://github.com/kagnes/prevmatrix> adatai alapján.

szerkesztettség, a szótagszám, a gyakoriság és a produktivitás kapcsolatával.



4. ábra. Korrelációk az összevont igekötői alakvariánsokat és abszolút gyakoriságokat tartalmazó mátrixon. A sorokban és az oszlopokban is a vizsgált jegyek láthatók, a cellákban pedig az a korrelációs együttható, amely az adott sor és az adott oszlop jegypárjára érvényes. A cellák színe a korreláció erősségét jelzi: minél nagyobb a korrelációs együttható, annál vörösebb a cella. A rövidítések feloldása felülről lefelé haladva: *szerk.* = szerkesztettség, *gyak.* = gyakoriság, *prod.* = (megvalósult) produktivitás, *redupl.* = reduplikáció, *-fele* = a *-fele/felé* morféma csatolhatósága, *fokoz.* = fokozhatóság, *sz.ragoz.* = személyragozhatóság, *szótagsz.* = szótagszám.

A szerkesztettség és a szótagszám között pozitív korreláció figyelhető meg, bár nem túl magas ($r = 0,36$). Ennek okát korábban a 6. táblázatban láthattuk: vannak szerkesztetlen, de több szótagú (pl. *össze*) és szerkesztett, mégis egy szótagú lexikai elemek (pl. *részt*), de ezek a jegykombinációk viszonylag ritkák.

A szótagszám és a szerkesztettség egyaránt negatívan korrelálnak a gyakorisággal (az előbbi esetében $r = -0,35$, az utóbbinál $r = -0,31$). A szótagszám és a gyakoriság kapcsolata a Zipf-törvény szerint alakul, amely kimondja, hogy a gyakran használt szavak jellemzően rövidebbek, mint a ritka szavak (Zipf 1932). Ennek az állításnak az általam ismert legszélesebb spektrumú vizsgálata 986 nyelv bevonásával történt, és mindegyik esetében kimutatta a Zipf-törvény érvényességét (Bentz és Ferrer-i Cancho 2016). A szerzők szerint ez olyan robusztus adat, hogy a nyelvek, vagy még

általánosabban az információfeldolgozás és -átadás univerzális tulajdonságának mondható.

Az ábrán látható az is, hogy a szerkesztettség és a szótagszám ugyancsak negatívan korrelál a produktivitással ($r = -0,47$ és $r = -0,41$). Tehát minél produktívabb egy igekötő, annál valószínűbb, hogy rövid és morfológiailag bonthatatlan. Végül nagyon erős pozitív korreláció ($r = 0,80$) figyelhető meg a produktivitás és a gyakoriság között. Bár a két jegy szorosan összefügg, mégsem lehet egymás szinonimájának tekinteni ezeket. Ezt jól példázza az *agyon*, amely a tokengyakoriság szerinti rangsorban a 46., a megvalósult produktivitását tekintve a 22. helyet foglalja el. A gyakoriság és a produktivitás kapcsolata egy 2x2-es mátrixszal írható le, amelyet a 8. táblázat szemléltet.

	produktív	nem produktív
gyakori	pl. <i>meg, el, be</i>	pl. <i>létre, egyet, észre</i>
nem gyakori	pl. <i>pofon, szénné, szarrá</i>	pl. <i>hajba, síkra, csonttá</i>

8. táblázat. A produktivitás és a gyakoriság kapcsolata, Pakerys (2017) alapján.

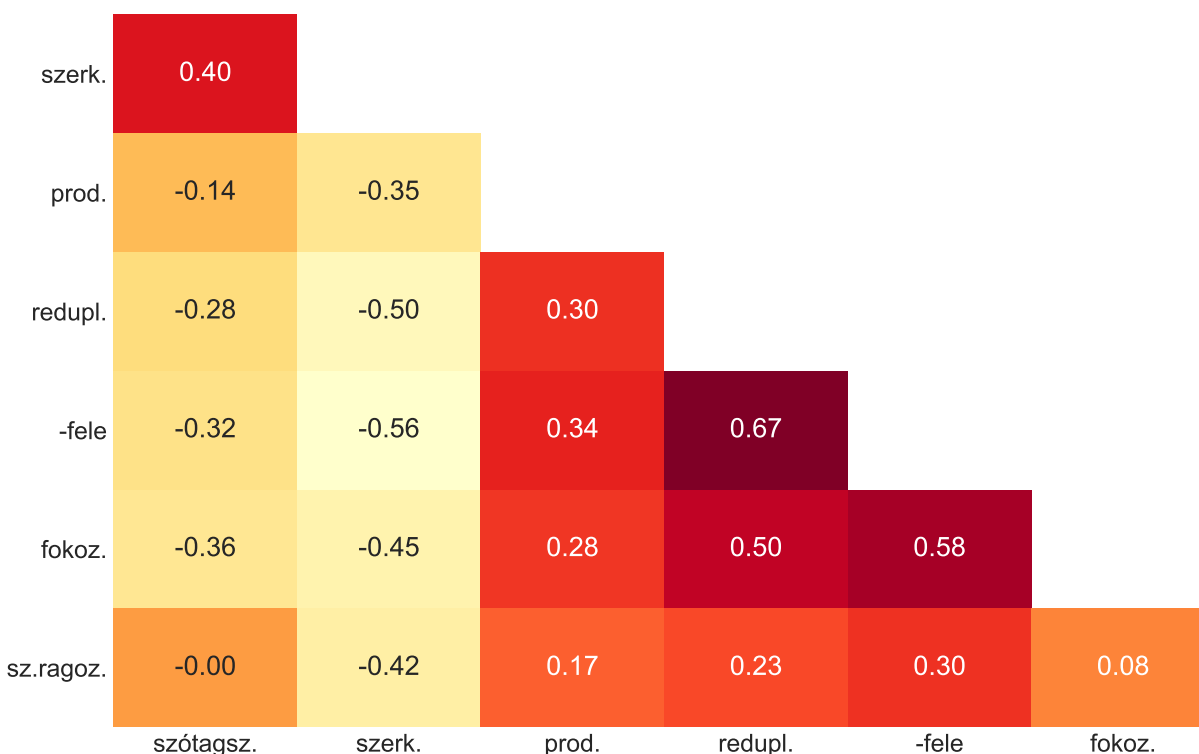
Ezek a korrelációs viszonyok a **grammatikalizáció** folyamatát tükrözik, amelynek során egy önálló lexikai elem morfológiai jelenséggé – többnyire affixummá – válik. Ezt jellegzetes változások kísérik: a szóalak rövidül és morfémákat is veszít, mígnem teljesen elveszíti a morfológiai szerkezetét, bonthatatlan lesz. Mindez szemantikai következményekkel is jár: a szó eredeti jelentése elhomályosul vagy más jelentéselemekkel egészül ki. Ezzel még meglehetősen leegyszerűsítve jellemeztük a folyamatot; a grammatikalizációról bővebben ld. többek között D. Máta (1989), Kiefer (2006), Dér (2008), É. Kiss (2014b) és Cser (2017) munkáit.

A 4. ábra középső harmadát tekintve a legszembetűnőbb talán a reduplikáció és a gyakoriság közötti kapcsolat. A 0,91 értékű korrelációs együttható olyan közel van az elvileg lehetséges maximumhoz (az 1-hez), hogy kijelenthetjük: a reduplikáció mértéke a gyakoriság függvényében változik. A további összefüggés az, hogy a reduplikáció leginkább a grammatikalizálódás folyamatában előrehaladott igekötőkre jellemző – amelyek gyakoriak, produktívak, rövidek és szerkesztetlenek.

Az alsó harmadot illetően a fokozhatóság és a *-fele/felé* csatolhatósága közti pozitív korrelációra ($r = 0,7$) érdemes kitérni. Ez leginkább a határozószói jellegből adódik: ennek a két jegynek az értékei az olyan igekötőknél magasak, amelyek irányjelölő funkcióval is bírnak. Ezek a felső harmadban látott jegyekkel lazább kapcsolatban állnak, akár pozitív, akár negatív korrelációkat vizsgálunk. Ez arra utal, hogy a fokozhatóság és a *-fele/felé* csatolhatósága nem lényegi jegyei a le-

xikai elemek vizsgált halmazának. Különösen igaz ez a személyragozhatóságra, amelynek minden korrelációs együtthatója 0 körül van, vagyis egyetlen más jeggyel sem korrelál.

Az eddigi eredményekkel szemben felmerülhet az a kritika, hogy a bizonyos jegyek között mért erős korreláció valamilyen harmadik változó hatására vezethető vissza, amely mindkét jeggyel szoros kapcsolatban áll. Gyanítható, hogy a gyakoriság egy ilyen változó. Ennek a hatását ki tudjuk szűrni úgy, hogy binárisra egyszerűsítjük az adatot: ekkor a *gyakoriság* jegy megszűnik, a többi jegynél pedig csak annyit vizsgálunk, hogy az adott igekötő esetében megjelenik-e vagy sem. Kísérletképpen létrehoztam egy ilyen bináris mátrixot, és ennek a jegypárjain is elvégeztem a korrelációszámítást. Az eredményt az 5. ábra szemlélteti.



5. ábra. Korrelációk az összevont igekötői alakvariánsokat és bináris jegyértékeket tartalmazó mátrixon. A sorokban és az oszlopokban is a vizsgált jegyek láthatók, a cellákban pedig az a korrelációs együttható, amely az adott sor és az adott oszlop jegypárjára érvényes. A cellák színe a korreláció erősségét jelzi: minél nagyobb a korrelációs együttható, annál vörösebb a cella. A rövidítések feloldása felülről lefelé haladva: *szerk.* = szerkesztettség, *prod.* = (megvalósult) produktivitás, *redupl.* = reduplikáció, *-fele* = a *-fele/felé* morféma csatolhatósága, *fokoz.* = fokozhatóság, *sz.ragoz.* = személyragozhatóság, *szótagsz.* = szótagszám.

Megfigyelhető, hogy a jegyek közti korrelációk mértéke szinte minden esetben csökkent, de az arányait tekintve nincs lényeges változás. Két nagyobb eltérést látunk a 4. ábrához viszonyítva. Az

egyik az, hogy a reduplikáció és a produktivitás közötti korreláció 0,8-ról 0,3-ra csökkent. Ennek az oka, hogy ha binárisá egyszerűsítjük az adatot, akkor az alacsony produktivitás felértékelődik (0,0005-ös értékből 1 lesz), így például az *éhen* elem (*éhenhal*, *éhenveszik*, *éhendöglik*, *éhenpusztul*, *éhengebed*) egyenrangú lesz az olyanokkal, mint a *meg*, *ki*, *be*, stb. Ugyanez az oka annak is, hogy a személyragozhatóság és a *-fele/felé* csatolhatósága között enyhe pozitív korreláció jelenik meg ($r = 0,3$). Ebben az esetben az olyan ritka alakok értékelődnek fel, mint a *hozzáfele*, *ráfele*.

Az abszolút gyakoriságokon és a bináris értékeken mért korrelációk összehasonlítása alapján két megfigyelés tehető. Először is azt látjuk, hogy a jegyek közti alapvető összefüggések nem változnak attól, ha a gyakoriságra vonatkozó információt elhagyjuk. Ennek viszont nem az az oka, hogy a gyakoriságnak nincs jelentősége. Épp ellenkezőleg: történetileg olyan szorosan együtt jár más jegyekkel – a grammatikalizációs folyamat révén –, hogy akkor is megjelenik bennük, ha expliciten, külön jegyként nem vesszük figyelembe. Ez alapján feltételezhető, hogy ha a kiindulásként szolgáló 235 lexikai elemet az itt vizsgált jegyértékeik mentén csoportosítjuk, akkor nagyjából hasonló eredmény várható akkor is, ha a nagyon részletes abszolút gyakoriságú, és akkor is, ha a végletekig egyszerűsített ternáris vagy bináris adatot választjuk kiindulópontként.

3.6. A prototípus és a prototipikus jegyek

Az előzőleg vizsgált jegyek 28 jellemzően igekötőnek minősített lexikai elem jellemzői voltak. Az, hogy melyikük mennyire releváns az ‘igekötő’ fogalom meghatározásában, már a prototipikus egyed tulajdonságaitól függ. Ezért elérkeztünk arra a pontra, ahol a 28 elemű halmazt még tovább kell szűkíteni egyetlen elemre, amelyet a továbbiakban a prototípusnak tekintünk.

A prototípus-elmélet a fogalmak pszichológiai modellje, és ennek megfelelően kísérletes módszerekkel együtt szokás alkalmazni. Így az itt bemutatott, korpuszvezérelt módszer esetében bizonyos lépések nem tudnak nem önkényesek lenni, mert végeredményben csak egy anyanyelvi beszélő döntésén alapulnak. Ez mindenképp egy gyenge pontja a megközelítésnek, és a dolgozat végén – ld. a 6. fejezetet – még visszatérek rá.

Prototipikus igekötőnek a *meg*-et választottam. Ez messze a leggyakoribb igekötő: ha csak a finit igével egybeírt előfordulásait számoljuk, akkor az 1,04 milliárd tokenes, módosított MNSZ2-ben 3,6 milliószor fordul elő. Az igekötői státuszát illetően általános a szakirodalmi egyetértés, a történeti adatoltsága alapján pedig feltehetően a legősibb ilyen funkciójú elem.

A *meg* tulajdonságai alapján a vizsgált jegyek és értékeik közül prototipikusnak tekinthető a magas morfológiai produktivitás és gyakoriság, a szerkesztettség hiánya, az egyszótagúság és a magas gyakoriságú reduplikáció. Ahogy a korreláció-vizsgálatból látni lehetett, ezek az értékek a prototípus választásától függetlenül is szoros kapcsolatban állnak egymással, ezért valószínűleg jó úton járunk, ha fontos jegyeknek ítéljük ezeket. Kevésbé releváns az, hogy a *-fele/felé* morféma csatolható-e a szóhoz, és nem releváns az, hogy a szó fokozható illetve személyragozható-e. Az utóbbi két jegy a korrelációi alapján meglehetősen független bármi mástól, és egyidejűleg nem is lehetnek jelen a lexikai elemek konkrét példányaiban.

3.7. Az igekötők jegyalapú osztályozása

A prototípus-elmélet központi gondolata mellett kitartva úgy gondolom, hogy az osztályozás – vagyis a kategóriahatárok behúzása – nincs összhangban a vizsgált jelenség természetével. Leginkább csak gyakorlati szempontból tartom ezt indokoltnak: jelen esetben azért van rá szükség, hogy megkönnyítse az egyes problémakörök tárgyalását a dolgozatban. Emellett látni kell azt is, hogy nem létezik olyan objektív mérce, amelyhez hasonlítva meg tudnánk állapítani az osztályozás eredményének a helyességét. Így ennek a fejezetnek nem lehet célja az objektív, hanem csak a minél kevésbé szubjektív kategorizálás. Elsőként egy olyan lehetőséget mutatok be, amely Smith et al. (1988) módszerén alapul, majd két klaszterező eljárást. Ezeket végül öt szempont alapján hasonlítom össze, és választok közülük.

3.7.1. Smith et al. (1988) módszere

Smith et al. (1988) a prototípus-elmélet keretében alakítottak ki egy olyan számítási módszert, amellyel meg lehet állapítani az egyes elemek tagsági értékét, majd ezeket az értékeket csökkenő sorrendbe állítva kialakítható egy kontinuum a legjellemzőbb kategóriatagoktól a legkevésbé jellemzőkig.¹⁸

A tagsági érték két változóból ered. Az egyik a diagnosztikus érték (*diagnosticity*), amely azt mutatja meg, hogy egy jegy mennyire releváns a kategóriába tartozás szempontjából. A másik a tulajdonságérték súlya (*saliency*), amely azt jelzi, hogy az adott jegynek mi a tipikus értéke. Ezt

¹⁸ Ennek rövid összefoglalását ld. Jakab (2014: 702-703), az ő magyar terminusait használom a módszer bemutatásánál.

a prototípusként választott elem – az igekötőkre vetítve tehát a *meg* – jegyértékei határozzák meg. A 9. táblázat azt szemlélteti, hogy eszerint hogyan lehet felvázolni az ‘igekötő’ kategória szerkezetét az eddig vizsgált jegyek alapján.

tulajdonság	diagnosztikus érték	tulajdonságérték	tul.érték súlya
morf. produktivitás	1	magas	1
		közepes	0,5
		alacsony	0
szerkesztettség	0,8	nincs	1
		van	0
szótagszám	0,8	1	1
		2	0,5
		3	0
gyakoriság	0,8	magas	1
		közepes	0,5
		alacsony	0
reduplikáció	0,6	előfordul	1
		ritkán fordul elő	0,5
		nem fordul elő	0
<i>-fele/felé</i>	0,2	előfordul	1
		ritkán fordul elő	0,5
		nem fordul elő	0
fokozhatóság	0	-	-
személyragozhatóság	0	-	-

9. táblázat. Az ‘igekötő’ kategória szerkezete Smith et al. (1988) módszeréből kiindulva. A *morf. produktivitás* a morfológiai produktivitás mindhárom típusát magába foglalja.

A diagnosztikus értékeket a 3.6. fejezetben leírtak mentén – tehát főként a prototípus tulajdonságai, a korreláció-vizsgálat és a szakirodalmi álláspontok alapján – alakítottam ki. A prototipikus tulajdonságok közül a minden szóra érvényes jegyeknek kisebb diagnosztikus értéket adtam (0,8), a reduplikációnak pedig szintén valamivel kisebbet (0,6), mert a hiányának sok elem esetében konceptuális és szemantikai megszorítások lehetnek az okai (ld. a 3.3.2. fejezetet).

Az egyes elemek tagsági értéke a diagnosztikus értékek és a tulajdonságértékek súlyainak lineáris kombinációja. Tehát az adott igekötőjelölt tulajdonságértékeinek a súlyát egyesével megszorozzuk az adott tulajdonság diagnosztikus értékével, majd összeadjuk a szorzatokat. Ez a ternáris

mátrixon valósítható meg két okból: (1) erre könnyen átfordíthatók a 9. táblázatban feltüntetett tulajdonságértékek, (2) az abszolút és relatív gyakoriságok esetében az eredmény nem sokban különbözne attól, mintha egyszerűen a gyakoriságuk alapján rendeznénk sorba az igekötőjelölteket.

A bemutatott módszerrel kapott igekötő-skála az I. számú mellékletben tekinthető meg. A legmagasabb tagsági értékű szavakat a szakirodalmak kivétel nélkül igekötőnek tekintik, a legalacsonyabb értékűeket pedig soha – ezek csak az automatikusan előállított igei vonzatkerettárakban vannak igekötőként címkézve.

3.7.2. Klaszteranalízis

A klaszterezés egy gyűjtőfogalom: olyan eljárás, amely adatpontokat csoportosít kizárólag olyan információ alapján, amely ezekből kinyerhető. A célja az, hogy az egy csoportba – klaszterbe – kerülő adatpontok egymáshoz hasonlóak legyenek, és más klaszterek adatpontjaitól különbözzenek (Tan et al. 2018: 528). A jegyeket itt dimenzióknak nevezzük, és az adatpontok hasonlósága vagy különbsége ezeknek a dimenzióknak a mentén határozható meg.

A klaszteranalízisnél használhatóvá válnak az abszolút gyakoriságokat tartalmazó mátrixok. Viszont ahogy láttuk, ezek az adatok Zipf-eloszlást mutatnak, a klaszterező eljárások pedig eredményesebbek, ha normál eloszlású adat a bemenetük: ekkor az adatpontok értéke kisebb skálán mozog, nincsenek nagyon kiugró értékek. Számos olyan skálázó eljárás létezik, amellyel az adat normál eloszlásúvá alakítható.¹⁹ Az Igekötő \times Jegy mátrixok esetében a Yeo–Johnson transzformáció bizonyult a legjobbnak, az eljárás matematikai háttérét ld. Yeo és Johnson (2000) tanulmányában. Előbb tehát ezt alkalmaztam az abszolút gyakoriságokat tartalmazó mátrixon, majd a transzformált adatot klasztereztem. A ternáris adaton is lefuttattam a klaszterezéseket (itt kicsi skálán mozognak az értékek, így nincs szükség transzformációra). A jegyeket minden esetben a 3.7.1. fejezetben meghatározott diagnosztikus értékek szerint súlyoztam.

Elsőként a k-közép (k-means) klaszterezést próbáltam ki (Pedregosa et al. 2011). Ennek az az előnye, hogy minden elemet csoportosít, ami a klaszterező eljárások esetében nem evidens tulajdonság. A hátránya viszont az, hogy előre meg kell határozni a klaszterek számát, bár a leoptimalisabb klaszterszám megállapítására vannak jól bevált módszerek.²⁰

¹⁹ Ezek jól áttekinthetők a scikit-learn (Pedregosa et al. 2011) weboldalan:

https://scikit-learn.org/stable/auto_examples/preprocessing/plot_all_scaling.html

²⁰ Én a „könyökmódszert” használtam, erről bővebben ld.: <https://www.geeksforgeeks.org/elbow-method-for-optimal->

A klaszterezés esetében csoportosítás történik, olyan sorbarendezés viszont nem, mint amit Smith et al. (1988) módszerével el lehetett érni. Ezért az eredményről úgy közlök mintát, hogy választok egy-egy elemet abból a három mezőnyből, amely az előző módszernél elkülönült (legmagasabb – közepes – legalacsonyabb tagsági értékek), és megmutatom, hogy a választott elem melyik más elemekkel került egy klaszterbe. A k-közép klaszterezés eredményét a 10. táblázat szemlélteti.

ternáris értékek	
meg	alá be bele el elő fel haza hozzá ide ki körül le meg neki oda odább rá szét tele tova vissza át össze
agyon	agyon egybe előre fejbe félbe félre helyre hátba hátra keresztbe keresztül ketté körbe közbe közre szembe tovább tönkre végig újjá újra
valóra	[...] semmibe sorban szemben szemre számon szárba színre szóba szót szóvá sárba síkra testet testre tetten tisztán tollba trónra táncra törbe utol valóra viszont végbe véghez végre zokon állást élen érvénybe észre
normalizált abszolút gyakoriság	
meg	alá be bele el elő fel haza hozzá ide ki körül le meg neki oda rá szét tele túl vissza át össze
agyon	agyon egybe előre félre helyre hátra keresztül körbe közbe szembe tovább utána végig újra
valóra	[...] semmibe sorba sorban sorra szemben szemre számba számon szárba színre szóba szót szóvá szörnyet sárba síkra talpra testet testre tetten tisztán tollba trónra táncra tönkre törbe valóra világra végbe véget véghez végre zsebre állást éhen életre érvénybe észre újjá útba útra

10. táblázat. Második módszer: Osztályozás k-közép klaszterezéssel. A felső sávban a ternáris, az alsó sávban a normalizált abszolút gyakoriságú jegyértékeken végzett klaszterezésből látható egy kis részlet. A klaszteren belül az elemek betűrendben követik egymást, tehát a sorrendnek elvi jelentősége nincs. A *valóra* klasztere esetében a felsorolás nem teljes.

A k-közép klaszterezés esetében azt látjuk, hogy a csoportosítás meglehetősen elnagyolt. A ternáris értékek és a normalizált abszolút gyakoriság között az *agyon*-féle klaszterben látszanak leginkább különbségek, vagyis az előző fejezetben bemutatott skála középső sávjában.

Ezután a HDBScan klaszterezéssel kísérleteztem (McInnes et al. 2017), ld. a 11. táblázatot. Ennek nagy előnye a k-középhez képest az, hogy nem kell előre megadni klaszterek számát. A

[value-of-k-in-kmeans/](#)

hátránya viszont az, hogy nem klaszterez minden elemet: ha egy adatpontnak egy vagy több dimenzióban kiugró értékei vannak (*outlier*), akkor azt kiszórja. Ez történik például a *ketté* vagy a *tönkre* esetében. A kiszórt elemek nem alkothatnak egy külön csoportot – csak annyi a közös bennük, hogy valamely jegyeik alapján nagyon eltérnek az átlagtól.

ternáris értékek

meg be el fel ki le meg rá szét át

agyon agyon egybe fejbe félbe ketté közre szembe tovább tönkre újjá újra

valóra hatályba küszöbön semmibe valóra érvénybe

normalizált abszolút gyakoriság

meg be el fel ki le meg rá szét át

agyon agyon félre körbe közbe végig

valóra hatályba igazat keresztbe kétségbe küszöbön lényegre oldalba piacra semmibe témába tükörbe valóra világgá világra zavarba életben életre érvénybe

11. táblázat. Harmadik módszer: Osztályozás HDBScan klaszterezéssel. A felső sávban a ternáris, az alsó sávban a normalizált abszolút gyakoriságú jegyértékeken végzett klaszterezésből látható egy kis részlet. A klaszteren belül az elemek betűrendben követik egymást, tehát a sorrendnek elvi jelentősége nincs.

3.7.3. A módszerek közötti választás

Az előző két fejezetben három módszert láttunk arra, hogy hogyan lehet meghatározni az igekötők állományát az Igekötő \times Jegy mátrixok alapján (és ez természetesen nem merítette ki az összes lehetőséget). Zárásképp érdemes sorra venni néhány olyan szempontot, amelyek mentén összehasonlíthatók ezek a módszerek. A szempontokat a 12. táblázatban tekintem át.

Az összehasonlításból világossá válik, hogy a három módszer egyike sem felel meg minden szempontnak. Ötből két szempont szerint gyengén vagy gyengébben teljesítenek. Innentől kezdve csak azon múlik a választás, hogy melyik szempontokat tartjuk feláldozhatóbbnak a többinél. Ha például ragaszkodunk ahhoz, hogy a normalizált abszolút gyakoriságot használjuk a ternáris értékekkel szemben, akkor valamelyik klaszteranalízist célszerű választani. Ha inkább azt tartjuk fontosnak, hogy tipikalitási skálát kapjunk, akkor egyértelműen az elsőként tárgyalt módszer jöhet

szempont	Smith et al. (1988)	k-közép	HDBScan
minden elem kategorizálható	✓	✓	✗
a kategóriák száma automatikusan alakul ki	✗	✗	✓
a tipikalitási skála automatikusan alakul ki	✓	✗	✗
a jegyek eltérő súlyokkal vehetők figyelembe	✓	✓	✓
normalizált abszolút gyakoriság használható	✗	✓	✓

12. táblázat. A három vizsgált módszer áttekintése öt szempont alapján. A ✓ azt jelenti, hogy a módszer megfelel az adott szempontnak, a ✗ pedig azt, hogy nem felel meg, vagy csak nagyon erőltetetten feleltethető meg neki.

szóba, amely eleve a prototípus-elmélet keretében alakult ki. Én ez utóbbi mellett döntöttem. Az I. mellékletben látható tipikalitási skálát 4 kategóriára egyszerűsítettem, amelyek a következők:

I. Prototipikus: *be el fel ki le meg*

II. Centrális: *át rá szét haza ide oda össze vissza alá bele hozzá neki elő elé körül tova túl odább tele széjjel szerte által alul felül hátra külön mellé*

III. Félperiférikus: *benn ott félre körbe közbe végig fölé utána agyon helyre előre ketté tönkre egybe fejbe félbe fenn közre szembe tovább újjá újra közé alább bent kinn keresztbe keresztül utol hanyatt helyt kívül együtt ellen nyilván rajta viszont hátba éhen helyben jót kölcsön rendbe szörnyet véget jól részt útra zsebre abba*

IV. Periférikus: *ágyba arcul békén célba csődbe csúcsra égbe észhez falra fejen férjhez főbe földhöz földre fülön füstbe hadba harcba hasba hasra házhoz helybe kezét kézre lábba mellbe nagyot nagyra nyakon partra pofán pofon porba rendre rosszul seggbe seggre sorba sorra szájon számba számot szárnyra szarrá szemem szemet szénné talpra tarkón térdre tükön torokon tűzbe útba vízre cserben egyet ellent észre jóvá karban közben közzé létre szemre szóba szót végbe véghez végre zokon*

A kategóriatagság küszöbértékét aszerint állapítottam meg, hogy mi az utolsó olyan tagsági érték, amely alatt már egy olyan elem sincs, amelyet legalább egy forrásmű igekötőnek jelöl Jakab (1976) adatai szerint. A küszöbérték eszerint a 2,7-es lett: az ennél kisebb tagsági értékű 73 elem így a kontinuum 'puszta névszói igemódosító' végpontját képviseli. A következő két fejezet korpuszmé-
réseinek ezek nem részei, de a dolgozat végén még vissza fogok térni rájuk.

3.8. Összegzés

Ez a fejezet azzal a kérdéssel foglalkozott, hogy mely szavak tekinthetők igekötőnek, és milyen szempontok alapján. A kiindulási pontom az volt, hogy a kérdés nem dönthető el világosan, az igekötők és más igemódosítók közé nem húzható éles határ, hanem fokozatos az átmenet. A feltevézésem az volt, hogy a prototípus-elmélet alkalmas lehet az igekötők állományának meghatározására. A megközelítésem szerint az igemódosítói szerepű lexikai elemek „igekötőségének a mértéke” aszerint határozható meg, hogy bizonyos jegyeik alapján mennyire hasonlítanak egy prototipikus elemre, amelyre egyezményesen azt tudjuk mondani, hogy ez egy igekötő.

A fejezet első részében a kutatás elméleti és módszertani háttéréről számoltam be. Röviden összefoglaltam az igekötő-fogalom meghatározásával kapcsolatos problémákat és nézetkülönbségeket, és kitértem arra, hogy az igekötő-állományt érintő bizonytalanság meglátszik a korpuszok meglehetősen következtelen annotációján is. Ezután bemutattam a prototípus-elméletet, amely lehetővé teszi az elmosódott kategóriahatárok kezelését, és megfogalmazható benne olyan állítás, hogy egy lexikai elem igekötősebb, mint más elemek. A prototípus-elméletet úgy kíséreltem meg átültetni a gyakorlatba, hogy meghatároztam és kimértem olyan morfológiai és általános jegyeket, amelyek (1) megfigyelhetők legalább egy lexikai elem esetében abból a 28 elemű halmazból, amelyet a Jakab (1976) által vizsgált összes forrásmű igekötőnek tekint, és (2) amelyek az MNSZ2 korpuszban jól kereshetők voltak.

A második részben, a kutatás adatgyűjtési szakaszában a következő jegyeket vizsgáltam:

- morfológiai produktivitás, amelynek Baayen (1989, 2009) nyomán három típusát különítettem el, ezek a megvalósult, a terjeszkedő és a lehetséges produktivitás
- a reduplikáció lehetősége (pl. *be-beoson*)
- a *-fele/felé* szócska csatolhatósága (pl. *nézzed megfele*)
- fokozhatóság (pl. *összébhhúz*)
- személyragozhatóság (pl. *rámnéz*)
- három általános jegy, amely minden szónál mérhető, de az igekötők esetében jellegzetes értékei vannak: szerkesztettség, szótagszám és gyakoriság

Az adatgyűjtés eredményéből Igekötő × Jegy mátrixokat hoztam létre, amelyek elsősorban a jegyértékek ábrázolásában különböznek. A legfontosabb a korpuszon mért abszolút gyakoriságokat tartalmazó mátrix, ebből vezettem le a relatív gyakoriságokat, valamint a ternáris (0 – 0,5 – 1 értékekre egyszerűsített) mátrixokat. Ezek nyílt hozzáférésű adatok, és további kutatásokhoz is felhasználhatók. Aki nem ért egyet az általam meghatározott igekötő-állománnyal, az ettől függetlenül felhasználhatja az Igekötő × Jegy mátrixokat a saját munkájához, mivel ezek egyszerű mérési eredmények – teljesen objektív adatok.

A fejezet harmadik, adatelemző szakaszában kiszámoltam minden jegypár korrelációját, először az abszolút gyakoriságokat tartalmazó mátrixon, majd a gyakorisági tényező kivonásával egy bináris mátrixon (itt tehát csak ‘van’ vagy ‘nincs’ értéke lehetett a jegyeknek). A korreláció-számítás legfontosabb eredménye a következő: Világosan kimutathatóvá vált a grammatikalizáció folyamata az igekötők szerkesztettségének, szótagszámának, gyakoriságának és produktivitásának az összefüggései alapján. A megvalósult és a terjeszkedő produktivitás erős pozitív korrelációt mutat a gyakorisággal, míg a szótagszám és a szerkesztettség ezekkel negatív korrelációban áll – a gyakori és produktív igekötők jellemzően rövidek és morfológiai szempontból bonthatatlanok.

A bináris adaton végzett korreláció-számításból az derült ki, hogy a jegyek közti alapvető összefüggések nem változnak attól, ha a gyakoriságra vonatkozó információt elhagyjuk. Ennek az az oka, hogy a gyakoriság történetileg olyan szorosan együtt jár más jegyekkel – éppen a grammatikalizációs folyamat révén –, hogy akkor is érezhető a hatása a vizsgált jegyekben, ha külön jegyként nem vesszük figyelembe. A korrelációs vizsgálat eredménye, a kapcsolódó szakirodalmak, és végül tagadhatatlanul a saját intuícióm alapján a *meg*-et jelöltem ki prototipikus igekötőnek, majd az igekötőség tipikus jellemzőit a *meg* tulajdonságai alapján határoztam meg.

Az adatelemző szakasz utolsó lépéseként arra tettem kísérletet, hogy a vizsgált lexikai elemeket a jegyértékeik alapján rangsoroljam egy tipikalitási skála mentén, és végül csoportosítsam ezeket. Ahogy jeleztem, ez nincs összhangban a jelenség természetével, ezért mindössze a tárgyalást megkönnyítő kategóriák megnevezéséről van szó. A jegyalapú osztályozás céljából három módszert próbáltam ki: Smith et al. (1988) prototípus-elmélet keretében kialakított számítási módszerét, valamint két klaszterelemzést. A módszereket ezután összehasonlítottam öt lényeges szempont szerint, amelyek a következők:

- Lehet-e bennük minden elemet kategorizálni?

- Automatikusan alakítják-e ki a kategóriák számát?
- Kiadnak-e tipikalitási skálát, vagy egymás mellé helyezik a csoportokat?
- Megoldható-e bennük a jegyek súlyozása?
- Használható-e bennük a (normalizált) abszolút gyakoriság, vagy egyszerűsített jegyértékekre van szükség?

Minden vizsgált módszerre igaz, hogy ebből az öt szempontból kettőnek nem felel meg. Így végül aszerint választottam közülük, hogy a kutatás célját tekintve melyik szempontokról lehet könnyebben lemondani, és az így fennmaradó lehetőségek közül melyik hozza a szakirodalomból ismert kategóriákhoz a közelebb eredményt. Az első módszert választottam, majd ennek alapján négy kategóriát neveztem meg: a prototipikus, a centrális, a félperiférikus és a periférikus igekötőket.

A prototipikus igekötők főbb jellemzői így határozhatók meg a vizsgált jegyek alapján: (1) Nagyon gyakoriak: az 1,04 milliárd tokenes, módosított MNSZ2-ben a *meg* adja a tokenek 3,5%-át, a *le* – amely a legkevésbé gyakori a prototipikusak közül – a 0,7%-át.²¹ (2) Szerkesztetlenek, azaz morfológiailag bonthatatlanok. (3) A fonológiai kopás következtében egyszótagúak. (4) A morfológiai produktivitásuk kimagasló: a terjeszkedő produktivitásuk 0,099 és 0,048 közötti. (5) Reduplikálhatók, és a -fele/felé morféma csatolható hozzájuk.

A következő két fejezetben az igekötőknek további jellemzőiről lesz szó – amelyeket disztribúciós és szemantikai jegyekként is értelmezhetünk –, ezért a dolgozat végén lehetőség nyílik az itt bemutatott megközelítés kiértékelésére az új adatok ismeretében. Ennek gyenge pontjait látva ismertetek egy lényegileg más megközelítést is, amelynek a középpontjában nem az egyes lexikai elemek, hanem az ezeket tartalmazó konstrukciók állnak.

²¹ Csak a közvetlenül a finit ige előtti előfordulásokat számolva.

4. Az igekötős szerkezetek szórendi mintázatai

„Ha elképzelni el tudja,
miért nem tudja megalkotni?”

(korpuszpélda)

Az igekötők mondattani viselkedésük alapján önálló szavak, szemben például a szláv nyelvek prefixumaival, amelyek sosem válhatnak el az igétől. A magyar nyelv igekötői nem pusztán elválni képesek, hanem több szónyi távolságba is kerülhetnek (pl. *meg ugyan nem vettem, adtak már itt el*). Ez a tulajdonságuk csökken, de nem szűnik meg azzal, ha az igéből igenevet képzünk (pl. *meg sem hallván, nem keverendő össze*). Emellett bőven találni példát arra is, hogy az igekötő az igenevekből képzett főnevek, melléknevek, határozószók esetében sem tapad szorosan az igei származékhoz (pl. *be nem következtenek, szét nem szakadós, el sem téveszthetően*). Ilyen és ehhez hasonló korpuszpéldák vizsgálatával a céloom két kérdés megválaszolása: Milyen mintázattípusokat mutatnak az igekötős szerkezetek? Mikor és mennyire távolodhat el az igekötő az igétől, igenévtől, igei származéktól?

Az igekötős szerkezetek szórendi mintázataival számos tanulmány foglalkozik.²² Azt viszont, hogy elválás esetén az igétől milyen távolságba kerülhet az igekötő, korábban nem vizsgálták, pedig egy ilyen kutatásnak gyakorlati és elméleti jelentősége is van. Az előbbire példa a nyelvtechnológiai felhasználás: gondoljunk egy szintaktikai elemző programra, amelynek az a feladata, hogy összekösse a mondatban az igekötőt azzal a szóval, amelytől elvált. Az utóbbi pedig abban nyilvánul meg, hogy az ‘igekötő – ige’ szórendet és az igekötő eltávolodását tekintve is kirajzolódnak olyan tendenciák, amelyeket csak adatvezérelt módszerrel lehet kimutatni.

A fejezet két nagy részből áll: egy szinkrón és egy diakrón korpuszvizsgálatból. A szinkrón korpuszvizsgálatban az igekötős szerkezetek széles skáláját mutatom be. A diakrón esetében egy szűkebb jelenségkört vizsgállok, a finit igék és a hozzájuk tartozó prototipikus igekötők viszonyát, de ezt nagyobb adathalmazzal végzem, az ómagyar kortól napjainkig terjedő szöveganyag bevonásával. A két kutatáshoz külön-külön módszertani alfejezetek tartoznak, mert ezek eltérnek a felhasznált korpuszokat, valamint az adatgyűjtés és adatfeldolgozás részleteit tekintve is. Végül összefoglalom a két vizsgálat legfontosabb eredményeit.

²²Ld. többek között J. Soltész (1959), Ackerman és LeSourd (1997), É. Kiss (2021).

4.1. Szinkrón vizsgálat: a mai magyar nyelvállapot

Ebben a fejezetben 20–21. századi szövegek vizsgálatán keresztül ismertetek számos igekötős szerkezetet, főként az olyan szerkezetekre fókuszálva, amelyekben az igekötő elválhat az igétől, igenevtől, vagy ritkábban egyéb igei származéktól. Ezeket elsősorban az igekötőik eloszlási mintái alapján jellemzem. Ahol csak lehet, kitérek az igekötő pozícióját befolyásoló tényezőkre, valamint a szakirodalomban eddig nem vagy kevésbé tárgyalt jelenségekre is.

4.1.1. A kutatás módszere

A szinkrón korpuszvizsgálathoz az MNSZ 2.0.4 módosított verzióját használtam (erről bővebben ld. a 2.2.1. fejezetet). Az adatgyűjtés első lépéseként lekértem a korpuszból minden olyan mondatot, amely valószínűleg tartalmaz igekötős igét, igenevet vagy egyéb igei származékot. Abból indultam ki, hogy az ilyen szerkezetek túlnyomó részében az igekötő egy tagmondatban jelenik meg azzal a szóval, amelyhez tartozik. Ez alól vannak kivételek – pl. az *el kell, hogy menjek* típusú szerkezet –, ezek kigyűjtésére utólag végeztem célzott kereséseket. Ezután a lekért mondatokból lokálisan külön alkorpuszokat hoztam létre a feltételezhető szerkezet típus szerint. Így egy mondat több alkorpusznak is a része lehetett, és egy alkorpuszon belül többször is megjelenhetett:

[...] , tehát minden egyes figyelembe vehető pont el kell legyen cseszve.

→ -hAtÓ végződésű melléknévi igenevek korpusza: *el+vehető*

→ finit igék korpusza: *el+kell*

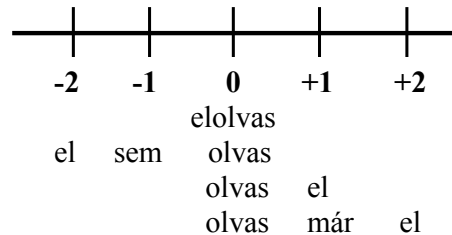
→ finit igék korpusza: *el+legyen*

→ határozói igenevek korpusza: *el+cseszve*

Az adatfeldolgozásnak egy jóval későbbi szakaszában döntöttem el, hogy a fenti példa négy lehetőségéből csak az utolsó érvényes, ezen a szinten még minden kombinációt szükséges volt feltételezni.

Az így létrejött összes alkorpusz minden mondatát szűrtem aszerint, hogy a feltételezett ‘igekötő + igei rész lemma’ kombináció szerepel-e a PREVLEX táblázatban (erről bővebben ld. a 2.2.2. fejezetet). Ez magával hozza jó találatok elvesztését is (pl. *elé+gördít, belé+gebed*), egyszerűen azért, mert ezek nincsenek feltüntetve a PREVLEX-ben. Sokkal nagyobb mértékű viszont a rossz kombinációk kiszűrése (pl. *ki+tud, bele+visel*).

Minden mondat esetében meghatároztam az igekötő pozícióját. Ez is automatikus módszerrel történt úgy, hogy a mondatokat szavanként egy számegyenesre képeztem le, ahol az origó az az igei rész volt, amelyhez az igekötő feltételezhetően tartozik. A negatív tartományba kerültek a preverbális, a pozitív tartományba a posztverbális igekötők, ld. a 6. ábrát.



6. ábra. Az igekötők pozíciójának meghatározása az igéhez vagy igenévhez képest.

A -1 pozíciót csak az olyan igekötőknél vettem figyelembe, amelyek gyakrabban jelennek meg az igétől különírvá akkor is, ha közvetlenül az ige előtti pozícióban vannak (tehát amelyeknek ez a helyesírási norma szerinti írásmódja). Itt főleg a periférikus csoportba sorolt igekötőkre kell gondolni, pl. *szénné, pofon*. A -1 pozíciót ilyenkor 0-nak tekintettem, mert nem akartam pusztán helyesírási tényezők miatt élesebben elválasztani a két esetet. Ugyanakkor felmerül egy olyan helyzet, amikor a -1 pozíció nem pusztán helyesírási kérdés, hanem tényleg van létjogosultsága. Ez akkor fordul elő, amikor az igekötő kontrasztív topik szerepű (és különösen akkor, ha a mondat egzisztenciális, vagyis egyszer már megtörtént eseményt hangsúlyoz), pl. *Be mentek már oda, de ki nem jöttek soha*. Ezek az itt bemutatott mérésekből kimaradtak, de tervezem megvizsgálni azt, hogy az eredeti -1 pozíciós adatok között mennyi ilyen szerkezet van.

Ezután következett a munka leghosszabb szakasza: minden alkorpusz minden pozícióját tanulmányozni kellett, majd külön szűrőszabályokat írni rájuk. A szabályok általában szófaji címkék sorozatára illeszkedő reguláris kifejezések. Egy gyakran használt szűrőszabály például az, hogy ha 'finit ige + igekötő + finit ige' sorozat szerepel a találatban, akkor az kiszűrendő, mert valószínűleg hibás annotációról van szó (pl. *erre ő csak állt meg bámult*, ahol a *meg* kötőszó tévesen igekötőnek elemzett). Ezeket a szabályokat aztán egyesével kellett futtatni az adott pozíció mondatain, és ellenőrizni, hogy valóban a hibás találatokat szűrik-e, és nem illeszkednek-e sok helyes találatra.

A fentiekből következik, hogy automatikus szűrésekkel nem lehet hibátlan kiinduló adatot előállítani, viszont lehet arra törekedni, hogy az adat minél tisztább legyen.²³ Ekkora korpuszméret

²³ Ezzel a problémakörrel több publikációmban foglalkozom, ld. Kalivoda (2016, 2017, 2018). Mivel ez inkább tech-

esetén a kézi ellenőrzés nem valósítható meg. Fontos tehát hangsúlyozni, hogy a továbbiakban bemutatásra kerülő eloszlási minták fenntartással kezelendők. A megadott darabszámok hibás találatokat is magukban foglalnak, de a vizsgált értékek egymáshoz való viszonya, arányai ezek ellenére is jól tükröződnek.

4.1.2. Finit ige

Az igekötős szerkezetek közül a leggyakoribbak a finit igei szerkezetek, ami egyáltalán nem meglepő, mivel a finit ige – amelynek ideje, száma, személye van – általában a mondat főigéje. Mielőtt ismertetném az igekötők disztribúciójára vonatkozó korpuszmérések eredményeit, érdemes tisztázni néhány fogalmat, amely szükséges az adatok megértéséhez.

Az első ilyen fogalompár a semleges és a nem-semleges mondat, mert a két típus különbsége az igekötő–ige szórendben is megmutatkozik. Alberti (2006: 37-38) három nyelvészeti terület irányából közelíti meg a semleges mondat fogalmát. Szemantikai szempontból azt mondhatjuk, hogy nincs benne szerkezeti többletjelentés, vagy csak minimális. Hangtanilag semleges intonáció jellemzi, vagyis a mondatban a tartalmas szavak az alapértelmezett kezdőszótagi hangsúlyukat viselik, a grammatikai funkciószavak hangsúlytalanok. Pragmatikai szempontból a semleges mondatban bizonyos szavak előfeltevést hordoznak, de a szórendi és intonációs sajátosságok nem hoznak további előfeltevéseket. A nem-semleges mondatban a hangsúly eloszlása nem egyenletes, és sajátos jelentéstöbbletet hordoz, amely szerkezetileg is megjelenhet (Alberti 2006: 54). Az igekötő semleges mondatban jellemzően preverbális,²⁴ nem-semleges mondatban lehet pre- és posztverbális, a mondat típusától és a szándékolt jelentéstől függően. A 13. táblázat áttekintést nyújt a kombinációs lehetőségekről.

A 13. táblázatban felsorolt mondattípusok közül a progresszív és egzisztenciális kijelentő mondatok igényelnek magyarázatot. Az előbbieket esetében azon van a hangsúly, hogy éppen történik az adott esemény, az utóbbiaknál azt emeljük ki, hogy az esemény már legalább egyszer előfordult. A táblázat utolsó sorában jelölt szerkezeti fókuszos mondattípus pedig más típusokkal is kombinálódik (pl. a kérdő és a tagadó mondatokkal). Az egyes mondattípusok gyakoriságát nem vizsgáltam.

nikai jellegű téma, nem tartom igazán ebbe a dolgozatba illőnek – az említett munkákban lehet erről bővebben tájékozódni.

²⁴ Létezik ige–igekötő sorrend semleges mondatban is, ld. Kádár (2007) és Peredy (2011) csángó nyelvjárási adatait. A magyar köznyelvre ez nem jellemző.

Mondattípus	Preverbális igekötő	Posztverbális igekötő
Kijelentő (semleges)	Lemegyek a közértbe.	-
Kijelentő (progresszív)	-	Verik le a képet.
Kijelentő (egzisztenciális)	-	Veszítettem már el könyvet.
Felszólító	Elhallgass végre!	Menj el hozzájuk!
Felkiáltó	Milyen korán felkelt!	Milyen későn jött meg!
Tagadó	El sem hiszem.	Nem hiszem én már el.
Tiltó	Meg ne kérdezd!	Ne mondd el!
Óhajtó	Bárcsak elhinné végre.	Csak szokna le Valenciáról!
Kérdő (eldöntendő)	Eljössz a találkozóra?	-
Kérdő (kiegészítendő)	-	Miért mennek el máris?
Szerkezeti fókuszos	-	Így került végül meg.

13. táblázat. Mondattípusok, amelyekben preverbálisan, posztverbálisan vagy mindkét módon megjelenik az igekötő, egy-egy példával.

Az igekötők finit igehez viszonyított, pozíció szerinti eloszlásának ismertetését²⁵ a **preverbális igekötőkkel** kezdem. Az igekötők tipikus helye a közvetlenül preverbális – „az igevel egybeírt” – pozíció. Az igétől elvált preverbális igekötőket intuitívan is jóval ritkábbnak vélhetjük, mert a magyar mondatok bal perifériáján az összetevők elrendeződése sokkal megszorítottabb, mint az ige utáni mondatszakaszban. A korpuszadatokban szélsőértéknek az igekötő -6-os pozíciója bizonyult, vagyis az igekötő és a finit ige között legfeljebb öt másik szó állt.²⁶

pozíció	-6	-5	-4	-3	-2	0
db	7	24	355	2 943	346 705	12 658 242
%	-	-	-	0,02	2,67	97,31

14. táblázat. Lehetséges igekötő pozíciók a finit ige bal oldalán.

A 14. táblázatban látható, hogy a közvetlenül preverbális helyzetű igekötők aránya kiemelkedően magas: 12,6 millió találattal a 0 pozíció fedi le a preverbális esetek 97,31%-át. A következő olyan pozíció, ahol a finit igehez tartozó igekötő áll, a -2. Ide leggyakrabban *is*-szerkezet (14a) vagy annak tagadása (14b, 14c), tagadó felszólítás, tiltás (14d) vagy tagadás (14e) esetén kerül az igekötő. Az igekötőt az igétől elválasztó szavak a *is, sem, se, nem, ne, sose, sosem, sohase, sohasem*.

²⁵ Ilyen méréseket korábbi munkáimban is bemutattam (Kalivoda 2016, 2018). Az itt közölt adatok számszerűen kissé eltérnek a korábbiaktól, aminek az az oka, hogy a publikációk óta a korpuszmérést nagyobb szöveganyagon és több lexikai elem bevonásával végeztem el.

²⁶ A disztribúciós adatok a 3.7.3. fejezetben megállapított igekötőcsoportok szerinti bontásban is elérhetők: https://github.com/kagnes/phd_thesis

- (14) a. 10-12 óra között **le** is **zárják** a teret.
- b. Pedig még **ki** sem **próbáltam**.
- c. **Vissza** se **gyere!**
- d. Aztán **fel** ne **fázz!**
- e. Egy 15 milliós terepjáróval **be** nem **mennék** a fák közé.

Az említetteken kívül azokat a tagmondatokat is a -2 pozícióhoz számoltam, ahol az igekötőhöz az -e simulószó járul, és ez választja el a finit igétől. Erre – a sztenderd nyelvhasználatban szokatlan – szerkezetre összesen 54 találat volt a korpuszban, amelyek közül hármat mutat be a (15) példasor.

- (15) a. És **meg-e találják** vajon a hetediket?
- b. [...] néztek engem és azon **gondolkoztak**, hogy **fel-e keltsenek**.
- c. A magyar területen **át-e jöttek**, azt én nem láttam.

Amíg -2-ben több mint 340 ezer igekötőt lehetett találni, -3 pozícióban mindössze 2 943 állt. A példák túlnyomó többsége tagadó szerkezet, és az igekötőt a leggyakrabban a következő szósorok választják el az igétől: *nem is, már nem, még nem, soha nem, én nem, senki sem, ugyan nem*, ld. a (16) példákat.

- (16) a. Azt feltettem, **le** már nem **szedem** potyára.
- b. **Le** sem nagyon **ütötte** a labdát szegény, csak csinálta, amit az edző kért tőle.
- c. **Meg** ugyan nem **vettem**, de elég alaposan beleolvastam.
- d. Ami viszont biztos, hogy **el** semmiképpen nem **vesz** a szórakozásból, [...]
- e. [...] egy hálátlan kutya, aki **meg** sem igen **érdemli** azt a nagy szerencsét, ami érte.

A -4 pozíciós igekötők már meglehetősen ritkák, ezekre mindössze 355 találat volt. Ezek leginkább centrális igekötők – különösen a *haza, vissza, ide, oda* –, ritkább esetben (fél)periférikusak, ld. a (17) példákat.

- (17) a. Még **haza** se nagyon szívesen **eresztik**.
 b. **Vissza** ezt már úgysem **szívják**...
 c. [...], hogy **oda** még véletlenül sem **tévedhet** a gyerek.
 d. De **tönkre** ebbe én nem **mentem**.
 e. Ha **egyet** gyakran nem is **értek** vele, de ezt azért becsülöm.

Elvétve még -5 és -6 pozícióban is találunk igekötőt, mindig kontrasztív topikként. Itt is a centrális és (fél)periférikus igekötők jellemzőek, ld. a (18) példasort.

- (18) a. [...] hét hónapot voltam otthon, **vissza** már nem az eredeti munkakörömbe **jöttem**, [...]
 b. [...], ketten mentünk, ám **haza** már hárman egy potyautassal **érkeztünk**.
 c. [...] mit **egybe** csak a súly maltera **köt**.

Az igekötők eloszlási mintáinak vizsgálatát a hátravetett, más szóval **posztverbális igekötőkkel** folytatom. A hátravetett igekötőt tartalmazó tagmondatok – a 24-es lábjegyzetben említett, nagyon ritka példák kivételével – nem-semlegesek. Érdekes adat, hogy +1 pozíciójú igekötőt több mint 8,1 millió mondatban találunk az MNSZ2-ben. Ez a 0 pozíció gyakoriságával összevetve azt jelenti, hogy az igekötők az esetek több mint egyharmadában posztverbálisak.

pozíció	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	+11
db	8 170 147	258 208	46 969	28 293	18 945	46 375	10	1	1	2	
%	95,86	3,03	0,55	0,33	0,22	-	-	-	-	-	

15. táblázat. Lehetséges igekötő pozíciók a finit ige jobb oldalán.

A 15. táblázatból az a tendencia olvasható ki, hogy az igekötők jóval nagyobb intervallumon helyezkednek el a finit ige utáni mondatszakaszban, mint az ige előtt, de hátravetéskor 95,86%-ban mégis közvetlenül az ige mögé kerülnek. A relatív gyakorisági adatok alapján azt is megállapíthatjuk, hogy a prototipikus igekötők távolodnak el legkevésbé az igtől (96,7%-ban közvetlenül követik), míg a másik három csoport tagjainál a +1 pozíció relatív gyakorisága 90 és 94,6% között alakul. Megfogalmazhatunk tehát egy olyan általánosítást, hogy minél igekötőszerűbb egy elem, annál valószínűbb, hogy posztverbális helyzetben közel marad a finit igehez.

A (19) példasor egy-egy mondattal illusztrál minden pozíciót:

- (19) a. Ennek nyomán 17 megyében **nyújtottak be** felülvizsgálati kérelmet [...] +1
- b. – **Heverj csak el** te is nyugodtan – mondta Pilátus. +2
- c. Akkor hogy **függnek** itt most **össze** a szándékok és a bekövetkező események? +3
- d. Én mindenesetre Dancsó úrral **értek** ebben a dologban **egyet**. +4
- e. Háromszor **veri** ezt kenden Lúdas Matyi **vissza!** +5
- f. Azért **mentem** egy kicsit a popzene felé **el**, [...] +6
- g. Tehát ezért **függnek** ezek a folyamatok egymással nagyon szorosan **össze** [...] +7
- h. Ezen leánnyal **utazott** ön azután egyszer később a héven estefelé **haza?** +8
- i. Azt mondják, száz napból tizenöt-húszban **sodor** a szél dögszagot a sárvári ATEV üzem felől **ki** az országútra. +9
- j. [...], hogy miért **teszi** megannyi egyébként oly sikeresnek látszó ember saját magát szánt-szándékkal **tönkre**. +10
- k. 27 gyereket **vitt** egy feltehetően részeg buszsofőr Szentesen még csütörtökön egy sportrendezvény után **vissza** az iskolába. +11

A (19i) és (19k) példákban egy helyhatározó áll közvetlenül az igekötő előtt illetve után, így fennáll annak a lehetősége, hogy az igekötő sokkal inkább a helyhatározóhoz, mintsem a finit igéhez kötődik. További példákhoz ld. Alberti (2006: 50).

A hátravetett igekötők eloszlási mintáinak ismeretében érdemes megvizsgálni **az ige–igekötő távolság lehetséges okait**. A 15. táblázat tanúsága szerint a posztverbális igekötők 98,9%-ban közvetlenül vagy egy szónyi távolságban követik a finit igét. Ennek a pszicholingvisztikai magyarázata elsősorban a távolsági hatásban keresendő, miszerint „minél nagyobb a távolság a szöveg alkotóelemei között, annál nagyobb a hiba lehetősége és annál kapacitásigényesebb az összekapcsolás” (Németh 2001: 94). Hasonló gondolatot fogalmaz meg Alberti (2006: 82) is: „Fontos forrása azonban a jelentés-kalkulációnak a lexikonból hozott jelentés is, amit viszont nehezebben felismerhetővé tesz például az, ha az igekötő(szerűen viselkedő elem) eltávolodik az igetőtől.”

Az igekötők szerepet játszanak az ige irányította mondatfeldolgozásban, argumentumok keresésében (Pléh 1998: 133-135), ezért intuitívan is számítani lehetett rá, hogy az ige elhangzása után minél előbb következnie kell az igekötőnek, amely módosítja az ige jelentését.

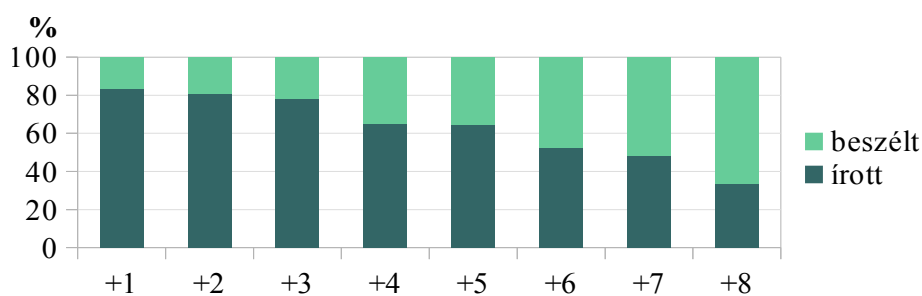
Fontos lenne magyarázatot adni az adatok 1,1%-ára is, tehát azokra, amelyekben az ige és az igekötő között kettő vagy több szó áll. Két szempontból²⁷ elemzem a korpusz adatait: (1) A metaadatok, főként írott és beszélt nyelvi oppozíció, valamint a stílusregiszter alapján. (2) Fonológiai szempontból, amelyet a szavak hosszának kimérésével lehetett megvalósítani.

Az első szempont, tehát az írott és a beszélt nyelvi stílus kapcsán abból indultam ki, hogy a szóbeli megnyilatkozások alapvetően tervezetlenek, spontának az írásbeliekhez képest, így szintaktikailag kevésbé szabatosak (Goody 1987: 258-289). Ezért arra lehet számítani, hogy a szélsőbb pozíciós igekötőt tartalmazó mondatok többsége eredetileg szóban hangzott el. Az MNSZ2 metaadatai segítségével viszonylag egyszerűen lehetett automatikus vizsgálatot végezni azzal kapcsolatban, hogy melyik igekötői pozícióra milyen szövegtípus jellemző. A szövegtípusra vonatkozó adatokat a következő kategóriákba rendeztem:

(1) *Írott nyelvi*: sajtó, személyes (ezek jellemzően internetes blogok és fórumhozzászólások), tudományos, szépirodalom (külön a próza és a vers), hivatalos (többnyire jogszabályok).

(2) *Beszélt nyelvi*: rádióban elhangzott szöveg, felszólalás.

A 7. ábra azt mutatja be, hogy az egyes pozícióknál milyen arányban reprezentáltak az írott illetve a beszélt nyelvi szövegek.



7. ábra. Írott és beszélt nyelvi szövegek százalékos aránya az egyes pozíciókon. A távolabbi pozíciókon már csak egy-egy találat volt, ezért ezeket nem vettem figyelembe.

Az adatok értelmezéséhez tudni kell, hogy a teljes MNSZ2-ben a beszélt nyelvi adat nagyon kevés

²⁷Érdemes volna szemantikai szempontból is megvizsgálni a posztverbális igekötős igéket, mert az is szerepet játszhat az igekötő eltávolodásának mértékében, hogy szemantikailag mennyire transzparens az adott igekötős ige. Ennek megvalósítása szisztematikusan, számítógépes módszerekkel a disszertáció kereteit meghaladó vállalkozásnak tűnt.

(Oravecz et al. 2014): a rádiós szövegek a teljes szövegállomány 5,4%-át alkotják, és ha a hivatalos szövegek felét szóban elhangzottak tekintjük, a végeredmény akkor sem több 9,4%-nál. A +1 pozíció szövegtípus szerinti megoszlása még nem tér el jelentősen a teljes MNSZ2-től: lényegesen több írott szöveg a forrása (83,51%), mint beszélt nyelvi (16,49%). A beszélt nyelvi szövegek aránya lassan növekszik: minél távolabbi igekötő-pozíciót vizsgálunk, annál valószínűbb, hogy szóban hangzott el a forrásmondat. A +7 pozíciónál már a mondatok több mint fele szóbeli szövegből származik. A (20) példásor első három tagmondata országgyűlési felszólalás során, az utolsó három pedig rádióműsorban hangzott el.

- (20) a. [...] **tette** a minisztérium ezt a pályázati felhívást **közzé** [...]
- b. [...] **került** talán ebben a jogharmonizációs folyamatban **elő** [...]
- c. [...] **támasztom** ezt a tőlem talán szokatlan megfogalmazást **alá** [...]
- d. [...] **dobja** valakinek az egyéniségét a hullámhegyre **fel** [...]
- e. [...] **merül** ez a kérdés ilyen sarkallatosan **föl** [...]
- f. [...] **jámulhatott** Ön szerint ehhez a szörnyű tethez **hozzá** [...]

Azt mondhatjuk tehát, hogy a szövegtípus és a posztverbális igekötő lehetséges eltávolodása összefüggenek: szerkesztetlen – főleg szóban elhangzott – szövegben nagyobb annak a valószínűsége, hogy az igekötő távol kerül a finit igétől, mint gondozott, szerkesztett szöveg esetében.

Térjünk rá a második szempontra, a fonológiai tényezőre. Behaghel (1932: 4-8) öt törvényt – pontosabban tendenciát – fogalmazott meg, amelyekkel magyarázatot kívánt adni a szavak és frázisok mondatbeli elhelyezkedésére. Ezek közül a negyedik azt mondja ki, hogy a rövidebb összetevő megelőzi a hosszabbat, ha ezt szintaktikai szabály nem gátolja.²⁸ É. Kiss (2007) amellet érvel, hogy ez a tendencia – a növekvő összetevők törvénye – a magyarban a finit ige utáni mondatszakaszra érvényes: az ige utáni, alapvetően szabad szórendet befolyásolja az, ha az igét követő összetevők fonológiailag különböző súlyúak. Ezt többek között a (21) példával szemlélteti. Az ideális sorrend a (21)-ban finit ige után az ‘igekötő – határozó – tárgy’, egyéb kombinációk legalábbis szokatlanul hangzanának.

²⁸ Ezt a jelenséget Pāṇini, a 4. században élt szanszkrit grammatikus is felismerte, a szakirodalomban Pāṇini törvényeként is hivatkoznak rá, ld. O'Connor (1978: 97-99) és Cooper és Ross (1975: 71). Ő a szanszkrit szóösszetételek vonatkozásában figyelte meg azt a tendenciát, hogy két összetételi tag közül jellemzően a rövidebb kerül előrébb.

(21) János sértette meg nagyon Éva nagyapját.

A bemutatott tendencia jól látszik a posztverbális igekötőt tartalmazó mondatok esetében. A mérést a következőképp végeztem el: Az olyan tagmondatokban, amelyek +1, +2 és +3 pozícióban tartalmaztak igekötőt, kimértem az ige után álló három szó hosszát (karakterszámban), és átlagoltam az egyes pozíciók hossz-értékeit. Az eredményeket a 16. táblázat foglalja össze.

alkorpusz	a finit ige utáni 3 szó átlagos hossza			%	a finit ige és az igekötő között álló leggyakoribb szavak
+1	2,9	4,1	4,7	95,86	-
+2	3,1	3,4	4,5	3,03	<i>majd, már, csak, -e, még, is, ezt, itt, most, ez, ...</i>
+3	3,5	5,4	4,2	0,55	<i>egy kicsit, én is, -e már, már csak, -e majd, ...</i>

16. táblázat. A növekvő összetevők törvénye – számokban. Az *alkorpusz* azt mutatja, hogy melyik igekötő-pozícióhoz tartozó tagmondatokról van szó, a 3. % oszlop pedig ezek aránya a posztverbális igekötők korpuszában.

A nagyon gyakori (95,86%-os), +1 pozíciójú igekötőt tartalmazó mondatokban az igekötő átlagosan 2,9 karakter hosszú, az ezt követő szavak 4,1, majd 4,7 hosszúak. A növekvő összetevők törvénye a +2 pozíciós igekötők esetében is érvényesnek látszik: az átlagosan 3,4 hosszú igekötőt annál valamivel rövidebb szó (3,1) előzi meg, és lényegesen hosszabb követi (4,5). A tendencia +3 pozíciós igekötők esetén már érvényét veszti: az átlagosan 4,2 hosszúságú igekötőt nála hosszabb szó előzi meg. Fontos észrevennünk, hogy ez a helyzet rendkívül kevés mondatban áll fenn, a +3 pozíciós igekötők a posztverbális eseteknek mindössze 0,55 százalékát alkotják.

Látni kell azt is, hogy ez a kísérlet nem a mondatösszetevők, hanem a szavak fonológiai súlyát vizsgálja, mert az MNSZ2 korpusz annotációja ezt teszi lehetővé. Ezzel tehát a növekvő összetevők törvénye még nem teljes körűen bizonyított, viszont az eredmények alapján nagyon valószínűnek tűnik, hogy a magyarban az ige utáni szabad szórend esetén a fonológia hat a szintaxisra, és az egyes mondatösszetevők sorrendjét befolyásolja a fonológiai súlyuk.

A 16. táblázatban látható tendenciára nemcsak az imént bemutatott fonológiai tényezők szolgálhatnak magyarázatként, hanem prozódiai okai is lehetnek. Varga (1981) azt a megfigyelést teszi, hogy a hangsúlytalan elemek jellemzően megelőzik a hangsúlyosakat a finit ige utáni mondatszakaszban. Ezt később Szalontai és Surányi (2020) kísérletes módszerrel ki is tudták mutatni. Mivel az igekötők – különösen az egy szótagú, prototipikus igekötők – gyakran hangsúlytalanok, ez is

hatással lehet arra, hogy a finit igét követően rendszerint azonnal vagy csak kis késleltetéssel hangzanak el.

4.1.3. Infinitívusz

A hagyományos grammatikák az infinitívuszt főnévi igenévnek hívják. Ez az elnevezés arra utal, hogy az infinitívusznak több olyan tulajdonsága van, amely a főnevekre jellemző, például azonos jelentésű az -*ás* képzős deverbális főnevekkel (vö. *írni fárasztó* → *az írás fárasztó*), és megjelenhet minden olyan mondatrészi szerepben, amelyben a főnevek, főnévi kifejezések is. Az itt tárgyalt téma szempontjából viszont lényegesebbek az igei tulajdonságai,²⁹ és ezeket jobban megragadja az ‘infinitívusz’ elnevezés, amely arra utal, hogy ezek az igeneves kifejezések értelmezhetők időjel nélküli tagmondatként is (vö. *kitörni készült* → *arra készült, hogy kitörjön*).

Az infinitívusz képzője a *-ni*, bár Kiefer (2003a: 175) megjegyzi, hogy ez a képzés korlátlanul termékeny, és a létrejövő szó jelentése kompozicionális, ezért a *-ni* inkább az inflexiók toldalékokhoz hasonlít. Az infinitívusz felvehet személyragokat, ha személytelen modális ige vagy névszói állítmány mellett áll, pl. *dolgoznom kell, érdemes folytatnod*.

Az igekötők eloszlási mintáinak vizsgálatokor szükségesnek tartottam két szerkezet típus elkülönítését. Az elsónél nem férkőzik be más ige a szerkezetbe, a másodiknál igen.³⁰ Mindkét esetben lehetséges az igekötő elválása az infinitívusztól a preverbális és a posztverbális tartományban is, de a mozgásterét tekintve nagy különbségek fedezhetők fel.

A tárgyalást azzal a szerkezetípussal kezdem, amelynek esetében nincs beférkőző ige vagy igenév, ld. a 17. táblázatot. Ekkor jellemzően névszói állítmány kerül a szerkezetbe, de nem feltétlenül: ide tartoznak az olyan esetek is, amikor például a *nem, sem, se, is* szavak valamelyike áll az igekötő és az infinitívusz között (pl. *nem lesz időm mindezt **be is fejezni***).

A (22) példasor illusztrál minden pozíciót.

²⁹ Az infinitívusz igei tulajdonsága például az, hogy lehet tárgya, viszont főnévi bővítményei – jelző, névelő – nem lehetnek.

³⁰ A vizsgált korpuszban a leggyakoribb beférkőző igék a következők: *kell, tud, lehet, fog, akar, kíván, szeret(ne), szokik* (múlt időben), *mer, tetszik* (mint udvariassági formula), *kezd, bír, próbál, szabad, talál, szándékozik, sikerül, merészel, óhajt, igyekszik*. Ezeket a továbbiakban segédigészerű elemeknek hívom (függetlenül attól, hogy beférkőznek-e a szerkezetbe vagy sem).

pozíció	-4	-3	-2	0	+1	+2
db	10	83	8 058	2 563 605	5 879	1 288
%	-	-	0,31	99,41	0,23	0,05

17. táblázat. Lehetséges igekötő pozíciók az infinitívusnál abban az esetben, ha a szerkezetbe nem férkőzik be finit ige vagy határozói igenév.

- (22) a. [...], hogy felsorolásukba **bele** sem is érdemes **fogni**. -4
- b. [...] a Windows Vista ennél kevesebb **el** sem hajlandó **indulni**. -3
- c. [...] és **el** képes **választani** a helyest a helytelentől. -2
- d. Nagyon nehéz **előjőnni** a sötét mélységből. 0
- e. [...], de egy vidéki városban, faluban nem feltétlenül **találni meg** eme számszámot. +1
- f. A műfordítás kétségtelenül úgy jó, ha nem **venni** rajta **észre**, hogy műfordítás. +2

Látjuk, hogy a hátravetett igekötős esetek ritkák: akkor fordulnak elő, ha az infinitívusz tagadott vagy saját fókuszja van. Az sem jellemző, hogy az igekötők preverbálisan elváljanak az igetőtől, az esetek 99,41%-ában közvetlenül előtte állnak. Ha mégis elválnak, akkor leggyakrabban egy névszói állítmány elé kerülnek. A gyűjtött anyagban a következő szavak fordulnak elő ilyen funkcióban: *szokás, képes(ek), szükséges, hajlandó(ak), kénytelen(ek), muszáj, érdemes, célszerű*. Bár ezek szófajlag nem igék, a beférkőzés képessége szempontjából a segédigeszerű elemek közé sorolhatók. Kálmán C. et al. (1989) ezek közül a *szokás*-t be is sorolják a segédigék csoportjába, és jelzik, hogy funkcionális szempontból más hasonló szavakra is tekinthetünk így.

Térjünk rá a második szerkezet típusra, amelyben finit ige vagy határozói igenév férkőzik be az infinitívuszi szerkezetbe. Az igekötők disztribúcióját ld. a 18. táblázatban.

pozíció	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	+2
db	5	7	31	110	262	569	5904	14533	96652	857010	959
%	-	-	-	0,01	0,03	0,06	0,6	1,49	9,9	87,8	0,1

18. táblázat. Lehetséges igekötő pozíciók az infinitívusnál, beférkőző szerkezet esetén.

A (23) példaszor illusztrálja az egyes pozíciókat. Két megfigyelést biztosan tehetünk: (1) Bár az igekötő preverbálisan messzire eltávolodhat az infinitívustól, mindig közel van a finit igehez, az

ahhoz tartozó igemódosítói pozíciót foglalja el. Közte és a finit ige között csak klitikumszerű szavak állnak (*is, se, sem, ne, nem*). (2) Segédigeszerű elem akkor is be tud férközni a szerkezetbe, ha az infinitívusz igekötője hátravetett, ld. (23k) példa.

- (23) a. **Meg** tudják az emberek ezt egy tulajdonosi közösségi gyűlés közben is **beszél**ni. -11
 b. [...] **el** tudnánk magunknak a saját szánk íze szerinti tökéletes édességet **készít**eni. -10
 c. [...] **fel** kell tudniuk az olívaolajtermelésre szánt támogatás egy részét **haszná**lni [...] -9
 d. [...], természetesen **hozzá** lehet majd az ön ügyrendi indítványához is **szól**ni. -8
 e. [...], ha **össze** lehetne őket többé-kevésbé objektív módon is **mér**ni, [...] -7
 f. **Át** kellett ezt a régi épületet **épít**eni, azóta is folynak még a munkálatok. -6
 g. Hát **fel** lehet ezt józan ésszel **fogn**i? -5
 h. [...] innen a cégtől már kifizelték, és **elő** kéne már otthon **fizet**niem, [...] -4
 i. Nekem nagyon kellemetlen, hogy állandóan **meg** kell önöket **szakít**anom. -3
 j. **Tönkre** akarja **tenn**i a bútoraimat? -2
 k. Közben már **húz**ni kezdi **be** a zsalut, aztán beteszi az ablakot is. +2

Az első esetet már többen leírták és magyarázták, legelőször Kálmán C. et al. (1989), akik rámutatnak, hogy bár az igemódosító az infinitívusszal alkot szótári egységet, szórendi és prozódiai szempontból szorosabban kapcsolódik a segédigeszerű elemhez. A méréseim alátámasztják ezt a megfigyelést, és kiegészítik azzal, hogy ha az 'igekötő + finit ige' egység elhangzott, az infinitívuszt akár az összes bővítmény is megelőzheti a további mondatszakaszban – minthogy a magyarban a finit igt követő tartományban a szórend kötetlen (É. Kiss 2007).

A második eset kapcsán azt figyelhetjük meg, hogy ez mindig nem-semleges mondat típusban, azon belül pedig progresszív mondatokban jelenik meg, tehát amelyek éppen zajló eseményt jelölnek. A beférköző ige szinte kizárólag a *kezd*.

A beférközés kapcsán megemlíteném még az, hogy a segédigeszerű elem határozói igenév is lehet. Erre összesen 183 példát találtam a korpuszban, néhányat a (24) példasor szemléltet. Érdekes az is, hogy ezek legnagyobb része (166 db) -vÁn képzős igenév, pedig ez sokkal ritkább, mint a -vA képzős párja.

- (24) a. [...] a látszatát is **el** akarván **kerülni** annak, hogy beavatkozik amerikai belügyekbe.
- b. [...], **el** kívánván **rettenteni** Milosevics jugoszláv elnököt attól, hogy [...]
- c. Aki gyámságról beszél, az a kritikát utasítja el kapásból, **meg** sem akarva **hallani** azt.
- d. [...], **rá** próbálva **ruházni** a felelősséget a legkülönbözőbb nehézségekért.

Megfigyelhetjük, hogy ilyen esetekben az igekötő a határozói igenév előtti pozíciót foglalja el, de az infinitívustól összességében nem kerül olyan messze, mint finit igék beférkőzésekor. Ilyen szempontból a határozói igenév a névszói állítmánnyal mutat párhuzamot, és további hasonlóság az is, hogy ezek egyike sem időjeles.

4.1.4. Határozói igenév

A határozói igenév képzője a -vA és a -vÁn. Az utóbbi az archaikusabb, és a mai nyelvhasználatban már lényegesen ritkább (a találatoknak mindössze 1,3%-át teszi ki). A két képzőnek nem pusztán a gyakorisága tér el, hanem a használati körük is, bővebben ld. Sárík (1998).

Akárcsak az infinitívusznál, úgy a határozói igenévnél is indokoltnak tartom két szerkezettípus elkülönítését. Az egyik esetben állapot- vagy módhatározóként áll önmagában (pl. *kihajolva integetett*), a másikban a létigével együtt passzív szerkezetet alkot (pl. *meg van oldva*). Látni fogjuk, hogy a határozói igenév igekötője is hátravethető, de csak akkor, ha az igenév állapot- vagy módhatározói szerepet tölt be. A passzív konstrukcióban az igekötő mindig preverbális.

A két típus elkülönítését indokolja az is, hogy a -vÁn képzős igenevek passzív szerkezetben nem állnak. Ezzel ellentmondani látszó példából mindössze három volt a korpuszban, de ezek figyelmen kívül hagyhatók, ugyanis az egyik példa egy versrészlet idézése, a másik kettő pedig 1500-as és 1600-as évekből származó mondatok idézése a mai magyar szövegen belül.

A 19. táblázat azt mutatja be, hogy hogyan alakul az igekötők mondatbeli eloszlása az első esetben, tehát amikor állapot- vagy módhatározói szerepű igenévhez kapcsolódnak.

pozíció	-3	-2	0	+1	+2	+3
db	49	2 326	514 634	10 785	111	4
%	0,01	0,44	97,48	2,04	0,02	-

19. táblázat. Lehetséges igekötő pozíciók az állapot- vagy módhatározói szerepű igenév esetében.

A (25) példaszor egy-egy mondattal illusztrál minden pozíciót.

- (25) a. Ez utóbbi törekvése **ki** ugyan nem **mondva**, sőt el is leplezve, de [...] -3
 b. Mindezt otthonról **ki** sem **mozdulva**, a számítógépünkön végezzük. -2
 c. Utólag **visszatekintve** ezt a középiskolások is tudják. 0
 d. Mese egy nőről, aki ablakvilágba zárta magát, onnan **szórva szét** üzeneteit [...] +1
 e. [...] gondozta titokban a sírt, állandó virágtakaróval **fedve** azt **be**. +2
 f. [...] ezzel **szoktatva** szépen lassan **át** az embereket az új rendszerre. +3

Ennél ritkább eset az, amikor a határozói igenév a létigével együtt passzív szerkezetben áll. A jelenség okát valószínűleg nyelvművelői törekvésekben kell keresnünk. Honti és H. Varga (2006), valamint H. Varga (2015) arról számolnak be, hogy a ‘létige + -vA képzős igenév’ típusú passzív szerkezet a 19. század második felétől kezdve „üldözött”, mert a kor nyelvművelői tévesen germanizmusnak ítélték, a közoktatásban tiltották, és irodalmi körökben is igyekeztek visszaszorítani. Az igekötők disztribúcióját a 20. táblázat szemlélteti.

pozíció	-9	-8	-7	-6	-5	-4	-3	-2	0
db	2	3	10	51	182	1 061	6 760	87 337	110 392
%	-	-	-	0,02	0,09	0,52	3,28	42,44	53,64

20. táblázat. Lehetséges igekötő pozíciók a passzív szerkezetben megjelenő igenév esetében.

Itt a finit igékhez viszonyítva szinte ellentétes tendenciát látunk az igekötők eltávolodásában: a periférikus igekötők közel maradnak az igenévhez, és a félperiférikusak között is csak elvétve van egy-egy eltávolodó példa, vannak lyukas pozíciók. Leginkább a centrális igekötőkre jellemző az eltávolodás, de prototipikusaknál sem ritka a jelenség. Talán arról lehet szó, hogy a (fél)periférikus igekötők kevés igével állnak, és ezek is olyan igék, amelyek szemantikai okokból nem vagy ritkán kerülnek passzív szerkezetbe. A (26) példaszor egy-egy mondattal illusztrál minden pozíciót.

- (26) a. Ez azért dühítő, mert az USA-ban **tele** van a közoktatás a bölcsitől kezdve olyan szakemberekkel **szórva**, akik [...] -9
 b. [...], hát teljesen **össze** van a keze a mai napig is **nyomorodva**, [...] -8

- c. [...] kirámoltak ugyan, csakumpak, de ezzel **el** is van a gond egy életre **intézve!** -7
- d. **Ki** vannak ezek a marketinges dolgok **találva.** -6
- e. **Meg** van az a hapsi **buggyanva.** -5
- f. **Meg** vannak ezek egyáltalán **fogalmazva?** -4
- g. Amerikában ezzel **el** is van **intézve** a dolog, Európában kicsit más a helyzet. -3
- h. A kocka **el** van **vetve.** -2
- i. Miért nincs minden **megoldva?** 0

A passzív szerkezetben álló határozói igenév szintaktikai viselkedése az infinitívusszal mutat rokonságot. Itt is megfigyelhető az, hogy az igekötő a finit ige előtti igemódosítói pozícióba törekszik, és a finit igével alkotott prozódiai egysége dominánsabb, mint az igenévvel alkotott szótári egysége. Minden nem közvetlenül preverbális pozícióra igaz, hogy a létige beférkőzik a szerkezetbe, néha egy másik, segédigyszerű elemmel együtt (pl. *el kell lennie küldve*).

4.1.5. Melléknévi igenév

A különféle igeneves szerkezetek közül a melléknévi igenevek azok, amelyek kapcsán a legnagyobb nézetkülönbségeket találjuk a szakirodalomban. Ezek jelentős része szófaji vita, amely abból a kérdésből indul ki, hogy minek tekintjük a melléknévi igenevet: igéből képzett melléknévnek, non-finit igének vagy önálló szófajnak? Ha az utóbbi két lehetőség egyikét fogadjuk el, akkor szükséges feltételeznünk, hogy a melléknévi igenév melléknévvé tud válni. Ez vezet el egy újabb szófaji kérdéshez: Hogyan állapítható meg, hogy egy adott mondatban a kérdéses szó melléknévi igenév vagy melléknév? A problémakör részletes tárgyalásához ld. Kenesei (2000) és Komlósy (1992) munkáit.

A szófaj-problémához kötődnek a melléknévi igenevek végződéseit érintő viták. Ha a melléknévi igenevet önálló szófajnak vagy melléknévnek tekintjük, akkor a toldaléka csak képző lehet, míg ha non-finit igeként elemezzük, a toldalék lehet inflexió is, bővebben ld. Kiefer (2003a). Ebben a kérdésben külön figyelmet kap a -hAtÓ státusza. Kiefer (2003a) amellet érvel, hogy ez melléknévképző, és a -hAtÓ toldalékos szavak az MNSZ2 korpuszban is melléknévként annotáltak. Ahogy a későbbiekben látni fogjuk, ez a megoldás gyakorlati szempontból előnytelen, ugyanis a

-hAtÓ végződésű igei származékok a szintaktikai viselkedésük tekintetében nagyon hasonlítanak a finit igékre. Az igekötőjük hátravetett is lehet, és messzire eltávolodhat az igei származéktól. Mindezt nehéz megmagyarázni – valamint a korpuszban megtalálni – akkor, ha ezeket a szavakat melléknévként címkézzük fel.³¹

Bőséges szakirodalma van a melléknévi igenevek alcsoportjait érintő kérdéseknek is. Ezek a rendezőelv-viták arra mutatnak rá, hogy a melléknévi igenevek hagyományos felosztása – folyamatos, befejezett, beálló – túláltalánosít, bővebben ld. Kiefer (1992) és Laczkó (2000b). Én úgy kerülöm meg ezt a problémát, hogy a különféle igenevekre nem a hagyományos terminusokkal, hanem egyszerűen a végzódéseikkel hivatkozok.

A fejezet további részében eltekintek az elméleti vitáktól, és rátérek a korpuszadatok ismertetésére. Ezek alapján meglehetősen unortodox módon fogom tárgyalni a melléknévi igenevek témáját. Először a -hAtÓ, majd az -AndÓ végződésű igenevekről lesz szó. Azért indokolt ezt a két típust egymáshoz közel bemutatni, mert szintaktikai szempontból nagyon hasonlóan viselkednek, ugyanis az igekötőjük hátravetett is lehet, ha állítmányi helyzetbe kerülnek (pl. *aligha oldható meg, azért ítélendő el*). Ezután az -Ó és -(O)(t)t végződésű igeneveket tárgyalom – ezeknek az igekötője csak preverbálisan válik el (pl. *el nem jövő korszak, vissza sem fizetett tartozás*).

Az első vizsgált típus, a **-hAtÓ végződésű igenév** tehát állhat és gyakran is áll hátravetett igekötővel, amikor állítmányi helyzetbe kerül. A 21. táblázat tanúsága szerint az igekötő egészen messzire kerülhet az igenévtől. A (27)-ban minden igekötő-pozició kapcsán láthatunk egy példát.

pozíció	-4	-3	-2	0	+1	+2	+3	+4	+5
db	5	189	9 085	508 962	125 172	1 585	45	6	2
%	-	0,03	1,41	78,9	19,4	0,25	0,01	-	-

21. táblázat. Lehetséges igekötő pozíciók a -hAtÓ végződésű melléknévi igenév esetében.

- (27) a. [...], keseregnek még ma is sokan a **vissza** már soha nem **fordítható** jelenségen. -4
 b. Így érthető, ha **el** nem is **fogadható**, hogy [...] -3
 c. Informatikával foglalkozni érdekes, soha **meg** nem **unható** foglalatosság. -2

³¹ Az igekötő még akkor is lehet posztverbális, ha a -hAtÓ képzőhöz a középfok jele járul (pl. *nem is tapasztalhatóbb meg teljesebben*), bár ez nagyon ritka. Az MNSZ2 korpuszban 2, a Webcorpusban 2.0-ban (Nemeskey 2020) 6 példát lehet rá találni.

- d. Ha a vizsgálat jellege szükségessé teszi, több szakértő is **kirendelhető**. 0
- e. A hidegneutronokkal sokféle elemzés **végezhető el**. +1
- f. Mennyi idő alatt **valósítható** ez **meg**? +2
- g. [...] ha van is, az mindig akkor **vonható** már csak **le**, amikor semmi értelme. +3
- h. [...], hogy gyors és nem **keverhető** már olyan könnyen **le** egy sporteseményen. +4
- i. Egyetértek még az eljárásban is, két szinten **indítható** szerintem ez a munka **el**: [...] +5

Ha az igekötő preverbálisan elvált, akkor rendszerint tagadó szerkezetet látunk, ritkábban *is*-est. Posztverbális igekötő esetében többnyire rövid szavak állnak az igenév és az igekötő között. Figyelemre méltó az is, hogy a -hAtÓ végződésű igenév nem csak az igekötők disztribúcióját, hanem a szerkezetbe gyakran beférkőző szavakat tekintve is hasonlít a finit ígére (ld. a 22. táblázatot) – hátravetett igekötője viszont csak akkor lehet, ha nincs finit ige a tagmondatban, így a -hAtÓ ennek a helyét foglalja el (Nádasdy 2001).

pozíció	köztes szavak
-3	már nem, nem is, még nem
-2	nem, is, sem
+2	majd, csak, ez, -e, még, az, már, itt, is, leginkább, jól, most, úgy, ezek, vele, ugyan, egyértelműen, csupán, talán, mindez, pontosan, újra

22. táblázat. Az igekötő és a -hAtÓ végződésű melléknévi igenév közti szavak, pozíciók szerinti bontásban. A táblázat csak a legalább ötször előforduló találatokat tartalmazza.

Az **-AndÓ végződésű igeneveket** a hagyományos nyelvtanok beálló melléknévi igenévnek nevezik. Komlósy (1992: 363) nem tartja szerencsésnek ezt az elnevezést, mert ezek elsősorban szükségességet fejeznek ki, nem annyira jövőt vagy utóidejűséget. A szakirodalmak rendszerint csak a -hAtÓ végű igenevek esetében említik azt, hogy hátravetett igekötőik lehetnek, az -AndÓ végűekkel kapcsolatban ez nem merül fel. A korpuszadatok mást mutatnak, ld. a 23. táblázatot. Az kétségtelenül igaz, hogy az -AndÓ esetében a posztverbális igekötők nem olyan gyakoriak, mint a -hAtÓ esetében, de nem is elhanyagolhatóan ritkák.

Az -AndÓ végződésű szerkezet esetében külön említést érdemelnek a centrális igekötők, amelyek az előfordulásaik közel egyharmadában (30,84%-ban) -2 pozícióban állnak. Ez komoly eltérés

a másik három csoporttól, amelyek relatív gyakorisága ugyanitt 0,12 és 0,92% közötti. Ennek az a magyarázata, hogy a centrális *vissza* igekötő nagyon sokszor megjelenik a következő kifejezésben: *vissza nem térítendő* (támogatás, hitel, kölcsön, stb.). Tehát tulajdonképpen ennek az egy igekötőnek a kollokációs viszonyai húzzák el az adott csoport egészét. Ugyanez történik a +1 pozíció esetében is, csak itt az *össze* igekötővel: a *nem tévesztendő össze* és a *nem keverendő össze* kifejezések gyakorisága eredményezi a centrális igekötők 7,55%-os értékét, míg a többi csoportnál ez 0,59 és 2,37% közötti.³²

pozíció	-3	-2	0	+1	+2
db	1	3 079	82 851	1 615	9
%	-	3,52	94,63	1,84	0,01

23. táblázat. Lehetséges igekötő pozíciók az -AndÓ végződésű melléknévi igenév esetében.

A (28) példaszor egy-egy mondat illusztrál minden pozíciót. Egyértelműen látszik, hogy a tárgyalt igenév állítmányként is szerepelhet a mondatban, hátravetett igekötővel.

- (28) a. [...] **tovább** már nem **bontandó** elemi műveletek vagy elemi utasítások [...] -3
- b. Ehhez ad az állam **vissza** nem **térítendő** támogatást. -2
- c. Ennyi potenciális, **leküzdendő** ellenség nincs. 0
- d. [...] egy vélemény nem **keverendő össze** az igazsággal a frankótutival. +1
- e. [...] a muzeális tárgyak védelmével foglalkozó rész **emelendő** még **ki**, [...] +2

A szerkezetben az igekötőt és az igenevet elválasztó szavak mind egy szótagúak, többségük klitikumszerű. -3 pozícióban a *már nem*, -2-ben *nem, is, sem*, +2 pozícióban pedig az *-e, még, csak* szavak fordulnak elő beférkőző elemként.

Az -Ó végződésű melléknévi igenévi kifejezésben az igekötő elválhat az igenévtől, de csak preverbálisan. Még fókusz vagy tagadás esetén sem lehet hátravetett. Az igekötők disztribúcióját a 24. táblázat szemlélteti.

Amint látjuk, ez egy nagyon megkötő szerkezet, az esetek 99,5%-ában az igekötő nem válik el az igenévtől. A centrális igekötőknél figyelhető meg a legnagyobb arányú elválás, de valószínűleg

³² Ez előrevetíti a kategóriák problematikusságát, amiről bővebben a 6. fejezetben lesz szó.

ennek is egy-egy kollokációs viszony az oka. A -2 pozícióhoz tartozó tagmondatokat vizsgálva az derül ki, hogy a leggyakoribbak között vannak a következő, centrális igekötőt tartalmazó szerkezetek: *oda nem illő, vissza nem térő, össze nem függő, össze nem illő*. Ezek szinte frazemaszerűek. A (29) példasor egy-egy mondattal illusztrálja az igekötő három pozícióját.

pozíció	-3	-2	0
db	6	9 012	1 811 393
%	-	0,5	99,5

24. táblázat. Lehetséges igekötő pozíciók az -Ó végződésű melléknévi igenév esetében.

- (29) a. [...] a Franciaországban élő, de **haza** soha nem **térő** Albert Dezsővel [...] -3
- b. Ha pedig egymással **össze** nem **függő** különféle jogszabályokat gyúrunk egybe, [...] -2
- c. A madarak beszéde éjszaka, hihetetlen, **megindító**, szinte emberi beszéd. 0

A beférkőző szavak -2 pozíciónál a *nem, is, sem*, ebben a gyakorisági sorrendben, -3 pozíciónál pedig a *soha sem, talán nem, ugyan nem, soha nem*.

Az **-(O)(t) végződésű igenevek** ugyanolyan szórendi mintákat követnek, mint az -Ó végűek. Az igekötő csak preverbálisan válhat el, de az esetek 99,29%-ában közvetlenül az igenév előtt marad. Tehát ilyen szempontból az -(O)(t) képzős melléknévi igeneveknek már jóval gyengébbek az igei, és erősebbek a melléknévi tulajdonságai.

Az igekötők disztribúcióját a 25. táblázat mutatja be. Ezek a gyakorisági adatok fenntartással kezelendők, mert az -(O)(t) végződésű igenév sokszor keveredik a múlt idejű finit igével az automatikus morfológiai elemzés és szófaji egyértelműsítés során. Sok esetben csak a kontextust ismerve lehet elkülöníteni a kettőt, amire ekkora, milliós nagyságrendű adat esetében nem lehet vállalkozni.

pozíció	-2	0
db	21 763	3 043 613
%	0,71	99,29

25. táblázat. Lehetséges igekötő pozíciók az -(O)(t) végződésű melléknévi igenév esetében.

Ha csoportokra bontva vizsgáljuk az igekötők eloszlási mintáit, olyan tendencia rajzolódik ki, hogy minél prototipikusabb az igekötő, annál könnyebben el tud válni az igenévtől – de valójában nem egészen erről van szó. A -2 pozíciós találatokat alaposan megvizsgálva inkább azt mondhatjuk, hogy minél prototipikusabb egy igekötő, annál több és gyakoribb kollokációs viszonya van, amely megszakított tagadó szórenddel rögzült. Például kiugróan gyakori a *meg nem nevezett* (forrás, személy), *meg nem erősített* (hír, információ), (18. életévét) *be nem töltött* (személy), *fel nem tett* (kérdés), *be nem tartott* (ígéret).

A (30) példásor egy-egy mondattal illusztrálja az igekötő két lehetséges pozícióját. Köztes szóként csak a *nem*, *is*, *sem* fordulnak elő, ebben a gyakorisági sorrendben.

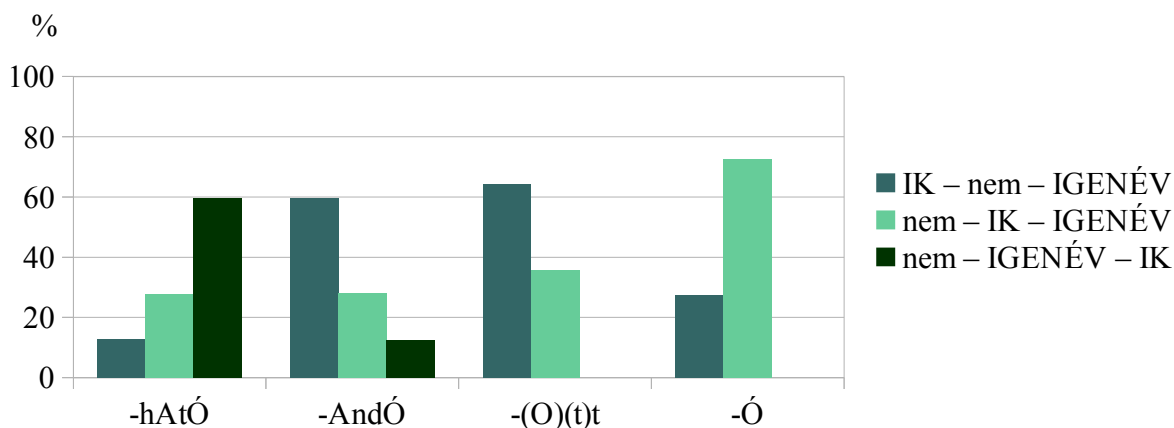
- (30) a. Az előző kormány által előírt és a kormányváltásig **be is tartott** menetrend szerint. -2
 b. Elég egy **elhamarkodott**, egy hirtelen gesztus, és mindenki félreérti. 0

Érdeemes még kitérni arra a kérdésre, hogy **tagadáskor milyen szórendi mintázatokat látunk** az igekötős melléknévi igenevek esetében. Arról már esett szó korábban, hogy hátravetett igekötője csak a -hAtÓ és -AndÓ végződésű igeneveknek lehet, és csakis akkor, ha ezek állítmányi helyzetben állnak – ekkor mind a három szórend előfordul, ld. a (31) példákat.

- (31) a. Ez az indoklás **el nem fogadható**.
 b. Ez az indoklás *nem elfogadható*.
 c. Ez az indoklás *nem fogadható el*.

Akkor is ingadozó szórendet figyelhetünk meg, ha a melléknévi igenevek jelzői helyzetben állnak: a tagadószó beférkőzhet a szerkezetbe, de ez nem szükségszerű. Kemény (2018) arról számol be, hogy a megszakított szórend (tehát ‘igekötő – tagadószó – igenév’) a régebbi – és a nyelvművelők körében elfogadottabb –, az új tendencia az, hogy az igekötő nem válik el az igenévtől.

A 8. ábra azt mutatja, hogy az egyes melléknévi igenév-típusok között jelentős eltérések vannak a különféle tagadó szórendek arányaiban. Ennek elsősorban nem mondattani okai vannak, hanem – ahogy korábban is láttuk – vannak gyakori szókapcsolatként rögzült szerkezetek, amelyeknek a tagadó szórendje is rögzült (pl. a *be nem tartott ígéret* bevett szófordulat, szemben a *nem betartott ígéret*-tel).



8. ábra. A lehetséges szórendek arányai a négy melléknévi igenév-típus tagadásakor.

4.1.6. A beférkőző *nem, sem, se, is* az igei származékokban

Az előző fejezet két nagyon megkötő szerkezettel, az -(O)(t)t és -Ó végződésű melléknévi igenevekkel zárult. Azt láttuk, hogy ezekben az igekötő csak preverbálisan válik el az igetőtől, és ekkor a *nem, sem, se, is* szavak valamelyike férkőzik be a kettő közé, ritkán a *soha, talán, ugyan* elemekkel együtt (pl. *véget soha nem érő*). Ez egy meglehetősen elterjedt mintázattípus. Az közismert és a szakirodalomban is tárgyalt jelenség, hogy ugyanilyen jellemzők mentén az -Ás végződésű főneveknek illetve gerundiumoknak³³ az esetében is lehetnek elvált igekötők – ld. többek között Laczkó (2000a), Kerekes (2011), Dékány (2014), Dékány és Hegedűs (2015) műveit –, de a korpuszadat tanúsága szerint még ennél is általánosabb jelenségről van szó: erre hívja fel a figyelmet ez a fejezet. A 26. táblázat áttekintést ad arról, hogy milyen típusú igei származékok esetében mennyire adatolt a négy említett elem beférkőzése a szerkezetbe.

Amint látjuk, a mintázattípus sok deverbális főnév, melléknév és határozószó esetében előfordul. Nem kiugró az a jelenség, hogy az -Ás végződésű szerkezet megszakított szórendű lehet: ha csak a kifejezetten gyakori – százezres előfordulás feletti – szerkezeteket nézzük, a -sÁg és az -(A)n esetében is kicsit magasabb a megszakított szórend relatív gyakorisága, de összességében minden esetben nagyon alacsony, a 0,2%-ot sem éri el. Szembetűnő még az írásmódok sokfélesége is. A korpuszpéldák szerzői többnyire külön írják a beférkőző elemet tartalmazó szerkezet egyes szavait,

³³ Szabolcsi és Laczkó (1992) az -Ás képzős alakulatokat két csoportra osztja a szintaktikai viselkedésük alapján. Az egyik csoport a „tény, cselekvés” jelentésű, komplex eseményt leíró szerkezeteké, amelyek az ige teljes vonatkoztatását megtartják, a másik az „eredmény” jelentésűeké, amelyeknek nincsenek kötelező vonzataik, így kevésbé igei jellegűek. A transzformációs generatív leírások az előbbi csoportot rendszerint igenévként, gerundiumként elemzik.

de nem ritka az sem, hogy kötőjelekkel választják el, vagy egy szóként írják le a teljes komplexumot.

végződés	példa	összes db	beférkőző példa	beférkőző db	%
- Ás	betanítására	7 884 148	vissza se nézés	5 361	0,07
-(A)t Al	megettélét	538 177	észre sem vétele	156	0,03
- i	felmentési	493 509	be nem avatkozási	291	0,06
-(A) n	megérdemelten	381 397	el is várhatóan	314	0,08
- sÁg	elhagyatottságot	181 124	meg nem értettség	280	0,15
- hAtAtlAn	elérhetetlen	160 435	be nem hozhatatlan	7	0,00
- ÓlAg	visszamenőleg	97 926	utol nem érhetőleg	1	0,00
- tA	beköszönte	49 645	létre nem jötte	21	0,04
-(A)t lAn	átlátszatlaná	37 501	ki nem adatlan	3	0,01
leg-...-bb	legelkeserítőbb	36 649	legmegnemfelelőbb	6	0,02
- Ul	felfoghatatlanul	24 163	ki nem mondatlanul	2	0,01
- Ós	kiülős	17 635	be nem vállalósok	14	0,08
- bb	összetettebb	13 720	megnemalkuvóbbak	2	0,01
- nivaló	elintéznivalóm	3 653	el-nem-felejténivaló	2	0,05
- sdi	kitúrósdi	685	észrenemvevősdit	1	0,15

26. táblázat. Igei származékok, amelyek megszakított szórenddel is adathatók. A beférkőző elemek relatív gyakorisága: *nem* 99,35%, *sem* 0,46%, *is* 0,15%, *se* 0,03%.

A táblázatban három fosztóképzőt látunk – -(A)t**lAn**, -**hAtAtlAn**, -**Ul** –, amelyek találkozása a beférkőző elemekkel némi magyarázatra szorul. Ha az *is* férkőzik a szerkezetbe, a létrejövő kifejezés elég szokatlanul hangzik, de jól átfordítható gyakoribb kifejezésbe, pl. *meg is valósíthatatlan* → *meg sem valósítható*, ld. a (32) példasort.

- (32) a. [...] és talán **meg is valósíthatatlan** társadalmi újításoktól várja minden üdvét, [...]
 b. [...] elég komoly elmaradásban vagyunk, és valahol **el is fogadhatatlan**, hogy [...]

A következő, *nem*-es példákban viszont hibázásokat látunk – (33) példasor –, egyszerűen nem a szándékolt jelentés adódik a szerkezetből. Arról van szó, hogy a két tagadás kioltja egymást: *le nem záratlan* → *lezárt*. Viszont a szerző nyilván arra gondolt, hogy *lezáratlan* vagy *le nem zárt*, ezek már ugyanazt fejezik ki.

- (33) a. [...] ebben a bizonyos **le nem záratlan** peres ügyünkben is [...]
 b. Elmosolyodtam, mikor felismertem a **le nem tagadhatatlan** hasonlóságait.
 c. [...] akkor szinte **be nem hozhatatlan** versenyelőnyre tett volna szert [...]

Megfigyelhető az is, hogy bizonyos szintaktikai környezetekben nem lehetséges a megszakított tagadó szórend, ekkor a *nem* az igei származék elé kerül. Ez a helyzet áll fenn fókusztagadásnál, ld. a (34) példásort. Ekkor ‘nem – hanem’ típusú szembeállítás van a mondatban, akár kimondva, akár csak odaértve.

- (34) a. [...], hogy **nem megérkezés** van, hanem út van.
 b. Ez már **nem elszámolódi**, hanem kiszámolódi.
 c. [...] ezt **nem elítélőleg** vagy akárhogyan írom, csak a tények kedvéért [...]

Amikor az igei származékot úgy tagadjuk, hogy nincs fókuszban, kétféle szórend lehetséges. A tagadószó beférkőzhet a szerkezetbe – ezt mutatja a 26. táblázat –, vagy állhat a szerkezet előtt, ld. a (35) példákat.

- (35) a. [...], amely egyelőre a jegyzőkönyv **nem aláírását** eredményezte.
 b. [...] a cigányságnak az elfogadottsága, vagy a **nem elfogadottsága**, [...]
 c. Úgy, mint az úszómedencét belevizelő és **nem belevizelő** részre?

Különösen érdekessé válik a helyzet akkor, ha a felsőfokú melléknévként funkcionáló igekötős szerkezet találkozik a tagadással.³⁴ Ekkor a *leg-* prefixum, az igekötő és a tagadószó (minden esetben a *nem*) sorrendje háromféle lehet: ‘leg – igekötő – *nem*’ (36), ‘leg – *nem* – igekötő’ (37),³⁵ ‘*nem* – leg – igekötő’ (38).

- (36) a. [...] a leggyűlöltebb, **legelnemfogadottabb**, leginkább árulónak tekintett párt [...]
 b. [...] a világ legdrágább és **legösszenemillőbb** giccseivel.
 (37) a. [...] a szó legszorosabb és **leg-nem-eltúlzottabb** értelmében a látásom a tét.
 b. [...] a **legnemmegfelelőbben** tartott kutyák gazdái vannak felháborodva [...]

³⁴ Elvileg a túlzófok is lehetséges, de erre csak állító szerkezetként lehetett példát találni a korpuszban.

³⁵ Ezek a példák a huTenTen12 korpuszból (Jakubíček et al. 2013) származnak.

- (38) a. [...] a kormányfő talán **nem legátgondoltabb** - bár felettebb becsületes megjegyzése [...]
 b. Mérlegelve azt, hogy az ingatlan a XVI. kerület **nem legfelkapottabb** területén van, [...]

Az itt bemutatott adatok rámutatnak tehát arra is, hogy egy adott jelenség morfológiai vagy szintaktikai voltát meghatározni nem kevésbé problematikus annál, mint amilyen problémákat az igekötők és az egyéb pusztá névszói igemódosítók közötti különbségtétel jelentett.

4.1.7. *El kell, hogy menjek – El kell menjek*

Ennek a fejezetnek a témája egy olyan szerkezet típus – vagy talán szerkezet típusok családja –, amelynek több, korábban tárgyalt szerkezet is az alkotórésze. Körülírás helyett egyszerűbb egy példasorral illusztrálni, ld. (39):

- | | | |
|------|--|--|
| (39) | fölül kell majd, hogy <i>múlja</i> | fel sem szabad <i>merüljön</i> |
| | el kéne, hogy <i>tudjunk jutni</i> | meg kell <i>akarjon halni</i> |
| | össze kell, hogy <i>legyen kötve</i> | rá kellett volna <i>legyen írva</i> |
| | meg kell, hogy <i>nyerhető legyen</i> | vissza kell <i>vezethető legyen</i> |

Három közös vonása biztosan van ezeknek a példáknak: (1) Az igekötő a – jellemzően személytelen – modális ige előtti igemódosítói pozícióban áll. (2) Ha az igekötővel lexikai egységet alkotó ige finit, akkor kötőmódban áll, ha non-finit vagy igei származék, akkor a hozzá tartozó modális vagy létige áll kötőmódban. (3) A *hogy* minden esetben elhagyható a szerkezetből.

A (39) példái közül a leggyakoribb és a szakirodalomból leginkább ismert mintázattípus az első, tehát amikor a kötőmódú finit igéhez tartozik az igekötő. A *hogy* kötőszót tartalmazó variánsnak 7 218, a *hogy* nélkülinek 7 451 korpuszbeli előfordulása adatolható.³⁶ Ennek a szerkezetnek – valamint további, szemantikailag hasonló szerkezeteknek, pl. *el kell mennem, el kellek menni* – az általam ismert legalaposabb leírását É. Kiss (2009) adja. Az ő elemzése szerint ebben a szerkezetben az igekötő kiemelése történik az időjeles mellékmondatból a fölérendelt tagmondatba. A továbbiakban É. Kiss egy-egy megállapításához kapcsolva mutatom be a korpuszadatokat.

³⁶ A korpuszlekérdezések során igyekeztem minden lehetséges írásmódot figyelembe venni: a *hogy*-ot nagyon sokszor *h*-nak rövidítik, és az is jellemző, hogy nem raknak elé vesszőt. Így tehát *hogy* kötőszót tartalmazó találatnak számoltam a következőket is: *el kell hogy menjek; el kell, h menjek; el kell h menjek*.

É. Kiss azt írja, hogy az igekötő kiemelésének a feltétele a modális ige és a *hogy* közvetlen szomszédossága. A korpuszadatok azonban azt mutatják, hogy ez nem mindig valósul meg: beférkőzhetnek rövid szavak, főként az *-e*, *majd*, *még*, *már*, ritkán a *most már*, *mindenki*, *nekem*, ld. a (40) példasort.

- (40) a. [...], és egyáltalán **ellent** kell-e hogy **mondjon** a kettő egymásnak.
 b. A szaktudásom alapján **fel** kellene majd, hogy **vegyenek**.
 c. [...] a felelősségtudata valóban **fel** kell most már hogy **ébredjen**, [...]

É. Kiss megállapítja azt is, hogy egy-két olyan kifejezésen, amely nem operátor-szerepű, az igekötő átemelhető. Én úgy találtam, hogy kifejezetten messzire is távolodhat az igekötő attól a szótól, amelyhez lexikailag tartozik. Ez párhuzamot mutat néhány korábban tárgyalt jelenséggel: hasonló volt megfigyelhető az olyan infinitívuszi szerkezetek esetében, amelyekben finit ige férközött be az igekötő és az infinit igealak közé, valamint a passzív, létigével álló határozói igeneves szerkezetekben. A *hogy* kötőszót is tartalmazó variánsokban jellemzően több kifejezés férközik be a kötőmódú igealak elé, de a *hogy* nélküliekben is találni hasonló példát, ld. (41).

- (41) a. [...], amit **el** kell hogy háromnegyed egy előtt néhány perccel **mondjak**: [...]
 b. [...], ahhoz tényleg **meg** kell, hogy az ember úgymond **bizonyosodjon** róla, [...]
 c. [...] igenis most már **meg** kellene valamilyen módon **tisztuljon** a XXI. század elején ez a vajúdo, szilánkos rendszerváltó politikai elit.

É. Kiss szerint ez a kiemelő szerkezet személytelen modálisok – *kell*, *szabad* – esetében jelenik meg. Az én korpuszmérésemben adatolható volt az *akar* is, bár erre nagyon kevés példa van, és csak a *hogy* kötőszós változatnál figyelhető meg, ld. (42). Ez feltehetően csak akkor fordul elő, ha a főige és az *akar* alanya különböző.

- (42) a. A Jutka **össze** akarja hogy **vesszünk**, de ez nem fog sikerülni.
 b. Legalábbis, ha **el** akarod hogy **fogadjam** az érveidet.

A szerkezetben reduplikált igekötő is megjelenhet, ld. (43):

- (43) a. [...] és bizony néha már magamnak is **meg-meg** kellett, hogy **toldjam** valamelyik történetet, hogy reggelig kitartson.
- b. Az állatok, mivel emlősök voltak, időnként **fel-fel** kellett **jöjjenek** a vízfelszínre, hogy levegőhöz jussanak.

A korpuszvizsgálat rávilágított arra is, hogy ez a mintázattípus sokkal általánosabb, mint ahogy a szakirodalom tárgyalja. Nemcsak a finit igék igekötőjét érinti, hanem non-finit igékét és igei származékokét is. Ezekről lesz szó a továbbiakban.

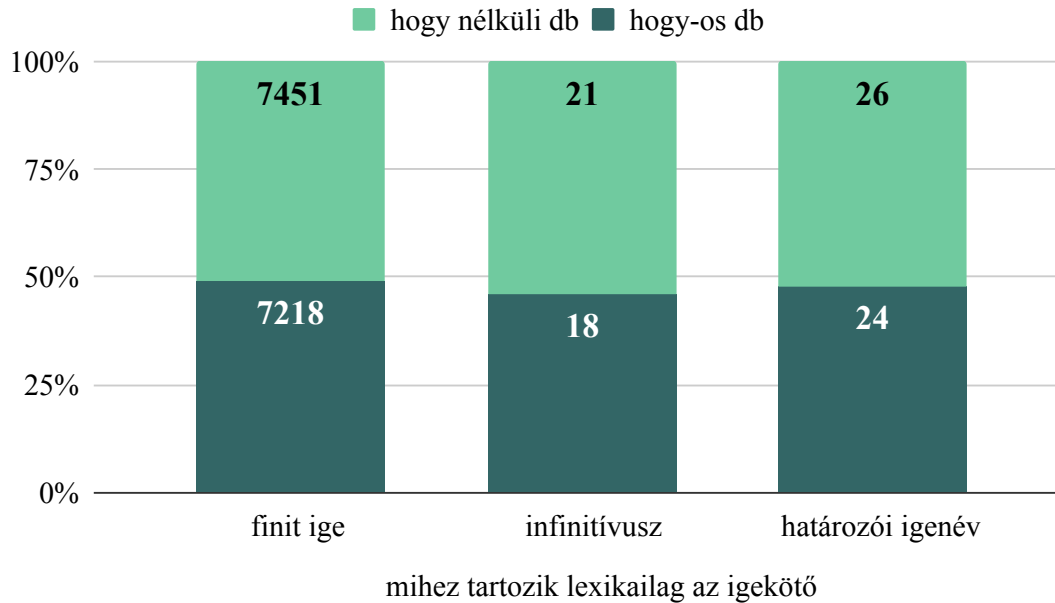
Míg a finit igés típusra a példák száma 7 000 fölötti volt mind a *hogy* kötőszót tartalmazó, mind a *hogy* nélküli esetekben, infinitívusra mindkét variációban csak huszas nagyságrendű példát lehetett találni. Az infinitívuszi alak előtt túlnyomórészt a *tud* ige jelent meg kötőmódban, ld. (44).

- (44) a. Egyébként is, mindent **meg** kell, hogy tudjon **oldani**, hiszen orvos.
- b. Valóban, hát **el** kéne, hogy tudjunk odáig **jutni**, hogy [...]
- c. Ebben a versenyben is **meg** kell tudjanak **állni** ezek a gazdaságok.

A 'létige + -vA' típusú passzív szerkezet – ld. (45) példasor – érdekessége az eddig látott mintázattípusokhoz képest az, hogy a -vA végződésű igenevet a kötőmódú ige egyaránt meg is előzheti vagy követheti.

- (45) a. De persze ehhez **ki** kell, hogy legyen **alakítva** egy megfelelő munkakultúra.
- b. Lélekben is **össze** kell legyél **forrva** mindenkivel.
- c. [...] általánosságban és konkrét ügyben is **fel** kell hogy **világosítva** legyen, [...]
- d. Ja, a kernelbe **be** kell **forгатva** legyen a PPP támogatás.

Ezen a ponton érdemes megvizsgálni az eddig látott mintázattípusok egymáshoz viszonyított arányait. Ezt mutatja be a 9. ábra. Látható, hogy minden esetben eléggé kiegyenlített a *hogy* kötőszót tartalmazó és nem tartalmazó korpuszpéldák aránya, de a *hogy* nélküli példák kicsivel gyakoribbak. Ezzel kapcsolatban érdemes majd olyan korpuszvizsgálatot is végezni, amely figyelembe veszi a forrásmondatok régióját: a *hogy* nélküli változat több szakirodalom szerint – ld. Posgay (2002),



9. ábra. A *hogya* kötőszót tartalmazó illetve nem tartalmazó változatok arányai az eddig vizsgált mintázattípusok esetében.

Dóra (2006), É. Kiss (2009) – Erdély felől kezdett terjedni a magyarországi régióba az utóbbi évtizedekben.

Az itt tárgyalt mintázattípus néhány egyéb igei származék esetében is megjelenhet. Ezek mindig megelőzik a kötőmódban álló igét, ezért gyanítható, hogy a ‘létiige + -vA’ szerkezet mintájára jöttek létre, ld. (46).

- (46) a. De ami értéket képvisel, **vissza** kell **vezethető** legyen akár sok áttételen keresztül [...]
- b. [...] pontosan **le** kell, hogy **szabályozott** legyen ebben a rendszerben.
- c. [...] **hogya** is mondjam, **meg** kell, hogy **akadályozásra** kerüljenek, [...]

A leggyakoribb, *el kell, hogy menjek* típusú szerkezet kapcsán É. Kiss (2009) azt írja, hogy az igekötőt – illetve bármilyen igemódosítót – mindig ki kell emelni a szerkezetből, kivéve akkor, ha a *hogya* kötőszóval bevezetett tagmondat tagadó vagy fókuszos. Érdekes módon arra is találunk néhány példát, hogy a tagmondat semleges, mégsem történik meg az igekötő kiemelése, ld. (47).

- (47) a. Kellett volna **elmenjek** nyugdíjba, de nem tudtam elmenni, [...]
- b. Kell, hogy a nemzeti emlékezet **megőrizze** a tragédiákat.
- c. Kell, hogy **lekössön** valami, kell a kuszaság, az unalmas csak árt neked.

4.1.8. Igekötős igék topikalizációja

Az utolsó jelenség, amellyel az igekötős szerkezetek szinkrón korpuszvizsgálata keretében foglalkozom, a topikalizáció. Ennek kapcsán először röviden ismertetek néhány fogalmat, amely a téma tárgyalásához szükséges.

A magyar mondat a kommunikációs szerepe alapján két szakaszra bontható: topikra és commentre. A topik a mondat elején elkülöníthető szakasz, általában azokat az elemeket tartalmazza, amelyekről a mondat szól. Az ezt követő szakasz a comment, amely a topikról tartalmaz állítást (pl. *A gyerekek*][*kirándulni mentek.*). A topikok egy speciális fajtája a kontrasztív topik: ekkor a mondat explicit vagy implicit formában azt is kifejezi, hogy van olyan elem, amelyre a szóban forgó predikátum nem teljesül (pl. *A levest*][*megette, a főzeléket viszont otthagya.*). A kontrasztív topikot követő comment első összetevőjét – amely kötelezően hangsúlyos –, a kontrasztív topik asszociáltjának nevezik (ez a példánkban a *megette* ige, lehet a fókusz is). Topikalizáció alatt azt a mondatszerkesztést értjük, amelynek során a topik szerepű elemek mondatkezdő helyzetbe kerülnek. A topik és kontrasztív topik szintaxisához és szemantikájához bővebben ld. Kálmán (2001) É. Kiss (2002), Gyuris (2002), a hangsúlyához és intonációjához ld. Kálmán és Nádasy (1994).

A továbbiakban olyan mintázattípusokat mutatok be, amelyekben az igekötős ige kontrasztív topik szerepbe kerül – leggyakrabban infinitívuszi alakban –, az asszociáltja pedig a finit ige, ld. a (48) példákat. Összesen 120 ilyen példát találni az MNSZ2-ben.

- (48) a. **Kitelni kitelik** tőle, azt már ugye tudom.
 b. De **feljelenteni** azért **feljelentik** – tette hozzá.

Érdekes variáció figyelhető meg akkor, amikor a finit igt tagadjuk. A fordított tagadó szórend érvényesül, viszont a hátravetett igekötő jelenléte opcionális: a korpuszadatban 109 esetből 25-ben elhagyták, ami az esetek több mint egyötöde (22,9%). A tagadásnál elhangzó igekötőhöz ld. a (49), a nem elhangzóhoz ld. az (50) példáit.

- (49) a. **Megbánni** nem **bántam meg** semmit.
 b. [...] de **összeveszni** még nem **vesztünk össze**.

- (50) a. Nem vagyok robot, **betanulni** nem **tanultam** semmit.
 b. **Nekiállni** nem **álltam** volna egy közepesen jónak sem, de ha már homár, akkor igen.

Amikor a komment szakaszban a finit ige segédigeszerű elem, egy elliptikus szerkezetet látunk, amely leegyszerűsítve ezzel a mintával írható le: ‘topikalizált igekötős infinitívusz + az infinitívusz igekötője + finit modális ige’. Tehát az infinitívuszt másodjára már nem ejtjük ki, csak az igekötőjét, ld. (51). Ez egy erős tendencia, amelyről mindössze 3 példa tér el a 201-ből (az eseteknek 1,5%-a), ld. (52).

- (51) a. **Betiltani be lehet**, de attól még létezik.
 b. **Elolvasni eltudom**, csak több próbálkozás után.³⁷
- (52) a. **Elpusztítani** ugyan **el lehet pusztítani** – volt rá példa –, de legyőzni nem.
 b. Meséltek nekem már minden féle mágia használatról így **felismerni feltudom ismerni** és itt éppen emberáldozatok bemutatása kellős közepére csöppentem.

Ha a segédigeszerű elem kevésbé hajlamos a beférkőzésre – a névszói állítmányok nagyrészt ilyenek –, akkor a kommentben már nem jelenik meg a topikalizált infinitívusznak sem az igekötője, sem az igei része (pl. *megelőzni célszerű, végigülni érdemes*). Ugyanez a helyzet tagadott komment esetén is (pl. *átgondolni nem tudta, végignézni nem érdemes*).

A létigével passzív szerkezetet alkotó határozói igenév sok párhuzamot mutat a segédigés infinitívuszi szerkezettel, ahogy azt több korábbi alfejezetben láttuk. A topikalizációt tekintve sincs ez másképp. A minta leegyszerűsítve így írható le: ‘topikalizált határozói igenév + az igenév igekötője + finit létige’. Mindössze 9 példát lehetett találni erre a korpuszban, néhányhoz ld. (53).

- (53) a. [...] **megverve megvolt**, az látszott kicsit az egyik szemén.
 b. **Kiírva ki van**, de úgysem kilóra mérik, hanem darabra, ahogy megszokták.

Az adatok alapján úgy tűnik, hogy az igekötőt a mondat komment szakaszában más, szintén a finit ige elé kívánczó elemek könnyen kiteszítják. Fentebb a *nem* tagadószt láttuk erre példaként,

³⁷ Az ilyen elliptikus szerkezetek esetében feltűnően gyakori az, hogy az igekötőt egybeírják az időjeles igével akkor is, ha lexikailag nem ahhoz tartozik. A korpuszpéldákat mindig az eredeti formájukban közlöm, függetlenül a helyesírási normáktól.

de fókuszált összetevő esetén is ugyanez a helyzet. Ekkor, ha a főige asszociáltja finit igeiként jelenik meg, az igeikötő megkettőzése opcionális (pl. *elmondani TE mondog el* vagy *elmondani TE mondog*), ha viszont a finit ige segédigeszerű elem, akkor a topikalizált infinitívusznak sem az igei, sem az igeikötői része nem ismétlődik meg (pl. *megoldani nekik kell, nem nekünk*).

A topikalizációval kapcsolatos megfigyelések tovább is általánosíthatók. Ha a mondatban már elhangzott egyszer a teljes igeikötős ige – nem feltétlenül topikként, hanem akár egy külön tagmondatban –, és a későbbi mondatszakaszban ez akár időjeles, akár időjel nélküli alakban megismétlődne, akkor az igeikötő és az igei rész is egyaránt könnyen elhagyható a finit ige elhangzása után. Az előbbire példa az (54), az utóbbira az (55).

- (54) a. [...] azt meg kell gondolni, hogy hát a kenyeret most **kidobom** vagy nem **dobom**.
 b. [...] akkor tudja eldönteni, hogy **megérte**, vagy nem **érte**.
- (55) a. Ez nyilván millió dolgot takar, lesz idő **kifilézni** és **ki is fogjuk**.
 b. [...] mi a magunk részéről **megettük**, amit **meg lehetett**.

4.2. Diakrón vizsgálat: az ómagyar kortól napjainkig

Az igeikötőzés, mint az aspektusjelölés alternatív módszere a 12. század végétől adatolható jelenség. Az ómagyar kor végére, tehát a 16. század elejére már stabilan működő rendszernek tekinthető (É. Kiss 2014a: 69-70). Az igeikötők állományát ekkor még kevés elem alkotja: számottevően a *meg, el, fel, ki, be, le* igeikötői használata mutatható ki a szövegelemekben, de az ómagyar kor későbbi szakaszában újabb lexikai elemek is egyre inkább igeikötőként funkcionálnak, bővebben ld. D. Máta (1989). Ebben a fejezetben az igeikötők legkorábbi rétegének, a hat prototipikus igeikötőnek a szórendi mintáit mutatom be az ó-, közép- és újmagyar kor korpuszai alapján. A kutatás a felsorolt igeikötők és a hozzájuk tartozó finit igeik egymáshoz viszonyított helyének, távolságának szisztematikus mérésén alapszik. Az eloszlási minták mellett azt is vizsgálom, hogy a tagadó mondatokra milyen arányban jellemző az ‘igeikötő – tagadószó – ige’, valamint a ‘tagadószó – ige – igeikötő’ szórend az egyes nyelvtörténeti korokban. Ez azért is érdekes kérdés, mert a szinkrón vizsgálat során láttuk, hogy a mai magyar nyelvállapotban sok igei származéktól csak tagadás ese-

tén válhat el az igekötő – ld. a 4.1.6. fejezetet –, mégpedig az elsőként említett, archaikus szórendet őrizve.

Először ismertetem a felhasznált korpuszokat, valamint az adatgyűjtés és az adatfeldolgozás módszerét. Ezután következnek a korpuszmérések eredményei, és végül egy olyan vizsgálat, amely a kétféle tagadó szórend arányának változását, valamint a fókuszos szerkezetek arányának növekedését mutatja be.

4.2.1. A kutatás módszere

A kutatás során négy korpuszt használtam, amelyek anyaga több mint 800 évet fed le. Az itt bemutatott mérések – az MNSZ2-n végzett mérés kivételével – 2017 májusában készültek, a korpuszok akkori állapotát tükrözik.

Az ómagyar kori adatok (896–1526) forrása az Ómagyar Korpusz – a továbbiakban ÓMK –, amely tartalmazza az összes fennmaradt ómagyar kori és néhány középmagyar kori szövegemléket (Simon és Sass 2012). Mérete 3,2 millió szövegszó. Ez a kutatás nem használja fel a teljes anyagot, csak azt az öt forrást, amelyhez morfológiai elemzés is tartozik.³⁸ Ezek a következők: *Jókai-kódex* (az eredeti szöveg 1370 körülről, mint másolat 1440 körülről), *Müncheni kódex* (1466), *Festetics-kódex* (1492–1494), *Guary-kódex* (1508 előtről), *Könyvecse az szent apostoloknak méltóságokról* (1521). A kutatás szempontjából releváns korpuszrész mérete 136 ezer szövegszó.

A középmagyar kori (1526–1772) szövegek vizsgálatához a Történeti Magánéleti Korpuszt – TMK – használtam (Dömötör et al. 2017). Itt az élőnyelvhez közel álló írásos anyagok érhetők el: főként boszorkányperek során készült jegyzőkönyvek, valamint misszilisek, elküldött levelek (főúri családi levelezések, jobbágylevelek). A korpuszvizsgálat idejében 850 ezer szövegszó alkotta a korpuszt,³⁹ amely teljes egészében szótövesített és morfológiaileg elemzett.

Az újmagyar kori (1772-től napjainkig tartó) szöveganyaghoz egyrészt a Magyar Történeti Szöveggyűjtő – MTSZ – használtam (Ittész 2009), másrészt a teljes disszertáció alapjául szolgáló MNSZ 2.0.4 módosított verzióját (erről bővebben ld. a 2.2.1 fejezetet). Az utóbbi 20–21. századi szövegeket tartalmaz, míg az MTSZ esetében az volt a cél, hogy a teljes újmagyar kor reprezentatív korpusza legyen. Az MTSZ 30 millió szövegszóból áll. Részlegesen feldolgozott, csak a szóalakok

³⁸ A 2017-es korpuszvizsgálat óta már elkészült a Jordánszky-kódex Újszövetség részének a morfológiaileg elemzett változata is.

³⁹ A TMK jelenlegi mérete 1 millió 50 ezer elemzett szövegszó.

érhetőek el benne. A morfológiai elemzés megvalósítása után a teljes korpusz részét képezhette a kutatásnak.

Az adatgyűjtésben a legfontosabb szempont az volt, hogy minden olyan mondat bekerüljön a vizsgált anyagba, amely egy tagmondaton belül tartalmaz egy finit igét és egy hozzá tartozó prototipikus igekötőt, bármilyen alakváltozatban (*meg~mög, fel~föl, be~bé*). Az ÓMK és TMK esetében viszonylag kevés anyaggal kellett számolni, így ezekben csak a prototipikus igekötőkre kerestem rá, és később saját scriptek segítségével szűrtem ki a finit igét is tartalmazó találatokat. Az MTSZ esetében – ahol csak a nyers szöveg állt rendelkezésre –, le kellett kérni minden olyan szövegrészt, amelyben a vizsgált igekötőkkel azonos írásképű szó szerepel (tehát ezen a szinten pl. a *mögötte* is érvényes találat volt). Csak a kapott találatok automatikus elemzése után lehetett tovább szűrni a mondatokat. Az MNSZ-hez kapcsolódó munkafolyamat a 4.1.1. fejezetben olvasható, itt a továbbiakban csak az ÓMK-t, TMK-t és MTSZ-t érintő adatfeldolgozási lépéseket ismertetem.

A lekérések után egységes formátumúvá kellett alakítani a három korpusz kimeneteit, hogy később mindet ugyanúgy lehessen kezelni. Az elemzett korpuszok esetében a gyűjtött anyagot olyan XML-be konvertáltam, amelyben egy sor egy mondatnak felel meg, és minden mondatnál elkülönülnek a metaadatok is, ha elérhetőek (pl. a pontos forrás, stílusregiszter). Az MTSZ-nél szükséges volt a betűhű szöveg normalizálása. Ez egyrészt a mai sztenderdtől eltérő karakterek átalakítását jelenti, másrészt a kötőjellel egybeírt szósorok elválasztását indokolt esetben (pl. *megbetstelenül* → *megbecstelenül*, *meg-sem-is-hallgat* → *meg sem is hallgat*). Ezután lehetett elvégezni a szófaji egyértelműsítést és morfológiai elemzést a magyarlanc 3.0 programmal (Zsibrita et al. 2013).

Minden korpusz esetében automatikus módszerrel szűrtem a tagmondatok halmazát. Csak olyan találatokat őriztem meg, ahol egy tagmondaton belül volt az igekötő és a finit ige. Azt is ellenőrizni kellett, hogy az igekötő valóban a finit igehez tartozik-e, nem pedig valamilyen igenévhez. Ezek a szűrések már tagmondatok szintjén történtek, de a teljes, eredeti mondatok elérhetőek maradtak az azonosítóik segítségével. A tagmondatokban az igekötő és a finit ige távolságát a 4.1.1. fejezetben bemutatott módszerrel vizsgáltam.

4.2.2. A prototipikus igekötők pozíció szerinti eloszlása

Az **ómagyar adatok** a már említett öt forrásból származnak. Ez összesen közel 8700 mondat, amelyből 4900 tartalmaz igekötős finit igét. A legnagyobb terjedelmű forrás a Münchener-kódex, amely a Huszita Biblia – az első magyar nyelvű bibliafordítás – részleteit tartalmazza. Németi György másolta a kódexet a moldvai Tatros településen 1466-ban (Nyíri et al. 1971). Ez a mű azért is érdemel külön említést, mert sok olyan nyelvi konstrukciót tartalmaz, amely a másik négy forrásban nem jelenik meg. A 27. táblázat szemlélteti az ómagyar anyagon végzett mérés eredményét. Látható, hogy az igekötők döntő többsége a preverbális tartományban van (83,3%), ennek fő oka az archaikus tagadó szerkezet fölénye az újabb, fordított szórendű tagadással szemben.

pozíció	-2	0	+1	+2	+3	+4
db	255	3 804	753	55	5	1
%	5,23	78,06	15,45	1,13	0,1	0,02

27. táblázat. Prototipikus igekötők mondatbeli eloszlása az ÓMK-ban.

A legszélső preverbális igekötői pozíció a -2, jellemzően tagadó mondatokkal, ld. az (56) példásort. Mindössze három olyan példa van, ahol a beférkőző szó a *sem*, és egy olyan, ahol az *is*.

- (56) a. [...] én lelkemet **el** nem **hagytad** sok veszedelmekben. (*Festetics-kódex, 1492–94*)
 b. [...] tehát önmagát **fel** nem **akasztja** vala. (*Guary-kódex, 1508 előtt*)

A posztverbális igekötőt tartalmazó mondatok – ld. az (57) példái – jellemzően felszólítások, vagy szerkezeti fókusz tartalmaznak. Valószínűleg a kódexek stílusából adódóan kevés a kérdő mondat. Az új tagadó szerkezet itt még elenyészően van jelen.

- (57) a. **Jelentette** vala az Úristen **meg** asszony Jacobának, [...] (*Jókai-kódex, 1440 körül*)
 b. Immár **bizonyíttassék** az **is meg**, [...] (*Guary-kódex, 1508 előtt*)

A találatok nagy része +1 pozícióban tartalmaz igekötőt, 55 esetben +2-ben. Ide kerültek a ‘finit ige – *vala/volna* – finit ige igekötője’ konstrukciók (mind nem-semleges mondat), főleg ezek alkotják a találatokat. A leggyakoribb beférkőző szavak: *azt, -e*.

A középmagyar adatokban (ld. 28. táblázat) nő a posztverbális igekötők aránya (16,7%-ról 23,8%-ra). A beszélt nyelvi jelleg miatt sok a felszólító mondat. A legszélső preverbális pozíció a -3, amely minden esetben tagadó mondatot jelöl. A beférkőző szavak egy tagadószó (*nem, sem, ne*) és az *is*, nem feltétlenül ebben a sorrendben, ld. (58).

pozíció	-3	-2	0	+1	+2	+3	+4	+5	+6	+7
db	5	3 098	16 833	5 675	475	48	14	7	2	3
%	0,02	11,84	64,35	21,69	1,82	0,18	0,05	0,03	0,01	0,01

28. táblázat. Prototipikus igekötők mondatbeli eloszlása a TMK-ban.

- (58) a. [...] hogy a szemeim héjai **ki** is ne **forduljanak** miatta. (*Kanizsai Orsolya, 1559*)
 b. [...] az ő kenése és segedelme nélkül **meg** sem is **csillapodtak**. (*boszorkányper, 1750*)

A TMK korpusz érdekessége, hogy viszonylag sok olyan mondatot tartalmaz, amelyben a hátravetett igekötő távol került a finit igétől. Az (59) példákban látható, hogy ez nem csak egy-két forráson belüli elszigetelt jelenség, hanem a teljes középmagyar korra jellemző, és különböző személyektől származó adat (ahol tulajdonnév van feltüntetve, ott az adott levelet diktáló személyről van szó). Figyelmet érdemel az is, hogy a példákban az igekötők után tagmondathatár van, tehát ezek nem elemezhetők valamilyen határozói frázis részeként (mint pl. *be a szobába*):

- (59) a. [...] gyakran **égeti** a kása a szánkat **meg**, [...] (*Kanizsai Orsolya, 1550*)
 b. [...] immár **adnáják** csak a hideg tetemet **meg**, [...] (*Nádasdy Anna, 1551*)
 c. [...] **menne** együtt az én [szol]gámmal **be**, [...] (*Telegdy János, 1593*)
 d. Úgy **küldém** Szokoly Albertet mindjára mintegy posta módon **el**, [...] (*Telegdy Kata, 1600*)
 e. [...] **tartson** az Isten jó egészségben **meg!** (*Poppel Éva, 1629*)
 f. [...] **jött** a száján olyan szózat **ki**, [...] (*boszorkányper, 1700*)
 g. [...] hogy soha bizony sem **gyógyul** ő abból a nyavalyából **meg**. (*boszorkányper, 1714*)
 h. [...] **jött** is egy kis munka németül róla **ki**, [...] (*Szilágyi Sámuel, 1730*)

Az **újmagyar adatok** esetében két korpusz is rendelkezésre állt, az MTSZ és az MNSZ. Az előbbi 1772 és 2010 között keletkezett forrásokat tartalmaz, ezek adatai láthatók a 29. táblázatban. A megszakított szórend itt is a tagadó mondatokra jellemző – ld. a (60) példákat –, de jól látszik, hogy a preverbális igekötők aránya tovább csökkent (70,9%).

pozíció	-3	-2	0	+1	+2	+3	+4	+5	+6	+7
db	61	21 237	237 643	102 200	3 604	127	14	12	2	2
%	0,02	5,82	65,13	28,01	0,99	0,03	-	-	-	-

29. táblázat. Prototipikus igekötők mondatbeli eloszlása az MTSZ-ben.

(60) a. Kin kívül más **meg** is nem **oltalmazhatja**. (*próza, 1781*)

b. Azért bizonyosan **meg** sem is **mondhatom**, [...] (*próza, 1781*)

A posztverbális igekötők esetében itt is észlelhető néhány nagyobb eltávolodás a finit igétől – ld. a (61) példákat –, de ez a korpusz méretéhez viszonyítva nem olyan jelentős, mint a TMK esetében.

(61) a. **Keljen** bár az egész világ **fel** ellened! (*próza, 1849*)

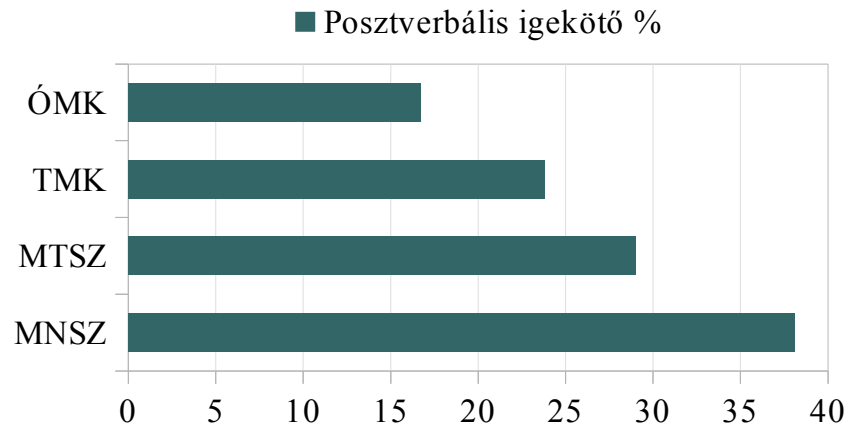
b. [...], nem **üledpedhetik** a víz a kávéval annyira **meg** a testben, [...] (*orvosi szöveg, 1791*)

Az MNSZ korpusz 20–21. századi szövegeiben a preverbális igekötők aránya csak 61,9% (ld. 30. táblázat). Ekkorra már teret nyert a fordított szórendű tagadó szerkezet. Az MNSZ-en végzett korpuszmérést részletesen tárgyalja a korábbi 4.1.2. fejezet.

pozíció	...	-3	-2	0	+1	+2	+3	+4	+5	...
db		1 753	258 301	10 048 828	6 123 146	153 361	25 583	17 506	13 299	
%		0,01	1,55	60,38	36,79	0,92	0,15	0,11	0,08	

30. táblázat. Prototipikus igekötők mondatbeli eloszlása az MNSZ-ben – részlet. Egy-két példa van a -5 és -4, valamint a +6, +7, +8, +9 pozíciókban való előfordulásra is.

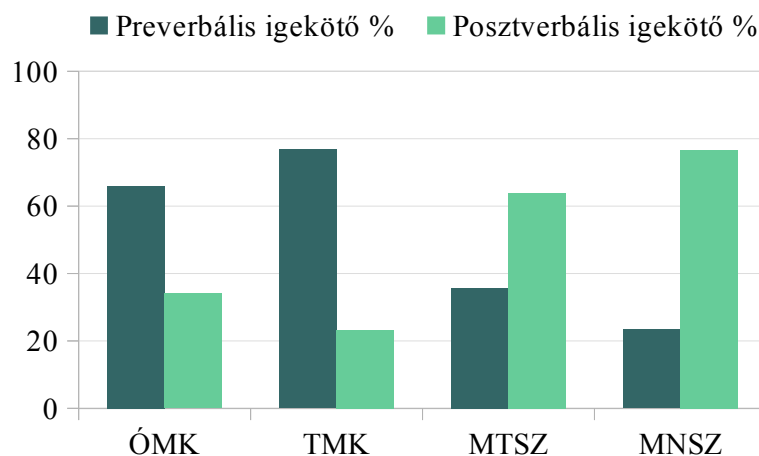
Az igekötők eloszlásában megfigyelhető az a tendencia, hogy az ómagyar kortól napjainkig folyamatosan növekszik a posztverbális igekötőt tartalmazó, nem-semleges mondatok aránya. Az Ómagyar Korpuszban még csak 16,7%-os a jelenlétük, míg az MNSZ-ben 38,1%, ld. a 10. ábrát. Ez összefüggésbe hozható a fordított szórendű tagadás térnyerésével, amelyről a következő fejezetben lesz szó.



10. ábra. A posztverbális igekötőt tartalmazó, nem-semleges mondatok arányának változása a vizsgált korpuszokban.

4.2.3. A kétféle tagadó szórend arányának változása

É. Kiss (2014c) részletesen foglalkozik a kétféle mondattagadó szerkezettel, amelyek közül az archaikusabbra az ‘igekötő – tagadószó – ige’ mintázat jellemző (megszakított szórend), az újabbra pedig a ‘tagadószó – ige – igekötő’ (fordított szórend). Az ezzel kapcsolatos megfigyeléseket árnyalja tovább Gugán (2015, 2017), aki a tagadó mondatok szórendjét ó-, közép- és korai újmagyar kori korpuszadatok alapján vizsgálja. Megállapítja, hogy a fordított szórend a kezdetektől jelen van és fokozatosan nyer teret: 1840-50 körülre – tehát az újmagyar kor középső harmadára – tehető az a szakasz, amikor ez a szerkezet válik gyakoribbá. Ennek kapcsán végeztem egy mérést, hogy látni lehessen, hogyan oszlik meg a tagadó mondatok aránya az egyes korpuszok preverbális és posztverbális igekötős mondatai között. Az eredményt a 11. ábra szemlélteti.

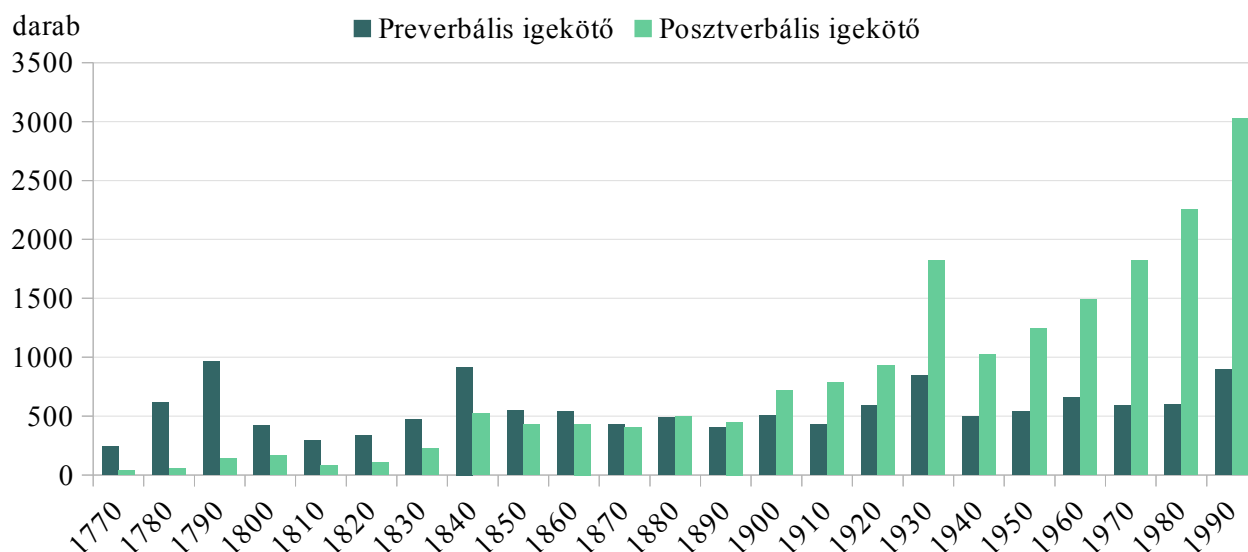


11. ábra. Megszakított és fordított szórendű tagadó mondatok százalékos aránya a vizsgált korpuszokban.

A megszakított szórendű tagadás jóval nagyobb arányú a fordítottnál az ó- és középmagyar adatokban, ahogy arra számítani lehetett. Kérdés viszont, hogy a TMK korpuszban miért gyakoribb (körülbelül 10%-kal), mint az ÓMK-ban. Ha S-görbét nem is, de monoton csökkenést várnánk. Elképzelhető, hogy az ÓMK adatait torzítják a Münchener-kódex mondatai. Ezekben általánosnak mondható még a közvetlenül preverbális igekötő is a tagadás esetén, ami így sem a megszakított, sem a fordított szórendhez nem sorolható be, ld. a (62) példásort. Ez azonban elszigetelt jelenségnek tűnik, a másik négy forrásra nem jellemző.

- (62) a. *nem megoldja*-e ő tehenét
 b. ki *nem megtöretik*
 c. ki *nem megtudatik*

Lehetséges magyarázat az is, hogy a beszélt nyelvben erősebb volt az archaikus változat jelenléte, ezért gyakoribb ez a TMK-ban. Az MTSZ korpuszt megvizsgáltam évtizedek szerinti bontásban is, hogy a fordított szórendű tagadás térnyerését pontosabban nyomon lehessen követni (12. ábra).

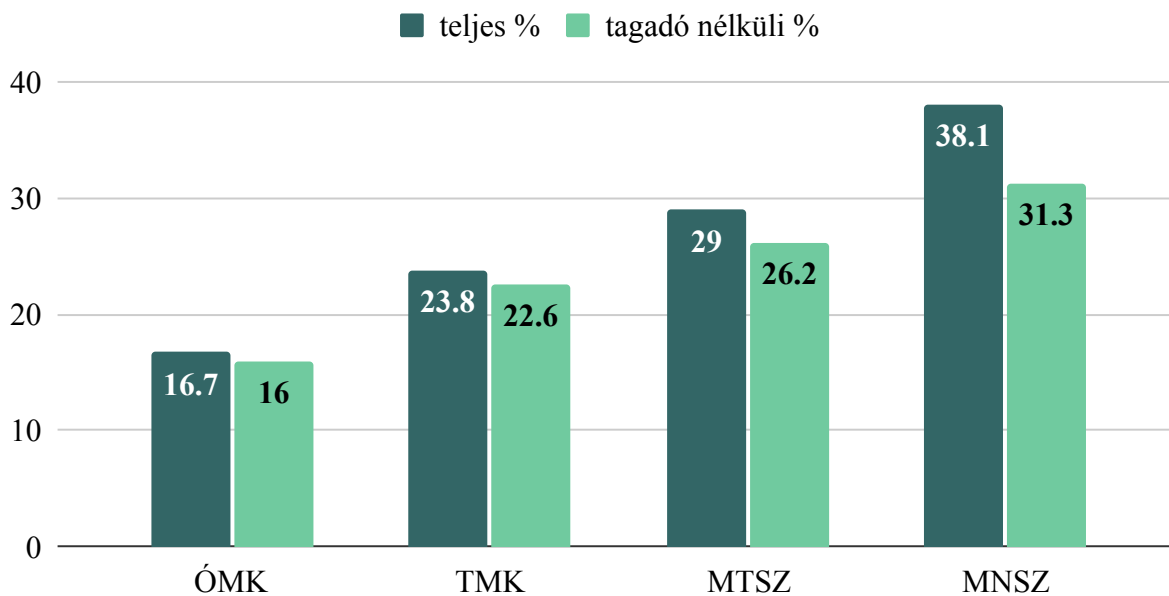


12. ábra. Megszakított és fordított szórendű tagadó mondatok arányának változása az MTSZ adatai alapján, évtizedes bontásban.

A fordulópont 1880 és 1890 közé esik, ettől kezdve a fordított szórendű tagadás növekedő tendenciát mutat. Ez kissé eltér Gugán (2017) eredményeitől: nála a fordulópont 40 évvel korábban látszik. A különbség oka valószínűleg az, hogy ő a *meg*, *el*, *fel* igekötőket vizsgálta közvetlenül a finit ige környezetében, míg ez a kutatás további prototipikus igekötőket is számba vett.

4.2.4. A fókuszos szerkezetek arányának növekedése

Az ige–igekötő sorrend arányának a növekedését nem egyedül a fordított szórendű tagadás ténye-rese magyarázza. É. Kiss Katalin azt a sejtést fogalmazza meg a disszertációm első változatára adott opponensi véleményében, hogy az ómagyar kor óta folyamatosan növekszik azoknak a szerkezeteknek az aránya, ahol a fókuszhasználat grammatikalizálódott. Ez a feltevés ellenőrizhető azáltal, hogy a tagadó mondatokat kivesszük az adatokból. Ennek az eredményét a 13. ábra mutatja be.



13. ábra. A posztverbális igekötőt tartalmazó mondatok arányának változása a korpuszokban aszerint, hogy minden mondattípust vizsgálunk-e (teljes %), vagy a tagadó mondatokat kihagyjuk a vizsgálatból (tagadó nélküli %).

A 13. ábrán látható, hogy a posztverbális igekötők mennyisége akkor is folyamatosan növekszik, ha a tagadó mondatokat nem vesszük figyelembe: az MNSZ 20–21. századi szövegein mért érték (31,3%) közel kétszerese az ÓMK-ban adatolhatónak (16%), tehát az adatok alátámasztják ezt a sejtést.

4.3. Összegzés

Ennek a fejezetnek az volt az elsődleges célja, hogy átfogó képet adjon az igekötők eloszlási mintáiról minden olyan szerkezet esetében, amelynél az igekötő el tud válni az igétől, igenévtől vagy egyéb igei származéktól, amellyel szótagi egységet alkot. Ezenfelül célja volt az is, hogy feltárja

azokat a tényezőket, amelyek az igekötő pozícióját befolyásolják, valamint hogy felhívja a figyelmet a ritkább mintázattípusokra.

A fejezet első részében egy szinkrón korpuszvizsgálatot mutattam be, amely az MNSZ2 korpusz 20–21. századi szövegein alapult. A kutatásom főbb eredményei a következők:

Kimutattam, hogy a prototipikus igekötők relatív gyakoriságukat tekintve közel maradnak a finit igéhez, míg a periférikusabb igekötők jobban eltávolodhatnak. Ehhez kapcsolódva megneveztem két olyan tényezőt, amely valószínűleg hatással van arra, hogy a hátravetett igekötő mennyire távolodik el a finit igétől. Az egyik a szöveg szerkesztettségi foka: szerkesztetlen – főleg szóban elhangzott – szövegben nagyobb annak a valószínűsége, hogy az igekötő távol kerül a finit igétől, mint gondozott, szerkesztett szöveg esetében. A másik az ige utáni mondatösszetevők, köztük a hátravetett igekötő fonológiai súlya. A méréseim alapján az a tendencia rajzolódik ki, amelyet a növekvő összetevők törvényének vagy Behaghel-törvénynek is neveznek, és É. Kiss (2007) szerint a magyarban a finit ige utáni mondatszakaszra érvényes: a rövidebb összetevő megelőzi a hosszabbat, ha ezt szintaktikai szabály nem gátolja. Ez azzal is egybevág, hogy az egy szótagú, prototipikus igekötők ritkábban távolodnak el a finit igétől, mint a több szótagú, periférikusabb igekötők.

Az infinitívusz kapcsán korpuszméréssel igazoltam, hogy az igekötője preverbálisan messzire eltávolodhat tőle, de csak akkor, ha segédigeszerű elem – elsősorban finit ige – férkőzik be a szerkezetbe, és az igekötő az ehhez tartozó igemódosítói pozíciót foglalja el. Ez alátámasztja azt a megfigyelést, hogy bár az igekötő az infinitívusszal alkot szótári egységet, szórendi és prozódiai szempontból szorosabban kapcsolódik a segédigeszerű elemhez (Kálmán C. et al. 1989). Megállapítottam, hogy ezzel rokon szintaktikai viselkedést mutat a létigével passzív szerkezetet alkotó határozói igenév is. Ennek az igekötője akkor kerül távoli preverbális helyzetbe, ha a létige előtti igemódosítói pozíciót foglalja el.

A határozói igenevet illetően még azt a megfigyelést tettem, hogy az igekötője hátravethető, de csak akkor, ha az igenév állapot- vagy módhatározói szerepet tölt be. A létigével alkotott passzív konstrukcióban az igekötő mindig preverbális.

A melléknévi igeneveket vizsgálva először is kiderült, hogy ha a -hAtÓ végződésű igenév állítmányi helyzetben van, és a tagmondatban nem hangzik el finit ige, akkor a -hAtÓ teljes mértékben átveszi ennek a helyét: az igekötője ugyanúgy elválhat, mint egy finit igéé, az igekötők disztribúciója is hasonlóan alakul, emellett a szerkezetbe férkőző szavakat tekintve is párhuzam látszik a

finít igei szerkezettel. Mindez az ellen szól, hogy a -hAtÓ-t melléknévképzőnek tekintjük – ahogy például Kiefer (2003a) teszi –, hiszen az ezzel képzett szó jellegzetes igei tulajdonságokat mutat.

Az általános szakirodalmi véleménnyel szemben kimutattam, hogy az -AndÓ végződésű melléknévi igenevek állhatnak állítmányi helyzetben, és ekkor az igekötőjük is hátravethető, bár a jelenség kétségtelenül ritka (az esetek 1,85%-ára jellemző, ez a vizsgált korpuszban 1624 példát jelent).

Az -Ó és az -(O)(t) végződésű igenevek kapcsán az vált láthatóvá, hogy az igekötőjük csak preverbálisan válik el, és úgy is meglehetősen ritkán: az -Ó végű igeneveknél 99,5%-ban, az -(O)(t) végűeknél 99,29%-ban közvetlenül az igenév előtt áll. Ugyanakkor az ezek esetében látott mintázat-típus nagyon elterjedt: deverbális főnevek, melléknevek és határozószók is állhatnak megszakított szórenddel, ekkor a *nem, sem, se, is* elemek valamelyike férkőzik a szerkezetbe.

Megvizsgáltam azokat a szerkezeteket, amelyekben egy kötómódú ige igekötője vagy egy kötómódú ige non-finít igei bővítményének az igekötője egy finít modális ige – tipikusan a *kell* – elé kerül. A fontosabb megfigyeléseim a következők: (1) A *hogy* kötőszót tartalmazó és nem tartalmazó változatok gyakorisága közel áll egymáshoz, függetlenül attól, hogy az igekötő finít kötómódú igehez, infinitívuszhoz vagy határozói igenévhez tartozik. (2) A finít modális ige és a kötőszó közé kerülhetnek rövid elemek, a finít modális ige és az igekötő igeje között álló kifejezések pedig hasonlóságot mutatnak az infinitívuszi beférkőzéses szerkezetekkel.

A szinkrón korpuszvizsgálat zárásaként megvizsgáltam azokat a szerkezeteket, amelyekben egy igekötős ige infinitívuszi vagy határozói igenévi alakban topikalizálva van, és emellett finít igeként is megjelenik. Megállapítottam, hogy tagadó, illetve fókuszos mondatokban a hátravetett igekötő jelenléte opcionális. A topikalizációs szerkezeteken belül külön vizsgáltam az olyan elliptikus szerkezetek tulajdonságait, amelyekben az igekötő újabb előfordulását egy segédigeszerű elem követi.

A fejezet második részében egy diakrón korpuszvizsgálat eredményeiről számoltam be. Ennek során a prototipikus igekötők mondatbeli helyét vizsgáltam a finít igehez képest ó-, közép- és újmagyar kori szöveganyagok bevonásával. A diakrón vizsgálat legfontosabb eredménye az, hogy kimutatta a hátravetett igekötőt tartalmazó, nem-semleges mondatok arányának növekedését az ómagyar kortól napjainkig. Ez a tendencia egyrészt azzal magyarázható, hogy a fordított szórendű tagadás fokozatosan teret nyert az archaikus, megszakított szórendű tagadással szemben. Másrészt látható az is, hogy az ómagyar kor óta folyamatosan növekszik azoknak a szerkezeteknek az aránya, ahol a fókuszhasználat grammatikalizálódott.

Ilyen jellegű disztribúciós vizsgálat korábban nem készült az igekötős szerkezetekről. Bár a kutatás középpontjában az igekötők álltak, bízom abban, hogy ez a munka hozzá tud járulni általában véve az igei szerkezetek adatvezérelt feltérképezéséhez is, és rá tudja irányítani a figyelmet olyan konstrukciókra, amelyek eddig elemzetlenül maradtak.

5. Az igekötők produktív kapcsolódási mintái

„[...] a bölcsélet nyelve valósággal elizélődött:
megízélték, tönkreizélték, agyonszöszölték
és elmaszlagolták a filozófiát.”

(Határ Győző: Darályvilág Buzdugániában)

Bár a fenti idézet nem bővelkedik tartalmas igékben, mégis tudjuk értelmezni: valakik hatására a filozófia megváltozott, és ez a változás nem volt előnyös. Ezt azért értjük, mert a mondatban szereplő ‘igekötő – képző – vonzatkeret’ hármassoknak (pl. *el-X-elődik valami, tönkre-X-el valamit*) önálló jelentése van, és ezt felismerjük akkor is, ha a szerkezetben szereplő ige az *izél*-hez hasonlóan semmitmondó. Ezeknek a hármassoknak a korpuszvezérelt vizsgálata közelebb visz minket ahhoz, hogy megismerjük az igekötők produktív kapcsolódási mintáit. A vizsgált jelenség azonban annyira szerteágazó, hogy ennél a közelebb vitelnél többre nem vállalkozhatok.

A produktív kapcsolódási minták azt adják meg, hogy az igekötő különféle jelentéseiben milyen igecsoportok tagjaihoz járulhat. Ehhez három tényezőt szükséges figyelembe venni (Ladányi 2012: 72-74): (1) Az igekötő szemantikája: Milyen jelentései vannak az igekötőnek? (2) Az igecsoport szemantikája: Mi a közös jelentéseleme azoknak az igéknek, amelyekhez az igekötő egy-egy jelentésében járulhat? (3) A létrejövő igekötős ige vonzatkerete. Amint láttuk, sok esetben már az igekötő és a vonzatkeret együttese elég ahhoz, hogy tudjuk, milyen – szemantikailag hasonló – igék lehetnek a szerkezetben.

Ebben a fejezetben az első tényezőt vizsgálom meg tüzetesen azáltal, hogy 56 igekötő produktív jelentéseit foglalom össze egy ontológiában. A második és a harmadik tényezőt, tehát az igecsoportok szemantikáját és az igekötős ige vonzatkeretét automatikus módszerekkel vizsgálom – az előbbi szóbeágyazással, az utóbbit vonzatkeret-kinyerő eljárással –, és az eredményeimet a PREVCONS nevű erőforrásban teszem közzé. A produktív kapcsolódási minták átfogó leírása nem része ennek a kutatásnak, de jövőbeli kutatási irányként egyértelműen következik belőle.

Elsőként a munkafázisokat és a felhasznált eszközöket ismertetem, majd az igealkotás három leggyakoribb módjáról lesz szó: a névszói és az igei bemenetű igeképzésről, valamint a hangzásséma-alapú igealkotásról. Ezt a két újonnan létrejött erőforrásnak, a PREVCONS-nak és az

igekötők ontológiájának bemutatása követi. Végül összegzem a legfontosabb eredményeket.⁴⁰

5.1. A kutatás módszere

A kutatás alapjául a PREVLEX szolgált (ld. 2.2.2. fejezet), de ebből csak a hapaxokat, tehát az egyszer előforduló igekötős igéket vettem figyelembe. Azért döntöttem így, mert a hapaxok jellemzően produktív szóalkotás eredményei, így a morfológiai szerkezetük és a jelentésük is kompozicionális – legalábbis nagyobb eséllyel az, mint a gyakori szavak esetében. A hapaxok közül is csak azokat őriztem meg, amelyeknél az igekötő terjeszkedő produktivitása 0-nál nagyobb volt (ld. 3.3.1. fejezet). A PREVLEX 122 ilyen igekötőt tartalmaz, ezek összesen 21 038 hapaxban jelennek meg.⁴¹

A kigyűjtött hapaxokat az emMorph (Novák et al. 2016, 2017) segítségével elemeztem. Ez egy nyílt forráskódú, automatikus morfológiai elemző rendszer, amely a szóalakot morfémákra és morfológiai kódokra bontja fel. Azért volt erre szükség, mert az MNSZ2-ben – amelyből a PREVLEX épült – az igekötős igék sokszor UNKNOWN elemzést kapnak, vagy ha jó az elemzésük, akkor sem látszik a morfológiai szerkezetük: sem az, hogy hol van az igekötő és ige közötti határ, sem az, hogy az ige adott esetben szóképzés eredménye-e. A (63) példa azt mutatja, hogy milyen elemzéssel jelenik meg az *elmamlaszkodik* szó az MNSZ2-ben, és milyen elemzés érhető el hozzá az emMorph-ban.⁴²

- (63) a. *elmamlaszkodik/elmamlaszkodik/IK.IGE.e3*
 b. *el[/Prev]=el+mamlasz[/N]=mamlasz+kodik[_Vbz:kOd/V]=kod+ik[Prs.NDef.3Sg]=ik*

Az emMorph nélkülözhetetlen ehhez a munkához, mivel manuálisan szinte lehetetlen ekkora adathalmazt feldolgozni. A fedése kiemelkedőnek számít: 92,6% körüli értékű (Novák et al. 2017: 73), és pontosság tekintetében is megbízható. Ugyanakkor generál néha olyan elemzést is, amely elvileg lehetséges, de nagyon valószínűtlen. Ilyen például a következő igék elemzése: *ereszkedik* → *eresz+kedik* ‘ereszként viselkedik’, *módosít* → *módos+ít* ‘módossá tesz’. Az ilyen jellegű hibák javítása csak manuális munkával lehetséges, amelyre az átnézendő adat nagy mennyisége miatt nem

⁴⁰ Ez a fejezet külön publikációként is megjelent, ld. Kalivoda (2021).

⁴¹ Előfordul, hogy lényegében azonos igelemmák tűnnek fel *-ik*-es és *-ik*-telen alakban is (pl. *ajándékoz* – *ajándékozik*, *akcióz* – *akciózik*). Emiatt „emberi szemmel” nézve kevesebb hapaxot kapnánk, mint automatikus módszerrel.

⁴² <https://emmorph.herokuapp.com/> (A morfológiai szerkezet az *analyze* opcióval nyerhető ki.)

vállalkoztam. Érdeemes tehát szem előtt tartani, hogy a kutatás eredményei nem hibátlan adathalmaz alapján álltak elő, ezeket kissé módosíthatja a szóban forgó javítási munka elvégzése.

Az emMorph a produktív képzőket ismeri fel aktív – azaz önállóan is használt – tövek mellett. A kutatás előrehaladtával világossá vált, hogy szükségem van olyan szerkezetek azonosítására is, amelyek az emMorph-ban nem jelennek meg. Ezek két okból maradhatnak ki: vagy azért, mert passzív – vagy más szóval fiktív – tő van bennük, vagy azért, mert túl szigorú az elemző szabályrendszere, például nem azonosítja a -(s)Ul képzőt, ha az több szótagú szóhoz kapcsolódik. Az ilyen esetekre saját szabályokat írtam, ezeknek a kimenete az emMorph-éval formailag pontosan egyező elemzés.

Az olyan hapaxok esetében, amelyeknek az igéi szóképzéssel jöttek létre, a képzés bemenetétől szolgáló szavak listáját – például névszói bemenetű igeképzésnél a névszók listáját – megvizsgáltam word2vec szóbeágyazással is (Siklósi és Novák 2016; Novák et al. 2016). Ennek az eljárásnak a lényege az, hogy a szavakat egy valós vektortér pontjaiként képezzük le, és ebben a sokdimenziós térben a hasonló szavak egymáshoz közel, az eltérőek egymástól távol rendeződnek el. A hasonlóság a választott modelltől függően szemantikai vagy szintaktikai jellegű lehet.

A szóbeágyazási modell használatára azért volt szükség, mert az igekötők produktív kapcsolódási mintáit csak az egyes jelentéseikre nézve tudjuk meghatározni (Ladányi 2012: 72). Ezzel az eszközzel lehetőség nyílik olyan objektív szemantikai osztályozásra, amelyet hagyományos módszerekkel hatalmas időbefektetés árán lehetne elérni, ráadásul a legjobb szándékkal sem tudna következetes maradni az osztályozás. Természetesen nem arról van szó, hogy kritika nélkül hagyatkozzunk az automatikusan előállított szemantikai osztályokra, mert ezek néha más jellemzők (pl. a szavak stílusregisztere) alapján állnak össze. A szóbeágyazás alkalmazása mégis olyan alapot ad, amelyet később esetleg csak korrigálni kell.

Végül lekértem az MNSZ2-ből azokat a mondatokat, amelyek a hapaxokat tartalmazzák. Ezeket Sass (2011) módszerével tagmondatokra bontottam, és a releváns tagmondatokból kinyertem a hapaxok feltételezett vonzatkeretét. Ezek az adatok – kiegészítve a fentebb tárgyalt morfológiai és szemantikai információval – alkotják a PREVCONS-ot, amelyet az 5.3. fejezet mutat be részletesen. A PREVCONS segítségével elemezhetővé váltak az igekötők produktív kapcsolódási mintái.

5.2. Az igealkotás produktív esetei

Mézesmázaskodik, markecolgat, succog. Korábban talán még nem hallottuk ezeket az igéket, mégis van elképzelésünk arról, hogy mit jelentenek. A *mézesmázaskodik* esetében a *-kodik* végződés miatt tudható, hogy az alany valamiként, valahogyan viselkedik. A *markecolgat* igében a *-gat* arra utal, hogy ismétlődve történik valami, még ha nem is tudjuk, mi az (a *markecol* jiddis eredetű szlengszó, ‘részeg embert kirabol’ jelentéssel). A *succog* esetében pedig szinte halljuk a hangot, amivel a cselekvés jár, bár a pontos cselekvést nehéz volna megnevezni. Ezeknek az igéknek a mintájára képesek vagyunk még számos hasonló szót alkotni. A *mézesmázás* helyére beilleszthetünk bármilyen melléknévi szerkezetet, a *markecol* helyére tetszőleges igét rakhatunk, a *succog* mintájára sok hangutánzó vagy hangulatfestő igét hozhatunk létre. Ez a magyar nyelvben működő, produktív igealkotási módokkal – jelen esetben a névszói és igei bemenetű igeképzéssel, valamint a hangzássémákkal – magyarázható.

Ez a fejezet az említett három igealkotási módot ismerteti. Ezeknek azért van jelentőségük a központi témánk szempontjából, mert a produktív szóalkotással létrejött igék csoportja nyitott, tehát amelyik igealkotó ilyen igecsoport tagjaihoz kapcsolható, annak minden bizonnyal a produktív kapcsolódási mintáit találjuk meg (Ladányi 2007: 201).

A kutatás kiinduló adathalmaza a 21 038 igealkotós igei hapax (ld. az 5.1. fejezetet). Ezeket úgy adtam meg az emMorph-nak, hogy az igealkotót elhagytam belőlük – így 11 602 igealkotó morfologiai elemzését kaptam meg. Az igealkotó elhagyásának csupán gyakorlati oka volt: az emMorph sok esetben ismeretlennek jelölte a szót, ha igealkotós volt, igealkotó nélküli formában viszont helyes elemzést rendelt hozzá. A sikeres elemzést követően természetesen összekapcsoltam a kimenetet az eredeti, igealkotós alakkal.

A továbbiakban külön alfejezetek foglalkoznak a három igealkotási móddal, a névszói és igei bemenetű igealkotás esetében a képzők, a hangzássémák esetében a különféle alsémák részletes bemutatásán keresztül. Végül arról lesz szó, hogy ezek milyen arányban jelennek meg az igealkotós hapaxok halmazában. A három igealkotási mód egyoldalas áttekintését tartalmazza a II. számú melléklet.

5.2.1. Névszó + igeképző

A névszói bemenetű igeképzés esetében a névszói bemenet nagyon sokféle lehet: leggyakrabban egyetlen főnév, melléknév vagy számnév (pl. *traktorozik, бүдösít, tizennyolcaz*), de nem ritka az sem, hogy komplex, több szóból álló szerkezet. Lehet például jelzős főnév (pl. *buta liba* → *butalibáskodik*), tulajdonnév (pl. *Sasvári Szilárd* → *sasváriszilárdoz*), sőt előfordul az is, hogy egy teljes üdvözlési forma válik igévé (pl. *Jó reggelt!* → *jóreggeltez*, *Kezit csókolom!* → *kezitcsókolomoz*).

A névszókhoz járuló igeképzők két csoportba sorolhatók annak alapján, hogy milyen tövekhez kapcsolódnak. Eszerint beszélhetünk valódi (aktív) tövekhez, valamint fiktív (passzív) tövekhez járuló igeképzőkről. Az előbbi esetben a fő önállóan is létezik a magyar nyelvben, míg az utóbbiban mindig képzővel együtt fordul elő. Sokan fiktív tövel működő igeképzésként elemzik a hangutánzó és hangulatfestő igéket is, én ettől el fogok térni (ennek részleteit ld. az 5.2.3. fejezetben).

A valódi névszói tövekhez járuló, produktív igeképzőkről nyújt áttekintést az 1. táblázat. Ezekről részletesen olvashatunk Kiefer és Ladányi (2000a), valamint Ladányi (2007) munkáiban, és a korpuszadatok mellett én is elsősorban ezt a két művet használom fel az egyes igeképzők bemutatásához.

A **valódi névszói tövekhez járuló, produktív igeképzőkről** nyújt áttekintést a 31. táblázat. Ezekről részletesen olvashatunk Kiefer és Ladányi (2000b), valamint Ladányi (2007) munkáiban, és a korpuszadatok mellett én is elsősorban ezt a két művet használom fel az egyes igeképzők bemutatásához.

képző	példák	db
-Vz(ik)	<i>zebracsíkoz, doktoroz, kávézik</i>	2558
-(s)Odik	<i>babásodik, jegesedik, vörösödik</i>	638
-Vl	<i>falfirkál, offtopikol, chatel</i>	525
-(V)(s)ít	<i>bénít, felhőkarcólósít, vidámít</i>	420
-(s)Ul	<i>halkul, szelídiül, zombul</i>	205
-(s)kOdik	<i>flegmáskodik, izmoskodik, trollkodik</i>	180

31. táblázat. Valódi névszói tövekhez járuló, produktív igeképzők néhány példával. A *db* az ige-típusok gyakoriságát mutatja.

A leggyakrabban megjelenő képző, a **-Vz(ik)** alapértelmezettnek tekinthető a denominális igeképzésben. Ladányi (2007: 76) a képző általános jelentését így fogalmazza meg: ‘N-nel kapcsolatban

tesz valamit, N-nel foglalkozik’ (ahol az N a névszói bemenetet jelöli). Ez egy annyira kiüresedett jelentés, hogy tulajdonképpen azt is mondhatjuk, a képzőnek csak funkciója van. Az emMorph különválasztja az *-ik*-re végződő variánst, amelyet intranszitivnak tekint (pl. *torrentezik*, tárgy nélkül), és az *-ik* nélkülit, amelyet tranzitívnek elemez (pl. *torrentez valamit*). A korpuszadatok vizsgálatakor úgy láttam, hogy ez a két eset nem válik el ilyen egyértelműen egymástól – intranszitivnak elemzett ige mellett is megjelenik tárgy –, ezért nem különítem el a két változatot.

A **-VI** a **-Vz(ik)** rivális képzője, mivel a jelentésében ugyanúgy kiüresedett, és elviekben tetszőleges névszóhoz járulhat. A két képző hatóköri konfliktusát jól mutatja az, hogy a vizsgált igtípusok közül 86 esetben egyezik a névszói bemenet, pl. *botoxol – botoxoz, szörföl – szörföz, filterel – filterez*. Ezeknek az ige pároknak a tagjai között nem találunk jelentésbeli különbséget, de a nyelvi rendszerben rögzültebb, gyakoribb igéknél már jellemzően igen (ilyen például a *házal* és a *házaz* közti jelentéskülönbség).

A 31. táblázat gyakorisági adatai jól mutatják, hogy a **-VI** képző sokkal ritkábban jelenik meg, mint a **-Vz(ik)**, ami arra utal, hogy az előbbinek a használati köre szűkebb. Ladányi (2007: 107) ezt a következőkkel magyarázza: a **-VI** beillesztő képző, amelyet jellemzően akkor választunk, ha a névszó (1) mássalhangzóra végződik (kivéve /l/ vagy /r/) és egy szótagból áll, vagy (2) magas vagy vegyes hangrendű fiktív tő, vagy (3) angol cselekvésigéből létrejött *-ing* végződésű főnév. Ladányi megfigyeléseit a korpuszadataim is alátámasztják.

Ugyancsak rivális képzők a **-(s)Odik** és a **-(s)Ul**, amelyek a következő általános jelentésen osztoznak: ‘N-né változik’. A használati körükben az az eltérés figyelhető meg, hogy egy szótagú alapszóhoz jellemzően a **-(s)Ul** járul, míg több szótagúhoz inkább a **-(s)Odik** (Ladányi 2007: 245). Természetesen ezeknél is előfordulhatnak párhuzamos képzéssel létrejött alakok, pl. *sötétedik – sötétül, népesedik – népesül*. Ennek a két képzőnek a műveltető párja a **-(V)(s)ít**, amelynek az általános jelentése: ‘N-né változtat’ (pl. *sötétít, népesít*).

Hangsúlyozni kell, hogy a használati körre vonatkozó megfigyelések tendencia-jellegűek. Ezért igaz ugyan, hogy a **-(s)Ul** általában egy szótagú tövek mellett jelenik meg, de jócskán találunk példát ennek az ellenkezőjére is (pl. *magányosul, kristályosul, anyagiassul, mémesül, kereszténydemokratul*). Az ilyen eseteket Ladányi (2007: 253-254) úgy írja le, hogy a vetélkedő képzők közül általános esetben is a speciálisat alkalmazzuk, a képző ezáltal túllépi a hatókörét. Szintén érdekes jelenség, hogy a **-(s)Ul** képző nem kizárólag melléknévhez, hanem néha főnévhez is kapcsolódhat,

pl. *szenttehenül, pépül*.

Több szakirodalomban – Kiefer és Ladányi (2000b: 154-155), Ladányi (2007: 246) – olvasható az az állítás, hogy a -(s)Ul képző nem produktív, mert a magyarban nem jöhetnek létre új, egy szótagú melléknevek, a nyelvbe bekerült idegen szavakkal pedig nem működik ez a képző. A korpuszadatok ezt cáfolják (pl. *punk* → *(be)punkul*), és a hatókör túllépése is elég gyakori ennek a képzőnek az esetében ahhoz, hogy – legalábbis egyelőre – produktívnak lehessen minősíteni.

Ennek a csoportnak a legkevesebb ige típusban megjelenő tagja a **-(s)kOdik** képző, amelynek általános jelentése ‘N-ként viselkedik’. A viszonylagos ritkaságának az lehet az oka, hogy az eddig látott képzőkkel ellentétben itt a bemeneten szemantikai korlát is van: a képző jellemzően élőlényre utaló névszóhoz kapcsolódik. Élettelen jelentő névszó is szolgálhat bemenetként (pl. *asztalkodik, póznáskodik*), de ez nagyon ritka jelenség.

Az **idegen eredetű, fiktív tövekhez járuló ige képzőket** a 32. táblázat foglalja össze. Mivel ezekben a névszói tő önmagában nem létezik a magyar nyelvben, joggal merülhet fel a kérdés, hogy van-e értelme a tő és a képző különválasztásának (az emMorph például nem jelez szóképzést az ilyen igeik esetében). Az adatok között találtam néhány példát arra, hogy magyar szavakban is működnek ezek a képzők (pl. *bolondizál, kántorizál*), ezért döntöttem a különválasztásuk mellett.

képző	példák	db
-ál	<i>annotál, fókuszál, transzponál</i>	303
-izál	<i>motorizál, szolmizál, tematizál</i>	109
-íroz	<i>lavíroz, morfondíroz, políroz</i>	39
-ifikál	<i>kodifikál, purifikál, identifikál</i>	7

32. táblázat. Idegen eredetű névszói tövekhez járuló, produktív ige képzők néhány példával. A *db* az ige típusok gyakoriságát mutatja.

A fejezetben tárgyalt tíz képző mind produktív, bár a produktivitásuk mértékében jelentős eltérések mutatkoznak. Zárásképp néhány olyan képzőt említek, amelyek többször felbukkantak ugyan a vizsgált hapaxokban, mégsem tűnnek produktívnak. Ide sorolható az -Ász(ik) képző (pl. *madarászik, egerészik, bolhászik*), valamint a -(V)gOl (pl. *szamaragol, kacsagol, elefántogol*).

5.2.2. Ige + igeképző

Az igei tőhöz kapcsolható képzők illeszkedhetnek eleve képzett igehez (pl. *[[magyar]ít]+gat*), következésképp egymáshoz is (pl. *[[[magyar]ít]gat]+ódik*). Ebbe az alosztályba négy produktív képző tartozik, amelyekről a 33. táblázat ad áttekintést.

A leggyakoribb a műveltetést kifejező **-(t)At** képző, ennek – és a műveltetés egyéb kifejezési módjainak – részletes elemzését adja Komlósy (2000). Ezt az **-Ódik** követi, amely mediális igeik képzésére alkalmas. A mediális elnevezés arra utal, hogy az ilyen igeik „középuton” vannak a cselekvő és szenvedő igenem között (az igeiben jelölt esemény megtörténik az alannal, annak akarata vagy aktív közreműködése nélkül).

képző	példák	db
-(t)At	<i>kocsikáztat, pötyögtet, oszlat</i>	1412
-Ódik	<i>duplázódik, hekkelődik, szövődik</i>	851
-(O)gAt	<i>hintáztat, fuvintgat, búsulgat</i>	510
-(t)Atik	<i>doboltatik, aranyoztatik, rostáltatik</i>	178

33. táblázat. Igékhez járuló, produktív igeképzők néhány példával. A *db* az ige típusok gyakoriságát mutatja.

Az **-(O)gAt** képzőnek két funkciója van: a gyakorítás és a kicsinyítés, vagyis az intenzitás csökkentése. A gyakorító funkció jellemzően az egyszerűséget kifejező igeiknél jelenik meg (pl. *ad* → *adogat*), a kicsinyítő pedig azoknál, amelyek hosszabban tartó folyamatot jelölnek (pl. *olvas* → *olvasgat*) (Kiefer és Ladányi 2000b: 199). A két funkció nem független egymástól, sokszor együttesen vannak jelen.

A passzivizáló funkciójú **-(t)Atik** képző jóval elmarad a többitől a gyakoriságát tekintve. Ennek az lehet az oka, hogy a mai nyelvhasználat inkább az összetett passzívet (létige + -vA képzős igealak), vagy az **-Ódik** képzős, mediális formát részesíti előnyben. Tehát például a *bearanyoztatik* helyett *be van/lesz aranyozva, bearanyozódik* a jellemzőbb.

Az említett négy eseten túl még sok olyan képzőt találunk az adatban, amely igei tőhöz kapcsolódik, de nem – vagy csak nagyon kevésbé – produktív. Nagy csoportot alkotnak azok, amelyek gyakorító-kicsinyítő funkcióval bírnak, tehát az **-(O)gAt** alternatívái. Ide sorolható a **-kOd** (pl. *lop-kod, tömköd*), a **-dOs** (pl. *vágdos, öldös*), a **-dOgÁl** (pl. *hulldogál, nézdegél*), a **-dAl** (pl. *rugdal, szökdel*), a **-dÁcsOl** (pl. *bukdácsol, nyögdecsel*), az **-Ong** (pl. *jajong, forrong*) és az **-igál** (pl. *hu-*

zigál, dobigál). Ezek a képzők paradigmákba is szerveződnek, erről bővebben ld. Ladányi (2007: 101-105) esettanulmányát.

5.2.3. Hangzásséma

Az eddig tárgyalt igealkotási módokat igeképzésnek neveztem, most azonban elhagyom ezt a terminust, és hangzásséma-alapú igealkotásról fogok beszélni. Azért látom indokoltnak a megkülönböztetést, mert az utóbbi esetben nemcsak a végződés kötött, hanem a szótagszám is, és még számos szótagszerkezeti hasonlóság figyelhető meg az így alkotott igéknél. Nem önmagában a szuffixum produktív, hanem a teljes hangzásséma.

A szakirodalom jelentős része igeképzésként tárgyalja ezeket az igealkotási módokat is, azzal a kitételrel, hogy a szótó mindig fiktív. Ilyen elemzést választ pl. Rebrus (2000: 778-779), Kiefer és Ladányi (2000b: 201) és Ladányi (2007: 244-245). A képzős elemzés helyett eddig kevesen választották a hangzásséma-modellt. A kevesek között az egyik legalaposabb munka Benő és Szilágyi N. (2015) cikke, amely az én munkámhoz is nagyban hozzájárult.

A hangzásséma-alapú elemzés meglehetősen eltér a hagyományosabb megközelítésektől, ezért nem meglepő, hogy nem nyerhető ki az emMorph-ból. A hangzássémába illeszkedő igéket csak akkor tudja igeként azonosítani és elemezni, ha szerepelnek a lexikonjában, egyébként ismeretlen szavakként kezeli őket. A hangzássémák azonosítására saját algoritmust írtam, a fejezet további részében ismertetett sémátípusok ennek az algoritmusnak, és nem az emMorph-nak a kimenetei.

A leggyakoribb és legtöbb altípust felmutató hangzásséma az **-Og végű séma** (ld. 34. táblázat). Tagjainak a szerkezeti hasonlóságukon túl közös szemantikai tulajdonságuk az, hogy hosszabb ideig tartó vagy ismétlődő eseményt fejeznek ki.

Az ide tartozó sémák állhatnak kettő vagy három szótagból. A két szótagúak között a leggyakoribbak azok, amelyekben a középső mássalhangzó rövid, a két magánhangzó pedig azonos (pl. *potyog, tötyög, mekeg*). A három szótagúak további alsémákra bonthatók az utolsó szótagkezdő mássalhangzójuk alapján. Ez ugyanis csak /r/, /l/ vagy /j/ lehet, ami azért érdekes, mert ezek a mássalhangzók egy csoportot alkotnak a fonetikai hangzósságuk (Kassai 1994: 631) és a fonológiai jellemzőik alapján is (Siptár 1994: 185).

Bizonyos -Og végű igék -OrÁsz(ik) végződéssel is működnek, ekkor a cselekvés nem pusztán hosszan tartó vagy ismétlődő, hanem alacsony intenzitású is (pl. *vihog* → *vihorász, röhög* → *röhö-*

általános séma és példák	db	megszorított séma és példák	db
(C)(C)VC:Og <i>battyog, hűmmög, billeg</i>	62	(C)CO ₁ C:O ₁ g <i>szöttyög, kerreg, csetteg</i>	6
(C)(C)VCOg <i>krárog, büfög, lifeg</i>	59	(C)CO ₁ CO ₁ g <i>kopog, fröcsög, cseveg</i>	70
(C)(C)VCCOg <i>harsog, csámcsog, prüntyög</i>	13	(C)CO ₁ CCO ₁ g <i>mormog, dörmög, csemcseg</i>	11
(C)(C)V(C)CVrOg <i>kacsmarog, csavarog, szivárog</i>	6	(C)(C)V(C)CO ₁ rO ₁ g <i>sutyorog, nyöszörög, sistereg</i>	59
(C)(C)V(C)CVIOg <i>andalog, sandalog, oldalog</i>	5	(C)(C)V(C)CO ₁ lO ₁ g <i>őgyeleg, édeleg, hízeleg</i>	13
(C)(C)V(C)CV/j/Og <i>kóvályog, nyavalyog, vihajog</i>	4	(C)(C)V(C)CO ₁ /j/O ₁ g <i>gomolyog, hömpölyög, tévelyeg</i>	10

34. táblázat. Az -Og végű hangzássémák, az általános sémák gyakorisága szerinti csökkenő sorrendben. A jobb oldali, megszorított sémák a bal oldali párjuk szigorúbb változatai. A fejléc alatti első sorban szereplő C: (hosszú mássalhangzó) is egyfajta megszorítás, a harmadik sorban látható CC-é (két mássalhangzó), mégis külön sémának tekintem a gyakorisága miatt.

rész, kacag → *kacarász, tötyög* → *tötyörész*). Ez azonban sokkal korlátozottabb, mint az -Og végű séma (például kérdésesnek tűnik a *fröcsög* → *?fröcsörész, krárog* → *?krákorász*).

Az -An, -Ant és -int végű sémákat érdemes együtt tárgyalni a szemantikai rokonságuk miatt (ezt a 35. táblázat példái illusztrálják). Mindháromra igaz, hogy mozzanatosak, azaz pillanatnyi eseményt fejeznek ki. Az -An ilyen tekintetben az -Og funkcionális párja (vö. *csattan* – *csattog, zöttyen* – *zötyög*). Az -An végű séma általános jelentése az, hogy ‘X-szel történik valami pillanatnyi’. Az -Ant végű séma ennek a kauzatív megfelelője, tehát alanyként megjelenik a történés okozója: ‘Y tesz X-szel valami pillanatnyit’ (pl. *meqlibben a függöny* – *a szél meqlibbenti a függönyt*). Az -int az -Ant-hoz hasonlóan kauzatív, és azzal a jelentéstöbblettel bír, hogy az esemény nagyon alacsony intenzitású.

A 35. táblázatban látható, hogy ezeknek a sémáknak minden tagja két szótagú. Az alsémák között csak a középső mássalhangzókat tekintve van különbség. Szembetűnő, hogy csak a /h/ jelenik meg egymagában középső mássalhangzóként. Ezt Benő és Szilágyi N. (2015: 48) úgy magyarázza, hogy a magyarban tulajdonképpen nincsen hosszú /h/, még teljes hasonulás esetén is röviden ejtjük – bár az írása ingadozó –, például *az+hoz* → /ahoz/, *ez+hez* → /ehez/.

-An		-Ant		-int	
séma és példák	db	séma és példák	db	séma és példák	db
(C)CVC:An <i>csattan, zöttyen</i>	127	(C)CVC:Ant <i>csattant, zöttyent</i>	138	(C)CVC:int <i>csattint, zöttyint</i>	34
(C)CVCCAn <i>sercen</i>	12	(C)CVCCAnt <i>sercent</i>	9	(C)CVCCint <i>sercint</i>	4
(C)CVhAn <i>suhan, zuhan</i>	2			(C)CVhint <i>suhint, zuhint</i>	3

35. táblázat. Az -An, -Ant és -int végű hangzássémák a gyakoriságuk szerinti csökkenő sorrendben.

Feltűnő a hosszú mássalhangzók gyakorisága is, szemben az -Og végű sémával, ahol éppen a rövid mássalhangzók gyakoribbak. Ennek az lehet az oka, hogy az -An végű sémák jellemzően nagyobb intenzitást jelölnek, és ezt az erősebb hanghatású hosszú mássalhangzók festik le jobban, míg az -Og végű sémákban a gyengébb hanghatású rövid mássalhangzó fejezi ki az esemény kisebb intenzitását (Benő és Szilágyi N. 2015: 48).

A harmadik nagy csoportot, az **-Ad és -Aszt végű sémákat** szintén a szemantikai rokonságuk miatt tárgyalom együtt (ld. 36. táblázat). Ez a rokonság abban nyilvánul meg, hogy az -Ad végűek általános jelentése ‘X-szel történik valami’, a kauzatív -Aszt végűeké pedig ‘Y tesz X-szel valamit’ (tehát ezek is funkcionális párok).

-Ad		-Aszt	
séma és példák	db	séma és példák	db
(C)(C)VC:Ad <i>fonnyad, pilled</i>	43	(C)(C)VC:Aszt <i>fonnyaszt, pilleszt</i>	31
(C)(C)VCAd <i>kókad, reked</i>	37	(C)(C)VCAszt <i>kókaszt, rekeszt</i>	30
(C)(C)VCCAd <i>sorvad, erjed</i>	25	(C)(C)VCCAszt <i>sorvaszt, erjeszt</i>	22

36. táblázat. Az -Ad és -Aszt végű hangzássémák a gyakoriságuk szerinti csökkenő sorrendben.

Az -Ad végű séma igéihez az -Ozik képző járulhat, amely a cselekvés tartós vagy ismétlődő jellegét fejezi ki, tehát az 5.2.2. fejezetben tárgyalt -(O)gAt képzőhöz hasonló szerepet tölt be (pl. *fullad+ozik, pukkad+ozik, püffed+ezik, reped+ezik*).

Míg az -Og, -An, -Ant és -int végű sémákba tartozó igéket a szakirodalom általában **hangutánzó vagy hangulatfestő** – összefoglaló néven onomatopoetikus, expresszív – szavaknak tekinti, az -Ad és -Aszt végű sémába illeszkedők ezen a kategórián kívül esnek. Én igyekeztem kerülni mind a ‘hangutánzó’, mind a ‘hangulatfestő’ kifejezést, mert túl szubjektívnek találom (különösen az utóbbit). A hangzásséma ennél konkrétabb, mivel pontosan formalizálható jellemzőkre – szuffixumra, szótagszámra, szótagszerkezetre⁴³ – illeszkedik.

5.2.4. A vizsgált igealkotási módok arányai az igekötős szerkezetekben

A korábbi alfejezetek csak az igékkal foglalkoztak, az igekötőket átmenetileg figyelmen kívül hagytam. Most visszatérek a dolgozat központi témájához, az igekötős szerkezetekhez.

Elsőként érdemesnek tartottam megvizsgálni azt, hogy a három produktív igealkotási mód milyen arányban van jelen az igekötős hapaxoknál, valamint az igekötő elhagyásával kapott ige típusoknál. Ennek az eredményét foglalja össze a 37. táblázat.

	igekötős hapaxok		igetípusok	
	db	%	db	%
névszó → ige	7401	35,2	4985	43,0
ige → ige	4492	21,4	2951	25,4
hangzásséma	1987	9,4	799	6,9
nincs / egyéb	7158	34,0	2867	24,7
	21 038	100	11 602	100

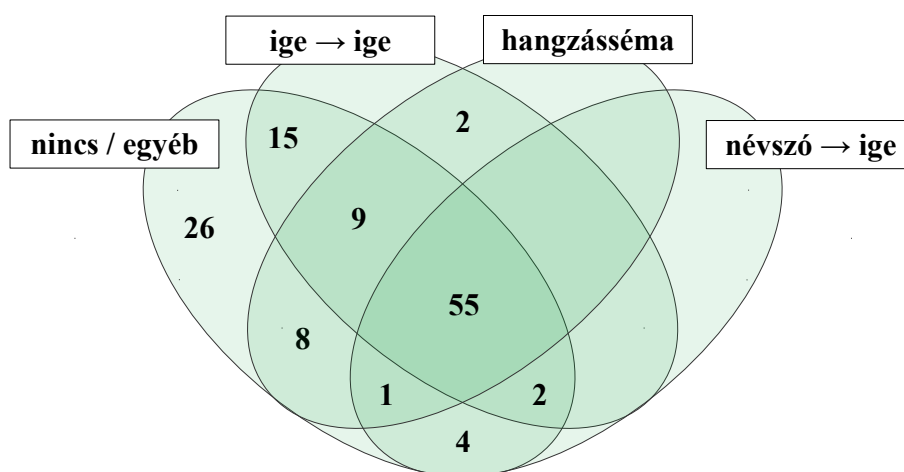
37. táblázat. Igealkotási módok gyakorisága az igekötős igei hapaxok, valamint az igekötő elhagyásával kapott ige típusok esetében. A *nincs / egyéb* azokat az eseteket jelzi, ahol nem képzett vagy hangzássémás a szó, vagy pedig nem produktív képzőt tartalmaz.

A legjelentősebbnek a névszói bemenetű ige képzés bizonyult, ez az igekötős hapaxok több mint egyharmadánál (35,2%) figyelhető meg. Ha igekötő nélkül nézzük csak az ige típusokat, akkor ez az adatnak majdnem a felét teszi ki. A hangzássémás ige aránya sem elhanyagolható: az igekötős hapaxok közel egytizedét (9,4%) ezek adják. Ezt fontosnak tartom hangsúlyozni, ugyanis úgy tűnik,

⁴³ A szótagszerkezettel kapcsolatban egy megjegyzés: a CV-sémákat nem úgy kell értelmezni, hogy a C helyén tetszőleges mássalhangzó állhat. A szótagnak jólformáltnak kell lennie, tehát a szótagmag felé a szegmentumok hangzóssága nő, onnan kifelé pedig csökken. Ennek nem megfelelő szavakat nem is igazán tudnánk kiejteni (pl. **rkákog*). Ez a törvényszerűség, amelynek a természetes nyelvek szótagjai legtöbbször megfelelnek, a szonoritási hierarchia – bővebben ld. Törkenczy (1994: 276).

hogy a nyelvészeti szakirodalom kevés jelentőséget tulajdonít az igealkotás ezen módjának. A létjogosultságát mostantól korpuszadatok igazolják.

Ezután azt vizsgáltam meg, hogy az egyes igekötők milyen igealkotási móddal keletkezett igékhez kapcsolhatók. Az eredményt a 14. ábra halmazai szemléltetik. Szinte minden igekötő hapaxai között találunk „egyszerű” igéket, amelyek se nem képzettek, se nem hangzássémára illeszkedők. Általánosnak mondható az igei bemenetű igeképzés megléte is. Ha egy igekötő esetében ez nem jelenik meg, annak csak az adathiány az oka. Erre példa a *hasba* és a *hátba* esete: a hapaxok között szerepel a *hasbabök* és a *hátbabök* is, de képzett ige csak a *hátba* esetében látszik (*hátbabököget*). Ez persze csak esetlegesség, és egy nagyobb korpuszban jó eséllyel a *hasbabököget* is felbukkant volna.



14. ábra. A 122 vizsgált igekötő megoszlása aszerint, hogy milyen igealkotási móddal keletkeztek az igék, amelyekhez kapcsolódnak. A halmazokban és azok metszeteiben megadott számok az igekötők darabszámát mutatják. A *nincs / egyéb* azokat az eseteket jelzi, ahol nem képzett vagy hangzássémás az ige, vagy pedig nem produktív képzőt tartalmaz.

Érdekes az is, hogy bár a névszói bemenetű igeképzés van jelen a legnagyobb arányban az igekötős hapaxok között, ez mindössze 62 igekötőnél mutatható ki (az alakváltozatokat összevonva ez a szám csak 56). Ez a kis halmaz nagyjából egybeesik azzal, amit a szakirodalmak többsége az igekötők állományaként jelöl ki, vannak viszont eltérések is. Valószínűleg keveseknek jutna eszébe, hogy igekötőnek minősítse a *szénné* és *szarrá* szavakat, pedig ezek a kapcsolódási mintáikat tekintve sokkal igekötősebbek, mint például a *közzé* vagy a *véghez*, amelyeket hagyományosan igekötőként sorolnak fel több munkában, ld. Jakab (1976: 9-10) adatait. Az utóbbiak inkább frazémajellegűek, csak egy-két igéhez kapcsolódnak.

A két mérés alapján belátható: az igealkotás módjára vonatkozó információnak van jelentősége abban, hogy az igekötők produktív kapcsolódási mintáit le tudjuk írni. Ezért döntöttem úgy, hogy ez az adat képezze szerves részét egy igekötős szerkezeteket tartalmazó erőforrásnak, a PREVCONS-nak, amelyet a következő fejezet mutat be.

5.3. PREVCONS: Igekötős igei hapaxok adatbázisa

A PREVCONS (*Preverb Constructions*) egy nyílt hozzáférésű adatbázis,⁴⁴ amely 21 038 igekötős hapaxot tartalmaz azok lehetséges vonzatkeretével együtt. Ezen felül számos metaadat érhető el benne a hapaxok morfológiai szerkezetéről, szemantikájáról és konkrét szövegbeli előfordulásáról.

Az erőforrásnak sok közös vonása van a Mazsola korpuszlekérdező rendszerrel (Sass 2011): a létrehozásának módjában és a vonzatkeretek leírásához használt formalizmusban is a Mazsolát követi. A lényeges különbség az, hogy a Mazsola igei szerkezetek, igék és bővítményeik tanulmányozására alkalmas, a PREVCONS pedig az igekötős szerkezetek vizsgálatára használható, akár úgy is, hogy a szerkezetben lévő igét nem vesszük figyelembe. Tehát kigyűjthetünk belőle ‘igekötő – képző – vonzatkeret’ mintákat gyakorisági adataikkal együtt (pl. *össze-N-Vz(ik) -t -vAl*, azaz ‘összevalamiz valamit valamivel’ vagy ugyanez élőlényre utaló vonzatokkal), és egy következő lépésben megnézhetjük, hogy a mintákban milyen igék állnak a ‘valamiz’ helyén. Ez a módszer lehetővé teszi az igekötős szerkezetek adatvezérelt és szisztematikus vizsgálatát.

A PREVCONS github repozitóriumában két forrásfájl található, egy TSV fájl és egy SQLite adatbázis. Ezek pusztán a formátumukat tekintve térnek el egymástól, tartalmilag megegyeznek. A két formátumot az indokolja, hogy a számítógépes feldolgozáshoz praktikusabb lehet a TSV, a kézi adatelemzéshez pedig az SQLite-alapú megoldás. Az adatbázis böngészhető a nyílt forráskódú Aszaló webes keresőfelület⁴⁵ segítségével is, amelyet Indig Balázs készített. A PREVCONS 16 oszlopból áll, amelyek a következők:

1. id: Az igekötős ige egyedi azonosítója, ennek csak technikai szerepe van.
2. prev: Igekötő.

⁴⁴ <https://github.com/kagnes/prevcons>

⁴⁵ A PREVCONS webes keresőfelülete: <https://aszalo.herokuapp.com/>
Az Aszaló github repozitórium: <https://github.com/dlazesz/aszalo>

3. verb: Igelemma.
4. actform: Az igekötős ige szóalakja.
5. argframe: Az igekötős ige lehetséges vonzatkerete. Ennek a kinyerése automatikus módszerrel történt, egy olyan algoritmussal, amely a Mazsola mintájára készült. A vonzatkeret megállapításához nem a teljes mondatot vettem figyelembe, hanem csak azt a tagmondatot, amelyben az adott igekötős ige szerepelt. A vonzatkeret összeállításával kapcsolatban fontos még tudni a következőket:

- Csak az olyan esetragokat nyertem ki a tagmondatból, amelyek maximális NP végén szerepelnek, tehát olyan névszói szerkezet végén, amely nem része egyetlen magasabb szintű névszói szerkezetnek sem. Ugyanezen az elven működik a Mazsola is, a különbség csak annyi, hogy a PREVCONS-nál a névszói szerkezetek azonosítása az emtsv emChunk moduljával történt (Indig et al. 2019; Váradi et al. 2018, 2017), a Mazsolában pedig egy szabályalapú algoritmussal (Sass 2011: 37-40).
- Amikor a kinyert esetragok bekerülnek a lehetséges vonzatkeretbe, a sorrendjük már nem a tagmondatbeli megjelenésüket követi, hanem abc-rendben szerepelnek.
- Az alanyesetet nem jelölöm.
- Egy speciális vonzat a *magát*, amelyet nem helyettesítek tárgyesetű raggal, hanem teljes szóként őrzök meg. Az 5.4.2. fejezetben látni fogjuk, hogy a *magát* vonzatos igekötős szerkezetek alapjelentése sokszor teljesen eltér azokétól, amelyek egyéb tárgyesetű vonzatot tartalmaznak.

Végül fontos látni, hogy az automatikus adatkinyerés miatt a vonzatkeretek több hibát is tartalmazhatnak. Például nem oldható meg a szabadhatározók elkülönítése a kötelező vonzatoktól, és a *magát* vonzat sem mindig azonosítható helyesen: a *magam*, *magad*, *magunk*, *magatok*, *maguk* alakok egyaránt lehetnek alany- és tárgyesetűek is. Továbbá az sem jósolható meg előre, hogy bizonyos esetragok mikor lennének kiválthatók egy irányhármast jelölő vonzattal és mikor nem. Például az *el* igekötő mellett megjelenő *-rA*, *-bA* vonzatok helyett *HOVÁ* is állhatna, de a *rA* melletti *-rA* vonzat tovább már nem általánosítható.

6. prev_prod: Az igekötő terjeszkedő produktivitása (erről ld. a 3.3.1. fejezetet).

7. *prev_hapaxes*: Az igekötőt tartalmazó hapaxok darabszáma.
8. *inputtype*: Az ige esetében azonosított szóalkotási mód. A *N* névszóból képzett, a *V* igéből képzett, a *PH* pedig hangzássémás igét jelöl. Amennyiben az ige nem produktív szóalkotással jött létre, ebben az oszlopban kötőjel (-) áll.
9. *suffixtype*: A képző vagy hangzásséma típusa. Itt is kötőjel (-) áll, ha az ige sem produktív képzővel, sem hangzássémával nem elemezhető. Egyéb esetekben az 5.2. fejezetben használt jelöléseket alkalmazom.
10. *phonotype*: Amennyiben az ige hangzássémába illeszkedik, itt látható a teljes CV-séma.
11. *style*: A forrásmondat MNSZ2-ből kinyert stílusregisztere. Mivel hapaxokról van szó, minden mondathoz pontosan egy stílusregiszter tartozik. Ez a következő értékeket veheti fel: *hivatalos*, *sajtó*, *spok* (beszélt nyelvi, általában rádióműsor leírata), *személyes* (fórum- és blogbejegyzések), *szépirodalom*, *tudományos*.
12. *region*: A régió, amelyből a forrásmondat származik (szintén az MNSZ2-ben elérhető adat). Lehetséges értékei: *erdélyi*, *kárpátaljai*, *magyarországi*, *szlovákiai*, *vajdasági*.
13. *w2v_input*: Az automatikus klaszterezés bemeneti szava. Ide szóképzéssel létrejött ige esetében az alapszó kerül (pl. *trollkodik* → *troll*), hangzássémás ige esetében pedig az igelemmával egyezik meg a mező tartalma.
14. *w2v_cluster*: Annak a klaszternek az azonosító száma, amelybe a *w2v_input* mező szava került a szóbeágyazás során (erről bővebben ld. az 5.1. fejezetet). Mivel külön körökben történt a különféle igealkotási módokkal létrejött igék klaszterezése, a számozás minden *inputtype* esetén 0-tól kezdődik. Az 1000 feletti értékek azt jelentik, hogy az adott ige nem produktív szóalkotással jött létre, vagy klaszteren kívül maradt.
15. *clause*: Az a tagmondat, amely az igekötős igét tartalmazza. A tagmondathatárok megállapítása Sass (2011: 34-37) módszerével történt.
16. *sent*: A teljes mondat, amely az igekötős igét tartalmazza. Ezen belül kétféle jelölés látható: az *<ncl>* (*new clause*, azaz tagmondathatár) és a *<q>* (*quotation mark*, azaz idézőjel). A

két szóból álló, nagybetűvel kezdődő szósorokra rákerestem, és amennyiben tulajdonnevek voltak, anonimizáltam őket *XY*-ra. Ez alól akkor tettem kivételt, ha az igekötős igének csak a tulajdonnév ismeretében volt igazán értelme, ilyen volt például a következő eset: *Ezek túladyendrézik magát Ady Endrét, [...]*

A PREVCONS adatbázissal számos kérdés vizsgálható, például az, hogy melyik igekötő milyen stílus-regiszterben van jelen a leginkább, vagy egyes igekötők mellett mennyire sokszínű vagy épp kötött a vonzatkeret. A legnagyobb jelentőségét mégis abban látom, amiről a fejezet elején is szó volt: segítségével az ‘igekötő – képző – vonzatkeret’ hármasok a konkrét igék kihagyásával vizsgálhatók. Erre látunk példát a 38. táblázatban.

igekötő	képző	vonzatkeret	db
<i>el</i>	-(s)Odik		196
<i>el</i>	-Vz(ik)	-t	107
<i>el</i>	-(V)(s)ít	-t	72
<i>el</i>	-Vz(ik)		69
<i>el</i>	-(s)Ul		41

38. táblázat. Az *el* igekötő mellett megjelenő öt leggyakoribb névszó → ige képző és vonzatkeret.

A hármasok lekérdezése után, egy következő lépésben meg lehet nézni az ezekhez tartozó igéket. Célszerű a találatokat az automatikus klaszterszámuk alapján rendezni, hogy a szemantikailag hasonló igék egymás mellé kerüljenek. Ezt a 39. táblázat mutatja be.

A legjobban akkor látható, hogy a szóbeágyazás milyen jól ragadja meg a szavak szemantikai hasonlóságát, ha a klaszterszám szerinti rendezést a hangzássémás igéken végezzük el. Ekkor kétféleképp csoportosulnak az igék. Az egyik lehetőség az, hogy tipikusan ki/mi adja ki az igével jellemzett hangot. Ezt látjuk az (64a) és az (64b) példában: az előbbi igéi folyadékok, az utóbbié állatok hangjára utalnak. A másik lehetőség az, hogy milyen hangja/hangulata van egy-egy cselekvésnek. Így például az (64c) igéi élőlény mozgását, az (64d) igéi élőlény nézését írják le.

- (64) a. bugyog csobog zubog csorog csurog csöpög csepeg ...
 b. csivog vartyog nyihog brekeg gágog mekeg rőfög ...
 c. szédeleg kóvályog biceg tipeg csoszog kocog cammog ...
 d. pillant pislog pillog pislant hunyorog mered ...

igekötő	ige	klaszter	tagmondat
<i>el</i>	<i>emdéefesedik</i>	1	és elemdéefesedett.
<i>el</i>	<i>balosodik</i>	2	ha teljesen elbalosodna ez a fórum.
<i>el</i>	<i>liberálisodik</i>	2	ugyanis reménytelenül elliberálisodtak,
<i>el</i>	<i>pártosodik</i>	2	s elpártosodott,
<i>el</i>	<i>ordasodik</i>	3	és elordasodik az ő hazája.
<i>el</i>	<i>simicskásodik</i>	3	Mielőtt még teljesen elsimicskásodnánk itt,
<i>el</i>	<i>tőkésedik</i>	6	ha a volt elvtársak eltőkésednek,
<i>el</i>	<i>magyarodik</i>	7	Óhatatlanul elmagyarodott,
<i>el</i>	<i>amerikaisodik</i>	7	24 óra alatt elamerikaisodtak.
<i>el</i>	<i>európaisodik</i>	7	Amely mostanra szépen eleurópaisodott,

39. táblázat. Az első tíz találat az *el-N-(s)Odik* szerkezetben szereplő igékre, az automatikus klaszterszámukkal és az igéket tartalmazó tagmondatokkal együtt.

A PREVCONS lekérdezéseivel betekintést nyerhetünk az igekötők produktív kapcsolódási mintáiba. Ezenfelül szemantikai vizsgálatokra is alkalmas, ugyanis az igekötők főbb jelentései elválnak benne a különböző vonzatkeretek, automatikus igei klaszterek mentén. Ezeknek a jelentéseknek a leírását, a köztük lévő szemantikai viszonyok feltérképezését mutatja be a következő fejezet.

5.4. Az igekötők ontológiája

Az igekötők jellemzően poliszém, azaz több jelentéssel bíró lexikai elemek. Az egyes jelentések rendszerint nem függetlenek egymástól, metaforikus kapcsolatok fedezhetők fel közöttük. Ilyen például az *agyon* történetileg első jelentése, a ‘megöl’ (pl. *agyonver*; *agyonlő*), és az ebből kialakult ‘túlzásba esve (károsodásig/kimerülésig)’ jelentés, ahol a cselekvés eredménye nem feltétlenül halál, de átvitt értelemben ahhoz közeli állapot (pl. *agyondőgönyöz*, *agyonsikál*). Nemcsak az egyes igekötők különféle jelentései jellemezhetők hálózatként, hanem az a kapcsolatrendszer is, amely különböző igekötőket köt össze az egyes jelentéseik mentén. Erre kézenfekvő példák az olyan igekötőpárok, amelyek irányjelentésükben ellentétesek, pl. *ki-be*, *fel-le*, vagy közel azonosak, pl. *be-bele*, *szét-szerte*. Ez a fejezet egy olyan kísérletet mutat be, amelynek célja az igekötők jelentéseinek és a jelentések közti viszonyoknak a hálózatszerű ábrázolása a PREVCONS alapján, egy

ontológia formájában.⁴⁶

Az ontológia informatikai értelemben olyan tudástár, amely entitásokból és azok relációiból áll. Entitásként jelennek meg az igekötők és a jelentések, relációként pedig három hagyományos jelentésviszony, a szinonímia, az antonímia, valamint a hiperonímia. Az általam létrehozott ontológia szigorúan véve nem teljesíti az ontológiák egyik alapvető kritériumát: nem használ formális leíró nyelvet.⁴⁷ Ezt azonban nem tartom nagy hiányosságnak, mert a célom nem a közvetlen nyelvtechnológiai alkalmazhatóság volt, hanem az, hogy nyelvészeti motivációval láttassam és magyarázzam az igekötők szemantikáját. Az ontológia reprezentációs formája egy síkba rajzolható gráf, amely PDF formátumban érhető el.⁴⁸

Bár ilyen jellegű ontológia korábban nem készült a magyar igekötőkről, az igekötők szemantikájának jelentős szakirodalma van. J. Soltész (1959) egy teljes könyvet szentel a hat ősi igekötőnek (*meg, el, ki, be, fel, le*), amelyben részletesen ismerteti a használati körüket, jelentésárnyalataikat és kialakulásuk történetét ún. „hagyományos nyelvészeti” szemlélettel. Ennek a hat igekötőnek a jelentésrendszerét tárgyalja Tolcsvai Nagy (2015: 72-86) is, de teljesen más megközelítéssel, a funkcionális kognitív szemantika keretében. Fontos még megemlíteni Kiefer és Ladányi (2000a: 483-513), valamint Ladányi (2007: 183-237) munkáit, amelyek a *be* ősi igekötő mellett részletesen elemzik két kevésbé régi (*szét, agyon*) és egy kifejezetten új (*tönkre*) igekötő jelentéseit és kapcsolódási mintáit. Az utóbbi mű kitér az igekötők épp formálódó, új jelentéseire is a neologizmusok vizsgálatán keresztül (Ladányi 2007: 258-283). Hegedűs (2020) az *újra, újjá* és *vissza* igekötők szintaktikai és szemantikai jellegzetességeit ismerteti. Az említett szakirodalmak közös vonása tehát az, hogy az igekötő-állományból néhány lexikai elemet választanak, és ezeket rendkívül alaposan elemzik, ügyelve a jelentésárnyalatok megkülönböztetésére, ugyanakkor bemutatva a jelentések közti összefüggéseket.

Az általam készített ontológia eltér a fentebb ismertetett munkáktól a vizsgált igekötők halmozát, az elemzés módszerét és az eredmény közzéadási módját tekintve is. Arra törekedtem, hogy minél jobban lefedjem az igekötő-állományt, így az ontológia 56 igekötőt tartalmaz (amennyiben a *fel-föl*

⁴⁶ Az a megoldás, hogy az ontológiában az igekötőkhöz kapcsolódnak a jelentések, mindenképp leegyszerűsítő. Ezek sok esetben egy szerkezethez tartoznak inkább, amelynek az igekötő egy alkotóeleme, és nem egymagában hordozza az adott jelentést.

⁴⁷ Angol nyelvre több olyan formális ontológia is készült, amelyben kitüntetett figyelmet szenteltek az igekötőszerű elemek, igei partikulák részletes szemantikai leírásának, erről bővebben ld. Bhatia et al. (2018), McShane et al. (2005).

⁴⁸ <https://github.com/kagnes/prevonto>

típusú alakváltozatokat külön számoljuk, ez a szám 62). Az igekötők kiválasztásának a szempontjait az 5.4.2. fejezet ismerteti. Az eltérő célkitűzés miatt az ontológia készítésekor nem törekedtem olyan részletgazdagságra, mint az említett szakirodalmak szerzői. Ami az elemzés módszerét illeti, minden következtetésem a PREVCONS adataira épül. Tehát csak produktív jelentéseket állapítok meg, és ezt elsősorban a leggyakrabban megjelenő vonzatkeretminták alapján, másodsorban az igék automatikus klaszterszáma alapján teszem. Végül az eredményt sematikus formában, gráffal ábrázolom. Az ontológia a mostani formájában „emberek számára” értelmezhető, de át lehet alakítani olyan formális leírássá, amellyel nyelvtechnológiai szempontból is használhatóvá válik.

A továbbiakban arról lesz szó, hogy milyen nehézségekbe ütközött az ontológia építése annak ellenére, hogy a munka jól strukturált adathalmazból indult ki. Ezután az ontológia részletes bemutatása következik, előbb az entitások (az igekötők és jelentéseik), majd a jelentések közötti szemantikai viszonyok leírásával.

5.4.1. A poliszémiából adódó nehézségek

A fejezet bevezetőjében már felmerült, hogy az igekötők jelentései között metaforikus kapcsolatok figyelhetők meg. Néha olyan erősek és kiszámíthatóak az ilyen típusú összefüggések, hogy nehéz eldönteni, indokolt-e egyáltalán a külön jelentés felvétele. Előre megjósolható esetek például azok, amikor az irányjelentések nem csak térbeli irányokat jelölnek (pl. *visszamegy a boltba*), hanem időbelieket is (pl. *visszatekint a gyermekkorra*). Ez az absztrakciós lépés minden irányjelentéssel bíró igekötő esetében megfigyelhető, ezért az ontológiában az időbeli irányt nem választottam el a térbelitől.

Ugyancsak tendenciaszerűnek tűnik az irányjelentésű igekötők kapcsolata a kommunikációs igékkel. Ilyenkor az igekötő azzal járul hozzá a szerkezet jelentéséhez, hogy meghatározza a kommunikáció irányát (pl. *hátraság, utánaszól*). Jellemző viszont az is, hogy egy gyakori, lexikalizált ige mintájára sok új, jelentésében nagyon hasonló ige képződik (pl. a *beleszól* ‘kéretlenül szól hozzá valamihez’ jelentése mintájára a *belejópofiz*, *beleoffol* egy topikba, chatbe, stb.). Itt azt az irányelvet követtem, hogy ha viszonylag sok – öt vagy annál több – hapax illeszkedett ugyanarra a mintára, akkor az igekötőhöz tartozó külön jelentésként adtam meg azt a lexikalizált igét, amelyet mintaadónak véltem. Így tehát a *bele* esetében megjelenik a konkrét irányjelentés és a ‘beleszól valamibe’ jelentés is.

További kérdéseket vet fel az igekötők tömörítő képessége is. Meglehetősen gyakori eset az, hogy a tagmondatban szereplő egyik bővítményből képzünk igét, és ezt illesztjük egy lexikalizálódott igekötős igei szerkezetben az eredeti ige helyére. Ezek jellemzően inkább egyedi, kreatív szóalkotások eredményei, és nem indokolt produktív jelentésként felvenni őket az ontológiába. A jelenség jobb megértését szolgálják a (65) példásor mondatai. Ezekben a következő tömörítéseket látjuk: a (65a)-ban a *lenéptáncolom* a ‘néptáncolással ledolgozom’, a (65b)-ben a *ledinnyéztem* a ‘dinnével lefizettem’, a (65c)-ben a *kiuhuskodtak* az ‘uhu szemeihez hasonlóan néztek kifelé’ rövidebb megfogalmazása.

- (65) a. Én inkább elmegyek, és **lenéptáncolom** a felesleget, és tanulok sokmindent.
 b. Azóta természetesen lekenyereztem, akarom mondani, **ledinnyéztem** őket.
 c. De a nagy merev szemek **kiuhuskodtak** az összetoldozott fényképből.

Nem kerültek be az ontológiába azok az esetek sem, amikor egy jelentést nem a konkrét igekötő szintjén, hanem inkább egy meghatározott igekötőcsoport szintjén volna érdemes megadni. Erre példa az ‘irányjelentéses igekötő + ige + *magát*’ konstrukció, amelynek a legáltalánosabb jelentése az, hogy ‘az igével jelzett cselekvéssel elérte, hogy az igekötővel jelzett helyre jusson’. Ezt a mintát figyelhetjük meg a (66) példásor mondataiban.

- (66) a. Sajnos mire kiértünk a boltból, a kisfiú **hazapityeregte magát**.
 b. Természetesen ép eszű ember nem kívánhatja, hogy az izraeli zsidókat tolják a tengerbe [...] ha már **odaerőszakoskodták magukat**, [...]
 c. [...] aki két év kihagyás után, tavaly ősszel állt újra munkába, „**visszaedzette magát** a csapatba”, [...]

Bizonyos igekötők esetében ez a konstrukció sajátos jelentést vehet fel, és ezt már a konkrét igekötő mellett szükséges feltüntetni. Ilyen például az ‘*el* + ige + *magát*’, amelynek a jelentése ‘nem szándékosan csinál valami hirtelent’ (pl. *elhorkantja magát*). Emellett természetesen létezik az általános jelentés is, így pl. a (66a) mondat mintájára működne az is, hogy *a kisfiú elpityeregte magát a cukrászdába*.

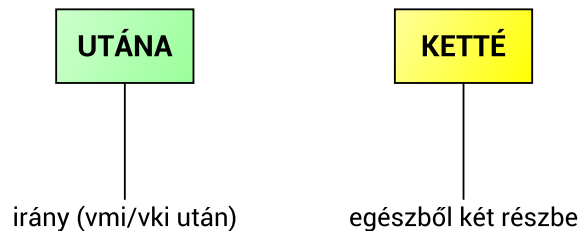
Végül fontos megemlíteni, hogy az aktuális jelentések mindig a szövegkörnyezettől függenek. Így például a (67a) mondatban a *letejz* jelentése az, hogy ‘tejjel véletlenül leönt’, míg a (67b)-ben látható *lespagettiz* azt jelenti, hogy ‘tévesen spagettinek nevez’, bár a *tej* és a *spagetti* szemantikai kategóriája közel azonos, és a vonzatkeret is egyezik.

- (67) a. Reggelijét befejezte // Nyakkendőjét **letejzte** // [...]
 b. [...] amit a hozzáértők simán „**lespagettiznének**”, hiszen ránézésre egyszerű paradicsomos tésztának látszik.

Látjuk tehát, hogy az egyes jelentések megállapítása nem egyszerű, és még az irányelvek gondos meghatározása mellett is befolyásoló tényező az egyéni intuíció. Ha több ember kapta volna a feladatot, hogy a PREVCONS alapján készítsen ontológiát, valószínűleg nem született volna két egyforma végeredmény.

5.4.2. Ontológiai entitások: igekötők és jelentések

Az ontológia kétféle entitásból áll – igekötőkből és ezek jelentéseiből –, amelyeknek a gráfban csomópontok felelnek meg (ld. 15. ábra).



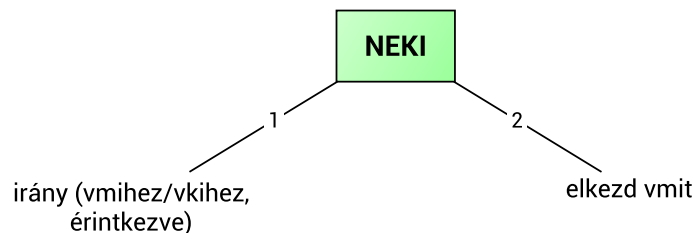
15. ábra. Két egyszerű példa az entításokra: az *utána* és a *ketté* igekötők a jelentéseikkel. A zöld háttérszín azt mutatja, hogy az igekötőnek van irányjelentése, a sárga háttér ennek a hiányát jelzi.

Csak azokkal az igekötőkkel foglalkoztam, amelyeknél a PREVCONS-ban megfigyelhető az, hogy névszóból képzett igéhez kapcsolódnak. A döntésemnek gyakorlati oka volt: a kutatás előrehaladtával egyértelművé vált, hogy a névszói bemenetek szemantikai kategóriái mondanak el a legtöbbet az igekötő lehetséges jelentéseiről, ezeknek a vizsgálatával lehet a leghatékonyabban haladni. Miután összevontam az alakváltozatokat,⁴⁹ 56 elemzendő igekötő maradt. Természetesen a későbbiekben további igekötőkkel is érdemes bővíteni az ontológiát.

⁴⁹ Az alakváltozatok a következők: *be – bé, bele – belé, fel – föl, felül – fölül, rá – reá, tele – teli*.

A jelentések kapcsán hangsúlyozni kell, hogy csak a PREVCONS alapján kimutatható, produktív jelentéseket vettem számba, ezeken belül pedig elsősorban a lexikai és metaforikus jelentésekre koncentráltam.⁵⁰ Ez az oka annak, hogy a *meg* jelentéskapcsolat nélkül szerepel az ontológiában. Az eredeti irányjelentése (‘hátra, vissza’) ugyanis ma már csak néhány igével mutatható ki, pl. *megad* ‘visszaad’, *megfordít* ‘visszafordít’, *megtér* ‘visszatér’.⁵¹ A *valami*, *valaki* stb. szavakat a lexikográfiában megszokott módon rövidítem: *vmi*, *vki*. A jelentések leírásánál törekedtem a minél általánosabb megfogalmazásra, és előfordult, hogy részben önmagával definiáltam egy szót, ha a gondosabb definíció túl bonyolult lett volna (pl. *utána* = *vmi/vki után*).

A 15. ábrán látható, hogy az igekötőt és a jelentést vonalak – a gráfban élek – kötik össze egymással. A legegyszerűbb és legritkább esetben az igekötőhöz egyetlen jelentés tartozik, ekkor az összekötő él nem kap semmilyen címkét. Amikor egy igekötőhöz több jelentés kapcsolódik, az élek számozottak (ld. 16. ábra). A számozási sorrend nem implikál semmit, tehát az 1-es nem jelenti azt, hogy a hozzá tartozó jelentés a legrégebbi vagy legtipikusabb. Ez csak a jelentésekre történő rövid hivatkozást szolgálja: $neki_1$ = ‘irány (valamihez/valakihez, érintkezve)’ (pl. *nekidöccen az ablaknak*), $neki_2$ = ‘elkezd valamit’ (pl. *nekidühödik az írásnak*). Az igekötők jelentéseit minden esetben példamondatok is illusztrálják, ezeknek a listája az ontológia github repozitóriumában érhető el.



16. ábra. A *neki* igekötő két produktív jelentése.

Az éleken nem csak számok szerepelhetnek, hanem szöveges tartalom is. A + magát címke azt jelenti, hogy az adott jelentés mindig együtt jár a *magát* vonzat megjelenésével. Ezt látjuk például a ‘*ki + ige + magát*’ szerkezetnél, amelynek jelentése ‘eleget csinál valamiből’, ld. a (68) példasort.

- (68) a. Na polgártársak, miután virtuálisan **kisöröztük magunkat**, döntsük el mi lesz ma este!
- b. S amikor mindenki **kiteniszte magát**, újra képbe jött a focimeccs, [...]

⁵⁰ Az igekötők kapcsolódási mintázataiban az aspektuális szerepük is lényeges lehet, ehhez ld. É. Kiss (2006a, b).

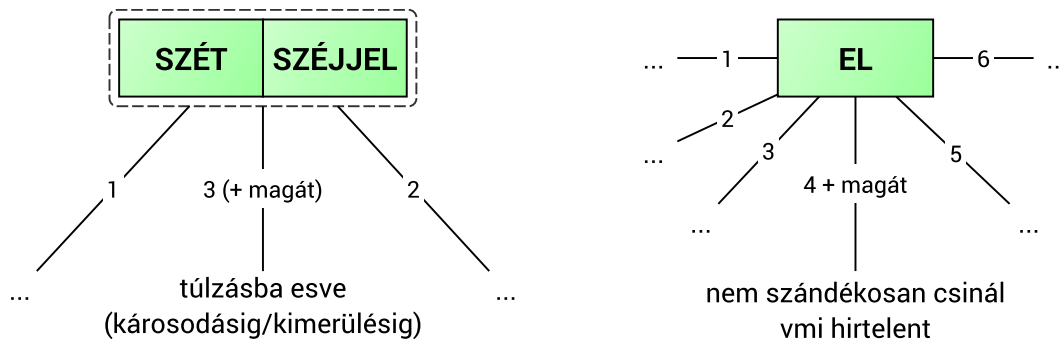
⁵¹ További példákhoz ld. J. Soltész (1959: 30-33).

- c. Külvárosi fiatalok, miután **kimolotovkoltélták magukat** Párizsban, egy tanyasi motelben húzzák meg magukat.

Ha a + magát címke zárójeles, akkor az adott jelentés gyakran, de nem kötelezően jár együtt a *magát* vonzattal. Ez figyelhető meg például a *szét*, *széjjel* igekötők ‘túlzásba esve (károsodásig/kimerülésig)’ jelentésénél, amelyet a (69) példaszor szemléltet.

- (69) a. Hosszú, hosszú hétvége következik, rengeteg idő lesz kirándulni, felállni a gép előtt - aki pedig nem akar az **szétkockulhatja magát**.
- b. Otthon mindenki **széjjelmulatja magát**, [...]
- c. Angolon meg nem hallok meg semmiccse, mer **szétwalkmanezem a fületem**.
- d. Egyébként ha **széjjeloffoljátok itt a topicot**, közétek lövötek.

Szólnak érvek mellett is, hogy itt nem egy, hanem két jelentésről van szó, mivel az akcióminőségek eltérők. A *magát* vonzat esetében exhaustív (‘kifáradásig, kimerülésig’), az egyéb tárgyeeset vonzat esetében intenzív (‘nagyon, túlságosan’) akcióminőség jelenik meg (Ladányi 2007: 221). Én nem választottam el ezt a két lehetőséget, hanem egy összevont, általánosabb jelentést fogalmaztam meg, mert azt tapasztaltam, hogy ha egy igekötőnél az egyik megvan, előkerül a másik is. A 17. ábra szemléltet minden lehetséges élcímketípust.



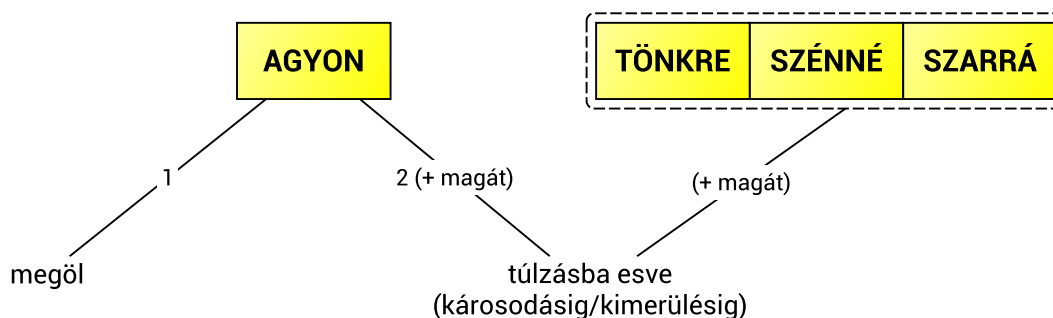
17. ábra. A *magát* vonzat megjelenése a *szét*, *széjjel* és *el* igekötőknél. A *szét* és *széjjel* körüli szaggatott vonal azt jelzi, hogy a két igekötő minden jelentésében szinonim, mindössze stílusbeli különbség van közöttük. Csak a tárgyalás szempontjából releváns jelentéseket tüntettem fel az ábrán.

5.4.3. Szemantikai relációk

Az ontológia háromféle jelentésviszonyt tartalmaz, ezek a következők: szinonímia (rokonértelműség), antonímia (ellentétes jelentés) és hiperonímia (hierarchikus viszony, amelyben az egyik szó – hiperonima – jelentése magába foglalja a másik szó – hiponima – jelentését). A gráfban ezek élekként jelennek meg, amelyek a jelentéseket kötik össze egymással. A továbbiakban mindhárom jelentésviszonyt részletesen, példákon keresztül veszem sorra.

A szinonímiát az jelzi, ha egy jelentésből több igekötőbe vezetnek élek. Ezt látjuk a 18. ábrán a ‘túlzásba esve (károsodásig/kimerülésig)’ jelentés esetében. Az ábrán szereplő *agyon*, *tönkre*, *szénné* és *szarrá* elemeken túl megtaláljuk ezt a jelentést a *szét*, *széjjel* és *össze* igekötőknél is.

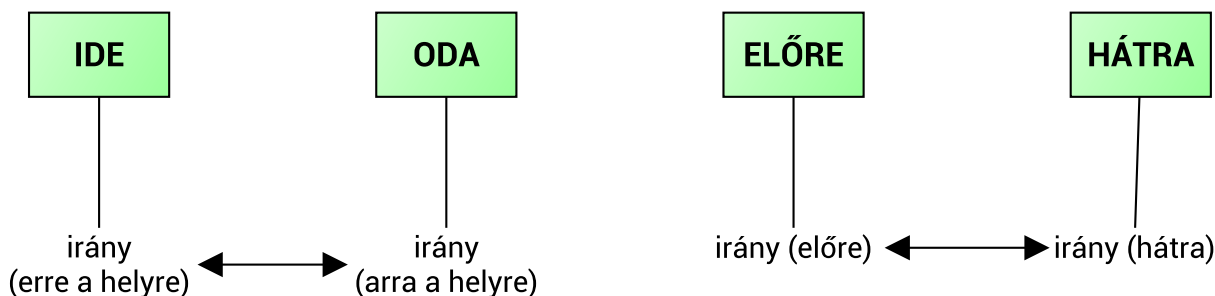
Ezt természetesen nem úgy kell érteni, hogy például a *szénné* bármikor kiválthatja az *agyon* igekötőt annak a második jelentésében anélkül, hogy a mondatban jelentésváltozás történne. Egy kis konnotációbeli vagy stílusbeli különbség mindig adódik, ehhez ld. a (70) példasort. A *szénné* és a *szarrá* szlengben fordulnak elő, a *tönkre* kissé informális – bár egyre gyakrabban jelenik meg hivatalos szövegekben is –, az *agyon* pedig semleges, bármilyen stílusregiszterben megtaláljuk.



18. ábra. A szinonímia megjelenítése az ontológia grájában: egy jelentésből több igekötőbe fut él.

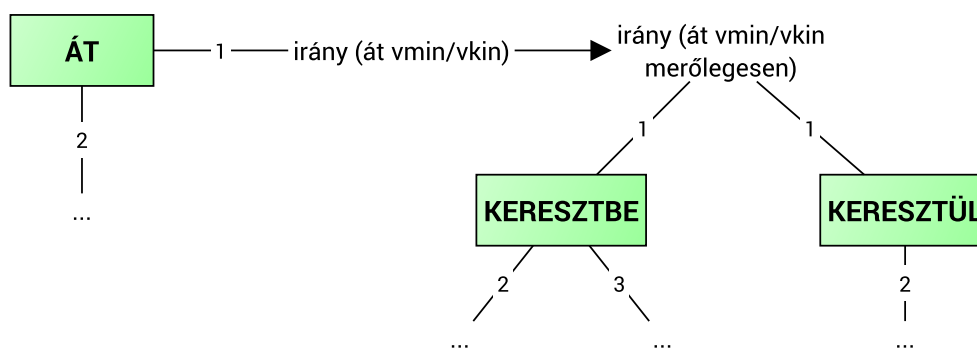
- (70) a. [...] bár a második itteni hetem végére az idegeim meglehetősen **agyoncincálódnak**.
- b. Én **tönkreröhögtem** magam ezen a hozzászóláson.
- c. [...] ő meg legalább **szénnéizgulja** magát az úton, vagy az autójában.
- d. Csórikám, ezzel **szarrákereshetné** magát a nyugati világban [...]

Az antonímiát a jelentések közé húzott kétirányú nyíl jelzi (ld. 19. ábra). Valószínűleg ez a legkönnyebben felismerhető szemantikai viszony az anyanyelvi beszélők többsége számára. Tipikusan a konkrét irányjelentéseknél jelenik meg.



19. ábra. Az *ide* és *oda*, valamint az *előre* és *hátra* jelentései közt fennálló antonim viszony, amelyet kétirányú nyíl jelez.

A hiperonímia jelzésére az egyirányú nyíl szolgál, amely a fölrendelt, általánosabb jelentésből az alárendelt, specifikusabb jelentés felé mutat (ld. 20. ábra). Ilyen viszony figyelhető meg például az *át* és a *keresztbe*, *keresztül* igekötők irányjelentései között, amelyeknek a különbségét a (71) példásor mutatja be.



20. ábra. A hiperonímiát egyirányú nyíl mutatja: az *át* irányjelentése magába foglalja a *keresztbe* és *keresztül* irányjelentését. Csak a tárgyalás szempontjából releváns jelentéseket tüntettem fel az ábrán.

A (71a)-ban az *áthajókáztak* ige alapján még nem tudjuk, hogy az alanyok a Bering-szorost szélteben szelték át, vagy hosszában hajókáztak rajta végig. Csak a második tagmondatból derül ki, hogy az utóbbi értelmezés a helyes. A *keresztbe*, *keresztül* irányjelentése specifikusabb az *át*-énál, mert azt is magukba foglalják, hogy a mozgás iránya a (71b)-ben az útra, a (71c)-ben az ágyra merőleges. Ez a különbség valószínűleg nem egyformán van jelen az anyanyelvi beszélők intuíciójában, mert néhány korpuszpélda azt sejteti, hogy az *át* és *keresztül* irányjelentése kvázi-szinonim, és csak a *keresztbe* fejezi ki mindig a merőlegességet.

(71) a. [...] 1995-ben **áthajókáztak** a Bering-szoroson, majd hosszú arktikus körút után 2000-ben Izland vizeire is eljutottak.

- b. [...] amikor hirtelen **keresztberohant** előtte az úton egy kondányi vaddisznó.
- c. Gyurica **keresztülvetődött** az ágyon, ruhástul.

Az *át* és *keresztül* esetéhez hasonló, apróbb bizonytalanságok ellenére törekedtem arra, hogy csak akkor jelöljek szemantikai relációt, ha a PREVCONS példái alapján mindig fennállni látszott a kapcsolat bizonyos jelentések között. Emiatt például nem került antonima-viszonyba az *egybe* és a *ketté*. Az utóbbi ugyanis azt jelenti, hogy egy egységből pontosan kettő lesz, és ennek az ellentéte az volna, hogy kettőből egy egység lesz. Az *egybe* ennél általánosabb: kettő vagy több rész eggyé válását jelenti, ld. a (72) példáit.

- (72) a. Néhány téglasor után jó bőven megöntözték a falakat, a borkősav felforrt, a mész megoltódott, és a téglasorok szinte **egybeégtek**.
- b. Összevissza kutyult levél lett belőle, ahogy újra **egybemarkoltam** széthulló gondolataimat, [...]

Zárásképpen fontosnak tartom egy példasorral illusztrálni, hogy a szemantikai viszonyok miatt az egyes jelentések szintjén értelmezhetők, és miért nem az igekötők szintjén. Létezik ugyanis olyan eset, amikor ugyanannál az igekötőpárnál megjelenik az antonímia és a szinonímia is, attól függően, hogy melyik jelentésüket vizsgáljuk – ilyen az *össze* és a *szét*. Ezek konkrét irányjelentésükben ellentétesek: az *össze* jelentése ‘több helyről egy helyre’, ld. a (73a) és (73b) példákat, a *szét* igekötőé pedig ‘egy helyről több helyre’, ld. a (73c) és (73d) példákat.

- (73) a. [...] sokkal erősebb kötelék, mint amivel a világirodalom nagyjait **összelasszózta**.
- b. A zsúfolt folyosón idegenül, de mégis egytetértő mosollyal pillantgattuk egymást, ahogy **összenyomakodtunk**.
- c. - **Szétlogatja** a bútorokat.
- d. Bizony, össze volt az már pöndörödve, de az ember szépen **szétsimítgatta**, [...]

Az *össze* és a *szét* ugyanakkor közel azonos értelmű egy absztraktabb jelentésüket tekintve: ‘túlzásba esve (károsodásig/kimerülésig)’. A (74) példasorban a két igekötő fel is cserélhető egymással úgy, hogy a mondat jelentése nem vagy csak árnyalatnyit változik.

- (74) a. Most jól **összeofftopikoztuk** XY rovatát.
- b. A névből sejtetődő, hogy a vírus **összebabrálja** a menüt, [...]
- c. **Szétgusztustalankodtad** az esküvőjét is, a sok dopeman meg tapsolt hozzá [...]
- d. **Szétkarcoltak** egy hajótestet, elégettek néhány kötelet, lökdösték egymást, [...]

Annak érdekében, hogy az egyes jelentések közötti kapcsolatok megállapítása minél objektívabb legyen, a jövőben érdemes lehet pszicholingvisztikai kísérletekkel, anyanyelvi beszélők nagy csoportján vizsgálni az ontológiában eddig rögzített jelentésviszonyok érvényességét.

5.5. Összegzés

Ebben a fejezetben egy sokoldalú témát vizsgáltam, az igekötők produktív kapcsolódási mintáit. Az adatvezérelt kutatás alapját 21 038 igekötős hapax képezte. Ahhoz, hogy ezzel a témával érdemben foglalkozni lehessen, több nyelvészeti részterületet is be kellett vonni a kutatásba. A produktív igealkotási módokhoz fonológiai és morfológiai, a vonzatkeretek kinyeréséhez szintaktikai, az igekötők és igecsoportok jelentéseihez szemantikai vizsgálatra volt szükség.

A fejezet első részében azzal a kérdéssel foglalkoztam, hogy hogyan hozunk létre új igéket. Az igealkotás számos módja közül azt a hármat vizsgáltam, amely számottevően volt jelen az adatban. Ezek a névszói és igei bemenetű igeképzés – és ezeken belül is a produktív képzők –, valamint a hangzásséma-alapú igealkotás. Az utóbbit Benő és Szilágyi N. (2015) nyomán elkülönítettem az igeképzésektől, és a leírását is máshogyan végeztem el (szótő–képző bontás helyett szótagszerkezeti sémákkal). Ennek a munkaszakasznak a jelentősége a meglévő ismeretek rendszerezésében, gyakorisági adatokkal való bővítésében, és pár ponton az átértékelésében mutatkozik meg (például annak kapcsán, hogy a -(s)Ul képző produktív-e vagy sem). Végül megvizsgáltam az igealkotási módok arányait az igekötős hapaxok halmazán, és kimutattam, hogy a legjelentősebb a névszói bemenetű igeképzés, amely a hapaxok 35,2%-át adja, viszont mindössze 62 – az alakváltozatok összevonása után 56 – igekötő kapcsolódik így képzett igéhez.

A fejezet második részében bemutattam a PREVCONS adatbázist, amelyet az igekötős szerkezetek vizsgálatára hoztam létre. Ez az erőforrás a kutatás alapjául szolgáló hapaxokat tartalmazza a morfológiai szerkezetükre, vonzatkeretükre, szemantikájukra és szövegbeli előfordulásukra vonatkozó információval együtt. A PREVCONS jelentőségét három dologban látom. Először is abban,

hogy már kiváló erőforrások készültek a gyakori igekötős igékre és ezek vonzatkereteire – ld. Sass et al. (2010) és Sass (2011) –, nem készült viszont olyan, amely hasonló információt adna a hapaxokról (ezek tokenszámukat tekintve kevesen, de típuszámukat tekintve nagyon sokan vannak). Másodsor, a PREVCONS-ban összegyűjtött minták lehetővé teszik a lexikai erőforrások hatékony bővítését. Végül, ami a legfontosabb: a PREVCONS a magyar nyelvet kutatók és tanítók számára jó áttekintést és érdekes adatokat szolgáltat az igekötős szerkezetekről.

A harmadik részben arra tettem kísérletet, hogy hálózat-szerűen, gráfként ábrázoljam az igekötők jelentéseit, és a jelentések közt fennálló viszonyokat. Ebből a célból létrehoztam egy ontológiát, amelyben entitásként jelennek meg az igekötők és a jelentések, relációként pedig három hagyományos jelentésviszony, a szinonímia, az antonímia és a hiperonímia. Az ontológia 56 igekötő szemantikáját írja le, és tervezem a további bővítését, javítását is. Számos intuitív vonása ellenére az ontológia a kutatás fontos eredményének tekinthető, főként a benne megjelenő új szemlélet miatt.

6. Visszatekintés, újragondolás

„Gondolkodom, tehát bajban vagyok.”

(korpuszpélda)

A disszertáció végéhez közeledve érdemes visszatérni ahhoz a hipotézishez, amelyből az egész gondolatmenet kiindult, miszerint az ‘igekötő’ fogalom prototípusos szerveződést mutat, így a prototípus-elmélet segítségével meghatározható. Azért indokolt ezt megtennünk ezen a ponton, mert immár lényegesen több adatunk van, mint amennyi a kiinduláskor rendelkezésre állt. Elsőként arról lesz szó, hogy a 3. fejezetben bemutatott igekötő-skála, és az ennek tagolásából előállt kategóriák mennyire illeszkednek az újabb adatokra. A válasz nem biztató: bár a két végpont („tipikus igekötők” – pusztá névszói igemódosítók) jól elkülönül, a kettő közötti szürke zóna egész máshogy alakul attól függően, hogy milyen további jellemzőket vizsgálunk. Ezért mindenképp indokolt, hogy más lehetőségeket is számításba vegyünk. Ennek a fejezetnek a központi témája egy ilyen lehetőség, amely Kálmán László javaslatából indul ki, és a lényege a nézőpontváltás: az egyes lexikai elemekről áthelyezi a fókusz az azokat tartalmazó szerkezetekre. Ez nem alternatív hipotézise a kezdetben megfogalmazottnak, mert nem az eredeti kérdés megválaszolása a célja – sőt, az a kérdés nem is releváns benne –, hanem egész más jellegű kérdéseket tesz fel. Azaz inkább egy alternatív szemléletnek lehetne nevezni – és ahogy látni fogjuk, ez a szemlélet számos előnnyel jár a korábbihoz képest. A fejezet végén kitérek a kutatás folytatási lehetőségeire, mert az itt bemutatott koncepció és adatok csak kezdeti lépéseknek tekinthetők az új megközelítés irányába.

6.1. Visszatekintés az igekötők állományára

A 3. fejezetben arra tettem kísérletet, hogy bizonyos – főként morfológiai és gyakorisági – jegyek alapján felállítsak egy skálát, amelynek az egyik végpontján a tipikus igekötők, a másik végpontján a pusztá névszói igemódosítók szerepelnek. Ezek a jegyek persze csak egy törtrészt alkotják a vizsgált elemek jellemzőinek. A későbbiekben bővült a jellemzők halmaza: a 4. fejezet az igekötők disztribúciós, az 5. a szemantikai tulajdonságairól nyújtott információt. Ezeknek az ismereteknek a birtokában érdemes visszatérnünk a dolgozat elejére, és megvizsgálunk a következő kérdést:

Változik-e lényegesen az igekötő-skála – és az ennek feldarabolásával létrejövő csoportok összetétele – akkor, ha egészen más jellegű jegyeket veszünk figyelembe?

6.1.1. Disztribúciós és szemantikai jegyek

Az igekötők disztribúciós jegyeinek tekinthető az arra vonatkozó összes információ, hogy milyen szerkezetekben állnak, és ezeken belül milyen szórendben és mennyire távolodnak el az igetőtől. A finit igei szerkezet esetében olyan általánosítást lehetett tenni az igekötők eloszlási mintái alapján, amely az előzőleg felállított skálával összhangban van (a prototipikus igekötők jellemzően közelebb állnak a finit igehez, mint a periférikusabbak). Ugyanakkor az igenevek és egyéb igei származékok esetében felszínre jöttek teljesen egyedi különbségek is. Itt az igekötő-skála és az ebből kinyert csoportosítás nem egyszerűen túláltalánosító, hanem megtévesztő is lehetett volna, ha nem vizsgáljuk meg alaposabban az adatot. Például úgy tűnhetne, mintha a periférikusabb igekötőkre lenne jellemzőbb, hogy -Ás képzős igei származékok tagadásakor preverbálisan elválnak az igetőtől, pedig valójában egyes lexikalizált esetek a kiugróak (pl. *egyed nem értés, végre nem hajítás*). Ha pedig nem az egyes szerkezeteket, hanem az összképet vizsgáljuk, az derül ki, hogy néhány elem, amely az eredeti skála szürke zónájában a puszta névszói igemódosítókhoz állt közelebb, a szerkezetekben való előfordulásai alapján éppen a tipikus igekötőkhöz közeledne (pl. *létre, észre*). Továbbá nincs olyan igekötő, amelyik minden igekötős szerkezetben adatolható lenne az MNSZ-ben, ami a prototípus kiválasztását – legalábbis ennek a korpusznak az alapján – újfent problémássá teszi.⁵²

Az igekötők szemantikai jegyeit illetően azt lehet megvizsgálni, hogy az ontológiában egy-egy igekötőhöz hány darab jelentés kapcsolódik, és ez az érték milyen kapcsolatban áll a kezdetben vizsgált jegyek értékeivel. Az ontológiába felvett jelentések darabszáma pozitív korrelációt mutat a morfológiai produktivitással ($r = 0,72$), és negatívan korrelál a szerkesztettséggel ($r = -0,53$) és a szótagszámmal ($r = -2,24$). Ez azt jelenti, hogy amelyik elem rendelkezik az igekötőség kezdetben megállapított, tipikus tulajdonságaival, az jellemzően polyszém is. Viszont fontos felidézni azt, hogy az ontológia lexikai és metaforikus jelentéseket tartalmaz, így a gráfban a *meg* igekötő izolált csúcsként jelenik (azaz nem csatlakozik hozzá él, nincs összekötve jelentéssel). Amennyi-

⁵² Az adatolhatóság hiánya természetesen nem feltétlenül jelenti azt, hogy ne lenne létrehozható, vagy ne lenne más korpuszban adatolható is. A *meg* például nem tűnik fel a -stUl/stÓl képzős társhatározói szerkezetben – míg a többi prototipikusnak nevezett igekötő igen –, de létrehozhatók olyan szavak, mint pl. a *megoldásostul*. Ezzel szemben a *véget* nem áll -hAtÓ igeneves szerkezetben, és valóban szokatlannak tűnne pl. a **?végetérhető* kifejezés.

ben grammatikai jelentéseket⁵³ is vizsgáltam volna, minden bizonnyal kisebb lenne az eltérés a hatósi igeikötő között, így azonban csak annyi látszik, hogy az ontológiában a prototípusnak választott elem minden más elemtől különbözik.

6.1.2. A prototípus-elméleti keret revíziója

Az eddig látottak alapján az körvonalazódik, hogy a sok különböző jegy mind önmagában egy skála, különböző értéktartományokkal. Egyértelműen csak azt tudjuk mérni, hogy egy adott jegy értéktartományában hol helyezkednek el bizonyos lexikai elemek, és ezt is csak akkor, ha nem szemantikai természetű a jegy. Ezáltal az kifejezhető, hogy egy-egy konkrét jegy mentén mennyire hasonló két elem. Viszont az változatlanul problémás kérdés, hogy ezt a sok skálát hogyan lehetne egyetlen tipikalitási skálába aggregálni, és hogy van-e ennek kognitív megalapozottsága az olyan tudományos fogalmak esetében, mint az ‘igeikötő’.

A továbbiakban félreteszem az eredeti hipotézist, miszerint az igeikötők prototípusos szerveződést mutatnak – nem vettem el a lehetőségét, de az itt alkalmazott módszer nem szolgál rá meggyőző bizonyítékkal –, és egy lényegileg más nézőpontba helyezkedek, amelyet a szerkezetek felőli megközelítésnek fogok nevezni.

6.2. A szerkezetek felőli megközelítés alapjai

Elsőként bemutatom azt a koncepciót, amely a lexikai elemek vizsgálatáról a szerkezetek vizsgálatára helyezi a hangsúlyt, majd ismertetem a koncepció megvalósításának a kezdeti lépéseit. A nulladik lépésnek – a további adatelemzések alapjának – a PREVDISTRO adatbázis tekinthető, amely 49 igeikötős szerkezet típus 41,5 millió korpuszbeli előfordulását tartalmazza. Végül az ezen végzett néhány egyszerű adatelemzéssel megpróbálok egy kis ízelítőt adni abból a rengeteg lehetőségéből, amelyet az új megközelítés nyújt.

6.2.1. Az elképzelés

Az a megközelítés, amely ennek a fejezetnek a központi témája, három forrás szintézisének tekinthető. Ötvözi Kálmán Lászlónak a disszertációm első verziójára adott bírálatát, a konstrukciós

⁵³Ladányi (2000) azt tartja az igeikötőség egyik kritériumának, hogy az adott elem a grammatikai jelentéseiben – aspektus, akcióminőség – produktív.

nyelvtanok általa bevezetett memória-, példány- vagy esetalapú változatát (Kálmán 2016), valamint az igekötős szerkezetekkel kapcsolatos kutatásaim eddigi tapasztalatait. Az alapgondolat az opponensi vélemény alább idézett soraiból származik:

„Arról van szó, hogy vannak a magyarban olyan szerkezetek (erről szól a 4. pont), amelyek bizonyos (sokféle, de egymással családi hasonlóságban álló, sok átfedést mutató) jelentéstani jellegzetességekkel rendelkeznek, és ennek következtében egyes összetevőikben ugyanazok a lexikai elemek is szívesen jelennek meg. Ezeknek a legjellegzetesebb összetevőjük az, amit sokszor (de, mint a szerző is megállapítja, nagyjából kiszámíthatatlan, hogy pontosan mikor) „igekötőnek” neveznek. Azok a tényezők, amiket az „igekötőség” legjellemzőbb sajátosságaiként felsorol, részben magukat ezeket a szerkezet típusokat jellemzik, részben pedig azt befolyásolják, hogy a korábbi szakirodalomban melyiket milyen valószínűséggel nevezik igekötőnek (minél kevésbé felismerhetően szerkesztettek, minél rövidebbek és minél gyakoribbak, annál nagyobb eséllyel).”

Tehát olyan jelenségről van szó, ahol a szerkezetek az elsődlegesek, és a kategóriák egyfajta melléktermékei a bizonyos szerkezetekben való hasonló előfordulási mintáknak. Így a kutatás célja sem az kell legyen – és főleg nem abból kell kiindulnia –, hogy ezeket az epifenomenális kategóriákat kísérelje meg azonosítani. Ha ezt elfogadjuk, akkor el lehet gondolkodni azon, hogy a korpuszvezérelt vizsgálat közvetlenül a szerkezetekkel kezdődjön. Jelen esetben ez azt jelenti, hogy a 4. fejezet tárgyköre és az ott alkalmazott módszer szolgálhat a kutatás alapjául, de nem egészen az eredeti formájában. A mérési eljárást át kell alakítani az itt vázolt koncepciónak megfelelően, ami azt jelenti, hogy bizonyos előfeltevéseket el kell hagyni belőle.

A 4. fejezet adottnak vette a korábban megállapított négy igekötő-kategóriát. Ezeket értelemszerűen félre kell tennünk, és ismét mind a 235 lexikai elemre vonatkozóan kell korpuszvizsgálatot végeznünk, ami tehát a PREVLEX-ből származó teljes halmazt magába foglalja. Persze már ez is leszűkíti a lehetőségeket, és továbbra is esetleges marad sok vizsgálatra érdemes elem kimaradása. Például az intenzifikáló jelentésű konstrukciókban hasonló gyakorisággal tűnik fel a *rommá*, mint a *szénné* (pl. *szénné/rommá keresi/töri/issza magát*), de a vizsgált halmazban csak az utóbbi szerepel. Ez a most még fennálló esetlegesség kiküszöbölhető lenne, ha a későbbiekben a korpusz összes szavára végeznénk disztribúciós vizsgálatot. Ez persze sokkal komplexebbé teszi a mérési eljárást, de a megvalósítása nem tűnik lehetetlennek.

Ahogy az igekötők, úgy az igekötős szerkezetek esetében is melléktermékszerűnek tekinthetők a kategóriák (vagyis a szófajok). A gyakorlatban ez a következőt jelenti: például ne tekintsük eleve

adottnak, hogy van -Ó végződésű melléknévi igenév, melléknév és főnév – a 4. fejezetben még ez történik, a korpusz annotációjára támaszkodva –, hanem inkább abból induljunk ki, hogy van egy -Ó végződésű elemet tartalmazó szerkezet, és ennek a disztribúciójából bontakozik ki a szófaj benyomása. Ennek jelentős szakirodalma van, pl. Langacker (1991: 59–100), Croft (2001: 63–107), Kálmán (2016), Diessel (2019: 142–171).

Végül el kell hagyni azt az előfeltevést, hogy van prototípus és vannak súlyozott jegyek. Ez utóbbi annyit jelent, hogy az absztrakt jegyek rangsorolását kiiktatjuk. Ettől függetlenül a modellben megjelennek súlyok, kapcsolati erősségek, viszont ezek az egyes példányok szintjén jelentkeznek, és a jelenlegi eszközeinkkel az előfordulási gyakoriságok alapján becsülhetők, ld. Kálmán (2016: 70).

A következő pontban egy olyan erőforrást mutatok be, amely az itt felsorolt szempontok figyelembe vételével készült.

6.2.2. A kiindulás: a PREVDISTRO adatbázis

A PREVDISTRO (*Preverb Distributions*) egy nyílt hozzáférésű adatbázis,⁵⁴ amely 49 igekötős szerkezettypusnak 41,5 millió korpuszbeli előfordulását tartalmazza. Az erőforrás 10 oszlopból áll, ezek a következők:

1. id: Az adatbázisrekord egyedi azonosítója, ennek csak technikai szerepe van.
2. constype: A szerkezettypus, amelynek az elnevezése rendszerint az igei részben azonosítható legutolsó toldalék⁵⁵ szerint alakul. Így például a *fel nem használendő* típusa -AndÓ, a *kiórlésű* típusa -Ú. Ettől a névadási konvenciótól csak a következő esetekben tértem el: (1) Felső- és túlzófok – pl. *legelsőprőbb, legeslegeslegmegoldottabb* –, amelyeket egyaránt *leg-...-bb* jelöl. (2) Finit igei szerkezetek (*FIN*). (3) A 4.1.7. fejezetben ismertetett *el kell, hogy menjek* – *el kell menjek* típusú szerkezetek (*climbing*). (4) A 4.1.8. fejezetben bemutatott topikalizáció (*topicalization*). A 49 szerkezettypus áttekintését a III. számú melléklet tartalmazza.

⁵⁴ <https://science-data.hu/dataset.xhtml?persistentId=doi:10.5072/FK2/TRSD50>

⁵⁵ A szakirodalomban ezek a toldalékok vagy egyértelműen képzők, vagy vitatott besorolású igenévi toldalékok. Az itt bemutatott megközelítés szempontjából ezeknek az elméleti vitáknak nincs közvetlen jelentősége, ezért mellőzöm a részletesebb tárgyalásukat.

3. subtype: Ez a mező tovább részletezi a szerkezet típusokat aszerint, hogy hogyan alakul bennük az igető és a lexikailag hozzá tartozó igekötő szórendje. A *prev* a preverbális, a *prefixed* a közvetlenül preverbális, a *post* a posztverbális igekötőre utal.
4. prevpos: Itt látható az igekötő pozíciója az igetőhöz képest. Ennek az értékét a 4.1.1. fejezetben ismertetett módszerrel határoztam meg. Az írásjeleket nem tekintettem külön pozíciónak, így pl. a *fel kellett volna, hogy tűnjön* és a *ki kellett volna hogy fizessék* igekötője egyaránt -4-et kapott. Az egybe- és különírás különbségeit sem vettem figyelembe, tehát pl. a *meg nem támadási* és a *megnem támadási* esetében is -2 pozícióban álló igekötőt azonosítottam. Az automatikus módszerből eredő hibákról a 4.1.1. fejezet nyújt tájékoztatást.
5. prev: Igekötő.
6. verb: Igelemma. Fontos látni, hogy az erre vonatkozó információ nem igazán megbízható, mert egy-egy igealakot több lemmára is vissza lehet vezetni, pl. *mentem* → *megy/ment*, *ért* → *ér/ért/érik*. Az ilyen igealakok esetében gyakori probléma, hogy az igelemma elvileg lehetséges, de az adott kontextusban hibás.
7. intervening: Az igető és a lexikailag hozzá tartozó igekötő között megjelenő szavak lemmái, pl. a *jönni akartunk haza* esetében az *akar*). Amennyiben nincsenek ilyen köztes elemek, a mező tartalma '-' (kötőjel). Az *-e* klitikum is különálló, köztes elemként jelenik meg.
8. actform: A szerkezet korpuszbeli előfordulása (szóalakok).
9. doc_id: A szerkezetet tartalmazó mondat azonosítója az MNSZ-ben. Ezáltal visszakereshető bizonyos metaadatok, és a mondat nagyobb kontextusban is vizsgálható.
10. sent: A szerkezetet tartalmazó mondat, amelyben a releváns részt < ... > jelek fogják közre. A mondatok tokenizáltak – tehát nem tapadnak az írásjelek –, de az MNSZ-ben elérhető lemmatizálást és morfoszintaktikai elemzést nem tartalmazzák, pl. *És mondják , mondják , < véget nem érően > .*

Az adatbázis számos kutatási irány előtt nyitja meg az utat. Egyfelől részletekbe menően vizsgálhatók benne az egyes szerkezet típusok – ennek az előképe volt a 4. fejezet –, másfelől alkalmas a

makroszintű elemzésekre, tehát akár az összes szerkezet típus együttes vizsgálatára. A következő fejezetben ennek az utóbbi irányoknak mutatom be néhány elemi lépését.

6.2.3. Adatelemzés a PREVDISTRO felhasználásával

A PREVDISTRO adatbázisból előállítható egy olyan mátrix, amelynek az oszlopait az egyes szerkezetek altípusai (a subtype-ok), a sorait pedig az igekötők adják. Az egyes cellák értéke az adott igekötő abszolút gyakorisága az adott szerkezetben. Ennek a mátrixnak egy kis részletét mutatja be a 21. ábra.⁵⁶

	prefixed_FIN	post_FIN	prefixed-Ás
meg	3653584	2067862	1834275
el	2386639	1389865	1090672
ki	1387195	1109826	1016347
fel/föl	957721	699244	858342
be/bé	930634	714179	775121
le	733055	351924	291222
át	287753	218125	218787
össze	206928	160548	264143
vissza	305613	160577	169377

21. ábra. Részlet az Igekötő × Szerkezet mátrixból. Az oszlopok és a sorok is gyakoriság szerinti csökkenő sorrendben állnak, így a mátrix bal felső szegmensében a leggyakoribb igekötőket látjuk a leggyakoribb szerkezetekben. A cellák színe az abszolút gyakoriság függvényében alakul: minél magasabb ez az érték, annál vörösebb a cella. Az oszlopnevek feloldása: ‘igekötő + finit ige’, ‘finit ige + igekötő’ és ‘igekötő + -Ás szerkezet’.

Elsőként érdemes ezt a mátrixot a szerkezetek felől megvizsgálni: egy-egy szerkezetben mennyire van jelen egy-egy igekötő? Ennek a legegyszerűbb módja az, ha kiszámítjuk az adott igekötő relatív gyakoriságát az adott szerkezetben. Ekkor két jelenség lesz szembevetendő. Az egyik az, amikor bár a konstrukció és a lexikai elem külön-külön gyakoriak, sosem állnak együtt (ilyen például az *egyed* vagy a *neki* teljes hiánya a -hAtÓ és -hAtAtAn konstrukciókban). A másik pedig a kiugró kapcsolatok jelenléte: bizonyos szerkezetekben elvileg bármelyik vizsgált elem állhatna, mégis csak egy-kettő jelenik meg bennük dominánsan, mert így használják a beszélők. Ilyen jellegzetes együttállásokat mutat be a 40. táblázat.⁵⁷

⁵⁶ A teljes mátrix a <https://github.com/kagnes/prevmatrix> címen érhető el.

⁵⁷ Kollokációs mintázatokat akkor is találunk, ha visszahelyezkedünk a korábbi nézőpontba, és a lexikai elemek felől

szerkezet altípus	igekötő	rel.gyak.	tipikus kifejezés(ek)
prev-AndÓ	vissza	96,05	vissza nem térítendő
prefixed-tAlAn	ki	91,27	kilátástalan, kifogástalan
prefixed-ÓlAg	ki	70,98	kizárólag
prefixed-nivaló	ki	47,82	kivetnivaló
prefixed-szerű	meg	47,00	meglepetésszerű
prefixed-tA	le	45,90	letelte, lejárta
prefixed-(A)tAl	fel/föl	44,08	feltétel
prefixed-hAtAtlAn	el	44,06	elfogadhatatlan, elengedhetetlen
prefixed-(A)n	meg	42,11	megfelelően
prefixed-sdi	ki	42,11	kiszorítósdí
prev-(O)(t)t	meg	40,41	meg nem nevezett
post-AndÓ	össze	38,92	nem tévesztendő össze
hogy_FIN	meg	38,70	meg kell, hogy mondjam
no_hogy_FIN	meg	35,90	meg kell mondjam
prefixed-(A)tAlAn	meg	35,83	megalapozatlan

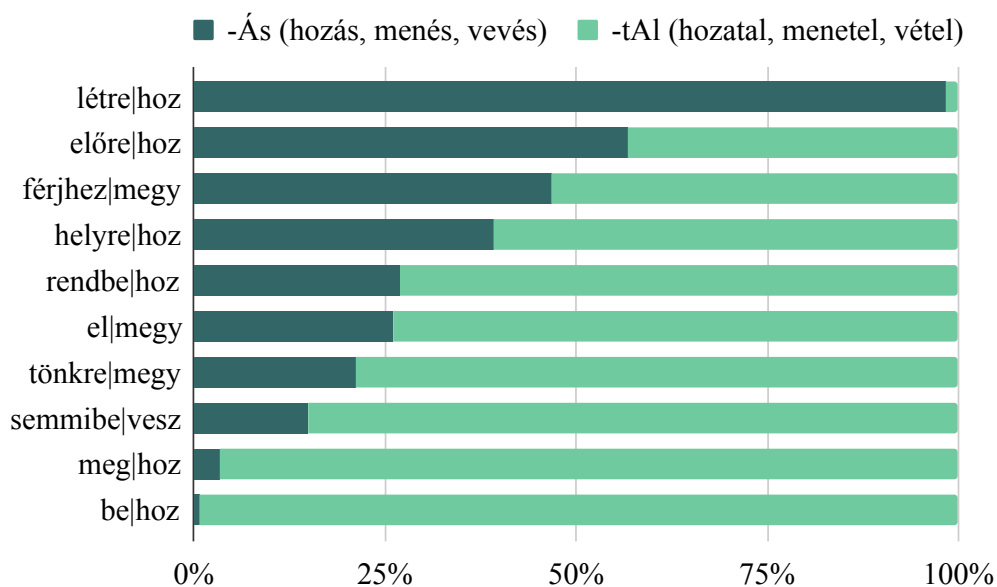
40. táblázat. Olyan konstrukciók, amelyek a korpuszbeli előfordulásaik minimum egyharmadában egyféle igekötőt tartalmaznak (500-as gyakorisági küszöb mellett). A PREVDISTRO adatait alaposabban szemügyre véve az is látható, hogy ilyenkor rendszerint nemcsak az igekötő a jellegzetes összetevő, hanem a teljes szerkezet ugyanazokkal az alkotóelemekkel ismétlődik.

A konstrukcióknak nemcsak bizonyos lexikai elemek válhatnak ilyen szerves részeivé. Például a 40. táblázatban is megjelenő -tA szerkezet jelentését hozzávetőleg úgy lehet meghatározni, hogy ‘az ige által kifejezett esemény időbeli kezdő- illetve végpontja’, emiatt a ragjai és névutói is jellegzetesek, ld. a (75) és a (76) példasort.

- (75) a. A tél **elmúltával** szinte biztosan újrakezdődnek a harcok, [...]
- b. Az új év **beköszöntével** úgy tűnik, csillapodnak az indulatok, [...]
- c. Négyéves mandátuma **lejártával** hazaindulni készül a szentszéki nagykövet.
- (76) a. A határidő **lejártáig** 8 pályázat érkezett.
- b. A rendelkezés a per **lezárultáig** érvényben marad.
- c. Egyébként jobb a másnaposság **elmúltáig** ágyban maradni.

vizsgáljuk az anyagot. Ez esetben az a kérdés, hogy melyik szerkezetet mennyire preferálja az adott elem? Azaz a szerkezetek relatív gyakoriságát számoljuk az egyes igekötők esetében. Itt a ritkább elemeknél fedezhetők fel kiugró kapcsolatok, pl. *közben|járás, részre|hajlás, lényegre|törő, témába|vágó, mennybe|menetel, nyomra|vezetői (díj).*

Kevésbé szembetűnő, de a mátrix adataiból kiolvasható jelenség az, amikor két szemantikailag közel álló szerkezetet vizsgálva sokkal gyakrabban adatható egy igekötő az egyikben, mint a másikban. Erre példa az -*As* és az -(A)tAl szerkezetek, amelyek valamilyen eseményre, történésre referálnak (pl. *jövés – jövetel, vivés – vitel*). Egy fontos különbségük az, hogy az -*As* produktív (magas típus- és tokengyakorisággal), az -(A)tAl pedig nem az (igéknek egy kicsi csoportjával áll, de a tokengyakorisága magas). A jelentéskülönbségük azonban minimális, már ha van egyáltalán. A 22. ábra olyan igekötős igét mutat be, amelyek mindkét szerkezetben legalább 50 előfordulással adathatók. Azt, hogy esetükben az -*As* vagy az -(A)tAl konstrukció kerekedik-e felül – vagy kiegyenlített maradnak-e az arányok –, nehéz lenne másra visszavezetni, mint a beszélők nyelvhasználati szokásaira. Kálmán (2020) ezt a precedensjog működéséhez tartja hasonlóknak, és kiterjeszti a nyelvi viselkedés egészére.



22. ábra. Néhány igekötős ige jelenlétének aránya az -*As* és az -(A)tAl konstrukciókban (mindkét esetben 50-es gyakorisági küszöbvel).

A 22. ábra azt szemlélteti, hogy az MNSZ adatai alapján egyes variánsok közül melyek a gyakoribbak. Például a *létrehoz* esetében a *létrehozatal* és a *létrehozás* is használt formák, de az esetek 98,3%-ában az utóbbit mondjuk. Ezzel ellentétes tendencia látható a *behoz* esetében: a *behozatal* alak tűnik fel 99,4%-ban – ez egyúttal szakszó is, amelyet közgazdasági kontextusban csak így szokás mondani –, a megegyező jelentésű *behozás* helyett. És arra is van példa, hogy a két forma közel azonos arányban fordul elő, például valami *előrehozása* vagy *előrehozatala*, valaki *férjhezmenése*

vagy *férjhezmenetele*.

Az említett jelenségek bemutatása elnagyolt és rövid volt, de feltételezésem szerint arra elég, hogy érzékeltesse a szerkezetek felőli megközelítésben rejlő lehetőségeket.

6.3. Hogyan tovább?

A kutatás folytatásának egy olyan területe biztosan van, amely elméleti modelltől függetlenül érdeklődésre tarthat számot: ez a PREVDISTRO adatbázis bővítése. Ennek három irányát látom. Először is, tovább lehet gazdagítani a mai magyar nyelvállapotra vonatkozó újabb adatokkal. Ide tartoznak a dolgozatban nem tárgyalt konstrukciók, például az inkorporáló szerkezetek – pl. *famegmun-káló* (gépek), (mamámra rájött a) *mélyhűtőkiüríthetnék* –, amelyek nyelvtipológiai szempontból is érdekesek. Emellett érdemes lenne belefoglalni az egyes szerkezet típusokon belül az igekötők komplexebb morfológiájú előfordulásait is. Ilyenek a reduplikált, *-fele/felé* morfémával ellátott, személyragos és fokozott alakok, illetve ezek kombinációi (ld. a 3.3. fejezetet).

Másodsor, diakrón korpuszok anyagát is fel lehet venni az adatbázisba. Az ilyen jellegű korpuszvizsgálatok előzménye volt a 4.2. fejezet, amely viszont több szempontból leszűkített volt: csak a hat ősi igekötő és csak a finit igei szerkezetek vizsgálata volt a célja. Ha a PREVDISTRO egységes formátumában elérhetőek a korábbi nyelvállapotok igekötős szerkezetei, az olyan kvantitatív vizsgálatokat tesz lehetővé, amelyek a történeti korpuszok eltérő elemzettségi szintjei és annotációs formátumai miatt jelenleg nagyon nehézkesek.

Harmadszor, a PREVDISTRO metaadatokkal is bővíthető. A korábbi nyelvtörténeti korok korpuszai részletgazdagok ebből a szempontból: legalább az évszámra, a szerzőre és a forrásmű címére vonatkozó adatok kinyerhetők az ÓMK-ból, a TMK-ból és az MTSZ-ből is. Az MNSZ ilyen tekintetben rosszabbul áll, de a forrásmondatok régiója és stílusregisztere elérhető benne. A metaadatok lényegesek ahhoz, hogy nyomon lehessen követni az egyes konstrukciók földrajzi és időbeli terjedését vagy visszaszorulását, valamint azokat a folyamatokat, amelyek során egyre több lexikai elem kerül igekötőszerű szerepbe.

Az adatbázishoz kapcsolódó, fontos fejlesztési lépés az is, hogy megteremtődjön a könnyebb kézi adatelemzés lehetősége. A PREVDISTRO jelenlegi formájában csak számítógépes feldolgozásra alkalmas, de tervezem az átalakítását olyan relációs adatbázissá, amely könnyű kereshetőséget biztosít a nyelvtechnológiában nem jártas felhasználók számára is.

A kutatás másik folytatási iránya a konstrukciós nyelvtanokhoz vezet, és – legalábbis ebben a kezdeti szakaszban – főként az adatelemzés fejlesztését érinti. A 6.2.3. fejezetben a legegyszerűbb módszert mutattam be, amely abszolút és relatív gyakoriságokon alapult, ugyanakkor az egyszerűsége okán még nem lebecsülendő, mert fontos jelenségekre hívta fel a figyelmet. Ezek szintagmatikus és paradigmaticus kapcsolatokat tükröztek, amelyek erősségének és irányának a vizsgálatára több módszer létezik. Ezek gyűjtőneve a kollostrukciós elemzés – bővebben ld. (Stefanowitsch és Gries 2003) –, amelyre PREVDISTRO a mérete és a részletgazdagsága folytán alkalmasnak tűnik. A későbbiekben érdemes lenne megvizsgálni, hogy a nyelvtechnológiában már általánosan elterjedt gépi tanulási módszerek hogyan hasznosíthatók a konstrukciós szemléletben, mert a hasznosíthatóságuknak világos jelei vannak, ld. (Gries 2013).

6.4. Összegzés

A fejezet első részében összeérték a korábbi témák szálai: az igekötőnek nevezett szóosztály morfológiai, gyakorisági, disztribúciós és szemantikai tulajdonságait együttesen vizsgáltam annak érdekében, hogy újabb szempontok alapján lehessen kiértékelni a dolgozat elején bevezetett prototípus-elméleti megközelítést. Bár az eszerint felállított skála két végpontja csak minimálisan változott a további jegyek függvényében, a köztes szakasza – ami egyúttal a legnagyobb elemszámú – jelentős eltéréseket mutatott, különösen a disztribúciós jellemzőket illetően.

A lexikai elemek felőli megközelítésben tapasztalt számos bizonytalanság vezetett el ahhoz a nézőpontváltáshoz, miszerint az igekötős szerkezetek vizsgálatát a szerkezetek felől is lehet – sőt úgy tűnik, hogy így érdemes – kezdeni. Ennek a megközelítésnek két nagy előnye biztosan van az eredetihez képest: (1) Nem kell bevezetni olyan, nem szilárd empirikus alapon nyugvó döntéseket, mint amelyek korábban szükségesek voltak. (2) Azáltal, hogy előzetesen nem kategorizáljuk a lexikai elemeket, sokkal kisebb az információvesztés, illetve annak a veszélye, hogy túláltalánosítunk.

Bemutattam az új megközelítés szellemében készült PREVDISTRO adatbázist, amely 49 szerkezet-típus korpuszbeli előfordulásait tartalmazza, minden előfordulás esetében feltüntetve külön az igekötőt és az igelemmát, az igekötő pozícióját az ige-tőhöz képest, valamint a szerkezetbe beférkőző egyéb szavakat. Emellett elérhető a szerkezetek nagyobb kontextusa – a teljes mondat – is. A 41,5 millió rekordból álló adatbázis nyílt hozzáférésű.

Ezután bemutattam PREVDISTRO-n végezhető adatelemzés legalapvetőbb módjait azzal a céllal,

hogy felhívjam a figyelmet a szerkezetek felőli megközelítésben rejlő lehetőségekre. Végül felvázoltam a kutatás folytatási lehetőségeit, amelyek elsősorban a PREVDISTRO adatbázis bővítését és az adatelemzés fejlesztését érintik, valamint az elméleti modell alakulására is hatással vannak.

7. Összefoglalás, új tudományos eredmények

„El kell, hogy engedj!”

(V'Moto-Rock)

A disszertációmban a magyar nyelv igekötős szerkezeteit vizsgáltam. A három fő témám a következő volt: (1) az igekötők állományának meghatározása, (2) az igekötős szerkezetek szórendi mintázatainak leírása, valamint (3) az igekötők produktív kapcsolódási mintáinak feltárása. A munkám során korpuszvezérelt megközelítést alkalmaztam, amelynek a lényege az, hogy a kutatás nagy mennyiségű szöveganyag automatikus elemzéséből indul ki, és elsődleges célja az introspekcióval nem feltárható jelenségek kimutatása.

A 2. fejezetben a kutatás módszertani háttérét mutattam be. Körüljártam a korpuszvezéreltség fogalmát, majd ismertettem azt a két erőforrást, amelyet a dolgozat egészében felhasználtam. Az első az MNSZ 2.0.4 korpusz (Oravecz et al. 2014) módosított verziója, amelyből kiszűrtem a verseket, a duplumokat és az értelmetlen elemzés nélküli mondatokat. A másik a PREVLEX táblázat, amelyhez az első tézisem kapcsolódik:

1. Az MNSZ2 korpusz felhasználásával létrehoztam a PREVLEX táblázatot, amely az igekötős igék jelenleg legbővebb (53 535 szavas), manuálisan ellenőrzött, nyílt hozzáférésű táblázata. Részét képezik a hapaxok – egyszer előforduló szavak –, valamint azok a szavak is, amelyek a korpuszban UNKNOWN címkével jelennek meg. Az igekötős igék mellett a korpuszban mért tokengyakoriságuk is szerepel.

A 3. fejezetben az igekötő-állomány meghatározásának problémájával foglalkoztam. A fő kérdésem az volt, hogy mely lexikai elemeket soroljuk az igekötők közé, és mi alapján végezzük a besorolást? Abból a megfigyelésből indultam ki, hogy az igekötők és egyéb, pusztán névszói igemódosítók között nem húzható éles határ, ezért olyan modellre van szükség, amely ezt az átmenetiséget képes megragadni. Feltételeztem, hogy a prototípus-elmélet megfelel erre a célra. Ennek keretében olyan automatikusan mérhető, morfológiai és gyakorisági jegyeket vettem sorra, amelyek segíthetnek körülhatárolni az igekötők csoportját. A jegyek értékét korpuszméréssel határoztam meg 235 igekötőjelölt esetében. Az adatgyűjtés eredményéből Igekötő × Jegy mátrixokat hoztam létre,

amelyek elsősorban a jegyértékek ábrázolásában különböznek. A mátrixok alapján megvizsgáltam az egyes jegypárok közötti korrelációkat. Ennek eredménye, a kapcsolódó szakirodalmak, és végül tagadhatatlanul a saját intuícióm alapján a *meg*-et jelöltem ki prototipikus igekötőnek, majd az igekötőség tipikus jellemzőit a *meg* tulajdonságai alapján határoztam meg. Végül bemutattam három módszert, amelyek alkalmasak lehetnek az igekötők jegyalapú osztályozására. A módszerek összehasonlítása után Smith et al. (1988) módszere mellett döntöttem, amely alapján négy, leginkább a tárgyalást segítő kategóriát neveztem meg: a prototipikus, a centrális, a félperiférikus és a periférikus igekötőket. Főbb eredményeim a következők:

2. Megállapítottam és korpuszon kimértem 10 olyan jegyet, amellyel az igekötők jellemezhetőek – ugyanakkor jeleztem, hogy nem minden jegy egyformán fontos. A korpuszmérések eredményét nyílt hozzáférésű Igekötő × Jegy mátrixokban rögzítettem.
3. Az abszolút gyakoriságot tartalmazó Igekötő × Jegy mátrixon kiszámoltam az egyes jegypárok közötti korrelációkat. Ezáltal mennyiségi mutatókkal jellemeztem a grammatikalizáció folyamatát az igekötők esetében. A produktivitás erős pozitív korrelációban áll a gyakorisággal, míg a szótagszám és a szerkesztettség ezekkel negatív korrelációt mutat – a gyakori és produktív igekötők jellemzően rövidek és morfológiai szempontból bonthatatlanok. A korreláció-számítást bináris adaton is elvégeztem, és kimutattam, hogy a jegyek közti alapvető összefüggések nem változnak attól, ha a gyakoriságra vonatkozó információt elhagyjuk. Ezt azzal magyaráztam, hogy a gyakoriság történetileg olyan szorosan együtt jár más jegyekkel – éppen a grammatikalizációs folyamat révén –, hogy akkor is érezhető a hatása a vizsgált jegyekben, ha külön jegyként nem vesszük figyelembe.
4. Smith et al. (1988) módszerét alapul véve felállítottam egy tipikalitási skálát, amelynek egyik végpontján a prototipikus igekötők, másik végpontján a pusztán névszói igemódosítók helyezhetők el.

A 4., mondattani tárgyú fejezetben a célom a következő kérdés megválaszolása volt: Milyen mintázattípusokat mutatnak az igekötős szerkezetek, mikor és mennyire távolodhat el az igekötő az igétől, igenévtől, igei származéktól? Először egy szinkrón korpuszvizsgálatot mutattam be, amely az MNSZ2 korpusz 20–21. századi szövegein alapult. Főként az olyan szerkezetekre fókuszáltam,

amelyekben az igekötő elválhat az igétől, igenévtől, vagy ritkábban egyéb igei származéktól. Ezeket elsősorban az igekötők eloszlási mintái alapján jellemeztem. Ezután egy diakrón korpuszvizsgálat eredményeiről számoltam be, amely során a prototipikus igekötők mondatbeli pozícióját mértem ki a finit igehez viszonyítva az ómagyar kortól napjainkig terjedő szöveganyagon. A kutatásom főbb eredményei a következők:

5. Kimutattam, hogy a prototipikus igekötők relatív gyakoriságukat tekintve közel maradnak a finit igehez, míg a periférikusabb igekötők jobban eltávolodhatnak. Ehhez kapcsolódva megneveztem két olyan tényezőt, amely valószínűleg hatással van arra, hogy a hátravetett igekötő mennyire távolodik el a finit igétől. Az egyik a szöveg szerkesztettségi foka: szerkesztetlen – főleg szóban elhangzott – szövegben nagyobb annak a valószínűsége, hogy az igekötő távol kerül a finit igétől, mint gondozott, szerkesztett szöveg esetében. A másik az ige utáni mondatösszetevők, köztük a hátravetett igekötő fonológiai súlya. A méréseim alapján az a tendencia rajzolódik ki, amelyet a növekvő összetevők törvényének vagy Behaghel-törvénynek is neveznek, és É. Kiss (2007) szerint a magyarban a finit ige utáni mondatzszakaszra érvényes: a rövidebb összetevő megelőzi a hosszabbat, ha ezt szintaktikai szabály nem gátolja. Ez azzal is egybevág, hogy az egy szótagú, prototipikus igekötők ritkábban távolodnak el a finit igétől, mint a több szótagú, periférikusabb igekötők.
6. Korpuszméréssel igazoltam, hogy az infinitívusztól az igekötője preverbálisan messzire kerülhet, de csak akkor, ha segédigeszerű elem – elsősorban finit ige – férkőzik be a szerkezetbe, és az igekötő az ehhez tartozó igemódosítói pozíciót foglalja el. Ez alátámasztja azt a megfigyelést, hogy bár az igekötő az infinitívusszal alkot szótári egységet, szórendi és prozódiai szempontból szorosabban kapcsolódik a segédigeszerű elemhez (Kálmán C. et al. 1989). Megállapítottam, hogy ezzel rokon szintaktikai viselkedést mutat a létigével passzív szerkezetet alkotó határozói igenév is. Ennek az igekötője akkor kerül távoli preverbális helyzetbe, ha a létige előtti igemódosítói pozíciót foglalja el.
7. Megállapítottam, hogy a határozói igenév igekötője csak akkor vethető hátra, ha az igenév állapot- vagy módhatározói szerepet tölt be. A létigével alkotott passzív konstrukcióban az igekötő mindig preverbális.

8. A melléknévi igenevek kapcsán a következő megfigyeléseket tettem: (1) Ha -hAtÓ végződésű igenév állítmányi helyzetben van, és a tagmondatban nem hangzik el finit ige, akkor a -hAtÓ teljes mértékben átveszi ennek a helyét. Az igekötője ugyanúgy elválhat, mint egy finit igéé, az igekötők disztribúciója is hasonlóan alakul, emellett a szerkezetbe férkőző szavakat tekintve is párhuzam látszik a finit igei szerkezettel. Mindez az ellen szól, hogy a -hAtÓ-t melléknévképzőnek tekintjük – ahogy például Kiefer (2003a) teszi –, mivel az ezzel képzett szónak jellegzetes igei tulajdonságai vannak. (2) Az általános szakirodalmi véleménnyel szemben kimutattam, hogy az -AndÓ végződésű melléknévi igenevek állhatnak állítmányi helyzetben, és ekkor az igekötőjük is hátravethető, bár a jelenség kétségtelenül ritka (az esetek 1,85%-ára jellemző, ez a vizsgált korpuszban 1624 példát jelent).
9. Kimutattam, hogy a megszakított szórend nagyon elterjedt mintázattípus, amely számos deverbális főnév, melléknév és határozószó esetében is megfigyelhető. Ekkor a *nem*, *sem*, *se*, *is* elemek valamelyike férkőzik a szerkezetbe.
10. Megvizsgáltam azokat a szerkezeteket, amelyekben egy kötőmódú ige igekötője vagy egy kötőmódú ige non-finit igei bővítményének az igekötője egy finit modális ige – tipikusan a *kell* – elé kerül. A fontosabb megfigyeléseim a következők: (1) A *hogy* kötőszót tartalmazó és nem tartalmazó változatok gyakorisága közel áll egymáshoz, függetlenül attól, hogy az igekötő finit kötőmódú igéhez, infinitívushoz vagy határozói igenévhez tartozik. (2) A finit modális ige és a kötőszó közé kerülhetnek rövid elemek, a finit modális ige és az igekötő igéje között álló kifejezések pedig hasonlóságot mutatnak az infinitívuszi beférkőzéses szerkezetekkel.
11. Megvizsgáltam azokat a szerkezeteket, amelyekben egy igekötős ige infinitívuszi vagy határozói igenévi alakban topikalizálva van, és emellett finit igeként is megjelenik. Megállapítottam, hogy tagadó, illetve fókuszos mondatokban a hátravetett igekötő jelenléte opcionális. A topikalizációs szerkezeteken belül külön vizsgáltam az olyan elliptikus szerkezetek tulajdonságait, amelyekben az igekötő újabb előfordulását egy segédigeszerű elem követi.
12. A diakrón vizsgálat keretében kimutattam a hátravetett igekötőt tartalmazó, nem-semleges mondatok arányának növekedését az ómagyar kortól napjainkig. Ez a tendencia egyrészt

azzal magyarázható, hogy a fordított szórendű tagadás fokozatosan teret nyert az archaikus, megszakított szórendű tagadással szemben. Másrészt látható az is, hogy az ómagyar kor óta folyamatosan növekszik azoknak a szerkezeteknek az aránya, ahol a fókuszhasználat grammatikalizálódott.

Az 5. fejezet központi kérdése az volt, hogy hogyan írhatjuk le az igekötők produktív kapcsolódási mintáit, és ez alapján milyen megállapítások tehetők az igekötők szemantikájáról? Ennek megválaszolására olyan módszert dolgoztam ki, amely ‘igekötő – képző – vonzatkeret’ hármassok (pl. *el-X-elődik valami, tönkre-X-el valamit*) korpuszvezérelt vizsgálatára épül. Ismertettem az igealkotás három leggyakoribb módját: a névszói és az igei bemenetű igeképzést, valamint a hangzásséma-alapú igealkotást. Ezután bemutattam a PREVCONS adatbázist, amely 21 038 igekötős igei hapaxot tartalmaz, és a fentebb említett hármassok kereshetőségével lehetővé teszi a produktív kapcsolódási minták feltárását. Végül beszámoltam egy olyan kísérletemről, amelynek célja az igekötők jelentéseinek és a jelentések közti viszonyoknak a hálózatszerű ábrázolása volt a PREVCONS alapján, egy ontológia formájában. A fejezethez kapcsolódó téziseim a következők:

13. Saját algoritmust írtam a hangzássémára illeszkedő igék azonosítására, amely egyúttal sématípusokba is rendezte az érintett igéket. Ezáltal kimutattam, hogy bár a nyelvészeti szakirodalom kevés jelentőséget tulajdonít az igealkotás ezen módjának, a hangzássémás igék aránya nem elhanyagolható: ezek jelentik az igekötős igei hapaxok közel egytizedét (9,4%).
14. Kimutattam, hogy az új igekötős igék létrehozásában a névszói bemenetű igeképzés a legjelentősebb. Ez a hapaxok 35,2%-a esetében mutatható ki. Ugyanakkor mindössze 62 – az alakváltozatok összevonása után 56 – igekötő kapcsolódik így képzett igéhez.
15. Létrehoztam a PREVCONS adatbázist, amely az igekötős szerkezetek vizsgálatára alkalmas, nyílt hozzáférésű kutatóeszköz. 21 038 igekötős igei hapaxot tartalmaz a morfológiai szerkezetükre, vonzatkeretükre, szemantikájukra és szövegbeli előfordulásukra vonatkozó információval együtt.
16. Létrehoztam egy nyílt hozzáférésű ontológiát, amely 56 igekötő jelentéseit és a jelentéseik közti viszonyokat jeleníti meg. Entitásként szerepelnek benne az igekötők és a jelentések,

relációként pedig három hagyományos jelentésviszony, a szinonímia, az antonímia és a hiperonímia. Az ontológiát síkba rajzolható gráfként ábrázoltam.

A disszertáció végén, a 6. fejezetben visszatértem a kiinduló feltevéséhez, miszerint az igekötőnek nevezett szóosztály a prototípus-elmélet segítségével jól megragadható. Megvizsgáltam, hogy a főként morfológiai és gyakorisági jegyek alapján felállított igekötő-kontinuum nem változik-e meg lényegesen akkor, ha disztribúciós és szemantikai jegyeket is figyelembe veszünk. A konklúzió az volt, hogy a skála két végpontja csak minimálisan változik, de a köztes mezőben jelentős eltérések mutatkoznak. Az eredeti megközelítésben tapasztalt bizonytalanság elvezetett egy új szemlélethez, amely Kálmán László opponensi véleményéből indul ki, és a lényege az, hogy az egyes lexikai elemekről áthelyezi a fókusz az azokat tartalmazó szerkezetekre. Főbb eredményeim a következők:

17. Felvázoltam a szerkezetek felőli megközelítés koncepcióját, és rámutattam két előnyére a lexikai elemek felől induló, prototípus-elméleti keretet használó megközelítéshez képest: (1) Nem kell bevezetni olyan, nem szilárd empirikus alapon nyugvó döntéseket, mint amelyek korábban szükségesek voltak. (2) Azáltal, hogy előzetesen nem kategorizáljuk a lexikai elemeket, sokkal kisebb az információvesztés, illetve annak a veszélye, hogy túláltalánosítunk.
18. Létrehoztam a PREVDISTRO adatbázist, amely 49 szerkezettypus korpuszbeli előfordulását tartalmazza, minden előfordulás esetében feltüntetve külön az igekötőt és az igelemmát, az igekötő pozícióját az ige-tőhöz képest, valamint a szerkezetbe beférkőző egyéb szavakat. Emellett elérhető a szerkezetek nagyobb kontextusa – a teljes mondat – is. A 41,5 millió rekordból álló adatbázis nyílt hozzáférésű.

A felsorolt tézispontok alapján látható, hogy a munkám elméleti és gyakorlati szempontból is hozott új tudományos eredményeket. A gyakorlati haszna elsősorban a PREVLEX, a PREVMATRIX, a PREVCONS, a PREVONTO és a PREVDISTRO létrejötte, amelyek önmagukban is értékes, nyílt hozzáférésű erőforrások. Ezek áttekintését ld. a IV. számú mellékletben.

A kutatás elméleti téren számos olyan tendenciát, jelenséget mutatott ki, amelyek korpuszvezérelt módszer hiányában észrevétlenül vagy sejtés szintjén maradtak volna. Emellett felhívta a figyelmet olyan jelenségekre, amelyek egyáltalán nem ritkák, mégis csak kevés nyelvész érdeklődésére tarthattak eddig számot (ilyenek például a hangzássémákra illeszkedő igék).

Bízom abban, hogy az adatvezérelt szemlélet egyre inkább teret tud nyerni a nyelvtudományban, és a dolgozatomban leírt módszerek, ötletek hasznosnak bizonyulnak nemcsak az igekötős szerkezetek, hanem más nyelvi jelenségek vizsgálatában is.

Hivatkozások

- Ackerman, Farrell – LeSourd, Philip 1997. Toward a Lexical Representation of Phrasal Predicates. In: Alsina, Alex – Bresnan, Joan – Sells, Peter (szerk.): *Complex predicates*. CSLI Publications. Stanford. 67–106.
- Ackerman, Farrell 2003. Lexeme derivation and multi-word predicates in Hungarian. *Acta Linguistica Hungarica* 50: 7–32.
- Ackerman, Farrell – Kalivoda, Ágnes – Malouf, Robert 2021. *Paradigmatic organization as a solution to Zipfian distributions in Hungarian grammar*. Kézirat, megjelenés előtt.
- Alberti Gábor 2006. *A magyar mondattan elmélete és gyakorlata generatív megközelítésben III. A háttérelmélet*. Pécsi Tudományegyetem. Pécs.
- Baayen, Harald 1989. *A Corpus-Based Approach to Morphological Productivity (Statistical Analysis and Psycholinguistic Interpretation)*. Doktori értekezés. Centrum voor Wiskunde en Informatica. Amsterdam.
- Baayen, Harald 2009. Corpus linguistics in morphology: morphological productivity. In: Lüdeling, Anke – Kytö, Merja (szerk.): *Corpus Linguistics. An international handbook, Volume 2*. Mouton De Gruyter. Berlin. 900–919.
- Behaghel, Otto 1932. *Deutsche Syntax IV*. Carl Winters. Heidelberg.
- Benő Attila – Szilágyi N. Sándor 2015. Hangzásséma és motiváltság a hangutánzó és hangulatfestő igéink körében. In: Kádár Edit – Szilágyi N. Sándor (szerk.): *Motiváltság és nyelvi ikonicitás*. Erdélyi Múzeum-Egyesület (EME). Kolozsvár. 43–57.
- Bentz, Christian – Ferrer-i Cancho, Ramon 2016. Zipf’s law of abbreviation as a language universal. In: Bentz, Christian – Jäger, Gerhard – Yanovich, Igor (szerk.): *Proceedings of the Leiden Workshop on Capturing Phylogenetic Algorithms for Linguistics*. University of Tübingen, online publication system. Tübingen.

- Bhatia, Archana – Teng, Choh Man – Allen, James F. 2018. Identifying senses of particles in verb-particle constructions. In: Markantonatou, Stella – Ramisch, Carlos – Savary, Agata – Vincze, Veronika (szerk.): *Multiword expressions at length and in depth: Extended papers from the MWE 2017 workshop*. Language Science Press. Berlin. 61–86.
- Cooper, William E. – Ross, John Robert 1975. World order. In: Grossman, Robin E. – San, James L. – Vance, Timothy J. (szerk.): *Papers from the Parasession on Functionalism*. Chicago Linguistic Society. Chicago. 63–111.
- Croft, William 2001. *Radical Construction Grammar*. Oxford University Press. Oxford.
- Cser András 2017. *Morfológia*. Pázmány Péter Katolikus Egyetem Bölcsész- és Társadalomtudományi Kar. Budapest – Piliscsaba.
- Dér Csilla Ilona 2008. *Grammatikalizáció*. Nyelvtudományi Értekezések 158. Akadémiai Kiadó. Budapest.
- Dékány Éva 2014. A nem véges alárendelés (az igenevek) története. In: É. Kiss Katalin (szerk.): *Magyar generatív történeti mondattan*. Akadémiai Kiadó. Budapest. 177–238.
- Dékány Éva – Hegedűs Veronika 2015. Word order variation in Hungarian PPs. In: É. Kiss Katalin – Surányi Balázs – Dékány Éva (szerk.): *Approaches to Hungarian 14: Papers from the Piliscsaba conference*. John Benjamins Publishing Company. Amsterdam. 95–120.
- Diessel, Holger 2019. *The Grammar Network: How Linguistic Structure Is Shaped by Language Use*. Cambridge University Press. Cambridge.
- D. Máta Mária 1989. *Igekötőrendszerünk történetéből*. A Magyar Nyelvtudományi Társaság Kiadványai 187. Magyar Nyelvtudományi Társaság. Budapest.
- Dóra Zoltán 2006. Kell mennem, kell menjek? *Magyar nyelvőr* 130/4: 413–421.
- Dömötör Adrienne – Gugán Katalin – Novák Attila – Varga Mónika 2017. Kiútkeresés a morfológiai labirintusból – korpuszépítés ó- és középmagyar kori magánéleti szövegekből. *Nyelvtudományi Közlemények* 113: 85–110.

- Dressler, Wolfgang U. 2003. Degrees of grammatical productivity in inflectional morphology. *Rivista di Linguistica (Italian Journal of Linguistics)* 15/1: 31–62.
- É. Kiss Katalin 1998. Verbal prefixes or postpositions? Postpositional aspectualizers in Hungarian. In: de Groot, Caspar – Kenesei, István (szerk.): *Approaches to Hungarian 6*. JATE. Szeged. 123–148.
- É. Kiss Katalin 2002. *The Syntax of Hungarian*. Cambridge University Press. Cambridge.
- É. Kiss Katalin 2003. Az egyszerű mondat szerkezete. In: É. Kiss Katalin – Siptár Péter – Kiefer Ferenc (szerk.): *Új magyar nyelvtan*. Osiris. Budapest. 74–160.
- É. Kiss Katalin 2006a. The Function and the Syntax of the Verbal Particle. In: É. Kiss Katalin (szerk.): *Event Structure and the Left Periphery. Studies on Hungarian*. Springer. Dordrecht. 17–56.
- É. Kiss Katalin 2006b. From the Grammaticalization of Viewpoint Aspect to the Grammaticalization of Situation Aspect. In: É. Kiss Katalin (szerk.): *Event Structure and the Left Periphery. Studies on Hungarian*. Springer. Dordrecht. 129–157.
- É. Kiss Katalin 2007. Az ige utáni szabad szórend magyarázata. *Nyelvtudományi Közlemények* 104: 124–152.
- É. Kiss Katalin 2009. Nekem el kell menni/el kell mennem/el kell, hogy menjek/el kell menjek/el kellek menni. In: É. Kiss Katalin – Hegedűs Attila (szerk.): *Nyelvelmélet és dialektológia*. PPKE BTK. Piliscsaba. 213–227.
- É. Kiss Katalin 2014a. Bevezetés. In: É. Kiss Katalin (szerk.): *Magyar generatív történeti mondattan*. Akadémiai Kiadó. Budapest. 11–13.
- É. Kiss Katalin 2014b. A tagadó és a kérdő mondatok változásai. In: É. Kiss Katalin (szerk.): *Magyar generatív történeti mondattan*. Akadémiai Kiadó. Budapest. 34–49.
- É. Kiss Katalin 2014c. Az ómagyar igeidőrendszer. In: É. Kiss Katalin (szerk.): *Magyar generatív történeti mondattan*. Akadémiai Kiadó. Budapest. 60–72.

- É. Kiss Katalin 2021. Predicative PPs. In: É. Kiss, Katalin – Hegedűs, Veronika (szerk.): *Syntax of Hungarian: Postpositions and Postpositional Phrases*. Amsterdam University Press. Amsterdam. 251–284.
- Forgács Tamás 2005. Grammatikalizálódás az igekötők körében. In: Oszkó Beatrix – Sipos Mária (szerk.): *Uráli grammatizáló*. MTA Nyelvtudományi Intézet. Budapest. 88–116.
- Francis, Gill 1993. A corpus-driven approach to grammar: Principles, methods and examples. In: Baker, Mona – Francis, Gill – Tognini-Bonelli, Elena (szerk.): *Text and Technology*. John Benjamins Publishing Company. Amsterdam. 137–156.
- Goody, Jack 1987. *The Interface Between the Written and the Oral*. Cambridge University Press. Cambridge.
- Gries, Stefan Th. 2013. Data in Construction Grammar. In: Hoffmann, Thomas – Trousdale, Graeme (szerk.): *The Oxford Handbook of Construction Grammar*. Oxford University Press. Oxford. 93–109.
- Gugán Katalin 2015. És mégis: mozog? Tagadás és igemódosítók az ómagyarban és a középmagyarban. *Általános Nyelvészeti Tanulmányok* 27: 153–178.
- Gugán Katalin 2017. A magyar tagadó mondatok szórendje és a konstansráta-hipotézis. In: É. Kiss Katalin – Hegedűs Attila – Pintér Lilla (szerk.): *Nyelvelmélet és diakrónia 3*. Pázmány Péter Katolikus Egyetem BTK, Szt. István Társulat. Budapest – Piliscsaba. 91–110.
- Gyuris Beáta 2002. *The semantics of contrastive topics in Hungarian*. Doktori értekezés. Eötvös Loránd Tudományegyetem. Budapest.
- Hegedűs Veronika 2013. Non-verbal predicates and predicate movement. LOT Dissertation Series 337. LOT. Utrecht. https://www.lotpublications.nl/Documents/337_fulltext.pdf
- Hegedűs Veronika 2020. Back to restitutives (again): A syntactic account of restitutive and counterdirectional verbal particles in Hungarian. *Acta Linguistica Academica* 67/3: 319–345.
<https://doi.org/10.1556/2062.2020.00017>

- Honti László – H. Varga Márta 2006. Meg van írva! A határozói igenév és a létige alkotta szerkezet funkciójáról és háttéréről. In: Mártonfi Attila – Papp Kornélia – Slíz Mariann (szerk.): *101 írás Pusztai Ferenc tiszteletére*. Argumentum. Budapest. 579–586.
- H. Varga Márta 2015. A ‘passzív’ jelentés és nyelvi kifejezőeszközei a magyarban. *Folia Uralica Debreceniensia* 22: 293–310.
- Indig Balázs – Sass Bálint – Simon Eszter – Mittelholcz Iván – Kundráth Péter – Vadász Noémi 2019. emtsv – Egy formátum mind felett. In: Berend Gábor – Gosztolya Gábor – Vincze Veronika (szerk.): *XV. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2019)* Szegedi Tudományegyetem, TTIK, Informatikai Intézet. Szeged. 235–247.
- Ittész Nóra 2009. A magyar nyelv nagyszótára. In: Fábíán Zsuzsanna (szerk.): *Szótárírás és szótárírók*. Lexikográfiai füzetek 4. Akadémiai Kiadó. Budapest. 65–80.
- Jakab István 1976. *A magyar igekötők állományi vizsgálata*. Nyelvtudományi Értekezések 91. Akadémiai Kiadó. Budapest.
- Jakab István 1982. *A magyar igekötő szófajtani útja*. Nyelvtudományi Értekezések 112. Akadémiai Kiadó. Budapest.
- Jakab Zoltán 2014. Jelentés: Referencia és fogalmak. In: Pléh Csaba – Lukács Ágnes (szerk.): *Pszicholingvisztika, 2. kötet*. Akadémiai Kiadó. Budapest. 659–742.
- Jakubiček, Miloš – Kilgarriff, Adam – Kovář, Vojtěch – Rychlý, Pavel – Suchomel, Vít 2013. The TenTen Corpus Family. *7th International Corpus Linguistics Conference CL 2013*. Online publikáció. Lancaster. 125–127.
- J. Soltész Katalin 1959. *Az ősi magyar igekötők: meg, el, ki, be, fel, le*. Akadémiai Kiadó. Budapest.
- Kádár Edit 2007. Az ige–igekötő szórend csángó beszélt nyelvi szövegekben. In: Maticsák Sándor – Jankovics József – Kolláth Anna – Nyerges Judit – Péntek János (szerk.): *Nyelv, nemzet, identitás. A VI. Nemzetközi Hungarológiai Kongresszus nyelvészeti előadásai*. Nemzetközi Magyarságtudományi Társaság. Budapest – Debrecen. 61–80.

- Kalivoda Ágnes 2016. *A magyar igei komplexumok vizsgálata*. Mesterszakos szakdolgozat. Pázmány Péter Katolikus Egyetem. Budapest.
- Kalivoda Ágnes 2017. Az igekötők gépi annotálásának problémái. In: Ludányi Zsófia (szerk.): *Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből 2017: XI. Alkalmazott Nyelvészeti Doktoranduszkonferencia*. MTA Nyelvtudományi Intézet. Budapest. 100–109.
- Kalivoda Ágnes 2018. Az igekötős igék szintaxisa korpuszvezérelt megközelítésben. In: Scheibl György (szerk.): *Lingdok 17.: Nyelvészdoktoranduszok dolgozatai*. Szegedi Tudományegyetem, Nyelvtudományi Doktori Iskola. Szeged. 159–176.
- Kalivoda Ágnes 2019. Véges erőforrás végtelen sok igekötős igére. In: Berend Gábor – Gosztolya Gábor – Vincze Veronika (szerk.): *XV. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2019)*. Szegedi Tudományegyetem, TTIK, Informatikai Intézet. Szeged. 331–344.
- Kalivoda Ágnes 2021. Az igekötők produktív kapcsolódási mintái. *Argumentum* 17: 56–82.
<https://doi.org/10.34103/ARGUMENTUM/2021/4>
- Kalivoda Ágnes – Vadász Noémi – Indig Balázs 2018. MANÓCSKA: A Unified Verb Frame Database for Hungarian. In: Sojka, Petr – Horák, Aleš – Kopeček, Ivan – Pala, Karel (szerk.): *Proceedings of the 21st International Conference on Text, Speech and Dialogue (TSD)*. Springer-Verlag. Brno. 135–143.
- Kálmán C. György – Kálmán László – Nádasy Ádám – Prószekey Gábor 1989. A magyar segédigék rendszere. *Általános Nyelvészeti Tanulmányok* XVII: 49–103.
- Kálmán László – Nádasy Ádám 1994. A hangsúly. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 2. Fonológia*. Akadémiai Kiadó. Budapest. 393–467.
- Kálmán László – Trón Viktor 2000. A magyar igekötő egyeztetése. In: Büky László – Maleczki Márta (szerk.): *A mai magyar nyelv leírásának újabb módszerei IV*. Szegedi Tudományegyetem. Szeged. 203–211.
- Kálmán László (szerk.) 2001. *Magyar leíró nyelvtan. Mondattan I*. Tinta Könyvkiadó. Budapest.

- Kálmán László 2013. *Egyre eljebb terjed*. Online publikáció a Nyelv és Tudomány ismeretterjesztő hírportálon. <https://www.nyest.hu/hirek/egyre-eljebb-terjed>
- Kálmán László 2016. Bővítménykeretek mint konstrukciók. In: Kas Bence (szerk.): „*Szavad ne feledd!*” *Tanulmányok Bánréti Zoltán tiszteletére*. MTA Nyelvtudományi Intézet. Budapest. 61–72.
- Kálmán László 2020. A nyelvészet és a precedensjog. *Qubit* 2020/09/19. Online publikáció. <https://qubit.hu/2020/09/19/a-nyelveszet-es-a-precedensjog>
- Kassai Ilona 1994. A fonetikai háttér. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 2. Fonológia*. Akadémiai Kiadó. Budapest. 581–666.
- Kemény Gábor 2018. Nem elváló igekötők. *Édes anyanyelvünk* 40/1: 7.
- Kenesei István 2000. Szavak, szófajok, toldalékok. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 3. Morfológia*. Akadémiai Kiadó. Budapest. 75–136.
- Kerekes Judit 2011. Az igekötők meghatározásának problémái. In: Gécszeg Zsuzsanna (szerk.): *LingDok10: Nyelvész-doktoranduszok dolgozatai*. Szegedi Tudományegyetem Nyelvtudományi Doktori Iskola. Szeged. 109–131.
- Kiefer Ferenc 1992. Az aspektus és a mondat szerkezete. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 1. Mondattan*. Akadémiai Kiadó. Budapest. 729–809.
- Kiefer Ferenc 1995. Prefix reduplication in Hungarian. *Acta Linguistica Hungarica* 43/1–2: 175–194.
- Kiefer Ferenc – Ladányi Mária 2000a. Az igekötők. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 3. Morfológia*. Akadémiai Kiadó. Budapest. 453–518.
- Kiefer Ferenc – Ladányi Mária 2000b. A szóképzés. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 3. Morfológia*. Akadémiai Kiadó. Budapest. 137–164.
- Kiefer Ferenc – Ladányi Mária 2000c. Morfoszintaktikailag semleges képzések. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 3. Morfológia*. Akadémiai Kiadó. Budapest. 165–214.

- Kiefer Ferenc 2003a. Alaktan. In: É. Kiss Katalin – Siptár Péter – Kiefer Ferenc (szerk.): *Új magyar nyelvtan*. Osiris. Budapest. 127–199.
- Kiefer Ferenc 2003b. A kétféle igemódosítóról. *Nyelvtudományi közlemények* 100: 177–186.
- Kiefer Ferenc 2006. *Aspektus és akcióminőség különös tekintettel a magyar nyelvre*. Akadémiai Kiadó. Budapest.
- Kiefer Ferenc 2007. *Jelentéselmélet*. 2. kiadás. Corvina. Budapest.
- Kiefer Ferenc – Németh Boglárka 2012. Amikor az igekötő nem telicizál. *Nyelvtudományi közlemények* 108: 267–273.
- Komlósy András 1992. Régensek és vonzatok. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 1. Mondattan*. Akadémiai Kiadó. Budapest. 299–527.
- Komlósy András 2000. A műveltetés. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 3. Morfológia*. Akadémiai Kiadó. Budapest. 215–292.
- Kornai András – Nemeskey Dávid Márk – Recski Gábor 2016. Detecting Optional Arguments of Verbs. In: Calzolari, Nicoletta – Choukri, Khalid – Declerck, Thierry – Goggi, Sara – Grobelnik, Marko – Maegaard, Bente – Mariani, Joseph – Mazo, Hélène – Moreno, Asunción – Odijk, Jan – Piperidis, Stelios (szerk.): *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. European Language Resources Association (ELRA). Portorož. 2815–2818.
- Laczkó Tibor 2000a. Az ige argumentumszerkezetét megőrző főnévképzés. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 3. Morfológia*. Akadémiai Kiadó. Budapest. 293–407.
- Laczkó Tibor 2000b. A melléknévi és határozói igenévképzők. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 3. Morfológia*. Akadémiai Kiadó. Budapest. 409–452.
- Ladányi Mária 2000. Productivity as a sign of category change: The case of Hungarian verbal prefixes. In: Dressler, Wolfgang U. – Pfeiffer, Oskar E. – Pöchtrager, Markus A. – Rennison, John

- R. (szerk.): *Morphological Analysis in Comparison*. John Benjamins Publishing Company. Amsterdam. 113–142.
- Ladányi Mária 2007. *Produktivitás és analógia a szóképzésben: elvek és esetek*. Tinta Könyvkiadó. Budapest.
- Ladányi Mária 2012. Igeközös igék kapcsolódási mintázatai. Vizsgálati lehetőségek. In: Tolcsvai Nagy Gábor – Tátrai Szilárd (szerk.): *Konstrukció és jelentés. Tanulmányok a magyar nyelv funkcionális kognitív leírására*. Eötvös Loránd Tudományegyetem. Budapest. 71–84.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar; 2. kötet*. Stanford University Press. Palo Alto, Kalifornia, USA.
- Langacker, Ronald W. 1991. *Concept, Image, and Symbol: The Cognitive Basis of Grammar*. Mouton de Gruyter. Berlin.
- McInnes, Leland – Healy, John – Astels, Steve 2017. hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software* 2/11: 205.
- McShane, Marjorie – Beale, Stephen – Nirenburg, Sergei 2005. *Disambiguating Homographous Prepositions and Verbal Particles In An Implemented Ontological Semantic Analyzer*. Technical Report 01-05. University of Maryland Baltimore County, Institute for Language and Information Technologies. Baltimore, USA.
- Nádasdy Ádám 2001. A két primadonna. *Magyar Narancs, Modern Talking* 2001/42. 10. 18. 15. Online publikáció. https://magyarnarancs.hu/egotripp/nadasdy_adam_modern_talking-55638
- Nádasdy Ádám 2005. A magyar folyamatos. *Magyar Narancs, Modern Talking* 2005/50. 12. 15. Online publikáció. https://magyarnarancs.hu/egotripp/modern_talking-64968
- Nemeskey Dávid Márk 2020. *Natural Language Processing methods for Language Modeling*. Doktori értekezés. Eötvös Loránd Tudományegyetem. Budapest.
- Németh Dezső 2001. A munkamemória szerepe a mondatmegértésben. In: Pléh Csaba – Lukács Ágnes (szerk.): *A magyar morfológia pszicholingvisztikája*. BIP – Osiris. Budapest. 83–118.

- Novák Attila – Rebrus Péter – Ludányi Zsófia 2017. Az emMorph morfológiai elemző annotációs formalizmusa. In: Vincze Veronika (szerk.): *XIII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2017)*. Szegedi Tudományegyetem Informatikai Intézet. Szeged. 70–78.
- Novák Attila – Siklósi Borbála – Oravecz Csaba 2016. A New Integrated Open-source Morphological Analyzer for Hungarian. In: Calzolari, Nicoletta – Choukri, Khalid – Declerck, Thierry – Goggi, Sara – Grobelnik, Marko – Maegaard, Bente – Mariani, Joseph – Mazo, Hélène – Moreno, Asunción – Odijk, Jan – Piperidis, Stelios (szerk.): *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. European Language Resources Association (ELRA). Portorož. 1315–1322.
- Nyíri Antal – Bodnár Ferenc – Pálfalvi Etelka – Rácz Endre – Velcsov Mártonné – Végh József Mihály (szerk.) 1971. *A Müncheneri Kódex 1466-ból. Kritikai szövegkiadás a latin megfelelővel együtt*. Codices Hungarici 7. Akadémiai Kiadó. Budapest.
- O’Connor, Michael Patrick 1978. *Hebrew Verse Structure, 1. kötet*. Eisenbrauns. Indiana, Winona Lake, USA.
- Oravecz Csaba – Váradi Tamás – Sass Bálint 2014. The Hungarian Gigaword Corpus. In: Calzolari, Nicoletta – Choukri, Khalid – Declerck, Thierry – Loftsson, Hrafn – Maegaard, Bente – Mariani, Joseph – Moreno, Asuncion – Odijk, Jan – Piperidis, Stelios (szerk.): *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*. European Language Resources Association (ELRA). Reykjavik. 1719–1723.
- Pakerys, Jurgis 2017. *Measuring morphological productivity*. Kiosztvány. Graduate School of Linguistics, Philosophy and Semiotics (GSLPS). Tartu. <https://tinyurl.com/y38xsxmm>
- Pedregosa, Fabian – Varoquaux, Gaël – Gramfort, Alexandre – Michel, Vincent – Thirion, Bertrand – Grisel, Olivier – Blondel, Mathieu – Müller, Andreas – Nothman, Joel – Louppe, Gilles – Prettenhofer, Peter – Weiss, Ron – Dubourg, Vincent – Vanderplas, Jake – Passos, Alexandre – Cournapeau, David – Brucher, Matthieu – Perrot, Matthieu – Duchesnay, Édouard 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12: 2825–2830.

- Peredy Márta 2011. Az ige–igekötő sorrend a Jókai-kódexben. In: Bakró-Nagy Marianne – Forgács Tamás (szerk.): *A nyelvtörténeti kutatások újabb eredményei VI.* Szegedi Tudományegyetem Magyar Nyelvészeti Tanszék. Szeged. 181–197.
- Piñón, Christopher 1991. *Falling in Paradise: Verbs, Preverbs and Reduplication in Hungarian.* Kiosztvány, Stanford University, 1991/05/21. <http://pinon.sdf-eu.org/covers/fp.html>
- Pléh Csaba 1998. *A mondatmegértés a magyar nyelvben. Pszicholingvisztikai kísérletek és modellek.* Osiris Kiadó. Budapest.
- Posgay Ildikó 2002. Kell tanítsuk? In: Balázs Géza – Adamikné Jászó Anna – Koltói Gábor (szerk.): *Éltető anyanyelvünk. Írások Grétsy László 70. születésnapjára.* Tinta Kiadó. Budapest. 392–395.
- Prószéky Gábor – Tihanyi László – Ugray Gábor 2004. Moose: a robust high-performance parser and generator. In: Hutchins, John (szerk.): *Proceedings of the 9th EAMT Conference.* Foundation for International Studies. La Valletta. 138–142.
- Rákosi György – Laczkó Tibor 2011. Inflecting spatial particles and shadows of the past in Hungarian. In: Butt, Miriam – King, Tracy Holloway (szerk.): *The proceedings of the LFG11 Conference.* CSLI Publications. Stanford. 440–460.
- Rákosi György 2014. A case of disagreement: On plural reduplicating particles in Hungarian. In: Kertész András – Rákosi Csilla (szerk.): *The Evidential Basis of Linguistic Argumentation.* Studies in Language Companion Series 153. John Benjamins Publishing Company. Amsterdam. 179–198.
- Rebrus Péter 2000. Morfofonológiai jelenségek. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 3. Morfológia.* Akadémiai Kiadó. Budapest. 763–948.
- Rosch, Eleanor H. 1973. Natural categories. *Cognitive Psychology* 4/3: 328–350.
- Sárik Pál 1998. A határozói igenevek néhány problémája. *Magyar Nyelv* 94/4: 423–435.
- Sass Bálint 2011. *Igei szerkezetek gyakorisági szótára – egy automatikus lexikai kinyerő eljárás és alkalmazása.* Doktori értekezés. Pázmány Péter Katolikus Egyetem. Budapest.

- Sass Bálint 2015. 28 millió szintaktikailag elemzett mondat és 500 000 igei szerkezet. In: Tanács Attila – Varga Viktor – Vincze Veronika (szerk.): *XI. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2015)*. Szegedi Tudományegyetem Informatikai Tanszékcsoport. Szeged. 399–403.
- Sass Bálint – Váradi Tamás – Pajzs Júlia – Kiss Margit 2010. *Magyar igei szerkezetek – A leggyakoribb vonzatok és szókapcsolatok szótára*. Tinta Könyvkiadó. Budapest.
- Siklósi Borbála – Novák Attila 2016. Beágyazási modellek alkalmazása lexikai kategorizációs feladatokra. In: Tanács Attila – Varga Viktor – Vincze Veronika (szerk.): *XII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2016)*. Szegedi Tudományegyetem, TTIK, Informatikai Intézet. Szeged. 3–14.
- Simon Eszter – Sass Bálint 2012. Nyelvtchnológia és kulturális örökség, avagy korpuszépítés ómagyar kódexekből. *Általános Nyelvészeti Tanulmányok* 24: 243–264.
- Siptár Péter 1994. A mássalhangzók. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 2. Fonológia*. Akadémiai Kiadó. Budapest. 183–272.
- Smith, Edward E. – Osherson, Daniel N. – Rips, Lance J. – Keane, Margaret 1988. Combining prototypes: A selective modification model. *Cognitive Science* 12: 485–527.
- Stefanowitsch, Anatol – Gries, Stefan Th. 2003. Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8/2: 209–243.
- Stubbs, Michael 2013. Sequence and order. The neo-Firthian tradition of corpus semantics. In: Hasselgård, Hilde – Ebeling, Jarle – Ebeling, Signe Oksefjell (szerk.): *Corpus Perspectives on Patterns of Lexis*. Studies in Corpus Linguistics 57. John Benjamins Publishing Company. Amsterdam/Philadelphia. 13–34.
- Surányi Balázs 2006. PP-hasadás és láncredukció. In: Kálmán László (szerk.): *KB 120: A titkos kötet: Nyelvészeti tanulmányok Bánréti Zoltán és Komlósy András tiszteletére*. Tinta Könyvkiadó. Budapest. 277–296.

- Surányi Balázs 2009a. Adpositional preverbs, chain reduction, and phases. In: den Dikken, Marcel – Vago, Robert M. (szerk.): *Approaches to Hungarian 11*. John Benjamins Publishing Company. Amsterdam. 217–250.
- Surányi Balázs 2009b. „Incorporated” locative adverbials in Hungarian. In: É. Kiss Katalin (szerk.): *Adverbs and Adverbial Adjuncts at the Interfaces*. De Gruyter. Berlin. 37–74.
- Surányi Balázs 2009c. Verbal particles inside and outside vP. *Acta Linguistica Hungarica* 56: 201–249.
- Szabolcsi Anna – Laczkó Tibor 1992. A főnévi csoport szerkezete. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 1. Mondattan*. Akadémiai Kiadó. Budapest. 179–298.
- Szalontai Ádám – Surányi Balázs 2020. Word order effects of givenness in Hungarian: syntax or prosody? In: Hegedűs Veronika – Vogel, Irene (szerk.): *Approaches to Hungarian 16. Papers from the 2017 Budapest Conference*. John Benjamins Publishing Company. Amsterdam/Philadelphia. 137–164.
- Tan, Pang-Ning – Steinbach, Michael – Karpatne, Anuj – Kumar, Vipin 2018. Cluster Analysis: Basic Concepts and Algorithms. In: Tan, Pang-Ning – Steinbach, Michael – Karpatne, Anuj – Kumar, Vipin (szerk.): *Introduction to Data Mining*. 2. kiadás. Pearson Education. London. 525–612.
- Teubert, Wolfgang 2005. My version of corpus linguistics. *International Journal of Corpus Linguistics* 10/1: 1–13.
- Tolcsvai Nagy Gábor 2015. *Az ige a magyar nyelvben. Funkcionális elemzés*. TINTA Könyvkiadó. Budapest.
- Törkenczy Miklós 1994. A szótag. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 2. Fonológia*. Akadémiai Kiadó. Budapest. 273–392.
- Váradai Tamás – Simon Eszter – Sass Bálint – Geröcs Mátyás – Mittelholcz Iván – Novák Attila – Indig Balázs – Prószéky Gábor – Farkas Richárd – Vincze Veronika 2017. Az e-magyar digitális

nyelvfeldolgozó rendszer. In: Vincze Veronika (szerk.): *XIII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2017)*. Szegedi Tudományegyetem, Informatikai Intézet. Szeged. 49–60.

Váradi Tamás – Simon Eszter – Sass Bálint – Mittelholcz Iván – Novák Attila – Indig Balázs 2018. E-magyar – A Digital Language Processing System. In: Calzolari, Nicoletta – Choukri, Khalid – Cieri, Christopher – Declerck, Thierry – Goggi, Sara – Hasida, Koiti – Isahara, Hitoshi – Maegaard, Bente – Mariani, Joseph – Mazo, Hélène – Moreno, Asuncion – Odijk, Jan – Piperidis, Stelios – Tokunaga, Takenobu (szerk.): *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA). Miyazaki. 1307–1312.

Varga László 1981. A topicról és a fókusz utáni elemek sorrendjéről (egy szintaktikai modell kapcsán). *Magyar nyelv* 77/2: 198–200.

Yeo, In-Kwon – Johnson, Richard A. 2000. A new family of power transformations to improve normality or symmetry. *Biometrika* 87/4: 954–959.

Zipf, George Kingsley 1932. *Selected studies of the principle of relative frequency in language*. Harvard University Press. Cambridge, Massachusetts, USA.

Zsibrita János – Vincze Veronika – Farkas Richárd 2013. magyarlanc: A Toolkit for Morphological and Dependency Parsing of Hungarian. In: Angelova, Galia – Bontcheva, Kalina – Mitkov, Ruslan (szerk.): *Recent Advances in Natural Language Processing (RANLP 2013)*. INCOMA Ltd. Sumen. 763–771.

Mellékletek

I. számú melléklet: Tipikalitási skála Smith et al. (1988) módszere alapján

6,2	be el fel ki le meg
6,1	át rá szét
5,8	haza ide oda össze vissza
5,7	alá bele hozzá neki
5,6	elő
5,4	elé körül tova túl
5,3	odább tele
5,1	széjjel szerte
5,0	által alul felül hátra külön mellé
4,9	benn ott
4,8	félre körbe közbe végig
4,6	föle utána
4,5	agyon helyre
4,3	előre ketté tönkre
4,2	egybe fejbe félbe fenn közre szembe tovább újjá újra
4,1	közé
4,0	alább bent kinn
3,8	keresztbe keresztül utol
3,6	hanyatt helyt kívül
3,5	együtt ellen nyilván rajta viszont
3,3	hátba
3,2	éhen helyben jót kölesön rendbe szörnyet véget
3,1	jól részt útra zsebre
3,0	abba
2,8	ágyba arcul békén célba csödbe csúcsra égbe észhez falra fejen férjhez főbe földhöz földre fülön füstbe hadba harcba hasba hasra házhoz helybe kezet kézre lábba mellbe nagyot nagyra nyakon partra pofán pofon porba rendre rosszul seggbe seggre sorba sorra szájon számba szá- mot szárnyra szarrá szemem szemet szénné talpra tarkón térdre tökön torkon tűzbe útba vízre
2,7	cserben egyet ellent észre jóvá karban közben közzé létre szemre szóba szót végbe véghez végre zokon
2,6	szárba
2,4	ajtot életben életre hegyet igazat irányt lábra lángba mennybe oldalba padlót piacra ringbe szemmel talpon törvényt utat vérig világgá világra zavarba
2,3	állást bérbe élen görcsbe gúzsba hajba kárba készen kétségbe kézben kézen kézhez kupán lángra latba lóvá nyélbe nyomon nyomra pórul rabul romba síkra sorban számon szemben színre szóvá tánra testet testre tetten tisztán tollba törbe trónra
2,0	lényegre témába tükörbe
1,9	bosszút csonttá földet hadat hatályba küszöbön kútba részre sárba semmibe valóra
1,5	érvénybe

II. számú melléklet: Produktív igeképzés gyakorisági adatokkal és példákkal

igeképzők és hangzás-sémák				
igeképzők és hangzás-sémák	névszói bemenet	-Vz(ik)	2558	<i>akármiz, puskázik, rántottesirkéz ...</i>
		-(s)Odik	638	<i>algásodik, britesedik, szótlanodik ...</i>
		-VI	525	<i>homokzsákol, klikkel, loginol ...</i>
		-(V)(s)ít	420	<i>esősít, titkosít, bozostásít ...</i>
		-ál	303	<i>dokumentál, konzervál, limitál ...</i>
		-(s)UI	205	<i>ritkul, hülyül, nácul ...</i>
		-(s)kOdik	180	<i>okoskodik, ökörködik, anyáskodik ...</i>
		-izál	109	<i>hellenizál, romantizál, moralizál ...</i>
		-íroz	39	<i>szortíroz, kopíroz, bázíroz ...</i>
	-ifikál	7	<i>szimplifikál, notifikál, pontifikál ...</i>	
	igei bemenet	-(t)At	1412	<i>cumiztat, pattogtat, véset ...</i>
		-Ódik	851	<i>érlelődik, shiftelődik, szövődik ...</i>
		-(O)gAt	510	<i>fűzőget, bütykölget, koktélozgat ...</i>
		-(t)Atik	178	<i>hívatik, guríttatik, bontatik ...</i>
	hangzás-séma	-Og	318	<i>mekeg, szörcsög, bugyborog ...</i>
		-Ant	147	<i>cuppant, fröccsent, pislant ...</i>
		-An	141	<i>csosszan, surran, fortyan ...</i>
		-Ad	105	<i>csoffad, görnyed, suvad ...</i>
		-Aszt	83	<i>lottyaszt, sarjaszt, ripeszt ...</i>
-int		41	<i>pöccint, hörpint, köhint ...</i>	

III. számú melléklet: A PREVDISTRO adatbázisban szereplő szerkezettypusok

szerkezet	db	preverbális igekötő	db	prefixált igekötő	db	posztv. ik.	db
finit ige	21 531 273	meg sem szólal	350 034	közbelép	12 658 242	edd csak meg	8 522 997
-Ás	7 884 148	egyet nem értés	5 361	végiggondolás	7 878 787	-	-
-(O)(t)t	3 066 219	be nem tartott	21 863	elhamarkodott	3 044 356	-	-
-ni₁	2 578 906	meg is tenni	8 151	előjönni	2 563 605	zární le	7 150
-Ó	1 836 875	össze nem függő	9 082	megindító	1 827 793	-	-
-ni₂	976 059	tönkre akarja tenni	975 083	-	-	jönni akar ki	976
-hAtÓ	645 051	le is emelhető	9 279	megvalósítható	508 962	tölthetők le	126 810
-vA₁	564 920	ki nem mondva	2 318	visszatekintve	552 261	szoktatva át	10 341
-(A)tAl	538 177	át nem vétel	156	rendbehozatal	538 021	-	-
-i	493 509	meg nem támadási	291	beruházási	493 218	-	-
-(A)n	381 397	el nem múlóan	314	megfelelően	381 083	-	-
-vA₂	207 559	el is van intézve	95 406	(van) megoldva	112 153	-	-
-sÁg	181 124	el nem ismertség	280	átláthatóság	180 844	-	-
-hAtAtlAn	160 435	el is fogadhatatlan	7	megoldhatatlan	160 428	-	-
-ÓlAg	97 926	utol nem érhetőleg	1	feltehetőleg	97 925	-	-
-AndÓ	87 565	meg nem írandó	3 090	megfontolandó	82 851	fejtendő meg	1 624
-Ú	65 345	-	-	elnevezésű	65 345	-	-
-tA	49 645	meg nem történte	21	létrejötté	49 624	-	-
-(A)tAn	37 501	le nem záratlan	3	megválaszolatlan	37 498	-	-
leg-...-bb	36 649	legösszenemillőbb	6	legelvakultabb	36 643	-	-
-(V)s	31 642	-	-	ellentmondásos	31 642	-	-
-Ul	24 163	ki nem mondatlanul	2	leírhatatlanul	24 161	-	-
-Ós	17 635	be nem vállalós	14	összebújós	17 621	-	-
-bb	13 720	oda nem illőbb	2	meggyőzőbb	13 716	viselhetőbb el	2
-vÁn₁	9 662	el sem olvasván	79	áttekintvén	9 024	dadogván el	559
-tAlAn	7 092	-	-	összefüggéstelen	7 092	-	-
-nivaló	3 653	el-nem-hinnivalók	2	megmásznivaló	3 651	-	-
-szerű	1 349	-	-	meglepetésszerű	1 349	-	-
-beli/béli	1 119	-	-	felkészültségbeli	1 119	-	-
-(O)nként	739	-	-	átutalásonként	739	-	-
-sdi	685	észrenemvevősdit	1	feljelentősdi	684	-	-
-ilAg	260	-	-	megvalósításilag	260	-	-
-féle	83	-	-	megérzésféle	83	-	-
-nyi	75	-	-	felragyogásnyi	75	-	-
-hAtnék	58	-	-	rágújthatnékja	58	-	-
-fajta	48	-	-	visszaélésfajta	48	-	-
-vÁst	22	-	-	végigrohanvást	22	-	-
-(V)cskA	22	-	-	fellángolásocska	22	-	-
leg-...-bbik	19	-	-	legelterjedtebbike	19	-	-
-szerűség	18	-	-	megoldásszerűség	18	-	-
-stUl/-stÓl	15	-	-	összevonásostul	15	-	-
-féleség	6	-	-	felvonulásféleség	6	-	-
-belileg	3	-	-	felfogásbelileg	3	-	-
-bbik	3	-	-	kihaltabbik	3	-	-
-vÁn₂	3	-	-	elrejtven (nincs)	3	-	-
-forma	2	-	-	elszólásforma	2	-	-

Az indexelt rövidítések feloldása: **-ni₁** = nincs beférkőző segédigeszerű elem, **-ni₂** = van beférkőző segédigeszerű elem; **-vA₁/-vÁn₁** = állapot- vagy módhatározói funkció, **-vA₂/-vÁn₂** = létigével passzív szerkezet. A felsoroltakon túl az *el kell*, *hogymenjek*, valamint *kell*, *hogymenjek* típusú szerkezetek és a topikalizáció is szerepelnek az adatbázisban.

IV. számú melléklet: A kutatás során létrehozott erőforrások

PREVLEX (*Preverb Lexicon*)

- 53 535 igekötős ige manuálisan ellenőrzött táblázata
- részletes ismertetés: 2.2.2. fejezet
- <https://github.com/kagnes/prevlex>

PREVMATRIX (*Preverb Matrices*)

- Igekötő × Jegy (morfológiai és általános jegyek) és Igekötő × Szerkezet mátrixok
- részletes ismertetés: 3.4. és 6.2.3. fejezet
- <https://github.com/kagnes/prevmatrix>

PREVCONS (*Preverb Constructions*)

- 21 038 igekötős igei hapax adatbázisa morfológiai, szemantikai és a vonzatkeretre vonatkozó információkkal, metaadatokkal
- részletes ismertetés: 5.3. fejezet
- <https://github.com/kagnes/prevcons>

PREVONTO (*Preverb Ontology*)

- 56 igekötő konkrét ill. metaforikus jelentései, és a jelentések közti viszonyok gráfként ábrázolva
- részletes ismertetés: 5.4. fejezet
- <https://github.com/kagnes/prevonto>

PREVDISTRO (*Preverb Distributions*)

- 49 igekötős szerkezet típus 41,5 millió korpuszbeli előfordulása az igekötő, az igető, a beférkőző szavak és az igekötő pozíciójának jelölésével
- részletes ismertetés: 6.2.2. fejezet
- <https://science-data.hu/dataset.xhtml?persistentId=doi:10.5072/FK2/TRSD50>

A disszertáció PDF verziója és egyéb kiegészítő anyagok: https://github.com/kagnes/phd_thesis

Magyar nyelvű összefoglaló

A disszertációmban a magyar igekötők állományát és az igekötős szerkezeteket vizsgálom korpuszvezérelt módszerrel. Fő kutatási kérdéseim a következők: (1) Mely lexikai elemeket soroljuk az igekötők közé, és mi alapján végezzük a besorolást? (2) Milyen mintázattípusokat mutatnak az igekötős szerkezetek, mikor és mennyire távolodhat el az igekötő az igétől, igenévtől, igei származéktól? (3) Hogyan írhatjuk le az igekötők produktív kapcsolódási mintáit, és ez alapján milyen megállapítások tehetők az igekötők szemantikájáról?

Az első kérdést illetően abból indulok ki, hogy az igekötők és egyéb, pusztán névszói igemódosítók között nem húzható éles határ, ezért olyan modellre van szükség, amely ezt az átmenetiséget képes megragadni. Feltételezem, hogy a prototípus-elmélet megfelel erre a célra. Ennek keretében olyan automatikusan mérhető, morfológiai és gyakorisági jegyeket veszek sorra, amelyek segíthetnek körülhatárolni az igekötők csoportját. A jegyek értékét korpuszméréssel határozom meg 235 igekötőjelölt esetében. A jegypárok közötti korrelációkat, valamint egyes szakirodalmak kritériumrendszerét tekintetbe véve egy kontinuumon helyezem el a vizsgált lexikai elemeket, majd ezt leegyszerűsítve négy csoportot nevezek meg: a prototipikus, a centrális, a félperiférikus és a periférikus igekötőket.

A második kérdés kapcsán először szinkrón korpuszvizsgálatot végzek, amelynek hangsúlyos része a finit igei szerkezetek feltérképezése. Kimutatom, hogy a hátravetett igekötők igétől való távolságát befolyásolhatják fonológiai tényezők és a szöveg szerkesztettsége is. Az igekötők disztribúcióját igenevek, igei származékok esetében is vizsgálom, és ennek során bemutatok néhány ritkán tárgyalt szerkezetet (pl. *szétválni szét lehet, a „be nem indulós Trabant”*). Ezután egy diakrón vizsgálatban mérem ki a prototipikus igekötők mondatbeli helyét az ómagyar kortól napjainkig terjedő szöveganyagon. Így láthatóvá válik a fókuszált összetevőt tartalmazó mondatok arányának növekedése, valamint a fordított szórendű tagadás térnyerése a megszokottal szemben.

A harmadik kérdés megválaszolására olyan módszert dolgozok ki, amely ‘igekötő – képző – vonzatkeret’ hármassok (pl. *el-X-elődik valami, tönkre-X-el valamit*) korpuszvezérelt vizsgálatán alapszik. Ismertetem az igealkotás három leggyakoribb módját: a névszói és az igei bemenetű igeképzést, valamint a hangzásséma-alapú igealkotást. Ezután bemutatom a PREVCONS adatbázist, amely 21 038 igekötős igei hapaxot tartalmaz, és a fentebb említett hármassok kereshetőségével

lehetővé teszi a produktív kapcsolódási minták feltárását. Végül egy olyan kísérletemet mutatom be, amelynek célja az igekötők jelentéseinek és a jelentések közti viszonyoknak a hálózatszerű ábrázolása a PREVCONS alapján, egy ontológia formájában.

A disszertáció végén visszatérek a kiinduló feltevésémhez, miszerint az igekötőnek nevezett szóosztály a prototípus-elmélet segítségével jól megragadható. Megvizsgálom, hogy a főként morfológiai és gyakorisági jegyek alapján felállított igekötő-kontinuum nem változik-e meg lényegesen akkor, ha disztribúciós és szemantikai jegyeket is figyelembe veszünk. A konklúzió az, hogy a skála két végpontja csak minimálisan változik, de a köztes mezőben jelentős eltérések mutatkoznak. Az eredeti megközelítésben tapasztalt bizonytalanság elvezet egy új szemlélethez, amely Kálmán László opponensi véleményéből indul ki, és a lényege az, hogy az egyes lexikai elemekről áthelyezi a fókusz az azokat tartalmazó szerkezetekre. A koncepción túl egy fontos hozadéka ennek a kutatási iránynak a nyílt hozzáférésű PREVDISTRO adatbázis, amely 49 igekötős szerkezet típus 41,5 millió korpuszbeli előfordulását tartalmazza.

Summary in English

The thesis investigates Hungarian preverbs – also called verbal particles, verbal prefixes – and preverb constructions by a corpus-driven method. Its main research questions are as follows: (1) Which lexical items can be regarded as preverbs, and what are the grounds of their classification? (2) What kinds of clausal orders do preverb constructions show, when and to what extent can a preverb be separated from a finite/non-finite verb or a deverbal element? (3) How can we describe the productive preverb–verb patterns, and – based on this – what conclusions can be made about the semantics of preverbs?

As for the first question, I assume that there is a fuzzy boundary between preverbs and other bare nominal verb modifiers. The prototype-theory seems to be suited for the graded categorization needed here. With this in mind, I collect morphological and frequency-related features of preverbs, measuring the value of each feature in the case of 235 preverb-like lexical items. Considering the correlations of each feature-pair as well as the standpoints made in a range of relevant literature, I set up a continuum ranging from the standard preverbs to the least preverbish elements. Finally, I decide to split the continuum into four categories: prototypical, central, semi-peripheral and peripheral preverbs.

Regarding the second question, I first perform a synchronic corpus study, putting emphasis on the finite constructions. I show that in inverted order constructions (finite verb – preverb), the distance of the finite verb and its preverb can be influenced by phonological factors and by an ‘oral versus written’ distinction. I also examine the distribution of preverbs in the case of non-finite verbs and deverbal elements. Doing so, I present some constructions rarely discussed in linguistic literature. I also conduct a diachronic study which aims to quantify the changes of the prototypical preverbs’ positions, from the Old Hungarian period to the present day. This study shows a monotonous increase in the proportion of sentences containing a focused constituent. Moreover, it shows how the negative sentences having ‘verb – negative particle – preverb’ order make headway against the ones having ‘preverb – negative particle – verb’ order.

To answer the third question, I develop a method based on the corpus-driven study of ‘preverb – derivational suffix – argument frame’ triplets. I present the three most common ways of word formation in Hungarian: verb formation from nouns and verbs, and thirdly, verb formation using sound

patterns. After that, I present the PREVCONS database containing 21 038 preverb–verb hapaxes. This resource makes it possible to explore the productive preverb–verb patterns by the accessibility of the triplets mentioned above. Finally, I present an experiment which aims to represent the different meanings associated with preverbs and the relationships between these meanings in a network-like structure based on PREVCONS, in the form of an ontology.

At the end of the thesis, I return to the concept which was my starting point, namely that the notion of preverb can be best captured by prototype-theoretical means. I check whether the original preverb continuum remains largely the same when considering distributional and semantic features of preverbs. The result shows that the two endpoints remain stable, while there is a considerable fluctuation in-between. The vagueness attested here leads to a viewpoint change from the study of lexical items to the study of constructions, largely based on László Kálmán’s review on the first version of this thesis. Beyond the introduction of this concept, a significant contribution of this short chapter is PREVDISTRO, an open-source database containing 41.5 million occurrences of 49 preverb construction types.