

AZ IGEKÖTŐS IGEK SZINTAXISA KORPUSZVEZÉRELT MEGKÖZELÍTÉSBEN

Kalivoda Ágnes

Bevezetés

Tanulmányomban rendhagyó módon, korpuszvezérelt módszerrel vizsgálom az igekötős igék szintaktikai viselkedését. A korpuszvezéreltség Sass (2011: 15–16) alapján azt jelenti, hogy nem volt kiinduló hipotézisem, kizárólag korpuszadatok alapján fogalmaztam meg a következtetéseimet. Több mint 21,5 millió mondatban mértem ki a finit ige és a hozzá tartozó igekötő távolságát. Jelen tanulmány ennek a korpuszmérésnek az eredményeit és az eredmények elméleti tanulságait ismerteti.

A kutatás mögött számítógépes nyelvészeti motiváció áll, a mérések az ANAGRAMMA gépi elemző számára készültek, l. Indig–Laki–Prószéky (2016). Az elemző balról jobbra haladva, szavanként – a szófaji annotációt is ismerve – dolgozza fel a mondatot. Egy ilyen modellben fontos tudni, hogy elvált igekötő esetén mire készüljön az elemző. Ha például beolvasott egy finit igét, meddig szükséges várnia a posztverbális igekötőre, várhat-e egyáltalán igekötőt a mondatban? Posztverbális igekötő esetén az elemző nem tudhatja biztosan a vonzatkeretet, amíg az igekötőt be nem olvassa, vagy annak lehetőségét el nem veti.

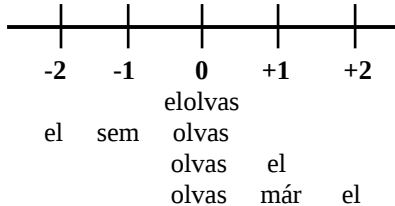
A tanulmány három kérdés megválaszolására törekszik: 1) Milyen mondat típusokra, szerkezetekre jellemző az igekötő-ige, illetve ige-igekötő sorrend, és milyen arányban szerepelnek a korpuszban a különféle szerkezetek? 2) Melyek a finit ige és a posztverbális igekötő távolságának lehetséges okai, és melyik közülük a legmeghatározóbb? 3) Mit mondanak a korpuszból nyert adatok a magyar igekötő-állomány összetételéről?

1. Módszertan

A korpuszmérést a Magyar Nemzeti Szövegtár bővített változatán (MNSZ2 2.0.3) végeztem. Ez a verzió írásjelek nélkül 785 millió szövegszót tartalmaz, található benne sajtószövegek, szépirodalom, tudományos, személyes és hivatalos írások, beszélt szövegek átírtai, l. Oravecz–Váradi–Sass (2014).

A korpuszból lekértem minden olyan mondatot, amely igekötős finit igét tartalmaz: az igekötő lehetett preverbális (az igétől balra elvált vagy az igével egybeírt), illetve posztverbális. A hibás találatok szűrése után több mint 21,5

millió mondat maradt, amelyen tovább lehetett dolgozni. Az igekötők pozícióját automatikusan határoztam meg úgy, hogy a mondatokat egy számegegyenesre képeztem le, ahol az origó a finit ige volt. A negatív tartományba kerültek a preverbális, a pozitív tartományba a posztverbális igekötők (a módszer alapjához l. 1. ábra).



1. ábra: Az igekötők pozíciójának meghatározása a finit igehez képest¹

Az MNSZ2-n végzett méréseim során nem bíraltam felül a korpusz annotációját, tehát az ott igekötőnek jelölt, a szakirodalom jelentős része szerint nem igekötő *létre, észre* szavak ugyanúgy bekerültek az elemzésbe, mint a *ki, le, fel* stb. 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 korpuszmérésben csak az egytagú igekötők szerepelnek (80 darab), a vizsgált 21,5 millió mondat pedig csak olyan igekötő+finít ige kombinációkat tartalmaz, ahol az igekötő és a finit ige egy tagmondaton belül szerepelnek, és összetartozásuk automatikus módszerrel eldönthető. Így például nem részei a vizsgálatnak az olyan komplex szerkezetek, mint a fölérendelt tagmondatba kiemelt igekötők (pl. *el kell, hogy mondjam*), valamint az infinitívuszos szerkezetekben szereplő „gazdátlan igekötők” (pl. *elképzelni bármit el lehet*).

2. Problémák

A korpuszmérés megvalósítását nehezítette az igekötők gyakran hibás gépi annotációja, valamint egy elméleti jellegű probléma: egyáltalán mely szavakat soroljuk az igekötők csoportjába?

¹ A -1 pozícióban csak akkor jelenik meg egytagú igekötő, ha a mondat írója hibásan gépelte a szót (pl. *agyonverte* helyett *agyon verte*), vagy ha a finit ige segédigei szerepet tölt be a mondatban, és az igekötő valamilyen igenévhez tartozik. Emiatt a -1 pozíció nem része a kiértékelésnek.

2.1. Technikai nehézségek

A módszertani fejezetben utaltam rá, hogy szűrni kellett a hibás találatokat, amely meglehetősen összetett feladat volt.² Egyrészt meg kellett találni az olyan mondatokat, ahol valójában igenévé az igekötő (ez nem triviális akkor, ha elvileg mindkét igével állhatna, pl. a *verni el nem mentek* esetében *el+ver* és *el+megy* is létező kombináció). Másrészt fel kellett ismerni a téves szófaji címkéket. Az automatikus szófaji egyértelműsítésben komoly hibaforrást jelentenek a homográf szavak és az igekötők esetében kiugróan magas a tévesztések aránya. Három esetben kaphatnak hibás szófaji címkét:

1) Elírások automatikus elemzésekor. Ilyen a felesleges szóköz (pl. *meg nézett*), ékezetek hiánya (pl. *fél* → *fel*) vagy internetes szövegekben gyakori rövidítések (pl. *megláttam* → *+láttam*).

2) Ha az igekötő más nyelvek szavaival homográf. Gyakori eset, hogy az angol *be* létige, a spanyol *el* és a francia *le* határozott névelők igekötőként jelennek meg a korpuszban.

3) Ha az igekötőnek homonim párja van a magyarban. Pl. *akkor csak lámpát kell vennem meg rácsot, az mennyibe fog kerülni és ki fogja rá adni a pénzt* – a *meg* és *ki* mindkét példában (tévesen) igekötőként annotáltak. A legtöbb igekötő átfedésben van egy hozzá közel álló szófajjal. Ilyenek az idiómáriszóból kialakult igekötők (pl. *tönkre*), amelyek sokszor esetragos főnévként annotáltak, valamint az olyanok, amelyek határozószóként, névutóként is funkcionálnak (pl. *együtt*). Számos ilyen szó igekötői státuszáról megoszlik a szakirodalom véleménye, a bizonytalanságot az annotáció következetlensége is tükrözi.

2.2. Az igekötő-állomány meghatározásának nehézségei

Komlósy (1992: 495–497) táblázata jól mutatja a bizonytalanságot abban a kérdésben, hogy mely szavak sorolhatók az igekötők közé. 90 szót listáz, amely esélyes az igekötői címre, és látható, hogy egyes források mennyire eltérően vélekednek ezekről a szavakról. Ahogy Komlósy is megfogalmazza, néhol még egy forráson belül sem követhető nyomon a besorolás logikája.

Forgács (2005: 11) szakirodalmi áttekintése két kritériumot talál 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 alkothatók vele), míg a *cserben*-t nem szokás annak tekinteni (csak a *cserbenhagy* igében fordul elő).

² A probléma és a kísérleti jellegű megoldások részletesebb ismertetéséhez l. Kalivoda (2016).

2. A szemantikailag transzparens szavakat, 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–Ladányi (2000) sem. Forgács ezt bírálja, arra hivatkozva, hogy a *haza*- előtag egyrészt nagyon produktív, másrészt átvitt értelmű használata is van (*hazavág* 'elront, megöl', *hazabeszél* 'saját maga vagy a magához közel állók érdemeit hangoztatja').

Számos szakirodalom mutat be tesztek, amelyek dönthetnek egy-egy szó igekötő-voltáról.³ Forgács (2005) a – hagyományosnak mondható – szükséges és elégséges feltételek megállapítása helyett 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. Az igekötők eredetét, nyelvtörténeti útját szem előtt tartva határozza meg az egyes szakaszokat. Ilyen prototípus-elméleti megközelítést alkalmaz Kerekes (2011) is.

Ahogy a későbbiekben látni fogjuk, a jelen kutatás ezt az elképzelést, tehát a lehetséges igekötők skálán való elhelyezését támasztja alá.

3. Preverbális igekötők

Az igekötők kanonikus helye a közvetlenül preverbális pozíció. Az igtől elvált preverbális igekötőket intuitívan is jóval ritkábbnak vélhetjük, mert a magyar mondatok bal perifériáján az összetevők elrendeződése sokkal megszorítottabb, mint az ige utáni mondatszakaszban.

3.1. A preverbális igekötők pozíció szerinti eloszlása

A korpuszadatokban szélsőértéknek az igekötő -4-es pozíciója bizonyult, vagyis az igekötők és a finit ige között legfeljebb három másik szó állt.

1. táblázat: Lehetséges igekötő pozíciók a finit ige bal oldalán

	-4	-3	-2	0
db	6	91	102768	13715093
%	-	-	0,74	99,26

³ Az igekötő-állomány meghatározásáról bővebben l. Pais (1959), Jakab (1976), Kiefer–Ladányi (2000).

Az 1. táblázatban látható, hogy a közvetlenül preverbális helyzetű igekötők aránya kiemelkedően magas: több mint 13,7 millió találattal a 0 pozíció fedile a preverbális esetek 99,26%-át. A következő olyan pozíció, ahol a finit igéhez tartozó igekötő állhat, a -2. Ide is-szerkezet (1a) vagy annak tagadása (1b, 1c), tagadó felszólítás, tiltás (1d) vagy tagadás (1e) esetén kerül az igekötő. A korpusz mondataiban az *is*, *sem*, *se*, *ne*, *nem* szavak választották el egymástól az igekötőt és a finit igét.

- (1) 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.*

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 bemutatott, öt elemű halmazból való: *nem is*, *már nem*, *még nem*, *soha nem*, *ugyan nem*, *sem nagyon*, *sem igen*, *se nagyon*, *biztosan nem*, *valóban nem...* Az MNSZ2-ben két olyan példa volt, amely az általánosításon kívül esik:

- (2) a. **Le** csak azért **taszították** mert Jézus áldozata a keresztfán megmentette a világegyetemet a további megkísértéstől [...]
b. **Le** a számokkal **írjunk** képleteket.⁴

A legszélső, -4 pozícióra mindössze hat példa akadt a korpuszban, és a (3e) kivételével mindegyik esetében kétséges, hogy a *haza* és *vissza* szavak igekötők vagy határozószók.

- (3) a. *Mentünk baška vodára nyaralni elég körülményesen **haza** is elég körülményesen **jöttünk** [...]*
b. *Nem tervezek itt maradni, de **haza** is csak látogatni **megyek** ezentúl.*
c. *Elcsörömpöl s **vissza** már egyre ritkábban **fordítja** a rudat...*
d. *[...] folyvást előretör **vissza** az istennek se **nézne** [...]*
e. *Vízont **vissza** most már nem **szívhatja** szavát!*
f. *A lapos papucscsónakkal ment, de **vissza** már nem egyedül **érkezett** a vendéglőtől [...]*

Ahogy az igekötő-állomány meghatározási nehézségeinél szó volt róla, a *haza* a szemantikai átlátszósága miatt több szakirodalomban kimarad az

⁴ A kontextus alapján valószínű, hogy itt egyszerű vesszőhibáról van szó, és a szerző szándéka ez lehetett: *Le a számokkal, írjunk képleteket*. Az ilyen típusú hibák kiszűrése egyelőre nem megoldott, és valószínűleg nem is sikerül általánosan jó megoldást találni rá.

igekötők listájából. Ezekben a példákban valóban felfogható hangsúlyos határozószóként is. A vissza jellemzően bekerül az igekötő-kategóriába.

3.2. Igekötő-ige szórendű mondatok

A közvetlenül preverbális igekötőt tartalmazó mondatok körülbelül⁵ 93%-át semleges mondatok alkotják, amelyeknek ez a szórend az egyik alapvető ismérve.⁶ Alberti (2006: 37–38) három nyelvészeti terület irányából közelíti meg a semleges mondat fogalmát. Szemantikai szempontból a semleges mondat „jelentése pusztán a lexikai jelentéstényezőkből áll össze, nincsen szerkezeti többletjelentés, vagy óvatosabban fogalmazva a szerkezetből adódó többletjelentés 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 hordozó elemek, „a szórendi és intonációs sajátosságok nem hoznak további előfeltevéseket”.

A nem-semleges mondatban ezzel szemben a hangsúly eloszlása nem egyenletes, és speciális jelentéstöbbletet hordoz, amely szerkezetileg is megjelenhet, l. Alberti (2006: 54). Ekkor sokkal inkább az ige-igekötő szórend jellemző, de találunk nem-semleges mondatípust közvetlenül preverbális igekötővel is. Ilyen az eldöntendő kérdés (a 0 pozíció kb. 6,4%-át alkotja) és az igei fókusz tartalmazó mondat.⁷

A felszólító mondatok egy stilisztikailag jelölt típusára is az igekötő-ige szórend jellemző. Ezek gyakran erős érzelmi többletet fejeznek ki (pl. *Elhallgass végre!*, *Eltakarodjatok a szemem elől!*). Összesen 4547 példát találunk erre a 0 pozíció adataiban, ez 0,03%-nak felel meg.

A -2 pozíció esetén *is*-szerkezetek, tiltó és tagadó mondatok jelennek meg, ahogy az előző alfejezetben láthattuk. A -3 és -4 pozícióban általában kontrasztív topik szerepű szavak állnak, vö. É. Kiss (2003: 6), amelyek esetében kérdéses és nem is feltétlenül eldönthető, hogy határozószók vagy valóban igekötők.

⁵ A mondatípusok meghatározása automatikus módszerrel, reguláris kifejezések használatával történt, amelyek a mondatípustól függően az ige morfológiai annotációjára, ritkábban szóalakra és/vagy szótőre illeszkedtek. Az automatizáltság miatt csak körülbelüli értékekről beszélhetünk.

⁶ Létezik „fordított”, ige-igekötő sorrend semleges mondatban is, l. a Peredy (2011) által prezentált mondatnak nevezett szerkezetek. Pl. *Akkor töszik föl az aszalóra, füstölik meg, vágod hilkára [szelet], s eszed meg.* (csángó nyelvjárási adat). Ilyen típusú mondatokat nem találtam az MNSZ2-ben.

⁷ Ennek a gyakorisága automatikus módszerrel nem állapítható meg (valószínűleg a semleges mondatoknál látható kb. 93%-ot csökkentené 1-2%-kal).

4. Posztverbális igekötők

Az igekötő gyakran elmozdul a privilegizált, közvetlenül ige előtti pozíciójából például tagadás, felszólítás vagy más összetevő fókuszba emelése esetén. Érdekes adat, hogy +1 pozíciójú igekötőt több mint 7,5 millió mondatban találunk az MNSZ2-ben, ez a 0 pozíció kétharmada, azaz figyelemre méltóan gyakori.

4.1. A posztverbális igekötők pozíció szerinti eloszlása

A 2. táblázat mutatja azt a tendenciát, hogy az igekötők jóval nagyobb intervallumon helyezkedhetnek el a finit ige utáni mondatszakaszban, mint az ige előtt, de a +1 pozíció az esetek 97,98%-át lefedi.

2. táblázat: *Lehetséges igekötő-pozíciók a finit ige jobb oldalán*

	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	+11
db	7527308	163993	5126	1193	267	101	27	5	3	2	1
%	97,98	2,13	0,07	0,02	-	-	-	-	-	-	-

A (4a-k) példásor egy-egy mondat illusztrál minden lehetséges pozíciót:

- (4) a. *Ennek nyomán 17 megyében **nyújtottak be** felülvizsgálati kérelmet a Legfelsőbb Bírósághoz.* +1
 b. – **Heverj csak el** te is nyugodtan – *mondta Pilátus.* +2
 c. *Akkor hogy **függenek** itt most **össze** a szándékok és a bekövetkező események?* +3
 d. *Én mindenesetre Dancsó úrral **értek** ebben a dologban **egyed**.* +4
 e. *Háromszor **veri** ezt kenden Lúdas Matyi **vissza!*** +5
 f. *Azért **mentem** egy kicsit a popzene felé **el**, mert szeretem a nívós, könnyed, jó popzenét, ami abszolút mindig keveredik a jazzal.* +6
 g. *Tehát ezért **függenek** ezek a folyamatok egymással nagyon szorosan **össze** és ezért adtuk ezt a nagyon provokatív címet a kötetnek.* +7
 h. *Miért **jár** egy olyan kicsi magyar falunak ezer hektár **vissza**, a sokkal nagyobbaknak meg alig valami, vagy éppen semmi?* +8
 i. *Azt mondják, száz nappól tizenöt-húszban **sodor** a szél dögszagot a sárvári ATEV üzem felől **ki** az országútra.* +9
 j. *Kerestem tehát az okát annak, hogy miért **teszi** megannyi egyébként oly sikeresnek látszó ember saját magát szántszándékkal **tönkre**.* +10
 k. *27 gyereket **vitt** egy feltehetően részeg buszsofőr Szentesen még csütörtökön egy sportrendezvény után **vissza** az iskolába.* +11

A (4i) és (4k) példákban egy helyhatározó áll közvetlenül az igekötő előtt, illetve után, így felmerülhet az a probléma, hogy az igekötő sokkal inkább a helyhatározóhoz, mintsem a finit igéhez kötődik. További példákhoz l. Alberti (2006: 50).

4.2. Ige-igekötő szórendű mondatok

Az ige-igekötő szórendű mondatok – a 6-os lábjegyzetben említett, nagyon ritka példák kivételével – nem-semlegesek. Alberti (2006)-ot alapul véve a jelen fejezet csak olyan nem-semleges mondat típusokat sorol föl, amelyekre az igekötő-ige sorrend felcserélődése jellemző, és viszonylag jól megoldható volt a mondat típus számítógépes azonosítása. Nyolc kategóriát különíték el, amelyek a következők:

1. Mondat- és mondatrésztagadások (beleértve a beágyazott tagadó mondatokat is). Számítógépesen ez nagyon jól formalizálható kategória. Átlagosan 15,7%-ban fordul elő a posztverbális igekötők korpuszában.

2. Felszólító mondatok (ide értve a beágyazott felszólítást is). Ez könnyen kereshető annak köszönhetően, hogy az imperatívusz morfológiailag is jelölt. Alberti (2006: 73) amellet érvel, hogy a felszólítás az aspektustól független, tehát progresszívben is kifejezhetünk felszólítást. Erre a következő példát hozza: *Épp másszunk fel a bordásfalra, amikor belép a tanár!* Kétségtelenül grammatikus szerkezet, de az MNSZ2-ben nincs példa hasonlóra. A felszólító mondatok gyakoriak, átlagosan 10,5%-ban szerepelnek az egyes pozíciókon.

3. Óhajtó mondatok, amelyekben a mondatszerkezet „egy reménytelenebb alapállást tükröz” (Imrényi 2011: 97) azzal kapcsolatban, hogy a kívánt esemény bekövetkezhet-e. Pl. *Csak hívna már fel!* Ezt a feltételes mód és a *csak* együttes jelenléte esetén lehetett automatikusan kiszűrni.

4. Tiltó mondatok. Ezt a kategóriát csak a könnyű számítógépes kereshetősége miatt emeltem ki külön. Imrényi (2011) szavaival: ez egy operátorláncot tartalmazó szerkezet, egy tagadó és felszólító operátor-projekció épül egymásra. Pozícióként átlagosan 2,3%-ban jelenik meg.

5. Kérdőszavas (kiegészítendő) kérdések és felkiáltó mondatok. Ezek nyelvészeti szempontból teljesen különváló kategóriák – jellegzetesen eltérő intonációs mintával –, de az írott nyelvi korpuszban egyformák. A kérdő-

illetve felkiáltójel természetesen segíthet a megkülönböztetésben (pl. *Milyen későn kelt fel!* → *Milyen későn kelt fel?*), de csak akkor, ha nem egy másik mondatba ágyazott kérdésről vagy felkiáltásról van szó.

6. Progresszív aspektusú mondatok, ahol azon van a hangsúly, hogy éppen történt az adott cselekvés. Ezt a következő mintával lehetett részlegesen kiszűrni: 'épp|éppen + kijelentő módú ige' vagy 'kijelentő módú ige + épp|éppen'. A +1 pozíció esetén 191, +2 esetén 3 progresszív mondatot lehetett találni. Néhány példa:

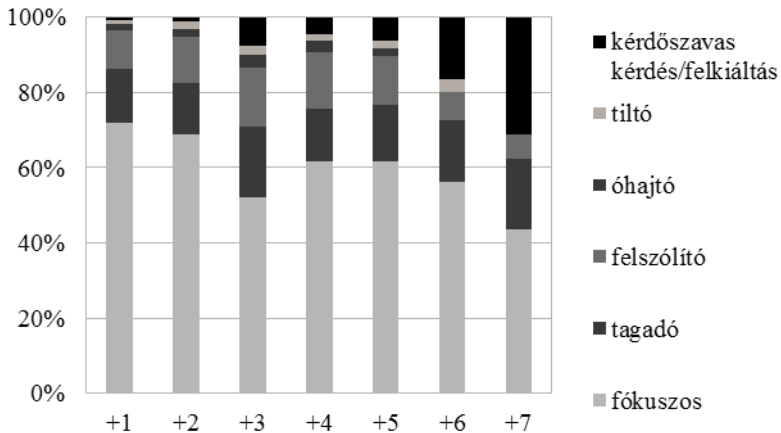
- (5) a. **Verik le a képet!**
b. **Rakom ki szőttesruhába, zászlórongyba, emlékiratba csomagolva ajándék könyveim s fegyvereim.**

7. Egzisztenciális mondatok, vö. É. Kiss (1992). A speciális jelentéstöbbletük az, hogy valamely esemény már legalább egyszer előfordult. Részleges automatikus azonosítását a 'mondatkezdet + kijelentő módú ige + már' minta teszi lehetővé. A korpuszban nagyon ritka szerkezet. Néhány példa:

- (6) a. **Rúgott már meg ló?**
b. **Vesztitettem már két saját könyvet el, igen, de még nem találta meg senki sem.**
c. **Kaptak már így el csoportokat, akik ebből élnek?**

8. Szerkezeti fókusz tartalmazó mondatok. Ez gyakoriságát tekintve kiemelkedő, számítógépes azonosíthatóságát tekintve a legproblémásabb mondat típus, mivel gyakorlatilag bármilyen szófajú szó fókuszba kerülhet. A kutatás jelen szakaszában minden olyan posztverbális igekötőt tartalmazó mondatot, amelyet a reguláris kifejezések az előző 7 kategória egyikébe sem tudnak besorolni, fókuszos mondatnak tekintek.

Az egyes mondat típusok arányát a posztverbális korpuszban a 2. ábra szemlélteti. Természetesen itt csak közelítő értékekről beszélhetünk. Érdekes, hogy a +1 pozíciójú igekötők korpuszában a kérdő és/vagy felkiáltó mondatok csak 0,6%-ban vannak jelen, míg +6 pozíció esetén már a mondatok 15,6%-át, +7 esetében pedig 33,3%-át teszik ki (fontos viszont az is, hogy itt már sokkal kevesebb vizsgálható mondat állt rendelkezésre). Ezekben a mondat típusokban gyakoriak az ige és az igekötő közé beszúrt diskurzuspartikulák és indulatszók.



2. ábra: Ige-igekötő szórendű mondatok aránya az egyes pozícióknál⁸

4.3. Jellemzően posztverbális igekötővel álló igék

A korpuszvezérelt vizsgálat felszínre hozta az igekötős igék egy olyan csoportját, amelynél az igekötő jellemzően posztverbális. Ennél a mérésnél olyan igéket vettem figyelembe, amelyek legalább 1000-szer fordultak elő az MNSZ2-ben, és az igekötő minimum 60%-kal gyakrabban állt az ige után, mint előtt (a mérésbe természetesen a 0 pozíció is beleszámított). A 3. táblázat tíz kimagasló találatot tartalmaz.

3. táblázat: *Igék, amelyek jellemzően posztverbális igekötővel állnak (a százalékjel a posztverbális igekötők arányát jelenti az adott ige esetén)*

Ige	%
1. dolgoz fel	98,6
2. tevődik össze	97,7
3. engedhet meg	96
4. áll hozzá	92,6
5. ér el	90,8
6. helyezkedik el	90,8
7. folytat tovább	90,7
8. könyvelhet el	90,7
9. indul ki	89,6
10. ítél meg	88,9

⁸ A progresszív és egzisztenciális mondatok 1% alatti gyakoriságúak, ezért nem szerepelnek az ábrán.

A 3. táblázat igéi között találunk hangsúlykerülő igéket (pl. *tevődik össze, helyezkedik el*), emellett sok esetben gyakori mondatsémáról van szó, amelyben az adott igekötős igét használjuk (pl. *jól/rosszul áll hozzá, abból indul ki*).

4.4. Ige-igekötő távolságok és lehetséges okaik

A 2. táblázat tanúsága szerint a posztverbális igekötők 99,9%-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)). Alberti (2006: 82) is hasonló gondolatot fogalmaz meg: „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 igétől.”

Az igekötők szerepet játszanak az ige irányította mondatfeldolgozásban, argumentumok keresésében, l. Pléh (1998: 133–135), ezért intuitíven is számítani lehetett rá, 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. A testesebb, részben idiomatikus igekötők (pl. *agyon, tönkre, ketté*) kevesebb igével állhatnak, az elhangzott kontextusból rendszerint kiszámítható, hogy ezek az igekötők fognak még elhangzani. Ezek távolabb kerülhetnek az igétől (a *tönkre*, ahogy a (4j) példában láttuk, +10 pozícióban is előfordul.)

Fontos lenne magyarázatot adni az adatok 0,1%-ára is, tehát azokra, amelyekben az ige és az igekötő között kettő vagy több szó áll. Két szempontból⁹ elemzem a korpusz adatait:

1) A meta-adatok, főként írott és beszélt nyelvi oppozíció, valamint a stílusregiszter alapján.

2) Fonológiai szempontból, amelyet a szavak hosszának kimérésével lehetett megvalósítani.

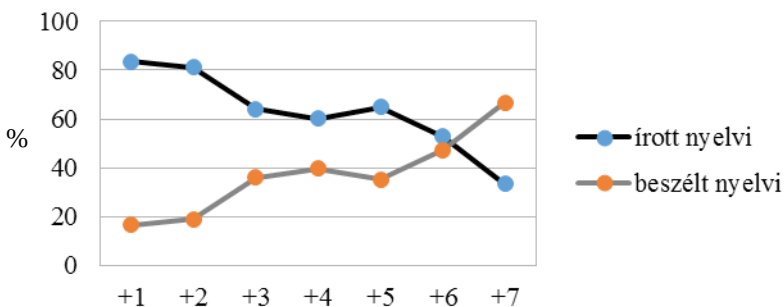
Írott és beszélt nyelvi stílus. A szóbeli megnyilatkozások alapvetően tervezetlenek, spontának az írásbeliekhez képest, így szintaktikailag kevésbé szabatosak, vö. Goody (1987: 258–289). Ezért arra számíthatunk, hogy a szélsőbb pozíciós igekötőt tartalmazó mondatok többsége eredetileg szóban hangzott

⁹ Érdemes lenne szemantikai szempontból is megvizsgálni a posztverbális igekötős igéket, mert intuitívan az is szerepet játszhat az igekötő eltávolodásának mértékében, hogy szemantikailag mennyire transzparens az adott igekötős ige. Ennek megvalósítása szisztematikusan, számítóképes módszerekkel egyelőre túl nagy vállalkozásnak tűnt.

el. Az MNSZ2 meta-adatai 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 3. ábra azt mutatja be, hogy az egyes pozícióknál milyen arányban reprezentáltak az írott, illetve beszélt nyelvi szövegek.



3. ábra: Írott és beszélt nyelvi szövegek százalékos aránya pozíciónként

Az adatok értelmezéséhez tudni kell, hogy a teljes MNSZ2-ben a beszélt nyelvi adat nagyon kevés, vö. Oravecz és mtsai (2014): a rádiós szövegek a teljes szövegállomány 5,4%-át alkotják, és ha a hivatalos szövegek felét szóban elhangzottnak tekintjük, a végeredmény akkor sem több 9,4%-nál. A +1 pozíció szövegtípus szerinti megoszlása még nem tér el nagyban a teljes MNSZ2-től: lényegesen több írott szöveg a forrása (83,55%), mint beszélt nyelvi (16,45%). A beszélt nyelvi szövegek aránya viszont folyamatosan nő: minél távolabbi igekötő-pozíciót vizsgálunk, annál valószínűbb, hogy szóban hangzott el a forrásmondat. A +7 pozíciónál már a mondatok kétharmada szóbeli, egyharmada írott szövegből származik. Alább a (7a-c) mondatrészek országgyűlési felszólalásban, a (7d-f) példák rádióműsorban hangzottak el:

- (7) 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. **járulhatott** Ön szerint ehhez a szörnyű tethez **hozzá**

Az írásbeli szövegek esetében a szépirodalom mutat monoton növekedő tendenciát: míg +1 pozícióban a mondatok 6,32%-a származik irodalmi műből, +6 pozíció esetén ez az érték már 51,72%. Azt mondhatjuk tehát, hogy a szöveg típus és a posztverbális igekötő lehetséges eltávolodása összefüggenek: szerkesztetlen, főleg szóban elhangzott szöveg, valamint művészien szerkesztett irodalmi szövegek esetén valószínűbb, hogy az igekötő eltávolodik a finit igtől.

Fonológiai tényezők. Otto Behaghel német nyelvész a *Deutsche Syntax* negyedik kötetében öt törvényt – pontosabban tendenciát – fogalmazott meg, amelyekkel magyarázatot kívánt adni a szavak és frázisok mondatbeli elhelyezkedésére. A tendenciák Behaghel (1932: 4–8) alapján röviden a következők:

1. Az értelmükben szorosan összetartozó összetevők a mondatban is egymás mellett állnak (pl. *sehr schön* 'nagyon szép').

2. Az új vagy fontosabb információ később hangzik el, mint ami kevésbé fontos vagy már ismert.

3. A megkülönböztető megelőzi a megkülönböztetettet (pl. a határozószó megelőzi a melléknévet, melléknévi igenevet, amelyre vonatkozik).

4. A rövidebb összetevő megelőzi a hosszabbat, ha ezt szintaktikai szabály nem gátolja.¹⁰

5. A hangsúlyos és hangsúlytalan szavak jellemzően váltakozva követik egymást (pl. *ein lieber alter Freund* 'egy kedves, régi barát' a németben hangsúlytalan–hangsúlyos–hangsúlytalan–hangsúlyos sémát követ.)

É. Kiss (2007) amellet érvel, hogy a 4. tendencia – a növekvő összetevők törvénye – a magyarban az ige utáni mondatszakaszra érvényes, az ige utáni, alapvetően szabad szórendet befolyásolja az, ha az igt követő összetevők fonológiailag különböző súlyúak. Ezt többek között a (8) példával szemlélteti. Az ideális sorrend a (8)-ban finit ige után az 'igekötő, határozó, tárgy', egyéb kombinációk legalábbis szokatlanul hangzanának.

(8) *János sértette meg nagyon Éva nagyapját.*

¹⁰ Ezt a jelenséget Pāṇini, a 4. században élt szanszkrit grammatikus is felismerte, a szakirodalomban Pāṇini törvényeként is hivatkoznak rá, vö. Cooper–Ross (1975: 71), O'Connor (1978: 97–99). Ő a szanszkrit szóösszetételek vonatkozásában figyelte meg azt a tendenciát, hogy két összetételi tag közül jellemzően a rövidebb kerül előrébb.

A bemutatott tendencia jól látszik a posztverbális igekötőt tartalmazó mondatokban. A mérést a következőképp végeztem el: Az olyan mondatokban, amelyek +1, +2 és +3 pozícióban tartalmaztak igekötőt, kimértem az ige után álló három szó hosszát (karakterzámban), és átlagoltam az egyes pozíciók hossz-értékeit. Az eredményeket a 4. táblázat foglalja össze.

4. táblázat: A növekvő összetevők törvénye – számokban

Igekötő helye	Szavak átlaghossza (+1-től +3-ig)			Gyakoriság (%)	A finit ige és az igekötő közt álló gyakori szavak
+1	2,8	4,1	4,9	97,98	∅
+2	3,0	3,2	4,5	2,13	-e, is, már, csak, még ...
+3	3,3	5,2	4,1	0,07	egy kicsit, én is, -e majd ...

A nagyon gyakori (97,98 százalékos), +1 pozíciójú igekötőt tartalmazó mondatokban az igekötő átlagosan 2,8 karakter hosszú, az ezt követő szavak 4,1, majd 4,9 hosszúak. A növekvő összetevők törvénye a +2 pozíciós igekötők esetében is érvényes: az átlagosan 3,2 hosszú igekötőt annál valamivel rövidebb szó (3,0) előzi meg, és lényegesen hosszabb követi (4,5). A tendencia +3 pozíciós igekötők esetén már érvényét veszti: az átlagosan 4,1 hosszúságú igekötőt nála hosszabb szó előzi meg. Fontos észrevennünk, hogy ez a helyzet csak nagyon kevés mondatban áll fenn, a +3 pozíciós igekötők a posztverbális eseteknek mindössze 0,07 százalékát alkotják.

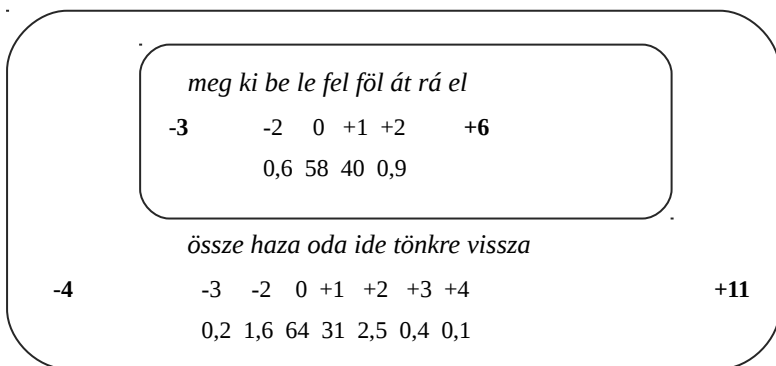
Igazolható tehát, hogy a magyarban az ige utáni szabad szórend esetén a fonológia hat a szintaxisra, és az egyes mondatösszetevők sorrendjét befolyásolja a fonológiai súlyuk.

5. Új szempont a centrum-periféria modellhez

A 2. fejezetben láthattuk, milyen nehézségeket okoz az igekötő-állomány meghatározása, és szó volt arról, hogy egy ilyen elmosódott határokkal bíró kategóriát célszerűbb lehet egy prototipikusági skálán értelmezni. Forgács (2005) nyelvtörténeti szempontból határozza meg az igekötők centrum és periféria csoportjait. A centrumban a prototipikus igekötők állnak, amelyek grammatikalizált, ősi igekötők (a csoport legtipikusabb példája a *meg*, amely már irányjelölő funkcióval sem bír). A külsőbb körökben az olyan szavak állnak, amelyek még kevésbé haladtak előre az igekötővé válás útján, nehezen különíthetők el egyes névutóktól, határozószóktól (pl. *haza, vissza*). Kerekes (2011) Forgácséhoz hasonlóan az igekötőség fokozatossága mellett érvel. E

két megközelítéshez szolgáltatnak újabb szempontot az itt bemutatott korpuszmérések.

A finit igéhez képest elfoglalt helyük alapján kirajzolódtak az igekötők különböző csoportjai, amelyek közül az igéhez legközelebbi halmazt (centrum) valamint a legtávolabbit mutatja be a 4. ábra.¹¹



4. ábra: Az igekötők centrum–periféria modellje a finit igéhez viszonyított pozíció szerint.

Az előjeles számok a pozíciót jelzik, az alattuk láthatók pedig azt, hogy milyen arányban (%) képviselteti magát az igekötőcsoport az adott pozícióban.

A centrum tagjait a *meg, ki, be, le, fel, föl, át, rá, el* alkotják. Esetükben azt találjuk, hogy vannak ugyan szélsőbb előfordulásai, de ezek egy-két példára korlátozódnak. Közös bennük az is, hogy ezek a leggyakoribb igekötők, a jelenlétük mégis szinte csak a preverbális pozícióra és a +1-re szorítkozik. Az eloszlásukat tekintve is összetartanak: 58%-ban közvetlenül preverbálisan, 40%-ban közvetlenül posztverbálisan állnak.

Az ígétől legtávolabb az *össze, haza, oda, ide, tönkre, vissza* igekötők kerülhetnek. Az eloszlásuk szintén nagy hasonlóságot mutat: a közvetlenül preverbális pozíció 64, a közvetlenül posztverbális 31 százalékot tesz ki.¹² A kapott eredmények egy egész más szempont alapján nagyon hasonló centrum–periféria elrendezést mutatnak, mint a nyelvtörténeti alapú skálázás.

¹¹ Mind a 80 igekötő részletes adatait l. Kalivoda (2016: 37–38).

¹² Kivételt képez az *ide*, amely főleg felszólító mondatokban fordul elő, emiatt a +1 pozíciója 56,8%-os.

Összegzés

A korpuszvezérelt módszerű kutatás elméleti és gyakorlati szempontból is hasznosnak bizonyult. Elméleti téren a posztverbális igekötők disztribúciója hozott újdonságot. Egyrészt láttuk, hogy szóbeli közlésekben jobban eltávolodik az igétől az igekötője, mint szerkesztett, írott szövegekben. Másrészt számszerűsíteni lehetett azt, hogy a növekvő összetevők törvénye mennyire érvényes a magyar mondatok posztverbális szakaszára. A kutatás új szempontot valósított meg az igekötők prototípus-elméleti vizsgálatában.

A korpuszmérés eredményei hozzájárultak az ANAGRAMMA elemző VFRAME keresőeljárásának kialakításához, amely az ige vonzatkeretének egyértelműsítését végzi, l. Indig–Vadász (2016), Vadász–Kalivoda–Indig (2017). Az igekötős igék gyakorisági adatai pedig a MANÓCSKA integrált igei vonzatkeret adatbázis részét képezik, l. Indig (2017)¹³.

Hivatkozások

- Alberti Gábor 2006. *A magyar mondattan elmélete és gyakorlata generatív megközelítésben III. A háttérelmélet*. Pécs, Pécsi Tudományegyetem.
- Behaghel, Otto 1932. *Deutsche Syntax IV*. Heidelberg, Carl Winters.
- Cooper, William E. – Ross, John Robert 1975. World order. In Robin E. Grossman – L. James San – Timothy J. Vance (eds.) *Papers from the Parasession on Functionalism*. Illinois, Chicago, Chicago Linguistic Society, 63–111.
- É. Kiss Katalin 1992. Az egyszerű mondat szerkezete. In Kiefer Ferenc (szerk.) *Strukturális magyar nyelvtan. I. Mondattan*. Budapest, Akadémiai Kiadó, 74–160.
- É. Kiss Katalin 2003. Mondattan. In É. Kiss Katalin – Kiefer Ferenc – Siptár Péter (szerk.) *Új magyar nyelvtan*. Budapest, Osiris Kiadó, 1–126.
- É. Kiss Katalin 2007. Az ige utáni szabad szórend magyarázata. *Nyelvtudományi Közlemények* 104: 124–152.
- Forgács Tamás 2005. Grammatikalizálódás az igekötők körében. In Oszkó Beatrix – Sipos Mária (szerk.) *Uráli grammatizáló*. Budapesti Uráli Műhely 4, Budapest, MTA Nyelvtudományi Intézet, 88–116.
- Goody, Jack 1987. *The Interface Between the Written and the Oral*. Cambridge, Cambridge University Press.
- Imrényi András 2011. *A magyar mondat viszonyhálózati modellje*. Ph.D. disszertáció, Budapest, Eötvös Loránd Tudományegyetem, Bölcsészettudományi Kar.
- Indig Balázs – Laki László – Prószekey Gábor 2016. Mozaik nyelvmodell az AnaGramma elemzőhöz. In Tanács Attila – Varga Viktor – Vincze Veronika (szerk.) *XII. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged, 2016. 01. 21–22., 260–270.

¹³ Az adatbázis a következő erőforrásokat használja: Sass–Váradi–Pajzs–Kiss (2010), Sass (2015), Kornai–Nemeskey–Recski (2016), Kalivoda (2016).

- Indig Balázs – Vadász Noémi 2016. Window in Human Parsing – How Far can a Preverb Go? *Tenth International Conference on Natural Language Proceedings (HrTAL2016)*. Dubrovnik, Horvátország, 2016. 09. 29-30. Elfogadva, nyomtatás alatt.
- Indig Balázs 2017. *Manócska – integrált igei vonzatkeret adatbázis*. Elérhető: <https://github.com/ppke-nlpg/manocska>.
- 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.
- Kalivoda Ágnes 2016. *A magyar igei komplexumok vizsgálata*. Mesterszakos szakdolgozat. Budapest, Pázmány Péter Katolikus Egyetem, Bölcsész- és Társadalomtudományi Kar.
- Kerekes Judit 2011. Az igekötők meghatározásának problémái. In Gécseg Zsuzsanna (szerk.) *LingDok 10. Nyelvészdoktoranduszok dolgozatai*. Szeged, JATEPress, 109–130.
- Kiefer Ferenc – Ladányi Mária 2000. Az igekötők. In Kiefer Ferenc (szerk.) *Strukturális magyar nyelvtan. III. Morfológia*. Budapest, Akadémiai Kiadó, 453–518.
- Komlósy András 1992. Régensek és vonzatok. In Kiefer Ferenc (szerk.) *Strukturális magyar nyelvtan. I. Mondattan*. Budapest, Akadémiai Kiadó, 299–527.
- Kornai András – Nemeskey Dávid Márk – Recski Gábor 2016. Detecting optional arguments of verbs. *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*.
- Németh Dezső 2001. A munkamemória szerepe a mondatmegértésben. In Pléh Csaba – Lukács Ágnes (szerk.) *A magyar morfológia pszicholingvisztikája*. Budapest, BIP – Osiris, 83–118.
- O’Connor, Michael Patrick 1978. *Hebrew Verse Structure, Volume 1*. Indiana, Winona Lake, Eisenbrauns.
- Oravecz Csaba – Várad Tamás – Sass Bálint 2014. The Hungarian Gigaword Corpus. *Proceedings of LREC 2014*, 1719–1723.
- Pais Dezső 1959. Az igekötők mivoltához és keletkezéséhez. *Magyar Nyelv 55*: 181–184.
- Peredy Márta 2011. Az ige-igekötő sorrend a Jókai-kódexben. In Bakró-Nagy Marianne – Forgács Tamás (szerk.) *A nyelvtörténeti kutatások újabb eredményei VI*. Szeged, Szegedi Tudományegyetem Magyar Nyelvészeti Tanszék, 181–197.
- Pléh Csaba 1998. *A mondatmegértés a magyar nyelvben. Pszicholingvisztikai kísérletek és modellek*. Budapest, Osiris Kiadó.
- Sass Bálint – Várad Tamás – Pajzs Júlia – Kiss Margit 2010. *Magyar igei szerkezetek – A leggyakoribb vonzatok és szókapcsolatok szótára*. Budapest, Tinta Kiadó.
- Sass Bálint 2011. *Igei szerkezetek gyakorisági szótára – egy automatikus lexikai kinyerő eljárás és alkalmazása*. Ph.D. disszertáció, Budapest, Pázmány Péter Katolikus Egyetem, Információs Technológiai és Bionikai Kar.

- Sass Bálint 2015. 28 millió szintaktikailag elemzett mondat és 500000 igei szerkezet.
In Tanács Attila – Varga Viktor – Vincze Veronika (szerk.) *XI. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY2015)*. Szeged, JATEPress, 303–308.
- Vadász Noémi – Kalivoda Ágnes – Indig Balázs 2017. Ablak által világosan – Vonzatkeret-egyértelműsítés az igekötők és az infinitívuszi vonzatok segítségével. *XIII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY2017)*. Szeged, 2017. 01. 26-27., Szeged, Szegedi Tudományegyetem, 3–12.