

A magyar igekötös szerkezetek korpuszvezérelt vizsgálata

Kalivoda Ágnes

2022. november 15.

A Magyar Nyelvtudományi Társaság
felolvasó ülése

Bevezetés

A kutatás célja:

- az igekötős szerkezetek minél teljesebb leírása korpuszvezérelt módszerrel
- szabadon elérhető adatbázisok létrehozása

Korpuszvezéreltség:

- egy hangsúlyozottan adatközpontú szemlélet
- induktív: a kutatás általában nagy mennyiségű szöveganyag automatikus elemzéséből indul ki (Sass 2011: 16)
- szorosan kötődik a használatalapú nyelvtanokhoz (Barth & Schnell 2022)

Igeközös szerkezetek: Elmosódó határok

Az igeekötők az igemódosítók közé tartoznak
→ hasonló disztribúció és szemantikai viselkedés (Kiefer 2007)

el

eltöpreng

nem különíthető **el**

el kell, hogy engedj

elszúrnivaló

el-elszöktem

haza

hazamegy

nem csábíthatók **haza**

haza kell, hogy menjek

hazalógnivalóm

haza-hazajárnak

zsebre

zsebre vág

nem tehető **zsebre**

zsebre kéne, hogy rakjon

zsebrevágnivalót

zsebre-zsebrevágott

Igeköötős szerkezetek: Elmosódó határok

Az igeköötők az igemódosítók közé tartoznak
→ hasonló disztribúció és szemantikai viselkedés (Kiefer 2007)

agyon
tönkre
szénné
szarrá
rommá
hülyére

+

izgulja
dolgozza
keresi
nyeri
röhögi
zabálja

+

magát

A vizsgált lexikai elemek köre

- kiinduló halmaz: 235 igekötő(szerűség) a PrevLex alapján (Kalivoda 2019)
- elv: inkább több legyen, mint kevesebb → sokszor épp a határesetek az érdekesek
- a “tipikus igekötők” és a “tipikus igemódosítók” jól elkülönülnek egymástól, de a határesetekből van a legtöbb

Az előadás főbb témái

1. Az igékötös szerkezetek szórendi mintázatai

Korpusz: Magyar Nemzeti Szövegtár 2.0.4 (Oravecz et al. 2014)

2. Az igékötők produktív kapcsolódási mintái

Korpusz: Magyar Nemzeti Szövegtár 2.0.4

3. Az igékötővé válás folyamata

Korpuszok:

Ómagyar Korpusz (Simon & Sass 2012)

Történeti Magánéleti Korpusz (Dömötör et al. 2017)

Magyar Történeti Szövegtár (Ittész 2009)

MNSz2 Universal Dependencies Treebank (Kalivoda et al. 2022)

Az igekötős szerkezetek szórendi mintázatai

Bevezetés

A vizsgálat tárgya:

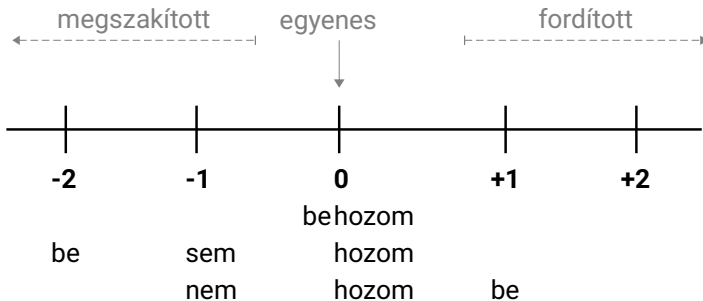
- az MNSz2 összes olyan szerkezettípusa, amelyben 'igekötő + igető' kapcsolat azonosítható:
finit igék, igenevek, igei származékok
- 49 szerkezettípus, 41,5 millió korpuszbeli előfordulás

A főbb kérdések:

- Milyen szórendi mintázatokat mutatnak az igekötős szerkezetek?
- Mikor és mennyire távolodhat el az igekötő az igétől, igenévtől, igei származéktól?

A módszer lényege

Az igeköötő pozíciójának meghatározása az ige-tőhöz képest:



Finít ige

db	pozíciók	prev. %	posztv. %
21 531 273	[-6; +11]	60,41	39,59

Az igeköötők 99,2%-ban közvetlenül az ige előtt vagy után állnak.

Az ősi igeköötők (*meg, el, le, fel, ki, be*) relatív gyakoriságukat tekintve kevésbé távolodnak el a finít igtől.

A hátravetett igeköötők eltávolodására hatással lehet:

- a szöveg szerkesztettségi foka
- a növekvő összetevők törvénye
Behaghel (1932), a magyar nyelvre É. Kiss (2007)

Infinitívusz

beférkőző ige(név)	db	pozíciók	prev. %	posztv. %
nincs	2 578 906	[-4; +2]	99,72	0,28
van	976 059	[-11; +2]	99,89	0,10

Ha nincs beférkőző ige(név), akkor jellemzően segédige-szerű névszó veszi át a helyét: *szokás, képes, érdemes, célszerű, ...*

Ha van, akkor az infinitívusz igekötője preverbálisan tőle távol kerülhet – a segédige-szerű elem elé, ld. Kálmán C. et al. (1989)

pl. **össze** lehetne őket többé-kevésbé objektív módon is **mérni**,

Határozói igenév

típus	db	pozíciók	prev. %	posztv. %
állapot- v. módhatározó	527 909	[-3; +3]	97,94	2,06
létigével passzív szerk.	205 798	[-9; 0]	100,00	0,00

A létigével alkotott passzív szerkezet párhuzamba állítható a beférkőző igés infinitívusszal:

pl. **Ki** vannak ezek a marketinges dolgok **találva**.

Hátravetett igeköötő csak akkor fordul elő, ha az igenév állapot- vagy módhatározói szerepet tölt be.

-hAtÓ és -AndÓ végződésű igenevek

végződés	db	pozíciók	prev. %	posztv. %
-hAtÓ	645 051	[-4; +5]	80,34	19,66
-AndÓ	87 555	[-3; +2]	98,15	1,85

-hAtÓ: ha állítmányi helyzetben van, és a tagmondatban nincs finit ige, akkor a -hAtÓ teljesen átveszi ennek a helyét

-AndÓ: állhat állítmányi helyzetben, és ekkor az igekötője is hátravethető

pl. *egy vélemény nem **keverendő össze** az igazsággal*

A három szórend áttekintése

szórend	példa	gyakoriság*	szerkezettípusok
egyenes	<i>összerak</i>	31,4	47
fordított	<i>nem mondható el</i>	8,7	8
megszakított	<i>át nem vétele</i>	1,5	34

*millió tokenben megadva

- fordított szórend: gyakori, de kevés szerkezettípusban
- megszakított szórend: ritka, de sokféle szerkezettípusban
igei származékokban a beférkőző elemek relatív gyakorisága:
nem 99,35%, *sem* 0,46%, *is* 0,15%, *se* 0,03%

Kapcsolódó témák

Még sok egyéb szerkezet – ld. Kalivoda (2021):

- igekötős igék topikalizációja:
pl. *Kimenni csak **ki** tudnánk, de hogy megyünk vissza?*
- *el kell, hogy menjek* – *el kell menjek* típusú szerkezetek
- melléknévi igenevek szórendi mintázatai tagadáskor:
pl. *el nem **mondható**, nem **elmondható**, nem **mondható el***
- ...

PrevDistro (Preverb Distributions): a szerkezettípusok további vizsgálatára alkalmas adatbázis

<https://zenodo.org/record/6349410>

Az igekötők produktív kapcsolódási mintái

Bevezetés

A **kapcsolódási minta** megmutatja, hogy egy igekötő

- milyen jelentéseiben milyen igecsoport tagjaihoz járulhat
- milyen hatással van az igei vonzatkeretre

(Ladányi 2012)

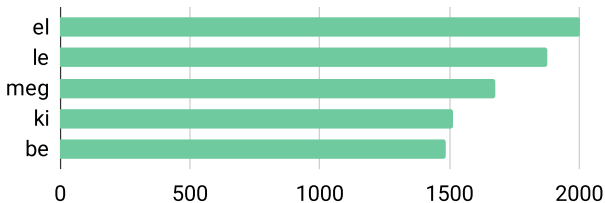
A kutatás célja:

A minták feltárása korpuszvezérelt módszerrel, azaz

1. 'igekötő – képző – vonzatkeret' hármassok kinyerésével
pl. *el-X-elődik valami, tönkre-X-el valamit*
2. a hármassokban szereplő igék vizsgálatával

A vizsgált adatok köre

- 21 038 finit igekötős igei hapax (= egyszer előforduló szó)
- jellemzően produktív szóalkotás eredményei, a morfológiai szerkezetük és a jelentésük többnyire kompozicionális
- 122 igekötő(szerűség) – a korábbi 235 elem fele nem fordul elő hapaxban, pl. *létre*, *cserben*, *nyilván*, ...
- a legtöbb hapaxszal adatolt igekötők:



Morfológiai és szemantikai információk kinyerése

*Ahhoz azért túlságosan is tiszteletre méltó az életműve, hogy **lekisöregezd**, nem gondolod?*

↓ 1. az igekötő és az ige különválasztása

le + kisöregezd

↓ 2. az ige emMorph elemzése (Novák et al. 2017)

kisöreg + -Vz(ik)

↓ 3. word2vec szóbeágyazás (Siklósi et al. 2016)

a 230-as klaszter tagja

(kisöreg, krapek, hapi, hapsi, csávó, muki, csaj, kiscsaj)

Szintaktikai információk kinyerése

Jó lenne, ha **elhisztizné** végre magát a teljes feledésbe.

↓ 1. tagmondathatárok azonosítása (Sass 2011)

Jó lenne, <ncl> ha **elhisztizné** végre magát a teljes feledésbe.

↓ 2. a tagmondat emtsv (emChunk) elemzése (Indig et al. 2019)

ha **elhisztizné** végre [magát] [a teljes feledésbe] .

↓ 3. vonzatkeret-kinyerés (Sass 2021)

-bA maga-t

Az igealkotás produktív esetei

1. Névszó + igeképző

pl. *mamlasz* → *mamlaszkodik*

Duracell-nyuszi → *duracellnyusziz*

2. Ige + igeképző

pl. *kuszál* → *kuszálódik*

jóreggeltez → *jóreggeltezget*

3. Hangzásséma

pl. *locsog* – *loccsan* – *loccsant* – *loccsint*

puffad – *puffaszt*

A hangzássémákról bővebben

- nemcsak a végződés kötött, hanem a fonológiai szerkezet is (Benő és Szilágyi N. 2015)
- a létrejövő ige jelentése nagyrészt megjósolható:
 - csattog, zötyög* ismétlődve, nem túl intenzíven (teszi X)
 - csattan, zöttyen* egyszer, intenzíven (teszi X)
 - csattant, zöttyent* egyszer, intenzíven (teszi Y X-szel)
 - csattint, zöttyint* egyszer, nem túl intenzíven (teszi Y X-szel)
- nem elhanyagolható: az igekötős igei hapaxok **9,4%**-a

Példa: -Og végű hangzássémák (részlet)

séma	példák	db
(C)CO ₁ CO ₁ g	<i>kopog, fröcsög, cseveg</i>	70
(C)CO ₁ CCO ₁ g	<i>mormog, dörmög, csemcseg</i>	11
(C)CO ₁ C:O ₁ g	<i>szöttyög, kerreg, csetteg</i>	6
(C)(C)V(C)CO ₁ rO ₁ g	<i>sutyorog, nyöszörög, sistereg</i>	59
(C)(C)V(C)CO ₁ lO ₁ g	<i>őgyeleg, édeleg, hízeleg</i>	13
(C)(C)V(C)CO ₁ /j/O ₁ g	<i>gomolyog, hömpölyög, tévelyeg</i>	10

Kapcsolódási minták: *el*

***el* - NÉVSZÓ -(s)Odik (196 db)**

'változik vmivé/vkivé'

elemdéeefesedik
elbalosodik
elliberálisodik
elpártosodik

w2v: 1, 2

elmagyarodik
elamerikaisodik
eleurópaisodik
elrómaiásodik

w2v: 7, 8

Kapcsolódási minták: *le*

le - NÉVSZÓ -z(ik) -t (443 db)

'nevez vkit vhogy'

lejenőz
lepéterez
leistvánoz

w2v: 291

'leönt vmivel'

lekakaóz
lepálinkáz
lekávéz

w2v: 99

'vmivel (majdnem) megöl'

ledárdáz
lefénykardoz
leguillotinoz

w2v: 140, 141

Az aktuális jelentés kontextusfüggő:

Ezt a kaját Lilu kétszer is **lerántottcsirkézte**.

Jól **levajaskenyereztem** ezt a papírt, [...]

[...] lekenyereztem, akarom mondani, **ledinnyéztem** őket.

Kapcsolódó témák

PrevCons (*Preverb Constructions*): a produktív kapcsolódási minták vizsgálatára alkalmas adatbázis

- forrásfájl: <https://github.com/kagnes/prevcons>
- webes keresőfelület (készítette: Indig Balázs):
<https://aszalo.herokuapp.com/>

A PrevCons alapján készült **ontológia**, amely hálózatszerűen ábrázolja 56 igekötvő jelentésrendszerét:

- forrásfájl: <https://github.com/kagnes/prevonto>

Az igekötővé válás folyamata

Bevezetés

- a különféle lexikai elemek igeköötővé válása komplex grammatikalizációs folyamat
 - különböző típusú változások kölcsönös összefüggése:
 - az elemet tartalmazó szerkezet jelentésváltozása
 - az elem jelentésváltozása (lexikai → grammatikai)
 - szerkezeti módosulások (pl. újraelemzés, kategóriaváltás)
 - növekvő produktivitás
- ld. D. Mátai (1989), Ladányi (2000), Forgács (2004), Dér (2008)
- kapcsolódó OTKA PD projekt (2022. 10. – 2025. 09.):
*Az igeköötős szerkezetek kialakulása a magyar nyelvben:
korpuszvezérelt megközelítés*

Pilot kutatás: az *agyon* grammatikalizációja

Bőséges szakirodalmi háttér: Ladányi (2000), Forgács (2004)

A kutatás kérdései:

- Alátámaszthatók-e kvantitatívan is a szakirodalomból ismert hipotézisek az *agyon* grammatikalizációs útja kapcsán?
- Kimutathatók-e újabb összefüggések?

A felhasznált korpuszok:

korpusz	korszak	regiszterek	méret*
ÓMK	ómagyar	formális	4,8
TMK	középmagyar	informális	1,5
MTSZ	újmagyar (1772-től)	formális és informális	30,0
MNSz2-UD	újmagyar (1920-tól)	formális és informális	928,1

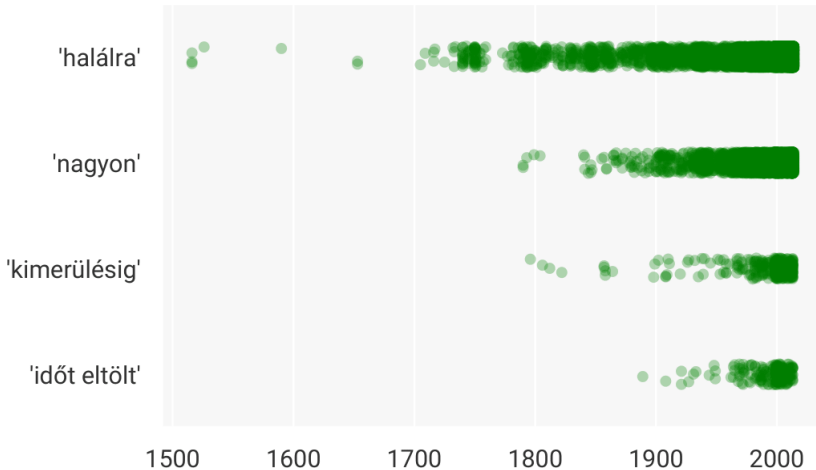
*millió tokenben megadva

Pilot kutatás: az *agyon* grammatikalizációja

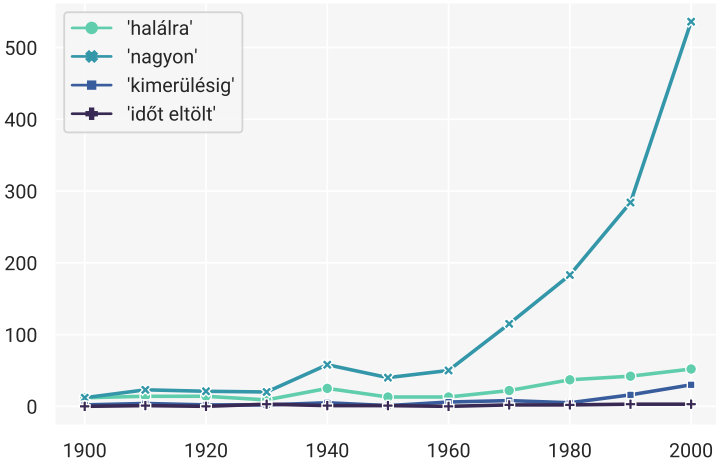
Minden adatpont (22 571 db) esetében kinyert információ:

- metaadatok: korpusz, dokumentum, regiszter, év
- kontextus: az *agyon*-t tartalmazó teljes mondat
- az *agyon* szóalakja (*agyon*, *agyonra*, *agyon-vissza*, ...)
- igelemma
- szerkezet típus
- az *agyon* pozíciója az ige tőhöz képest
- vonzatkeret
- jelentés ('halálra', 'nagyon', 'kimerülésig', 'időt eltölt')
→ félautomatikus módszerrel, manuálisan ellenőrizve

Agyon: az egyes jelentések időbeli adatoltsága



Az *agyon*+IGE típusgyakoriságok változása



Agyon: a jelentések és a szerkezetek kapcsolata

Az öt leggyakoribb szerkezet típus százalékos eloszlása:

	'halálra'	'nagyon'	'kimerülésig'	'időt eltölt'
FIN	83,5	32,2	92,7	65,4
INF	7,6	6	7,3	28,8
-(O)(t)t	4,6	51,5	0	2,6
-Ás	3,6	3,5	0	3,3
-vA	0,7	6,9	0	0

A 'nagyon' jelentés 51,5%-ban befejezett melléknévi igenévhez kapcsolódik. Főleg jelzős szerkezetek, pl.:

agyontuningolt járgány, **agyonnyúzott** póló, **agyoncukrozott** szörp, ...

Agyon: a jelentések és a szórendek kapcsolata

A három szórendi típus százalékos eloszlása:

	'halálra'	'nagyon'	'kimerülésig'	'időt eltölt'
egyenes	70,4	89,6	85	40
fordított	24,7	6,3	11,3	52,5
megszakított	4,9	4,1	3,7	7,5

Az 'időt eltölt' jelentés 52,5%-ban fordított szórenddel jelenik meg. A -VAL ragos bővítmény sokszor fókuszban, pl.:

*S ilyen haszontalan **merengéssel ütöm agyon** az időt.*

*[...] majd zenei **kvízzjátékkal ütik agyon** a délutáni órákat.*

*[...] a ráérő idejüket ilyen **légyottokkal csapják agyon**.*

További tervek

- 1. Rövid távon:** az *agyon* adatainak részletes elemzése, további szempontok figyelembe vétele (pl. a regiszter kölcsönhatása a már bemutatott változókkal)
- 2. Hosszú távon:** a pilot vizsgálat kiterjesztése a további (potenciális) igekötőkre és igekötős szerkezetekre

Ennek előfeltétele a gondosan előkészített korpuszadat:

- ÓMK: morfológiai elemzés a normalizált szövegekhez
- MTSZ: gépi normalizálás és morfológiai elemzés
- adattudományos “népi bölcsesség”: a munkafolyamat 80%-a az adatgyűjtés és -tisztítás, 20%-a az elemzés

Összegzés: a kutatás eredményei

- olyan tendenciák, jelenségek kimutatása, amelyek korpuszvezérelt módszer hiányában észrevétlenül vagy sejtés szintjén maradtak volna
- új módszerek → más nyelvi jelenségek vizsgálatában is hasznosíthatók
- nyílt hozzáférésű adatbázisok → kutatóeszközként szolgálhatnak mások számára is

Köszönöm a figyelmet!

kalivoda.agnes@nytud.hu