

PÁZMÁNY PÉTER KATOLIKUS EGYETEM  
Bölcsészettudományi Kar

# MESTERSZAKOS SZAKDOLGOZAT

*A magyar igei komplexumok vizsgálata*

**Dr. Prószéky Gábor**  
az MTA doktora, egyetemi tanár  
PPKE ITK

**Kalivoda Ágnes**  
digitális bölcsészet  
számítógépes nyelvészet szakirány

**2016**

## NYILATKOZAT

Alulírott Kalivoda Ágnes (Neptun kód: OBQV70) jelen nyilatkozat aláírásával kijelentem, hogy *A magyar igei komplexum vizsgálata* című szakdolgozat saját szellemi termékem, azt más szakon szakdolgozatként nem nyújtottam be, és csak a megjelölt segédeszközöket használtam. A szakdolgozat elkészítésekor betartottam a szerzői jogról szóló 1999. LXXVI törvény szabályait, valamint az egyetem által előírt, a dolgozat készítésére vonatkozó szabályokat, különösen a hivatkozások és idézések tekintetében.

Budapest, 2016. április 8.

*Kalivoda Ágnes*  
.....  
aláírás

## Tartalomjegyzék

<b>1. Bevezetés.....</b>	<b>4</b>
<b>2. Nyelvészeti háttér.....</b>	<b>5</b>
<b>3. A kutatás módszere.....</b>	<b>8</b>
<b>4. Igekötők.....</b>	<b>9</b>
4.1. Az igekötők állománya.....	9
4.2. Formai nehézségek.....	12
4.3. Igekötők az AnaGrammában.....	13
4.4. Az igekötők helye a mondatban.....	14
4.4.1. A finit ige bal oldala.....	15
4.4.2. A finit ige jobb oldala.....	17
4.5. Kategorizálási javaslatok.....	21
4.6. Az igék preferenciái.....	22
4.7. Speciális esetek.....	26
4.7.1. Igekötő kiemelése fölrendelt tagmondatba.....	26
4.7.2. Passzív szerkezet.....	28
4.7.3. Főnévi igeneves szerkezet.....	29
4.8. Az ideális n-gram.....	29
<b>5. Pusztai névszói igemódosítók.....</b>	<b>30</b>
5.1. Személyragos főnévi igenevek preverbális pozícióban.....	31
5.2. Dativuszi végződésű melléknevek.....	32
<b>6. Összegzés.....</b>	<b>33</b>
<b>7. Irodalomjegyzék.....</b>	<b>34</b>
<b>8. Mellékletek.....</b>	<b>37</b>

## 1. Bevezetés

Szakedolgozatomban az igei komplexum összetételére, a finit ige és az igemódosítók mondatbeli viselkedésére vonatkozó megfigyeléseimet ismertetem. A témát egyrészt nyelvtechnológiai, másrészt pszicholingvisztikai szemszögből közelítem meg. Ennek legfőbb oka az, hogy a kutatásom egy olyan számítógépes elemzőrendszert hivatott segíteni, amely működésével az emberi nyelvfeldolgozást modellezi.

A szóban forgó elemzőrendszer az AnaGrammar<sup>1</sup>, amelyet a MTA–PPKE Magyar Nyelvtechnológiai Kutatócsoport fejleszt. Az AnaGrammar főbb jellemzői:

1. Performancia-alapú, célja tehát a létező, leírt szövegek feldolgozása. Az elvileg lehetséges, de a gyakorlatban csak szórványosan előforduló szerkezetek elemzése háttérbe szorul. Ugyanakkor megkísérli feldolgozni a korpuszok sokszor hibás, normától eltérő mondatait is.

2. A szöveget balról jobbra haladva, szavanként dolgozza fel. Ha a beérkezett szó adott pillanatban nem elég az elemzéshez, a rendszer továbbhalad, és több információ birtokában végzi majd el az elemzést.

3. Több független, párhuzamosan futó szál felelős az elemzésért, és ezek kommunikálnak egymással. Így például a grammatikai jelenségeket felismerő szálak és a korpuszbéli gyakoriságot vizsgáló szál egyszerre érik el a beérkező szót.

Az AnaGrammar számára kiemelten fontos az igei komplexum hatékony felismerése és elemzése, mivel a vonzatokat (tematikus szerepeket) az ige osztja ki, az ige vonzatkeretét viszont az igemódosítók befolyásolják. Alapvető példa az igekötő mint igemódosító esete:

*ad* < valaki valakinek valamit >

*túl*ad < valaki valamin >

Ha az igekötő elvált az igétől (például más mondatrész fókuszba emelése miatt), az elemző nem tudhatja biztosan a vonzatkeretet, így azt sem, hogy a többi mondatrészt minek elemezze:

*Ő még nem adott...* < *túl*: valamin >

< Ø: valakinek valamit >

< *össze*: valakit/valamit valakivel/valamivel >

stb.

Itt érdemesnek tartom kiemelni az AnaGrammar korpuszgyakorisági szálának jelentőségét. A rendszer nagyban támaszkodik arra, hogy vannak bizonyos gyakori minták a nyelvhasználatban. Ezeket nem elemezzük teljesen, amikor egy mondatot értelmezzünk, hanem

---

1 Az AnaGrammáról bővebben ld. Prószyék–Indig 2015, Indig–Laki–Prószyék 2016.

egységként hívjuk elő őket a memóriánkból. Ez emlékeztet az informatikából ismert gyorsítótárazásra (Indig–Laki–Prószéky 2016: 261).

Számos igei szerkezet viselkedik ilyen egységként, kollokációként. Pl. *vállat von, számba vette* – itt a *vállat* és *számba* igemódosítók (a fogalmat bővebben tárgyalom a következő fejezetben). Ugyanez érvényes az igekötő esetében is, pl. az *állapít* nem állhat *meg* nélkül.

A fentieket figyelembe véve az a kiinduló feltevés, hogy az igemódosítók, legfőképpen az igekötők típusa, pozíciója alapján megjósolható a finit ige várt helye, esetleg a konkrét ige is – illetve fordítva, az igéből kiszámíthatók lehetnek az igemódosítók.

Ha ilyen szabályszerűségek fedezhetők fel a finit ige és módosítói között, azok nagy mértékben hozzájárulnak a gyorsítótárazáshoz. Mindez egybevág azzal a céllal, hogy az AnaGramma olyan nyelvi modell legyen, amely „[...] képes lesz számszerű becslést adni a soron következő szóra és/vagy kategóriára.” (Indig–Laki–Prószéky 2016: 262).

A kutatás fő kérdése tehát a következő: Milyen mintázatokkal ragadhatók meg az igei komplexumok? Magyarázhatók-e ezek a mintázatok az emberi nyelvfeldolgozás jellemzőivel?

## 2. Nyelvészeti háttér

Az igei komplexum<sup>2</sup> alatt olyan szerkezet értendő, amelynek része egy finit ige (legtöbbször a mondat főigéje) és a hozzá tartozó igemódosító. A finit ige jellemzője, hogy ideje, száma, személye van. Az AnaGramma ezeket az igéket FIN jeggyel látja el az elemzés során.<sup>3</sup>

Az igemódosítót É. Kiss olyan megnevezésre alkalmatlan igevonzatként definiálja, amely nem bővítő, hanem módosító szerepű. Más helyet foglal el a VP-ben, mint a bővítő: míg a bővítő követi az igét, a módosító jellemzően megelőzi. Az igemódosítók két típusa az igekötő és a pusztá névszói igemódosító (É. Kiss 2003: 12–13).

A 'pusztá névszói igemódosító + ige' szerkezetet Komlósy példáival (1992: 500) illusztrálom: *szénné éget, pirosra fest, okosnak tart, újságot olvas, úszni akar*. Ahogy a példákból is látszik, a pusztá névszó ismerve az, hogy nem egy konkrét dolgot vagy élőlényt jelöl, hanem általános értelmű.

Komlósy (1992: 500) úgy látja, hogy a ragos vagy névutós DP-k is igemódosítóként viselkedhetnek, amennyiben hangsúlykerülő igék lexikonban kijelölt direkcionális (1) vagy lokális (2) vonzatai.

2 Az igei komplexum lényegében ugyanannak a szerkezetnek az elnevezése, amelyre Kiefer (2003) *komplex ige*, Ackermann (1984) és Komlósy (1989) *komplex predikátum* névvel utal.

3 Az MNSZ2 korpuszban (Oravecz–Váradi–Sass 2014) a finit igékre úgy lehet pontosan szűrni, ha CQL-kifejezéssel olyan igét keresünk, amely nem valamilyen igenév: [msd=".\*IGE.\*" & msd!=".\*IGE\.\_?(HIN|OKEP|MIA|MIB|INF).\*"]. Ezt a meglehetősen bonyolult mintát váltaná ki az AnaGrammában a FIN jegy.

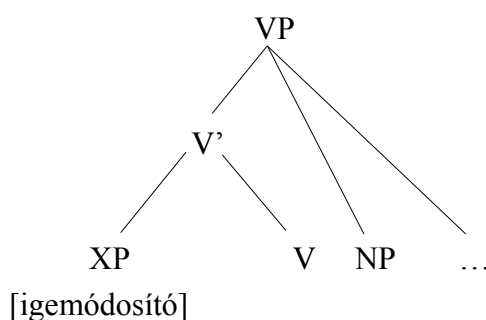
(1) Péter *az asztalra / az ágy alá* tette a könyvet.

(2) Péter *a kamrában / az ágy alatt* tartja a könyveit.

Ezeket a kategóriákat Pléh (1998: 121) is átveszi. Kétségtelen, hogy a finit ige előtti pozíciót foglalják el, amely jellegzetes igemódosítói hely, és az ige hangsúlykerülő volta miatt nem értelmezzük őket fókuszként. Azonban számítógépes felismerésük és igemódosítóként azonosításuk a munka jelen szakaszában nem lényeges, ezért a dolgozat nem tárgyalja őket részletesen.

É. Kiss (1992, 2003) generatív nyelvelméleti keretben három megszorítást fogalmaz meg az igemódosítók kapcsán:

1. Az igemódosító az igét közvetlenül megelőző, azzal V' összetevőt alkotó kifejezés.
2. Az [igemódosító] jegyű vonzatnak, amennyiben lehetséges, igemódosítói helyzetbe kell kerülnie. Ez érvényét veszti két ilyen elem esetén az egyikre nézve, pl. *elment iskolába*. Komlósy (1992: 512) hangsúlyozza, hogy a „kiszorított” elem el is veszti az igemódosítói jellegét, egyszerű vonzatként kezelendő.
3. Igemódosítói helyzetbe csak [igemódosító] jegyű összetevő kerülhet.



1. ábra. Az igemódosító mondat szerkezeti helye (É. Kiss 2003: 14)

Kiefer *A kétféle igemódosítóról* c. tanulmányában (Kiefer 2003) leírja, hogy a két típus szintaktikai műveletek segítségével nem választható el egymástól, de különböznek a vonzatosság és a továbbképezhetőség tekintetében, tehát szintaktikai és morfológiai szempontból egyaránt. A pusztai névszói igemódosító mindig vonzata az igének, míg az igekező sosem (pl. a *kimegy a kertbe* szerkezetben az *a kertbe* köti le az ige direkcionális vonzatát, a *ki* szerepe Kiefer szerint az irányjelölés explicitté és a direkcionális vonzat fakultatívvá tétele). Morfológiai szempontból a nominalizálhatóság és a megkettőzhetőség terén is ragadhatók meg különbségek:

1. Az igekezős ige mindig nominalizálható (*felugrik* → *felugrás*), a pusztai névszói

igemódosító esetében csak az ige nominalizált, a módosító vonzatként kerül mellé (*újságot olvas* → *újságot olvasás*, nem pedig *\*újságot olvasás*).

2. Az igekötő megkettőzhető (*fel-felugrott*), a puszta névszói igemódosító nem vehet fel ilyen frekventatív akcióminőséget (*\*vérré-vérré válik*).

A kétféle igemódosító kérdésköre foglalkoztatja a magyar pszicholingvisztikát is. Ezzel kapcsolatosak Pléh (1998) következő kísérletei:

Az intuitív összetartozási kísérletben felnőttek vettek részt, a feladatuk pedig az volt, hogy különböző mondatokban a szavak összetartozását ítélik meg. Az eredmények tanúsága szerint a különböző igemódosítók nagyon hasonlóan viselkednek.

- (3) *újságot olvas* a kertben
- (4) *vízbe esett* a kiránduláson
- (5) *le + szakadt* a gombom
- (6) *falnak fordult* az ágyban

A (3–6) példák vastagon szedett szavait szinte ugyanannyira összetartozónak ítélték meg a résztvevők olyan tesztmondatokban is, ahol több más szó szerepelt közöttük, tehát a felszíni szerkezetben távol állt egymástól az ige és az igemódosító.

Egy másik kísérletben három és nyolc év közötti gyerekekkel mondatokat idéztettek fel. Megfigyelhető volt, hogy 'igemódosító + ige' sorrend esetén jobban vissza tudták adni az eredeti mondatot. Ha az igemódosító eltávolodott az igétől, a gyerekek sokszor változtattak a szórenden.

(7) A kis krokodil a part mellett *melegnek érezte* a tó vizét.

(8) A vad cápa *hidegnek* a szikla alatt *érezte* a tenger vizét.

Míg a (7)-t a megfelelő sorrendben tudták felidézni, a (8) esetében nagyobb arányban volt pontatlan a visszaadás. A preverbális módosítók tehát könnyebbnek bizonyultak.

Egy harmadik, szintén gyerekekkel végzett kísérletben agrammatikus mondatokat használtak, megduplázott igemódosítókkal vagy fölösleges igekötővel, pl. *\*megivott teát, \*kerítést befestett*. A gyerekek érezték a konfliktust, és a feloldási módszereik arra utalnak, hogy mindig a lineárisan előrébb álló módosítót tekintették érvényesnek: *\*megivott teát* → *megitta a teát*, *\*kerítést befestett* → *kerítést festett*. Egy mondatban csak egy igemódosítót fogadtak el.

A kísérleti eredmények alapján Pléh (1998: 133–135) a következő megállapításokat teszi:

1. Megértés szempontjából az igekötők kitüntetett helye közvetlenül az ige előtti pozíció.

2. Az igekötők ilyen szempontból a többi igemódosítóhoz hasonlóan működnek.
3. Az igekötők szerepet játszanak az ige irányította mondatfeldolgozásban, argumentumok keresésében.
4. Alakilag többnyire a dekompozíciós elveknek megfelelően viselkednek.

Bővebb magyarázatot igényel a dekompozíciós elv fogalma. A pszicholingvisztikában régóta vitatott kérdés, hogy a szótári elemek vajon egészlegesen tárolódnak és férhetők hozzá a mentális szótárban (holisztikus modell), vagy a szótövek tárolódnak olyan morfológiai szabályokkal, amelyek a belőlük képezhető és ragozható alakokra vonatkoznak (dekompozíciós modell). A kísérletek alapján egy vegyes modell tűnik valószínűnek (bővebben ld. Thuma–Pléh 2001).

Az igekötők dekompozíciós viselkedését olyan lexikális döntési kísérlet igazolja (Juhász–Pléh 2001), amelyben a résztvevőknek azt kellett eldönteni, hogy a képernyőre kivetített szó létezik-e a magyarban. Az eredmények azt mutatják, hogy a rossz igekötő–ige kombináció elvetése tartott a legtovább: a *meg-intézésében* szó hibásságát lassabban ismerték fel a résztvevők, mint azt olyan szavakat, ahol nem létező igekötő volt (*mag-bízóimnak*), rossz tő (*be-súd-óinkat*) stb. Részben ez a „kombinatorikai probléma” ad majd magyarázatot a jelen kutatás egyik fő témájára, a különböző igekötő – finit ige távolságokra a mondaton belül.

Összefoglalva tehát a következőket mondhatjuk el: a különböző nyelvelméletek egyetértenek abban, hogy az igekötők és a pusztá főnevek működése nagyon hasonló, valamint abban is, hogy az igemódosítók preverbális helyzete az általános, tipikusabb jelenség. Ezek a megállapítások fontosak a számítógépes elemzés számára is.

### 3. A kutatás módszere

A kutatásomban korpuszvezérelt megközelítést<sup>4</sup> alkalmaztam, korpuszokból kinyert nyelvi adatok alapján fogalmaztam meg a következtetéseimet. A látott mintázatokra, szabályszerűségekre utólag kerestem pszicholingvisztikai magyarázatot, és ha volt ilyen, akkor a mintát igyekeztem algoritmikusan is értelmezhetővé tenni.

Két korpuszt használtam az adatgyűjtéshez. Az egyik a Magyar Nemzeti Szövegtár bővített változata (MNSZ2), a másik a most készülő Pázmány Korpusz.

Az MNSZ2-t 500 millió token (szövegszó) alkotja, egyaránt tartalmaz sajtószövegeket, szépirodalmat, tudományos, személyes és hivatalos írásokat, valamint beszélt szövegek

---

4 A korpuszvezéreltség fogalmáról bővebben ld. Sass 2011: 15–16.



(rádióműsorok) átiratát. Részletes ismertetéséről bővebben ld. Oravecz–Váradi–Sass 2014.<sup>5</sup>

A Pázmány Korpusz 1,2 milliárd szavas, jelenleg a legnagyobb magyar korpusz. Az MNSZ2-höz hasonlóan sok szövegtípus alkotja. A lekérdezésnél az első tízmillió találat érhető el, az innen vett adatokat tehát ehhez viszonyítva kell megadni (Endrédi–Prószéky 2016).

A korpuszokból komplex (CQL-ben megfogalmazott) lekérdezésekkel gyűjtöttem ki az összes finit ige szöveggörnyezetét, és a saját scriptjeimmel olyan statisztikai vizsgálatokat végeztem, amelyek külön szűrték a finit igtől számított „bal oldalt” (-1, -2, ... pozíciók), külön a „jobb oldalt” (+1, +2, ... pozíciók). Ezeknek az írásakor figyelembe vettem három fontos információt, amelyeket mindkét korpusz feltüntet minden egyes szónál: a tényleges szóalakot, a lemmát (szótő, szótári címszó) és a morfológiai elemzést.

A scriptek nagyobb része Unix Coreutils parancsok sorozata, kisebb része Python 3.4-ben íródott. A programok és az elkészült statisztikák elérhetők a CD-mellékleten.

## 4. Igekötők

A Bevezetésben láthattuk, miért alapvető fontosságúak az igekötők az igei komplexum (így tulajdonképpen a teljes mondat) elemzésében. Az igekötők befolyásolják az ige vonzatkeretét. Ha az elemző balról jobbra haladva már elérte az igt, de még nem kapta meg az igekötőt, illetve nem tudta eldönteni, hogy számíthat-e igekötőre vagy sem, addig a többi mondatrész tematikus szerepe is bizonytalan. Fontos az is, hogy sok igekötős ige gyorsítótárazható. Ha például az elemző elérte a *valósul* igt, akkor biztosan számíthat még igekötőre, ez pedig a *meg*.

Technikai szempontból is praktikus az igemódosítók vizsgálatát az igekötőkkel kezdeni: ezek véges halmazt alkotnak, amely viszonylag kevés elemből áll. Az viszont, hogy pontosan hány elemről van szó, vitatott téma.

### 4.1. Az igekötők állománya

Komlósy táblázata (1992: 495–497) jól mutatja a bizonytalanságot abban a kérdésben, hogy milyen szavak sorolhatók az igekötők közé. 90 szót listáz, amely „esélyes” az igekötői címre, és látható, hogy egyes források mennyire eltérően ítélik meg ezeket. Ahogy Komlósy is megfogalmazza, a besorolás véletlenszerűnek tűnik, olykor még egy forráson belül sem követhető nyomon a besorolás logikája.

---

5 A korpusz elérhető itt: <http://mnsz2.nytud.hu>

Az igekötő-kategória határa elmosódott, különösen az olyan esetekben, amikor határozószóként is funkcionálhat (pl. *ott+hagy*), vagy idióma jellegű (pl. *tönkre+tesz*). Kiefer (2003) például az *ott*-ot igekötőre hasonlító elemnek nevezi, és amellet érvel, hogy a lokatív jelentésű helyhatározók nem igemódosítók; az *otthag*y ige szerinte lexikalizálódott szó. Az igekötők állománya, úgy tűnik, szerzőnként változik annak a függvényében, hogy ki milyen definíciót ad az igekötőknek.

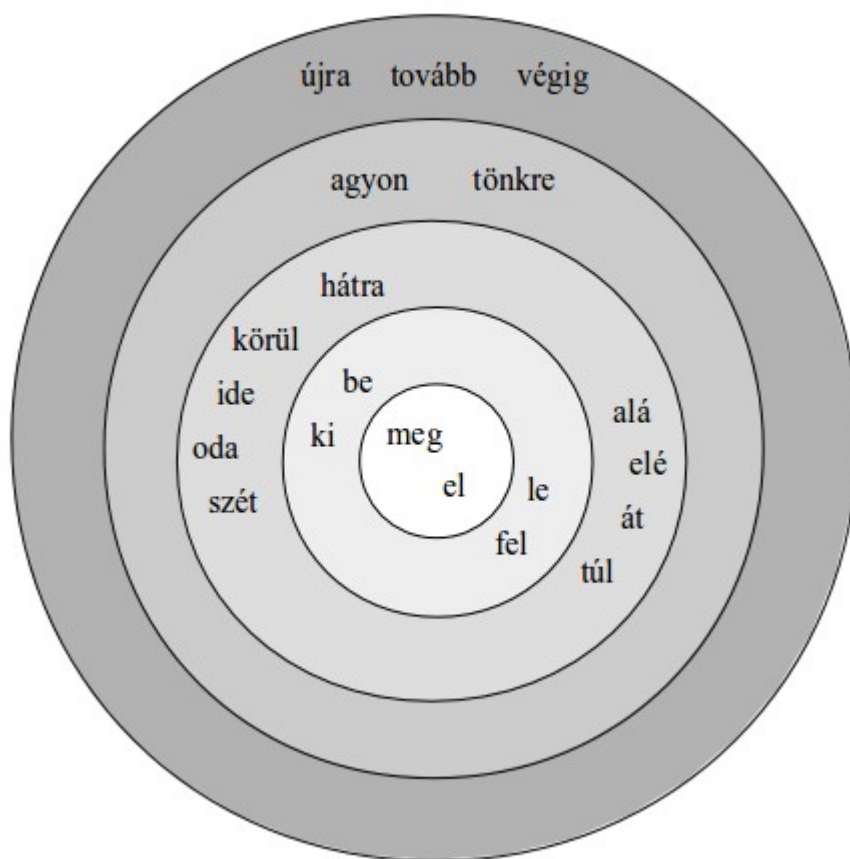
Forgács (2005: 11) szakirodalmi áttekintése az alábbi két kritériumot találja közösnek az egyes szerzőknél:

1. Az igekötő produktív, új igék hozhatók létre vele. Ez indokolja azt, hogy a *tönkre* a legtöbb forrásban igekötő (*tönkre|megy, -tesz, -vág, -zúz, -ver* stb. igék alkotók vele), míg a *cserben*-t nem szokás annak tekinteni (csak a *cserbenhagy* igében fordul elő).
2. A szemantikailag átlátszó elemeket, ahol az igekötős ige jelentése nem több, mint az igekötő és az ige külön-külön vett jelentései, általában kizárják az igekötő-kategóriából. Így például a *haza* szót nem tekinti igekötőnek J. Soltész (1959: 15), valamint Kiefer és Ladányi (2000) sem. Forgács ezt bírálja, arra hivatkozva, hogy a *haza*- előtag egyrészt nagyon produktív, másrészt átvitt értelmű használata is van (*hazaszidóznak valakit, 'övéit zsidónak titulálják'* értelemben, de gondolhatunk a szlenges *hazavág* szóra is: 'elront, megöl'). Jakab (1976: 12) szerint a szemantikai átlátszóság azért is problémás szempont, mert ilyen alapon a *bemegy* előtagja nem igekötő, a *berúg* szóé viszont igen ('lerészegedik' jelentésben).

Sok szerző mutat be különféle tesztek, amelyek elvégezhetősége dönt egy-egy szó igekötő-voltáról.<sup>6</sup> Forgács (2005) azonban más szemléletet javasol. Szerinte a prágai iskola centrum – periféria modelljét volna célszerű alkalmazni az igekötők esetében. A centrumban olyan szavak állnak, amelyek minden kétséget kizáróan igekötők, a periféria felé haladva csökken az „igekötőség” erőssége. A 2. ábra szemlélteti Forgács (2005) elképzelését.

---

6 Az igekötő-állomány meghatározásáról bővebben ld. Jakab (1976), Kiefer–Ladányi (2000), Pais (1959).



2. ábra. Az igekötők centrum – periféria modellje (Forgács 2005: 21 alapján)

Ő az igekötők eredetét, nyelvtörténeti útját szem előtt tartva határozza meg az egyes köröket. A későbbiekben látni fogjuk, hogy az elképzelése nagyon tanulságos a jelenlegi kutatás szempontjából, mert az igekötőnek az igétől való lehetséges eltávolodásában is felfedezhető hasonló mintázat.

A kutatásom korpuszvezérelt módszert követett az igekötők állományát illetően is. Nem volt céлом felülbírálni a korpuszok annotációját, és az [IK]-nak jelölt, a szakirodalmak többsége szerint viszont nem igekötő *létre*, *észre* szavakat ugyanúgy bevettem az elemzésbe, mint a *ki*, *be*, *fel* stb. elemeket. Az igekötő állomány meghatározásához lekértem az MNSZ2-ből minden igekötős finit igét. Egy script elválasztotta az igekötőt a lemmától, és aszerint, hogy volt-e benne kötőjel, két halmazba rendezte az eredményeket: egytagú és összetett igekötők halmazába. A végeredmény 80 egytagú igekötő lett, amelyek gyakorisági sorrendben (kezdve a leggyakoribbal) a következők:

*meg, el, ki, be, fel, le, vissza, át, össze, hozzá, rá, elő, bele, föl, oda, egyet, létre, haza, szét, túl, alá, végig, fenn, tovább, neki, újra, felül, közre, észre, félre, szembe, jóvá, ide, körül, előre, tönkre, abba, végre, helyre, közzé, együtt, egybe, ott, körbe,*

*hátra, agyon, utána, közbe, ellent, alul, újjá, ketté, ellen, utol, tele, félbe, keresztül, által, kölcsön, belé, rajta, mellé, végbe, külön, benn, fönn, szerte, nyilván, fölül, alább, viszont, véghez, tova, közben, széjjel, keresztbe, teli, közé, reá, kinn*

Az igekötős finit igék listáját ezután átnéztem, és több mint 24 ezer lemmát kaptam eredményül, megőrizve a tokenek gyakorisági adatát is (pl. *el+mond* 292019, *el+fogad* 145257, *ki+derül* 141467, ...).<sup>7</sup> Ezzel a listával a későbbiekben átszűrtem minden korpuszbéli adatot, és csak azt az igekötő–ige párt fogadtam el, amely a lista szerint létező kombináció. Erre azért volt szükség, hogy hatékonyan ki lehessen szűrni a hibásan annotált szavakat.

#### 4.2. Formai nehézségek

Az automatikus annotálástól nem várhatunk el maximális pontosságot, különösen akkor, ha homográf szóalakokról kell eldönteni, hogy milyen morfológiai elemzés lesz esetükben a megfelelő. Hibás annotálás gyakran fordul elő az igekötők esetében. A legtipikusabb tévesztések:

- a) *meg, viszont*: ezek a szavak általában akkor is [IK], azaz igekötő címkét kapnak, amikor mellérendelő kötőszóként szerepelnek a mondatban.
- b) *be*: melléknevek előtt többnyire a régies 'milyen, mennyire' értelmében szerepel (pl. *Be szép ez a lány!*), de igekötőként annotált.
- c) *el*: sokszor a spanyol határozott névelő hímnemű alakja (pl. *a spanyol El País*).
- d) *le*: sokszor a francia határozott névelő hímnemű alakja (pl. *a francia Le Poème*).
- e) *ki*: gyakran keveredik a homográf vonatkozó illetve kérdő névmással.
- f) *haza, alul, felül, ...* – A morfológiai elemző tanácstalan, amikor el kell döntenie, hogy ezek igekötők-e vagy határozószók. Így például a *nem működnek együtt* és a *nem írtak együtt* konstrukciók egyformán 'határozószó + ige + igekötő' elemzést kapnak. (Ilyenkor gyakran az ember is bizonytalan volna a döntésben.)
- g) Gyakori a lemmák hibás meghatározása. Néhány jellegzetes példa:

(9) Tönkrement a laptopom.

Előfordul, hogy a morfológiai elemző nem a *tönkremegy* alakot azonosítja lemmaként, hanem a *tönkrement-et* (*tönkre+ment*, az *el+ment* < *valamit* > analógiájára).

Ugyanígy:

---

<sup>7</sup> Küszöbértéknek a minimum öt előfordulást választottam. Az ennél ritkábban szereplő lemmákat nem vettem figyelembe. A műveltető és ható igék külön lemmának számítanak közös tő esetében is, tehát pl. *működik, működtet, működhet* három lemma.

(10) Kergetett és utolért.

Az *utolért* lemmájának az *utolér* szót várnának, olykor mégis *utolért* marad (*utol+ért*, a *félre+ért* mintájára) .

(11) A nő már régen elvált a férjétől.

Az *elvált* lemmája nem *elválik*, hanem *elvált* marad (*el+vált*, a *fel+vált* mintájára).

h) Egyszerű ige tévesen igekötős igeként szerepel (ez csak az MNSZ2-re érvényes, a Pázmány Korpuszra nem):

(12) szembesül → szembe+sül

(13) teleportál → tele+portál

i) A *kell*, *lehet*, *lesz* igékre jellemző, hogy az emberek sokszor elgépelik ezeket. A Pázmány Korpuszban például több mint 400 találatot kapunk a *megkell* lemmára.

Az ilyen típusú szavak akár ezres nagyságrenddel növelik a találati listát, és rontják a pozíciókra vonatkozó elemzés minőségét. Ezért döntöttem úgy, hogy a vizsgálataimban csak olyan igekötő–ige párokat fogadok el, amelyeket a korábban bemutatott módszerrel ellenőriztem.

### 4.3. Igekötők az AnaGrammában

A kutatási eredmények ismertetése előtt fontosnak tartom bemutatni azt, hogy az AnaGamma hogyan kezeli majd az igekötőket, és miért fontos a pozíciójukra vonatkozó információ, előrejelzés. Az elemző egyszerre egy szót kap bemenetként, viszont alapvetően trigramokat (háromszavas egységeket) „lát”, ezt a háromszavas ablakot csúsztatja előrébb, ahogy halad a mondatban:

Bemenet: *Sose fogom befejezni a szakdolgozatomat.*

0. lépés:	[ _ _ _ ]	← <i>Sose</i>
1. lépés:	[ _ _ Sose ]	← <i>fogom</i>
2. lépés:	[ _ Sose fogom ]	← <i>befejezni</i>
3. lépés:	[ Sose fogom befejezni ]	← <i>a</i>
4. lépés:	Sose [fogom befejezni a]	← <i>szakdolgozatomat</i>
5. lépés:	Sose fogom [befejezni a szakdolgozatomat]	← .
6. lépés:	Sose fogom befejezni [a szakdolgozatomat .]	

3. ábra. Trigramok az AnaGrammában

A gyakorlat azt mutatja, hogy a háromszavas ablak általában elég egy összetartozó szerkezet, mondatrész befogására. Emellett pszicholingvisztikai szempontból is jó imitálása lehet az emberi olvasásnak: egyszerre körülbelül három szón tartjuk a szemünket. A rendszer képes vissza is lépni az elemzésben, ha szükséges, és lehetnek predikciói a következő egy-két szóra, vagy legalábbis szófaji kategóriára – ahogy az ember is sokszor előre tudja, sejtí, mi következik a mondatban (Elman 2004).

A finit ige FIN címkét kap, az igekötő pedig PreV-et (*preverb*, az alapvetően preverbális pozícióból). Ha azonban nem preverbális az igekötő pozíciója, bizonytalanná válik az, hogy a trigram be tudja-e fogni az igei komplexum egészét. Pl. ilyen az [*ítéltek -e már*] *oda*, vagy számtalan olyan eset, amikor az igekötőt a főnévi igenévvel kellene összekapcsolni: *el [sem tudom hinni]*.

#### 4.4. Az igekötők helye a mondatban

A következő fejezetek az igekötő pozíciók statisztikai vizsgálatát mutatják be. Ez olyan adattáblákra épül, amelyeket egy script külön elkészít minden finit ige lemmájához.

Az adattábla oszlopai az igekötő igétől számított pozícióját<sup>8</sup> jelzik, a sorok tartalmazzák magát az igekötőt, a cellák értéke pedig az adott igekötő adott pozíción mért előfordulásainak száma. A két szélső oszlop a bal- és jobboldali „közbeeső szavak” gyakorisági listáját tartalmazza. Ha egy feldolgozandó mondat 90% feletti egyezést mutat egy már feldolgozott mondattal, az kimarad a statisztikából (különben bizonyos igekötők felülreprezentáltak lehetnek egy-egy híres idézet miatt).

<i>derül</i>									
IK	-2	0	+1	+2	+3	+4	+5	bal	jobb
<i>ki</i>	148	141467	50704	1078	12	2	1	Ø 141467 is 75 nem 51 ne 14 sem 7 se 1	Ø 50704 majd 495 csak 272 ez 73 még 69 az 59 ...
<i>fel</i>	1	398	74	2	0	0	0	Ø 398 is 1	Ø 74 másképpen 1 csak 1
<i>föl</i>	0	179	30	0	0	0	0	Ø 179	Ø 30

1. táblázat. A *derül* lemma reprezentációja

<sup>8</sup> A -1 pozíció hiányáról ld. a következő fejezetet.

Mi olvasható ki egy-egy ilyen igeablából? Először is az igével álló igekötők és azok gyakorisága: a *derül* igekötője többnyire a *ki*, ritkábban a *fel*, nagyon ritkán a *föl*. Látszik a kitüntetett preverbális (0) pozíció: a *ki* az esetek 73%-ában egybe van írva a *derül*-lel. Ha az ige mögé kerül (fókusz, tagadás stb. hatására), akkor közvetlenül az ige után áll (98%-ban), vagy jellemzően nagyon rövid szavakat enged maga és az ige közé: *derül majd ki*, *derül csak ki*, *derül ez ki* stb.

A több mint 24 ezer igekötő–ige kombináció adattáblái a CD-mellékleten érhetők el.

#### 4.4.1. A finit ige bal oldala

A korpuszadatokból jól látszik – és generatív nyelvelméleti megfontolásból számítani is lehetett rá –, hogy a finit ige bal oldala megszorított. Szélsőértéknek az igekötő -4-es pozíciója bizonyult, vagyis az igekötők és a finit ige között legfeljebb három másik szó szerepel, ez azonban nagyon ritka.

IK pozíciók	-4	-3	-2	(-1)	0 (FIN)
<i>példák</i>					elmond
			meg	is	nyeri
		el	sem	nagyon	jutnánk
	vissza	most	már	nem	szívhatom
<b>találatszám</b>	6	91	102768		13715093

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

A **preverbális helyzetű igekötők (0)** aránya kiemelkedően magas, ahogy azt a korábbi példákon is láttuk. A számítógépes elemzés szempontjából ez a legkönnyebb eset, a trigramban csak egy helyet tölt ki az igekötős ige. A korpuszokban itt a legkisebb az annotálási hibák aránya, és ez az egyetlen pozíció, amelyben minden [IK]-ként annotált szótípus megtalálható.

A **-1 pozíció** annyira speciális, hogy nem tartottam célszerűnek bevenni az adattáblákba. Itt csak a következő esetekben jelenik meg igekötő:

1. A mondat írója hibásan gépelte a szót, vagy nem érezte igekötőnek azt a szót, amelyet [IK]-ként annotál a morfológiai elemző (pl. *agyonverte* helyett *agyon verte*).
2. A *van* lemmánál elképzelhető olyan igekötő, amely a lemmától külön írandó: *tele van*.

3. A finit ige valójában segédigei szemantikájú (*kell, lehet* stb.), és a -1 pozícióban álló igekötőt egy későbbi (+1, +2, ...) főnévi igenévhez kell kapcsolni. Erről később lesz szó.

A **-2 pozíció** jól elemezhető: is-szerkezet (14) vagy annak tagadása (15, 16), tagadó felszólítás (17) vagy tagadás (18) esetén kerül ide az igekötő.

(14) 10-12 óra között *le is zárják* a teret.

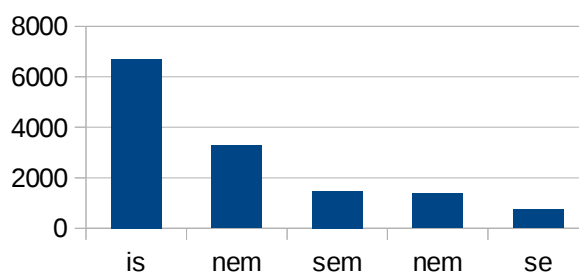
(15) Pedig még *ki sem próbáltam*.

(16) *Vissza se gyere!*

(17) Aztán *fel ne fázz!*

(18) Egy 15 milliós terepjárával *be nem mennék* a fák közé.

Ezek a példamondatok egyúttal tartalmazzák az összes olyan szót, amely a korpusz alapján az igekötő és a finit ige közé kerülhet:



4. ábra. A -2 pozíciójú igekötő és a finit ige közti szavak

Amíg -2-ben több mint 102 ezer igekötőt lehetett találni, **-3 pozícióban** csak 91 szerepelt. Ezekről elmondható, hogy a közbeeső két szó egyike a fentebb felsorolt öt elemből való (a másik pedig többnyire, de nem feltétlenül határozószó).

*sem is, nem is, már nem, még nem, soha nem, ugyan nem, sem nagyon, sem igen, se nagyon, biztosan nem, valóban nem, sosem is, soha ne, senkit sem, senki sem, semmiképpen nem, sem igazán, őt nem, ma sem, is ugyanígy, is nem, én nem, egy se, csak nem, az se, azért sosem*

Az MNSZ2-ben három olyan példa volt, amelyre nem érvényes ez az általánosítás:

(19) *le csak azért taszította*

(20) *ide talán hamarabb ér*

(21) *le a számokkal írjunk képleteket*

A (19, 20) példa szórendje magyarázható azzal, hogy az igekötő került fókuszba. A (21) esetében inkább az ige előtt álló főnévi frázis, *a számokkal* tűnik fókuszni.



A legszélsőbb, **-4 pozícióra** csak hat példa akadt a korpuszban:

- (21) *haza* is elég körülményesen *jön*
- (22) *haza* is csak látogatni *megy*
- (23) *vissza* már egyre ritkábban *fordít*
- (24) *vissza* az istennek se *néz*
- (25) *vissza* most már nem *szívhatja*
- (26) *vissza* már nem egyedül *érkezik*

Két igekötő, a *haza* és a *vissza* került bal oldalon ilyen távolra az igétől. Ahogy az igekötő állomány meghatározási nehézségeinél szó volt róla, a *haza* a szemantikai átlátszósága miatt több szakirodalomban kimarad az igekötők listájából. Ezekben a példákban valóban felfogható hangsúlyos határozószóként is. A *vissza* jellemzően bekerül az igekötő-kategóriába, és talán a (25)-ös példában tűnik a leginkább egyértelműnek az igével való összetartozás, a *visszaszív* igekötős ige esetében. A *haza* és a *vissza* lesznek azok az elemek, amelyek jobb oldalon is a legtávolabb képesek állni az igétől.

#### 4.4.2. A finit ige jobb oldala

Mitől kerülhet ki az igekötő a preverbális pozícióból, és mozdulhat el jobbra? Röviden: attól, hogy valami más foglalja el a privilegizált helyét (Pléh–Juhász 2001: 202). Ez történik egy mondatrész fókuszba emelésekor (27), felszólító mondatoknál (28), tagadásnál (29).

- (27) Tegnap 'én *mentem el* a boltba, nem te.
- (28) *Nézd csak meg!*
- (29) Ilyesmi nem *fordult még elő*.

A bal- és jobb oldal lehetséges igekötő pozíciói jól mutatják, hogy „[...] a magyar mondat ige előtti szakasza kötött szórendű, az ige utáni mondatszakaszban viszont szabad a fő összetevők sorrendje” (É. Kiss 2006: 2). Az igekötők jóval nagyobb intervallumon tudnak elhelyezkedni az ige utáni szakaszban, mint a bal oldalon.

FIN	+1	+2	+3	+4	+5	+6	+7	IK pozíciók
jelenik	még							<i>példák</i>
lát	majd	el						
fordulhat	még	egyszer	elő					
adok	én	most	itt	elő				
beszélék	még	mindig	egy	kicsit	haza			
fut	a	HÉV	a	vidám	néppel	ki		
csábítson	téged	a	nagy	retyezáti	barna	medve	oda	
	7527308	163993	5126	1193	267	101	27	<b>találatszám</b>

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

Látunk példát arra is, hogy az igekötő hét vagy akár tíz szót is maga elé engedjen. Ezek statisztikailag nem releváns, de érdekes esetek:

*bújnak az ég felhői mögül a csillagok se ki*

+8 pozíció, 5 találat

*kapnak erre a célra a Volán-társaságok az ÁPV Rt.-től vissza*

+9 pozíció, 3 találat

*teszi megannyi egyébként oly sikeresnek látszó ember saját magát szántszándékkal tönkre*

+10 pozíció, 2 találat

*visz egy feltehetően részeg buszsofőr Szentesen még csütörtökön egy sportrendezvény után vissza*

+11 pozíció, 1 találat

A statisztika a bal oldaléhoz hasonló olyan szempontból, hogy a finit igtől való távolság függvényében csökken a találatok száma. Átlagosan az igekötők nem engednek maguk és a finit ige közé kettőnél több (rendszerint rövid) szót, nagyobb eltávolodást az eseteknek mindössze 0,1%-ában látunk.

A leggyakoribb „közbeeső szavak” az igekötő +2 pozíciója esetén:

*majd, -e, már, csak, is, még, volna, ő, most, ez, az, én, itt, ők, hát*

Ugyanígy a leggyakoribb tizenöt elem +3 pozíciónál:

*egy kicsit, én is, -e majd, -e már, most már, csak úgy, már végre, én ezt, -e még, végül is, még egyszer, azt is, még mindig, már csak, majd csak*

Az utolsó olyan igekötő pozíció, amely 1000 feletti találatot hozott, a +4. Itt már nagyobb számban jelennek meg főnévi csoportok, de a gyakoribb köztes szó szerkezetek itt is a lehető legrövidebb elemekből állnak:

*még ma is, már eddig is, ez a dolog, még egy kicsit, már egy kicsit, ezt a kérdést, ezek a dolgok, ebbe a dologba, csak egy kicsit, még azt is, ezt az egészet, ez a kettő, ez a kérdés, egy kicsit jobban, most egy kicsit*

Érdekes mintázat fedezhető fel abban is, hogy mely igekötők távolodnak el jobban az igétől és melyek kevésbé (ld. 8.2. melléklet). A *vissza, haza, tönkre, össze, ide, oda* előtagok megjelennek a legszélsőbb pozíciókban is. A *keresztül, keresztbe* igekötők ugyan nem kerülnek a korpusz alapján ilyen távolra, de a szélső pozícióik telítettebbek, mint sok más igekötőé. A *keresztbe* esetében látható szám például azt jelenti, hogy az összes előfordulásainak 0,2%-a (3 token) a +5 pozíción volt, ez a többi +5 pozícióhoz viszonyítva a legmagasabb érték.

Eloszlás szempontjából hasonlóan viselkedik a *tova, kölcsön, körbe, közbe, ketté*, illetve hasonlítanak a *félre, félbe, egyet, egyre, utána, hátra, abba, újra, agyon, túl* igekötők is. A *meg, ki, be, fel, le, el* esetében azt látjuk, hogy vannak ugyan szélsőbb előfordulásaik, de ezek 1-1 példára korlátozódnak (a 0% rendszerint 1 példát jelent). Közös bennük az is, hogy ezek a leggyakoribb igekötők (ld. 8.3. melléklet), a jelenlétük mégis szinte csak a preverbális pozícióra és a +1-re korlátozódik. A 8.1. mellékletben láthatók ugyan ennél szűkebb intervallumot lefedő igekötők, de ennek oka lehet a jóval kevesebb korpuszbeli előfordulás is.

A kapott eredmények generatív keretben a Behaghel-törvénnyel magyarázhatók. É. Kiss (2007) amellet érvel, hogy az ige utáni, alapvetően szabad szórendet befolyásolja az, ha az igét követő összetevők fonológiaiailag különböző súlyúak. Ez egybevág Otto Behaghel (1932) általánosításával. Ő azt fogalmazta meg, hogy a világ nyelveiben a mondatösszetevők fonológiai súlyuknak megfelelően, növekvő sorrendben rendeződnek – a magyarban ez az ige utáni mondatrészre érvényes (É. Kiss 2007). Ezt többek között az alábbi példával

szemlélteti:

(30) János sértette meg nagyon Éva nagyapját.

Itt az ideális sorrend finit ige után az 'igekötő, határozó, tárgy', egyéb kombinációk legalábbis szokatlanul hangzanának.

Úgy gondolom, ez magyarázatot ad arra, hogy a testesebb igekötők miért távolodhatnak el jobban az igétől, és arra is, hogy az ige és igekötő közé kerülő elemek jellemzően miért rövidek (szófajuk pedig sokszor határozószó).

A látottak pszicholingvisztikai magyarázatát elsősorban a távolsági hatásban látom, eszerint „minél nagyobb a távolság a szöveg alkotóelemei között, annál nagyobb a hiba lehetősége és annál kapacitásigényesebb az összekapcsolás” (Németh 2001: 94). Fontosak még annak a lexikális döntési kísérletnek az eredményei is, amely a Nyelvészeti háttérnél került bemutatásra. Ha az igekötős igék feldolgozása dekompozíciós jellegű, akkor logikusnak tűnik, hogy az ige elhangzása után minél előbb meg kell érkeznie az igekötőnek, amely módosítja az ige jelentését. Különösen igaz ez a *meg, ki, be, fel, le, el* igekötőkre, amelyek nagyon sok igével kombinálhatók. Fontos az is, hogy ezek az igekötők irányt illetve aspektust jelölnek, így szorosan kapcsolódnak az ige jelentéséhez. A *haza, vissza* esetében gyengébbnek tűnik ez a kapcsolat. A testesebb jelentésű, részben idiomatikus igekötők (pl. *agyon, tönkre, ketté*) kevesebb igével kombinálhatók, az elhangzott kontextusból az ember rendszerint „számít rá”, hogy ezek az igekötők fognak még elhangzani. A (31) példában már korábban is két dolgról volt szó; taoizmus mint vallás, taoizmus mint filozófia:

(31) „[...] a kínai gondolkodás azonban soha nem választotta ezt a két fogalmat élesen ketté.” (+6 pozíció)

A (32) példában is előre sejthető, hogy a *tesz* lemma 35 lehetséges igekötője közül éppen a *tönkre* szerepel majd:

(32) „[...] a pap nem papi család lányát úgyse veszi el soha, nem akar a gyerekének rosszat; dehogyan *teszik* ők azt a csinos papot *tönkre* [...]” (+6 pozíció)

'A pap nem akar rosszat a gyerekének.' → 'A pap nem akarja, hogy nem papi család lányai tönkretegyék a gyereket.' A *tönkre* igekötő negatív jelentést hordoz, amelyet előfeszít a *rosszat* szó szemantikája.

Számítógépes elemzés során nagyon nehéz volna ilyen összetett kontextushatást vizsgálni. A javaslatom az, hogy az igekötőket a már bemutatott, pozíció szerinti eloszlás alapján kategorizáljuk. Ha az AnaGrammar beolvasott egy igekötő nélküli finit igét, rendelkezésére áll az információ az igelemma lehetséges igekötőiről, és ezek kategóriája alapján lehet predikciója arról, hogy „meddig kell még menni” az igekötőért.

#### 4.5. Kategorizálási javaslatok

Az igekötők kategóriába sorolása két szempont alapján történt: 1) Mekkora intervallumot fednek le, milyen pozíciókban állhatnak? 2) Hogyan alakul az egyes pozícióik telítettsége (vagyis milyen arányban oszlanak meg az előfordulások az egyes pozíciók között)?

A kategorizálás hangsúlyozottan a gépi elemzés számára készült, az alapötlet felfogható úgy is, mint Forgács (2005) centrum – periféria modellje számítógépes kivitelezésben. A kategóriák összetétele természetesen máshogy alakul, mert nyelvtörténeti vagy bármilyen nyelvelméleti szempontok itt nem játszanak szerepet. A „centrum” igekötői kizárólag preverbális és közvetlenül az ige utáni pozícióban állnak a korpuszban (1), a „periféria” igekötői szélsőségesen eltávolodhatnak (8). Értékes információ az is, ha egy igekötő gyakrabban áll a finit igtől jobbra, mint preverbális pozícióban. Ennek az az oka, hogy olyan igékhez kapcsolódik, amelyek jellemzően fókuszot várnak maguk elé, vagy felszólításban szerepelnek (ezekről bővebben lesz szó a következő fejezetben). Az ilyen sajátos helyzeteket az alábbi táblázatban ’+’ jelöli.

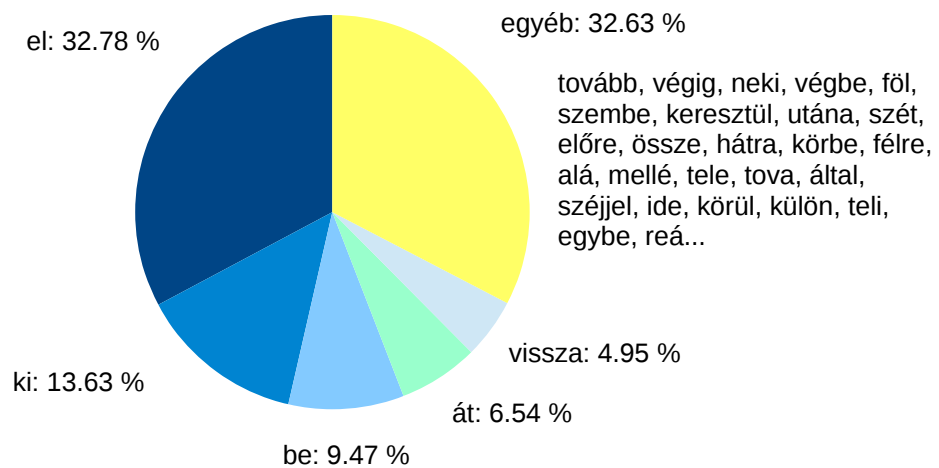
Kategória	Elemei	Jellemzése
1	<i>által nyilván ellen</i>	Ezek a legritkább igekötők (a szakirodalom többsége a <i>nyilván</i> és <i>ellen</i> szót nem tartja igekötőnek, az <i>által</i> pedig régies, mára elvesztette a produktivitását. Több mint 90%-ban preverbális pozícióban állnak, ritkábban +1-ben.
2	<i>teli, benn, fönn, fölül szerte, alul, ott, rajta + reá +, közé +, véghez +</i>	Szintén nagyon ritka igekötők. 0, +1, +2 pozícióban állnak, olykor -2-ben is, kivéve: <i>teli, benn, fönn</i> .
3	<i>mellé, alább, észre végre +, előre +, helyre utol, együtt, ellent, külön</i>	Az 1 és 2 kategória elemeinél valamivel gyakoribbak. 0, +1, +2, +3 pozícióban állnak, és -2-ben is, kivéve: <i>mellé, alább</i> .
4	<i>neki, fenn, tovább + végbe +, szembe, körül újjá, viszont, belé tele, széjjel, jóvá</i>	Vegyes csoport mind az igekötő gyakoriságát, mind a produktivitását nézve. Közös vonásuk, hogy +4 pozíciónál messzebb egyetlen korpuszbeli példában sem álltak az igtől.
5	<i>közre, túl, félre, félbe szét, egybe, kölcsön hátra, utána, agyon</i>	Jobbra +5 pozícióig fordulnak elő.
6	<i>egyet, újra, abba körbe, közbe, ketté tova, elő</i>	Jobbra +6 pozícióig fordulnak elő.
7	<i>közzé +, hozzá, létre + alá, bele</i>	Jobbra +7 pozícióig fordulnak elő, de a szélső értékek nagyon ritkák.

8	<i>ide +, oda össze, tönkre keresztbe +, keresztül + végig, haza, vissza</i>	A korpusz alapján ezek az igekötők távolodnak el legjobban a finit igtől úgy, hogy a szélsőbb értékeik nem csak egy-két találatból állnak.
0	<i>meg, ki, be, le fel, föl, el, át, rá</i>	A tizenöt leggyakoribb igekötő között vannak. Gyakori előfordulásuk ellenére mégis elég kötötten, egységesen viselkednek, azonos pozícióban és hasonló arányban fordulnak elő: -2    0    +1    +2    +3 0,5   58,5   40    1    0,01 %

4. táblázat. Igekötő kategorizálási javaslatok az AnaGamma számára

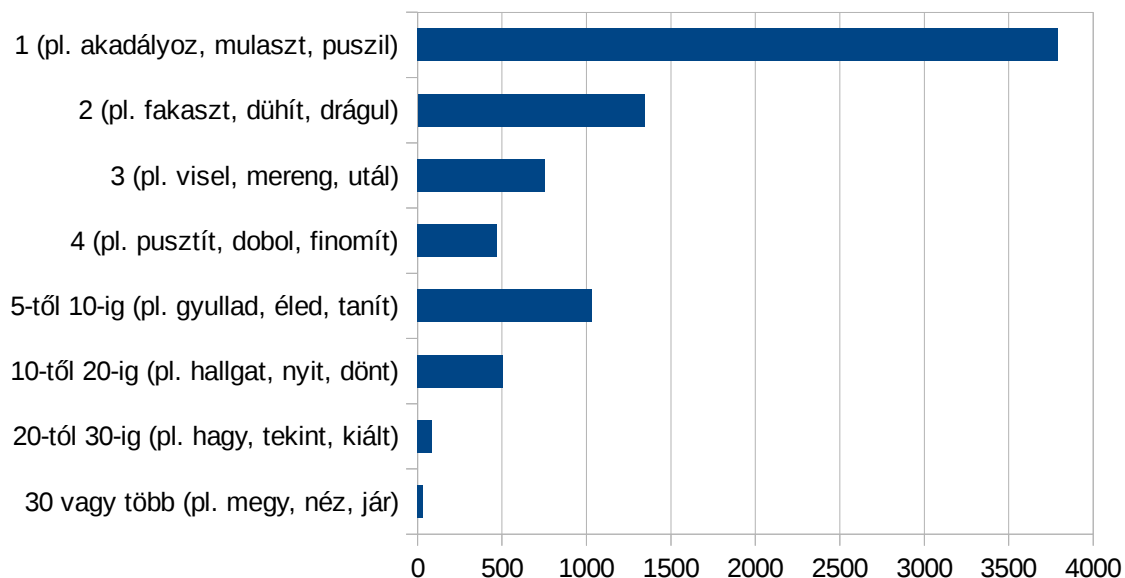
#### 4.6. Az igék preferenciái

A korpuszadatok alapján arról is számot adhatunk, hogy egyes igék mely igekötőkkel és milyen gyakran fordulnak elő. Ez támogatás lehet az AnaGamma korpuszgyakorisági szálának. Még az olyan igéknél is, amelyek több tucat igekötővel kombinálhatók, van néhány kiugróan gyakori igekötő, így például a *megy* esetében:



5. ábra. A *megy* lemmával álló igekötők százalékos eloszlása az MNSZ2-ben

Ugyanakkor sokkal több igére jellemző az, hogy csak egy-két igeekötőjük van:



6. ábra. Igek százalékos eloszlása az MNSZ2-ben aszerint, hogy hányféle igeekötővel állhatnak

Sok igenél jellegzetes minta látható olyan szempontból is, hogy magához képest balra vagy jobbra várja az igeekötőjét (ezt a szóhasználatot a számítógépes szemlélet indokolja). Mint látni fogjuk, ez elsősorban az ige szemantikájából adódik. Bizonyos igeeket például jellemzően tagadunk, ekkor a tagadószó elmozdítja az igeekötőt a preverbális helyzetből.

Az ilyen igeek összegyűjtésére írtam a *right\_pref.py* scriptet. Ez a korábban bemutatott adattáblák (*db\_prev.txt*) alapján szűri ki az olyan igeekötős igeeket, amelyek legalább 1000-szer előfordultak a korpuszban, és az igeekötő minimum 60%-kal gyakrabban állt a jobb oldalon, mint a balon. Természetesen a bal oldalba a 0 pozíció is beleszámít. Az 5. táblázat 20 kimagasló találatot tartalmaz (a teljes lista és a script elérhetők a CD-mellékleten).

	Ige	Igekötő	Jobb oldal aránya (%)	Magyarázat (az IK-mozgatás leggyakoribb kiváltó oka)
1.	dolgoz	fel	98,6	Fókusz: < valahogyan >
2.	tevődik	össze	97,7	Fókusz: < valamiből >
3.	engedett	meg	96	<i>Engedtesseék meg, hogy...</i>
4.	áll	hozzá	92,6	Fókusz: < valahogyan >
5.	ér	el	90,8	Tagadás: <i>nem ér el</i>
6.	helyezkedik	el	90,8	Fókusz: < valahol >
7.	folytat	tovább	90,7	Fókusz: < valahogyan >
8.	könyvelhet	el	90,7	Fókusz: < valaminek >

9.	indul	ki	89,6	Fókusz: < valamiből >
10.	ítél	meg	88,9	Fókusz: < valaminek >
11.	mehet	tovább	87,7	Tagadás: <i>nem mehet tovább</i>
12.	riad	vissza	87,2	Tagadás: <i>nem riad vissza</i>
13.	terül	el	86,7	Fókusz: < valahol >
14.	jelenik	meg	86,5	Tagadás: <i>nem jelenik meg</i>
15.	nyúlik	vissza	87,2	Fókusz: < valamire >
16.	él	tovább	85,6	Fókusz: < valamiként >
17.	nevez	el	85	Fókusz: < valaminek/valakinek >
18.	van	túl	85	Tagadás: <i>nem vagyok túl</i> < valamin >
19.	dolgoz	ki	84,7	Fókusz: < valahogyan >
20.	idéz	elő	84	Fókusz: < valamit >

5. táblázat. Igék, amelyek jellemzően *nem* a preverbális pozícióban tartják az igekötőt

Fentebb az igekötő-preferencia kérdését a finit ige felől közelítettük meg: melyik igék azok, amelyek jellemzően jobb oldalon tartják az igekötőt? Érdekes azonban máshogy is gondolkodni: melyik igekötők azok, amelyek jellemzően jobb oldalon állnak? Az előző fejezetben, a kategóriák egyes igekötői után írt '+' jelek ezt a tulajdonságot jelzik. A 6. táblázatban láthatók összegyűjtve ezek az igekötők és az igék, amelyekkel állhatnak.

Igekötő	+1 pozíció aránya (%)	Az igekötővel alkotott igék (az MNSZ2 alapján)
<i>reá</i>	69,3	<i>reá talál, -néz, -száll, -bizik, -nyom, -megy, -figyel, -borul</i>
<i>végbe</i>	68,3	<i>végbe megy, -visz</i>
<i>rajta</i>	67,9	<i>rajta kap, -üt, -veszt, -csíp, -hagy, -marad, -ér, -ragad</i>
<i>végre</i>	61,7	<i>végre hajt</i>
<i>keresztbe</i>	61,6	<i>keresztbe fordul, -tesz, -áll, -fekszik, -rak, -vet, -megy stb.</i>
<i>tovább</i>	59,5	<i>tovább jut, -megy, -ad, -lép, -visz, -fejleszt, -áll stb.</i>
<i>keresztül</i>	58,3	<i>keresztül megy, -visz, -húz, -halad, -vág, -néz stb.</i>
<i>ide</i>	56,8	<i>ide jön, -hoz, -tartozik, -kerül, -ad, -ér, -ír, -jár stb.</i>
<i>közzé</i>	55,6	<i>közzé tesz</i>
<i>létre</i>	53,3	<i>létre hoz, -jön, -hív</i>
<i>előre</i>	51,3	<i>előre vetít, -lép, -hoz, -bocsát, -hajol, -tör, -megy stb.</i>
<i>véghez</i>	51	<i>véghez visz, -megy</i>

6. táblázat. Igekötők, amelyek jellemzően +1 pozícióban követik az igét



Két esetet látunk:

1. Az igekötőnek alacsony a produktivitása vagy csak egy igéhez illeszthető (tehát a legtöbb szakirodalom szerint nem igekötő, hanem igekötő-szerű elem).
2. Az igekötő produktív (és jellemzően egy mozgásigével áll).

Valójában itt is az ige preferenciájáról van szó: sokszor fókuszot vagy tagadást vár el, ez mozdítja el az igekötőt a preverbális helyzetből. Gyakori a felszólító módú használat: *menj tovább, gyere ide, add ide* stb.

A preferenciák vizsgálatához két script készült (*prev\_sort.py* és *fin\_sort.py*). Az első igekötőkhöz rendeli a lehetséges igéket, és százalékos értékkel mutatja, hogy az adott igekötővel mennyire kapcsolódik az ige. A kimenete tehát így néz ki:

```
szembe
==> néz      31.97
==> száll    12.42
==> fordul   9.54
==> kerül    9.1
==> megy     6.54
...
```

A másik igelemmákhoz rendeli a lehetséges igekötőket, kezdve a leggyakoribbal (ezt itt is százalékos érték jelzi). Példa a kimenetre:

```
szabadul
==> meg      39.39
==> ki       24.69
==> el       17.68
==> fel      15.09
==> föl      1.59
==> be       0.76
==> rá       0.66
==> neki     0.1
==> le       0.05
```

Ez a két script az AnaGramma korpuszgyakorisági szálához nyújthat támogatást. A *prev\_sort.py* akkor lehet hasznos, ha az elemző már beolvasta az igekötőt, de finit ige még nincsen, a *fin\_sort.py* pedig akkor, ha finit ige már került a trigramba, de igekötő még nem. Ezzel részben imitálható az a predikciós képesség, amelyet az emberi nyelvfeldolgozásban a kontextus hív elő.

#### 4.7. Speciális esetek

Az eddig bemutatott igei szerkezetekben az volt a közös, hogy az igekötő a finit igéhez tartozott, attól általában nagyon kevés szó választotta el. Ebben a fejezetben olyan esetekről lesz szó, amelyek eltérnek minden eddigitől. Az eltérés okai a következők lehetnek:

- (1) az igekötő úgy találja meg a finit igét, hogy tagmondathatárt lép át (igekötő kiemelése fölérendelt tagmondatba, pl. *meg kell, hogy mondjam*);
- (2) az igekötő határozói igenévhez tartozik (passzív szerkezet, pl. *el van vetve*);
- (3) az igekötő főnévi igenévhez tartozik (segédigés szerkezet, pl. *meg kell értened*).

Az AnaGramma szempontjából azért is fontosak ezek a szerkezetek, mert csak a legrövidebb tagjaik foghatók be háromszavas ablakkal, egyetlen „közbeeső szó” jelenléte már kiszoríthatja akár az igekötőt, akár az igét:

trigrammal befogható elemek: [be kell menned]

trigramon kívül eső elemek: [be sem kell] menned                      be [sem kell menned]

##### 4.7.1. Igekötő kiemelése fölérendelt tagmondatba

É. Kiss két munkájában (2003, 2007) hívja fel a figyelmet arra a szerkezetre, amelynél a kötőmódba került ige igekötője kiemelhető (sőt, a *kell* esetén kötelező kiemelni) a fölérendelt tagmondatba.

(33) *el* kell, hogy *engedj*

(34) \**kell*, hogy *elengedj*

A *szabad* esetében opcionális az igekötő kiemelése:

(35) *el* szabad, hogy *vezessen*

(36) *szabad*, hogy *elvezessen*

Ezek és minden további megállapítás érvényes a puszta névszói igemódosítóra is (É. Kiss 2003: 13), így például lehetséges a (37) szerkezet:

(37) János *pirosra* kell, hogy *fesse* a kerítést.

É. Kiss (2007) legfontosabb észrevételei a következők: 1) Ha a *hogy*-kötőszós tagmondat tagadó vagy fókuszos, nem lehet kiemelni az igemódosítót. 2) Egy-két nem operátor-szerepű kifejezésen az igemódosító átemelhető. 3) A kiemelés feltétele a *kell/szabad* és a *hogy* közvetlen szomszédossága.

Ezek alapján írtam a Pázmány Korpuszhoz a következő CQL-lekérést:

```
[humor="IK"] [humor!="(PUNCT|KOT|.*IGE.*)"]{0,2} [lemma="(kell|szabad|szeret|akar)"] [lemma=","] [lemma="hogy"] [lemma=".*"]{0,1} [humor="IGE.*" & humor!="IGE\+_?(HIN|OKEP|MIA|MIB|INF).*"]
```

Vagyis a várt minta: 'igekötő + 0-2 szó, amely nem központozás/kötőszó/ige + a *kell, szabad, szeret, akar* lemmák egyike + vessző + *hogy* + 0-1 bármilyen szó + finit ige'. Több mint 3000 releváns találatot kaptam, ezek szinte mindig a *kell* igét tartalmazzák. A *szabad* 2, az *akar* 4 esetben szerepelt a főlérendelt tagmondatban. A *szeret*, pontosabban *szeretne* is hozott egy találatot (erre a szakirodalom alapján nem számítottam):

(38) A routertől elég messze *el* szeretném, hogy *érjen* a wifi jel.

Három példa volt arra, hogy az alárendelt mondatban a *hogy* és a finit ige közé más szó került:

(39) minden ember *fel* kell, hogy magát *áldozza*

(40) *meg* kell, hogy valaki *öleljen* engem

(41) *fel* kell, hogy *töltve* legyen (!)

A (41) azért is érdekes, mert passzív szerkezetet látunk, tehát az igekötőt a határozói ige-névhez kell kapcsolni.

A főlérendelt mondatról elmondható, hogy egy szó állhat az igekötő és a segédige között, ez a szó pedig nem lehet más, mint az *is, se* vagy *sem*.

(42) *együtt* is kell, hogy *működjenek*

(43) *fel* sem kell, hogy *álljon*

(44) *ki* se kell, hogy *kapjon*

A gyakori esetek egységesen illeszkednek a következő mintára:

igekötő + kell + vessző + *hogy* + kötőmódú finit ige

A 7. táblázat a nyolc leggyakoribb esetet mutatja (az igelemma természetesen mindig kötőmódban értendő (pl. *van* → *legyen*)).

A szerkezet lemmatizálva			Gyakoriság
meg	kell , hogy	mond	188
el	kell , hogy	mond	111
be	kell , hogy	vall	91
ki	kell , hogy	ábrándít	72
meg	kell , hogy	felel	69
meg	kell , hogy	van	53
el	kell , hogy	keserít	45
el	kell , hogy	fogad	41

7. táblázat. A Pázmány Korpusz leggyakoribb 'IK kell, hogy FIN' szerkezetei

#### 4.7.2. Passzív szerkezet

A passzív szerkezet általában így épül fel: 'igekötő + *van* (lemma) + határozói igenév'.

Ahhoz, hogy pontos képet lehessen kapni a szerkezet határaitól, széles spektrumú keresőkifejezést adtam meg a Pázmány Korpusznak:

```
[humor="IK"] [humor!="(IK|PUNCT|KOT|IGE\+\S+[123])"]{0,10} [lemma="(van|nincs|sincs|lehet|lesz)"] [humor!="(IK|PUNCT|KOT|IGE\+\S+[123])"]{0,10}
[humor="IGE\+_HIN.*"]
```

Tehát a várt minta: 'igekötő + 0-10 szó, amely nem igekötő/központosítás/kötőszó/ige + a *van*, *nincs*, *sincs*, *lehet*, *lesz* lemmák egyike + 0-10 szó, amely nem igekötő/központosítás/kötőszó/ige + határozói igenév'. Azért éppen 10 köztes szót engedélyez a keresőkifejezés, mert a finit ige és a *vissza* igekötő közt ez volt a maximum.

A passzív szerkezet megszorítottan mondható: A *van*-tól balra és jobbra maximum két közbeeső szó áll (ezek gyakoriság és hossz tekintetében is azokat a tendenciákat mutatják, amelyeket az igekötő pozíciók elemzésekor láttunk).

A passzív szerkezet gépi felismerésében az okoz némi nehézséget, hogy a *van* igének lehetnek igekötői, pl. *megvan* < *valami* >, *elvan* < *valaki* >, *kivan* < *valaki (valamitől)* >. Ilyenkor a balról jobbra haladó elemzésben előállhat az alábbi helyzet:

Már beolvasott mondatszakasz: [Az egyik fiú] [meg is van

Lehetőségek:

Lezárás: [Az egyik fiú] [meg is van] ('előkerült')

Várakozás a határozói igenévre: [Az egyik fiú] [meg is van

|--> keresztelve]

|--> magával |--> elégedve]

Bár van rá példa, hogy az igekötőt a *van* lemmához kell kapcsolni, jóval gyakoribbak az olyan esetek, amikor az igekötő a határozói igenévhez tartozik. Célszerű lehet előre készülni a passzív szerkezetre, és néha utólag elvetni a lehetőséget.

### 4.7.3. Főnévi igeneves szerkezet

A főnévi igenes szerkezet leggyakrabban így épül fel: 'igekötő + segédige jellegű finit ige + főnévi igenév'. Az itt felmerülő problémák nagyon hasonlóak ahhoz, amit a passzív szerkezetnél lehetett látni: egyes segédige jellegű finit igék állhatnak igekötővel is. Ilyen például a *fog* (*elfog, befog* stb.), a *szokik* (*megszokik*), a *tud* (*betud, letud, megtud*). Ugyanakkor gyakoribb a segédige jellegű használatuk, amikor igekötős főnévi igenevet vágnak ketté, és az igekötő nem a sajátjuk, pl. *el sem fogod hinni*. Ezeket az igéket célszerű lenne külön annotációval is ellátni. Mindenképp hasznos külön annotálni az olyan igéket is, amelyeknek biztosan nincs igekötője, és jellemzően főnévi igenevet vonzanak. Ilyen például a *kell, szabad, akar, sikerül, illik* stb.

Ki kell még térnünk arra jelenségre, amelyet Alberti (2001: 157) a semleges mondatban „felmászó” igekötőnek nevez. Itt azt látjuk, hogy az igekötőt egy segédige jellegű finit ige követi, majd halmozva főnévi igenevek következnek. A Pázmány Korpuszban sok olyan főnévi igenévi csoportosulást lehet találni, ahol gondot jelent az igekötő hovatartozása:

- (45) *el* fogják tudni *dalolni*
- (46) *el* fogod tudni *dönteni*
- (47) *be* tetszik tudni *szállni*
- (48) *el* kell tudni *válni*
- (49) *meg* kellett *tanulni* kaszálni
- (50) *ki* akartak *szállni* gombászni
- (51) *el* kell *kezdeni* gondolkozni
- (52) *ki* akartak *vonulni* tüntetni
- (53) *le* kell *ülni* tárgyalni
- (54) *el* kell *kezdeni* keringőzni tanulni

A példákban az a minta fedezhető fel, hogy ha a finit ige utáni főnévi igenév nem segédige jellegű, akkor az igekötő ehhez tartozik, ha viszont segédige jellegű (itt: *tudni*), akkor „tovább kell adnia” az igekötőt a következő főnévi igenévnek. Ehhez szükségesek a külön annotációk.

### 4.8. Az ideális n-gram

Összefoglalva elmondható, hogy az AnaGrammában alapértelmezett trigram le tudja fedni a leggyakoribb eseteket, tehát az igekötő -2, 0, +1 és +2 pozícióját a finit igéhez képest:

-2: [el sem jöttek]      0: [\_ \_ eljöttek]      +1: [\_ jöttek el]      +2: [jöttek volna el]

Van azonban néhány rendszeresnek mondható kivétel:

- 1) a fölérendelt tagmondatba kiemelt igekötő;
- 2) a passzív szerkezet;
- 3) a főnévi igeneves szerkezet;
- 4) a 4–8 kategóriába sorolt igekötők.

Növeli a trigramon kívül maradó igekötők vagy finit ige valószínűségét az is, ha rövid köztes szavak érkeznek (egyik rövid szó „görgeti maga után” a másikat). Ilyeneket a +3 pozíció előtt lehet leginkább látni: *-e már, én is* stb. Ezeknél és a négy rendszeres kivétel esetében hasznos lehet az ideiglenes ablaknövelés.

## 5. Pusztai névszói igemódosítók

A Nyelvészeti háttérben már felmerült az a probléma, hogy a pusztai névszói igemódosítók osztálya szófaji szempontból sokszínű, és az igekötők zárt, véges elemszámú halmazával szemben ezeket gyakorlatilag lehetetlen felsorolni.

Az ilyen összetételek hatékony azonosításában segít az AnaGamma rendelkezésére álló vonzatkeret-adatbázis (bővebben ld. Prószék–Indig: 2015). Mivel a pusztai névszói igemódosítók speciális igevonzatok, a vonzatkeret-adatbázis segítségével jól lefedhetők.

Mikor lehet mégis hasznos, ha a gépi elemző el tudja különíteni az igemódosítót az egyéb vonzatoktól? Elsősorban akkor, ha az elemzőnek arról kell döntenie, hogy az igekötő nélküli finit ige előtt álló mondatrész fókuszban van-e. Az AnaGamma a fókusz egyelőre úgy kezeli, hogy ha a finit, igekötő nélküli ige megjelent, az kijelöli a közvetlenül előtte lévő helyet a fókuszban. A jellemzően -1 pozícióban lévő pusztai névszói igemódosító viszont nem feltétlenül fókusz.<sup>9</sup> Az elemző fejlesztésének egy későbbi szakaszában ez még jelentkezhet problémaként.

A pusztai névszói igemódosítók részletes vizsgálatára ez a dolgozat nem vállalkozik. A következő fejezetekkel az a célom, hogy gyorsítótárazási lehetőségeket mutassak be két igemódosító típus, a személyragos főnévi igenév és a dativuszi végződésű melléknév esetében.

---

<sup>9</sup> A fókusz kérdésköre nagyon összetett. Lokális szinten nem lehet véglegesen eldönteni, hogy a -1 pozíció adott esetben tényleg fókusz-e, úgy gondolom, ez inkább az egész mondat beolvasása után dőlhet el. A fókusz és igemódosító megkülönböztetése az emberi nyelvfeldolgozásban főként a prozódia alapján történik, ilyen információ a gépi elemzésben nem érhető el.

## 5.1. Személyragos főnévi igenevek preverbális pozícióban

Ha az elemző jobbra haladva még nem olvasott be finit igét, de ragozott főnévi igenevet (INR) már igen, a rákövetkező szót majdnem biztosan előre tudja vetíteni: *kell*. Sokkal ritkábban a *szabad* követi, és egy-két példát találunk más, nem ennyire segédige jellegű szavakra. Ezek régi vagy szónoki szövegekben fordulnak elő:

(55) *ennem adtatok*

(56) *látnom engedett*

(57) *igazodnia illet*

A *szabad* jellemzően akkor követi a ragozott főnévi igenevet, ha előtte a *ha* vagy *csak* szó szerepelt:

(58) *Megvárjuk, míg felforr, lefedjük, a legeslegkisebb lángra vesszük a gázt, csak pittyegnie szabad, és három órán át tartjuk így [...]*

Egyéb esetek régi vagy emelkedett stílusú szövegeknél várhatók:

(59) *Tudnunk kell, tudnunk szabad.*

(60) *De amíg élnünk lehet és élnünk szabad [...]*

Az 'INR + finit ige' szerkezetet megszakíthatják az *is*, *se*, *sem*, *ne*, *nem* szavak. Ezek közül a *ne* jelenléte egy még inkább megszorított szerkezetet implikál: '(csak|hogy) INR ne kelljen'.

(61) *[...] szépen teszi a dolgát, csak kapkodnia ne kelljen.*

(62) *[...] ezek a rómaiak Varus seregének legkülönb futói, akik, hogy háborúzniuk ne kelljen, zendülést színleltek.*

Az észrevételeket formálisan így összegezhetjük (kezdve a valószínűbb rákövetkezésekkel):

< nem-FIN elemek > INR	--->	kell	
	--->	(is se sem nem)	---> kell
			---> szabad
	--->	szabad	
< nem-FIN elemek > (csak ha hogy) INR	--->	kell	
	--->	szabad	
	--->	ne	---> kelljen
	--->	(is se sem nem)	---> kell
			---> szabad

8. ábra. Az 'INR + finit ige' szerkezet rákövetkezési modellje

## 5.2. Dativuszi végződésű melléknevek

Azért döntöttem a 'melléknév+nak/nek (MN+DAT) + finit ige' szerkezetek vizsgálata mellett, mert esetükben nem csak a szerkezeti megszorítottság látszik jól, hanem szemantikai csoportok is elkülönülnek az igék halmazában. Ezek nagy vonalakban, egy-egy példával a következők:

- 1) véleményt kifejező igék: *jónak tart, követendőnek minősít, fejletlennek ítél, kockázatmentesnek vél, kihaltnak hisz, túlzónak érez ...*
- 2) kinyilatkoztató igék: *könyörtelennek nevez, elsődlegesnek nyilvánít, bűnösnek vall[ja magát] ...*
- 3) látszatot kifejező igék: *jelentősnek tűnik, ügyesnek látszik, átmenetinek bizonyul, alacsonynak számít ...*

Az AnaGrammarra figyelembe tud venni szemantikus szelekciós megszorításokat is az elemzéskor, így elképzelhető, hogy ezeket a kategóriákat érdemes lesz kidolgozni a jövőben. Egyelőre kézzelfoghatóbb információ az, hogy az 'MN+DAT + finit ige' szerkezet kizárja az igekötő lehetőségét, amennyiben az MN+DAT igemódosító, nem pedig fókusz szerepű. Ez összecseng azzal, amit Pléh (1998) ír: egy mondatban nem tűrünk meg két igemódosítót.

Az MNSZ2-ből lekértem az 'MN+DAT + finit ige + 0-2 szó + igekötő' sémára illeszkedő találatokat. Az MN+DAT ezekben valóban fókusz, és a találatok 80%-át tíz igekötős ige tette ki (ld. 8. táblázat; a teljes lista elérhető a CD-mellékleten).

< valamilyennek >	ítél meg	25,5 %
	mond ki	17,9 %
	ismer el	9,8 %
	fogad el	6,8 %
	ír le	4,3 %
	néz ki	3,9 %
	tüntet fel	3,7 %
	állít be	3,6 %
	kiált ki	2,7 %
	tűnik fel	1,7 %

8. táblázat. Az MN+DAT kategóriájú szavakat leggyakrabban követő igekötős igék



## 6. Összegzés

A korpuszvezérelt módszerű kutatás biztató eredményeket hozott azzal kapcsolatban, hogy az igei komplexumok gépileg felismerhetők-e az igemódosítók finit igéhez viszonyított pozíciói alapján. Az igekötők részletes vizsgálata olyan mintázatokat mutatott ki, amelyekre magyarázatot lehet adni mind generatív nyelvelméleti, mind pszicholingvisztikai keretben. A kategorizálási javaslatok és az igék, igekötők preferencialistái pedig gyakorlati segítséget jelentenek az AnaGamma számára.

A preverbális helyzetű, puszta névszói igemódosítók közül két típust vizsgáltam meg: a személyragos főnévi igenevet és a dativuszi végződésű melléknevet. Ezek olyan megszorított szerkezetet alkotnak a finit igével, amelyet egyrészt a gyorsítótárazásnál, másrészt a predikciók modellezésében lehet használni. Ha a rendszer beolvasta ezeket a speciális igemódosítókat, jó eséllyel „meg tudja jósolni” a következő szót.

A bemutatott téma két szempontból is kapcsolódik a fókusz kérdésköréhez. Egyrészt felmerült az a probléma, hogy a puszta névszói igemódosítót nehéz megkülönböztetni a fókuszba emelt puszta névszótól. Másrészt az igekötővizsgálat példáiból láttuk, hogy nagyon sokszor a fókuszba emelt összetevő váltja ki az igekötő-mozgatást. Az igei komplexumok mintázatai hozzájárulhatnak az AnaGamma fókusz-felismerő szálának a fejlesztéséhez is.

## 7. Irodalomjegyzék

- Ackermann, Farrel (1984): Verbal modifiers as argument taking predicates: complex verbs as predicate complexes in Hungarian. In: *Groninger Arbeiten zur Germanistischen Linguistik* 25. 23–71.
- Alberti Gábor (2001): Az aspektus szintaxisa a magyarban. In: Bakró-Nagy Marianne – Bánréti Zoltán – É. Kiss Katalin (szerk.): *Újabb tanulmányok a strukturális magyar nyelvtan és nyelvtörténet köréből*. Budapest, Osiris Kiadó. 145–164.
- Behaghel, Otto (1932): *Deutsche Syntax IV*. Heidelberg, Carl Winters.
- Elman, Jeffrey L. (2004): An alternative view of the mental lexicon. In: *TRENDS in Cognitive Sciences* 8/7. 301–306.  
[http://crl.ucsd.edu/~elman/Papers/elman\\_tics\\_opinion\\_2004.pdf](http://crl.ucsd.edu/~elman/Papers/elman_tics_opinion_2004.pdf)
- Endrédi István – Prószéky Gábor (2016): *A Pázmány Korpusz*. megjelenés előtt
- É. Kiss Katalin (1992): Az egyszerű mondat szerkezete. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan. 1. Mondattan*. Budapest, Akadémiai Kiadó. 79–178.
- É. Kiss Katalin (2003): Mondattan. In: É. Kiss Katalin – Kiefer Ferenc – Siptár Péter, 2003. *Új magyar nyelvtan*. Budapest, Osiris Kiadó. 1–126.
- É. Kiss Katalin (2007): Az ige utáni szabad szórend magyarázata. In: *Nyelvtudományi Közlemények* 104. 124–152.  
[http://www.nytud.hu/oszt/elmnyelv/ekiss/publ/az\\_ige\\_utani\\_szabad\\_szorend.pdf](http://www.nytud.hu/oszt/elmnyelv/ekiss/publ/az_ige_utani_szabad_szorend.pdf)
- 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*. Piliscsaba, Elméleti Nyelvészeti Tanszék – Magyar Nyelvészeti Tanszék. 213–228.
- Forgács Tamás (2005): Grammatikalizálódás az igezők körében. In: Oszkó Beatrix – Sipos Mária (szerk.): *Uráli grammatizáló*. BUM 4. Budapest, MTA Nyelvtudományi Intézet. 88–116.  
[http://www2.arts.u-szeged.hu/magyar/magyarny/oktatok\\_tematika/forgacs/10\\_11\\_2/grammatika.pdf](http://www2.arts.u-szeged.hu/magyar/magyarny/oktatok_tematika/forgacs/10_11_2/grammatika.pdf)

- Indig Balázs – Laki László – Prószéky Gábor (2016): Mozaik nyelvmodell az AnaGamma elemzőhöz. In: Tanács Attila – Varga Viktor – Vincze Veronika (szerk.): *XII. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged, 2016. január 21-22. 260–270.  
[http://real.mtak.hu/34384/7/vegl\\_MSZNY2016\\_web\\_ISO\\_B5.pdf](http://real.mtak.hu/34384/7/vegl_MSZNY2016_web_ISO_B5.pdf)
- Jakab István (1976): *A magyar igekötők állományi vizsgálata. Nyelvtudományi Értekezések 91*. Budapest, Akadémiai Kiadó.
- J. Soltész Katalin (1959): *Az ősi magyar igekötők*. Budapest, Akadémiai Kiadó.
- Juhász Levente – Pléh Csaba (2001): Többmorfémás szavak megértése a magyarban. In: Pléh Csaba – Lukács Ágnes (szerk.): *A magyar morfológia pszicholingvisztikája*. Budapest, BIP – Osiris Kiadó. 11–37.
- Kiefer Ferenc – Ladányi Mária (2000): Az igekötők. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan. 3. Morfológia*. Budapest, Akadémiai Kiadó. 453–518.
- Kiefer Ferenc (2003): A kétféle igemódosítóról. In: *Nyelvtudományi Közlemények 100*. 177–186. <http://www.nytud.hu/nyk/100/kiefer.pdf>
- Komlósy András (1989): Fókuszban az igék. In: Telegdi Zsigmond – Kiefer Ferenc (szerk.): *Általános Nyelvészeti Tanulmányok 17*. Budapest, Akadémiai Kiadó. 171–182.
- Komlósy András (1992): Régensek és vonzatok. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan. 1. Mondattan*. Budapest, Akadémiai Kiadó. 299–527.
- Németh Dezső (2001): A munkamemória szerepe a mondatmegértésben. In: Pléh Csaba – Lukács Ágnes (szerk.): *A magyar morfológia pszicholingvisztikája*. Budapest, BIP – Osiris Kiadó. 83–118.
- Oravecz Csaba – Váradi Tamás – Sass Bálint (2014): The Hungarian Gigaword Corpus. In: *Proceedings of LREC 2014*. Reykjavík, 2014. május 26-31. 1719–1723.  
[http://www.lrec-conf.org/proceedings/lrec2014/pdf/681\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/681_Paper.pdf)
- Pais Dezső (1959): Az igekötők mivoltához és keletkezéséhez. In: *Magyar Nyelv 55*. 181–184.
- Pléh Csaba (1998): *A mondatmegértés a magyar nyelvben. Pszicholingvisztikai kísérletek és modellek*. Budapest, Osiris Kiadó.

- Pléh Csaba – Juhász Levente Zsolt (2001): Az igekötők pszicholingvisztikájáról. In: Bakró-Nagy Marianne – Bánréti Zoltán – É. Kiss Katalin (szerk.): *Újabb tanulmányok a strukturális magyar nyelvtan és nyelvtörténet köréből*. Budapest, Osiris Kiadó. 199–211.
- Prószéky Gábor – Indig Balázs (2015): Magyar szövegek pszicholingvisztikai indíttatású elemzése számítógéppel. In: *Alkalmazott Nyelvtudomány*. 15/1–2. 29–44.  
[http://real.mtak.hu/34383/1/PROSZEKY\\_GABOR\\_INDIG\\_BALAZS\\_Magyar\\_szovegek\\_pszicholingvisztikai\\_indC3ADttatasu\\_elemzese\\_szamitogeppele\\_u.pdf](http://real.mtak.hu/34383/1/PROSZEKY_GABOR_INDIG_BALAZS_Magyar_szovegek_pszicholingvisztikai_indC3ADttatasu_elemzese_szamitogeppele_u.pdf)
- Sass Bálint (2011): *Igei szerkezetek gyakorisági szótára – egy automatikus lexikai kinyerő eljárás és alkalmazása*.  
[http://digitus.itk.ppke.hu/~sass/phd/sass\\_dolgozat.pdf](http://digitus.itk.ppke.hu/~sass/phd/sass_dolgozat.pdf)
- Thuma Orsolya – Pléh Csaba (2001): Kétértelműség és dekompozíció a magyar nyelvben. In: Pléh Csaba – Lukács Ágnes (szerk.): *A magyar morfológia pszicholingvisztikája*. Budapest, BIP – Osiris Kiadó. 39–53.

## 8. Mellékletek

### 8.1. Igekötők százalékos eloszlása a finit igéhez viszonyított pozíciók szerint I.

Igekötő	-4	-3	-2	0	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	+11
által:	-	-	-	99.6	0.4	-	-	-	-	-	-	-	-	-	-
nyilván:	-	-	-	96	4	-	-	-	-	-	-	-	-	-	-
ellen:	-	-	-	93.5	6.5	-	-	-	-	-	-	-	-	-	-
teli:	-	-	-	78.3	21.4	0.4	-	-	-	-	-	-	-	-	-
benn:	-	-	-	65.1	34.8	0	-	-	-	-	-	-	-	-	-
fönn:	-	-	-	62.2	37.6	0.2	-	-	-	-	-	-	-	-	-
fölül:	-	-	0.5	62.9	36	0.6	-	-	-	-	-	-	-	-	-
szerte:	-	-	0.2	90.2	8.8	0.8	-	-	-	-	-	-	-	-	-
alul:	-	-	0	74.3	25.5	0.2	-	-	-	-	-	-	-	-	-
ott:	-	-	0	80.5	19.3	0.1	-	-	-	-	-	-	-	-	-
rajta:	-	-	0.1	29	67.9	3	-	-	-	-	-	-	-	-	-
reá:	-	-	0.5	26.8	69.3	3.4	-	-	-	-	-	-	-	-	-
rajta:	-	-	0.1	29	67.9	3	-	-	-	-	-	-	-	-	-
közé:	-	-	0.6	29.3	67.9	2.3	-	-	-	-	-	-	-	-	-
véghez:	-	-	1.1	47	51	0.9	-	-	-	-	-	-	-	-	-
mellé:	-	-	-	63.7	35	1.3	0	-	-	-	-	-	-	-	-
alább:	-	-	-	70.5	28.7	0.8	0.1	-	-	-	-	-	-	-	-
észre:	-	-	0	88.9	11	0.1	0	-	-	-	-	-	-	-	-
végre:	-	-	0.1	37.4	61.7	0.7	0	-	-	-	-	-	-	-	-
előre:	-	-	0	48.3	51.4	0.4	0	-	-	-	-	-	-	-	-
helyre:	-	-	0.4	76	22.4	1.1	0	-	-	-	-	-	-	-	-
utol:	-	-	1.2	72.5	26.1	0.2	0	-	-	-	-	-	-	-	-
együtt:	-	-	0	79.4	20	0.7	0	-	-	-	-	-	-	-	-
ellent:	-	-	0.2	80.7	17.1	1.9	0.2	-	-	-	-	-	-	-	-
külön:	-	-	0.1	79.3	19.7	0.7	0.1	-	-	-	-	-	-	-	-
neki:	-	-	0.1	57.5	41.7	0.7	-	0	-	-	-	-	-	-	-
fenn:	-	-	0.1	61.8	37.5	0.6	0	0	-	-	-	-	-	-	-
tovább:	-	-	0	39.9	59.5	0.6	0	0	-	-	-	-	-	-	-
végbe:	-	-	0.2	28.8	68.3	2.5	0.1	0.1	-	-	-	-	-	-	-
szembe:	-	-	0.3	65.3	31.9	2.4	0.1	0	-	-	-	-	-	-	-
körül:	-	-	0.1	53.1	44.9	1.9	0.1	0	-	-	-	-	-	-	-
újjá:	-	-	0.1	68.1	31.1	0.7	0	0	-	-	-	-	-	-	-
viszont:	-	-	1.7	56	38.6	3.4	0.2	0.1	-	-	-	-	-	-	-
belé:	-	-	0.2	53.9	43.1	2.5	0.3	0	-	-	-	-	-	-	-
tele:	-	-	0.1	51.3	47.8	0.8	0	0	-	-	-	-	-	-	-
széjjel:	-	-	0.2	55.6	36.9	5.8	1.2	0.2	-	-	-	-	-	-	-
jóvá:	-	0	0.6	70	28.6	0.8	0.1	0	-	-	-	-	-	-	-

A méréseket az MNSZ2 korpuszon végeztem.

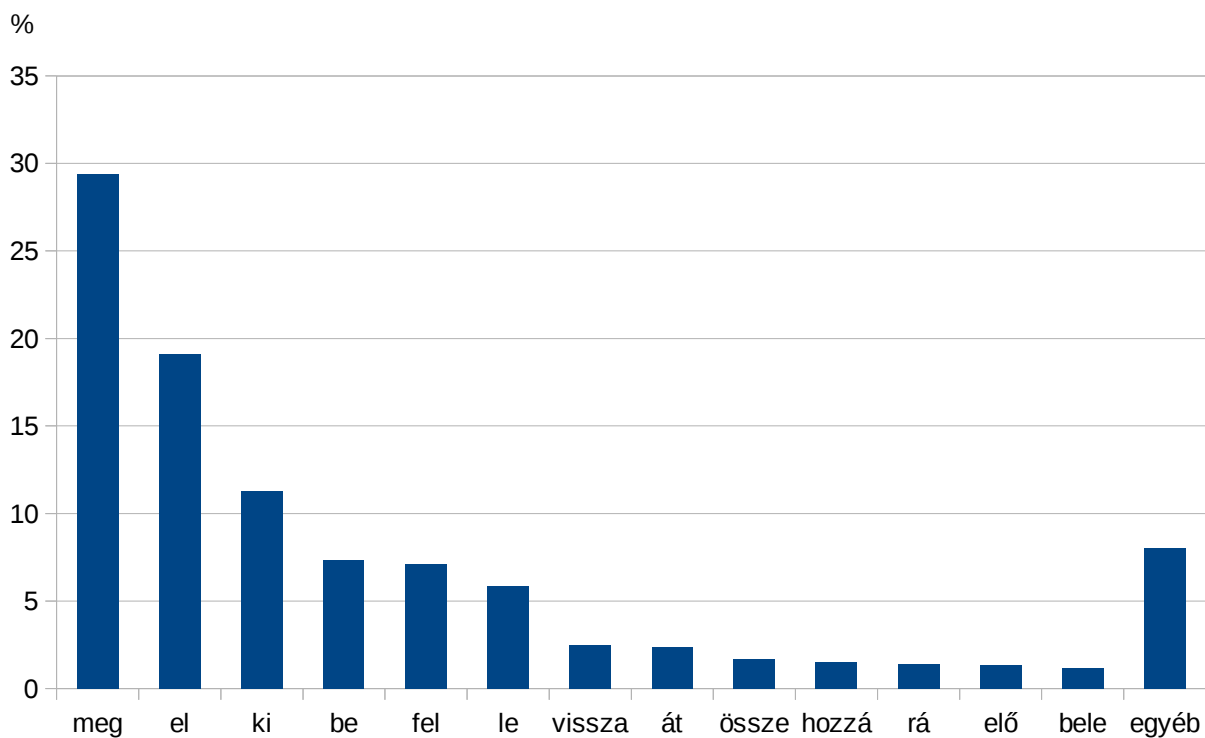
## 8.2. Igekötékek százalékos eloszlása a finit igehez viszonyított pozíciók szerint II.

Igeköté	-4	-3	-2	0	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	+11
közre:	-	-	0.1	80	19.5	0.5	0	-	0	-	-	-	-	-	-
túl:	-	-	0.1	71.1	28.2	0.5	0	0	0	-	-	-	-	-	-
félre:	-	-	0.3	79.2	19.2	1.2	0.1	0	0	-	-	-	-	-	-
félbe:	-	-	0.6	75.9	22.5	0.9	0.1	0	0	-	-	-	-	-	-
szét:	-	0	0.4	69.1	29.5	0.9	0.1	0	0	-	-	-	-	-	-
egybe:	-	-	0.2	73.6	24.7	1.3	0.1	0	0	-	-	-	-	-	-
kölcsön:	-	-	0.7	62.4	35.2	1.5	0.2	0.1	0	-	-	-	-	-	-
hátra:	-	-	0	52.4	44.7	2.8	0	0	0	-	-	-	-	-	-
utána:	-	-	0	57.9	40.9	1.1	0.1	0	0	-	-	-	-	-	-
agyon:	-	-	0.3	73	25.4	1.3	0	0	0	-	-	-	-	-	-
egyet:	-	-	0	87.1	12.1	0.7	0	0	-	0	-	-	-	-	-
újra:	-	-	0	69.2	30.5	0.3	0	0	-	0	-	-	-	-	-
abba:	-	-	0.7	66.6	30	2.6	0.1	0	0	0	-	-	-	-	-
körbe:	-	-	0.3	70.1	28.3	1.1	0.2	0.1	-	0	-	-	-	-	-
közbe:	-	-	0.2	62.2	36.4	1.1	0.1	0	0	0	-	-	-	-	-
ketté:	-	-	0.4	75	23.2	1.2	0.2	0.1	0	0	-	-	-	-	-
tova:	-	-	-	71.6	24.9	2.5	0.5	0.4	0.1	0.1	-	-	-	-	-
elő:	-	0	0.2	52.2	46.3	1.3	0.1	0	0	0	-	-	-	-	-
közze:	-	-	0.5	42.6	55.7	1.2	0	0	-	-	0	-	-	-	-
hozzá:	-	-	0.2	59.9	38.9	1	0.1	0	0	0	0	-	-	-	-
létre:	-	-	0.1	46.2	53.3	0.4	0	-	-	-	0	-	-	-	-
alá:	-	-	0.4	51.7	46.9	1	0	0	0	0	0	-	-	-	-
bele:	-	-	0.5	69.9	28.4	1.1	0.1	0	0	0	0	-	-	-	-
meg:	-	0	0.5	66.8	32.1	0.6	0	0	0	-	-	-	-	-	-
ki:	-	0	0.4	58.7	40.2	0.7	0	0	0	0	-	-	-	-	-
be:	-	0	0.6	58.3	40.4	0.7	0	0	0	0	-	-	-	-	-
le:	-	0	0.6	69.4	29.3	0.6	0	0	0	0	-	-	-	-	-
fel:	-	0	0.5	60.2	38.7	0.7	0	0	0	0	-	-	-	-	-
föl:	-	-	0.4	61.9	36.8	0.9	0	0	0	0	-	-	-	-	-
át:	-	0	0.5	58.5	39.9	1	0	0	0	0	-	-	-	-	-
rá:	-	-	0.2	65.9	32.9	0.9	0	0	0	0	-	-	-	-	-
el:	-	0	0.6	65.7	33	0.7	0	0	0	0	0	-	-	-	-
keresztül:	-	-	0.3	39.5	58.3	1.7	0.1	0	0	-	0	-	-	-	-
keresztbe:	-	-	1	30.1	61.6	6.3	0.6	0.1	0.2	0.1	-	-	-	-	-
végig:	-	-	0.1	66.9	31.9	1.1	0.1	0	0	0	0	-	-	-	-
össze:	-	0	0.4	58.2	39.9	1.5	0.1	0	0	0	0	-	-	-	-
haza:	0	0	0.4	57.8	39.1	2.4	0.2	0.1	0	0	0	-	-	-	-
oda:	-	0	0.6	65.4	31.5	2.3	0.2	0.1	0	0	0	0	-	-	-
ide:	-	0	0.2	36.6	56.8	5.5	0.5	0.2	0.1	0	0	-	0	-	-
tönkre:	-	-	0.5	74.6	23.5	1.2	0.1	0.1	0	0	-	-	-	0	-
vissza:	0	0	0.5	67.4	30.6	1.4	0.1	0	0	0	0	0	0	0	0

A méréseket az MNSZ2 korpuszon végeztem.

### 8.3. Az MNSZ2 igekötő állománya

(csak a preverbális pozícióban lévő elemek)



#### Egyéb:

föl	szembe	közbe	benn
oda	jóvá	ellent	fönn
egyet	ide	alul	szerte
létre	körül	újjá	nyilván
haza	előre	ketté	fölül
szét	tönkre	ellen	alább
túl	abba	utol	viszont
alá	végre	tele	véghez
végig	helyre	félbe	tova
fenn	közzé	keresztül	közben
tovább	együtt	által	széjjel
neki	egybe	kölcsön	keresztbe
újra	ott	belé	teli
felül	körbe	rajta	közé
közre	hátra	mellé	reá
észre	agyon	végbe	kinn
félre	utána	külön	

## 8.4. Dativuszi végződésű melléknevek a finit ige előtt és után

A méréseket az MNSZ2 korpuszon végeztem.

### A finit ige előtt:

I.			MN.DAT	IGE	gyakoriság: 271590
II.		MN.DAT	HA	IGE	2688
III.	MN.DAT	KOT	MN.DAT	IGE	3520

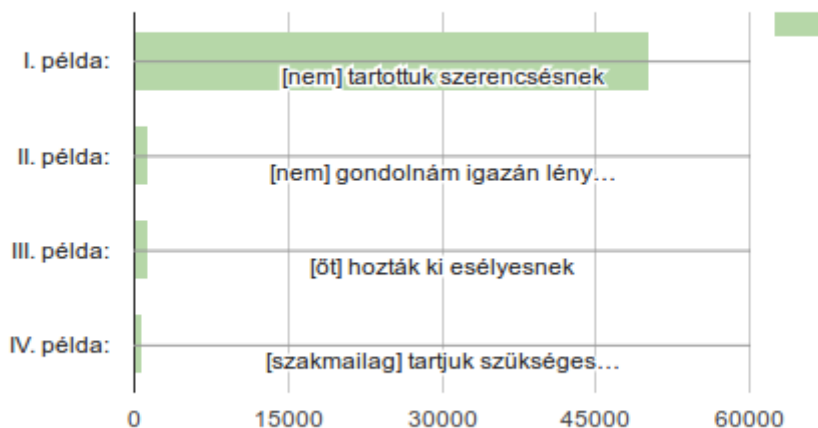
- I. példa: okosnak tartom  
 II. példa: őszintének aligha nevezném  
 III. példa: erősnek és akaratosnak gondolják



### A finit ige után:

I.	IGE	MN.DAT			gyakoriság: 50146
II.	IGE	HA	MN.DAT		1409
III.	IGE	IK	MN.DAT		1281
IV.	IGE	MN.DAT	KOT	MN.DAT	754

- I. példa: [nem] tartottuk szerencsésnek  
 II. példa: [nem] gondolnám igazán lényegesnek  
 III. példa: [őt] hozták ki esélyesnek  
 IV. példa: [szakmailag] tartjuk szükségesnek és indokoltnak





## 8.5. A bal oldali igekötő pozíciók vizsgálatához írt CQL-kifejezések

0	igekötős finit ige
MNSZ2	[msd="IK\.\IGE.*" & msd!="IK\.\IGE\._?(HIN OKEP MIA MIB INF).*"]
-2	nem finit ige + igekötő + [nem (igekötő   írásjel   kötőszó   finit ige)] + finit ige + nem igekötő
MNSZ2	[msd!="IGE\.\S+[123]" [msd="IK" [msd!="(IK SPUNCT WPUNCT KOT IGE\.\S+[123])" [msd="IGE.*" & msd!="IGE\._?(HIN OKEP MIA MIB INF).*" [msd!="IK"
Pázmány	[humor!="IGE\+\S+[123]" [humor="IK" [humor!="(IK PUNCT KOT IGE\+\S+[123])" [humor="IGE.*" & humor!="IGE\+_?(HIN OKEP MIA MIB INF).*" [humor!="IK"
-3	nem finit ige + igekötő + [nem (igekötő   írásjel   kötőszó   finit ige), 2 szó] + finit ige + nem igekötő
MNSZ2	[msd!="IGE\.\S+[123]" [msd="IK" [msd!="(IK SPUNCT WPUNCT KOT IGE\.\S+[123])"]{2} [msd="IGE.*" & msd!="IGE\._?(HIN OKEP MIA MIB INF).*" [msd!="IK"
Pázmány	[humor!="(IK PUNCT KOT IGE\+\S+[123])" [humor="IK" [humor!="(IK PUNCT KOT IGE\+\S+[123])"]{2} [humor="IGE.*" & humor!="IGE\+_?(HIN OKEP MIA MIB INF).*" [humor!="IK"
bal maximum	nem ige + igekötő + [nem (igekötő   írásjel   kötőszó   finit ige), 3-12 szó] + finit ige + nem igekötő
MNSZ2	[msd!="IGE.*" [msd="IK" [msd!="(IK SPUNCT WPUNCT KOT IGE\.\S+[123])"]{3,12} [msd="IGE.*" & msd!="IGE\._?(HIN OKEP MIA MIB INF).*" [msd!="IK"
Pázmány	[humor!="IGE.*" [humor="IK" [humor!="(IK PUNCT KOT IGE\+\S+[123])"]{3,12} [humor="IGE.*" & humor!="IGE\+_?(HIN OKEP MIA MIB INF).*" [humor!="IK"

## 8.6. A jobb oldali igeekötő pozíciók vizsgálatához írt CQL-kifejezések

+1	<p>finit ige + igeekötő</p> <p>MNSZ2 [msd="IGE.*" &amp; msd!="IGE\._?(HIN OKEP MIA MIB INF).*"] [msd="IK"]</p> <p>Pázmány [humor="IGE.*" &amp; humor!="IGE\._?(HIN OKEP MIA MIB INF).*"] [humor="IK"]</p>
+2	<p>finit ige + [nem (igeekötő   írásjel   kötőszó   finit ige) és a szó nem 'a' (határozott névelő)] + igeekötő</p> <p>MNSZ2 [msd="IGE.*" &amp; msd!="IGE\._?(HIN OKEP MIA MIB INF).*"] [msd!="(IK SPUNCT WPUNCT KOT IGE\.\S+[123])" &amp; word!="a"] [msd="IK"]</p> <p>Pázmány [humor="IGE.*" &amp; humor!="IGE\._?(HIN OKEP MIA MIB INF).*"] [humor!="(IK PUNCT KOT IGE\.\S+[123])" &amp; word!="a"] [humor="IK"]</p>
+3	<p>finit ige + [nem (igeekötő   írásjel   kötőszó   finit ige)] + [nem (igeekötő   írásjel   kötőszó   finit ige) és a szó nem 'a' (határozott névelő)] + igeekötő</p> <p>MNSZ2 [msd="IGE.*" &amp; msd!="IGE\._?(HIN OKEP MIA MIB INF).*"] [msd!="(IK SPUNCT WPUNCT KOT IGE\.\S+[123])"] [msd!="(IK SPUNCT WPUNCT KOT IGE\.\S+[123])" &amp; word!="a"] [msd="IK"]</p> <p>Pázmány [humor="IGE.*" &amp; humor!="IGE\._?(HIN OKEP MIA MIB INF).*"] [humor!="(IK PUNCT KOT IGE\.\S+[123])"] [humor!="(IK PUNCT KOT IGE\.\S+[123])" &amp; word!="a"] [humor="IK"]</p>
jobb maximum 12 szó	<p>finit ige + [nem (igeekötő   írásjel   kötőszó   finit ige), 2-12 szó] + [nem (igeekötő   írásjel   kötőszó   finit ige) és a szó nem 'a' (határozott névelő)] + igeekötő</p> <p>MNSZ2 [msd="IGE.*" &amp; msd!="IGE\._?(HIN OKEP MIA MIB INF).*"] [msd!="(IK SPUNCT WPUNCT KOT IGE\.\S+[123])"]{2,12} [msd!="(IK SPUNCT WPUNCT KOT IGE\.\S+[123])" &amp; word!="a"] [msd="IK"]</p> <p>Pázmány [humor="IGE.*" &amp; humor!="IGE\._?(HIN OKEP MIA MIB INF).*"] [humor!="(IK PUNCT KOT IGE\.\S+[123])"]{2,12} [humor!="(IK PUNCT KOT IGE\.\S+[123])" &amp; word!="a"] [humor="IK"]</p>