

Igekötős szerkezetek a magyarban

Kalivoda Ágnes

2018. június 26., Budapest

PPKE BTK Nyelvtudományi Doktori Iskola
Nyelvtechnológia Műhely
Témavezető: Prószéky Gábor

A kutatás célja

- az igekötős szerkezetek minél teljesebb leírása
- korpuszvezérelt módszerrel, a performancia oldaláról
- a Manócska igei vonzatkeret adatbázis fejlesztése (Indig et al. 2017)
- igekötős szerkezetek kezelése az AnaGramma elemzőben (Prószéky et al. 2016)

I. Az igekötők mondatbeli pozícióinak feltérképezése

Kapcsolódó publikációk:

(Kalivoda 2017a), (Kalivoda 2018)

Motiváció és módszer

Probléma:

Az elvált igekötő és az ige összekapcsolása → mire készítsük fel a gépi elemzőt?

Megoldás:

Igekötő – finit ige távolságok kimérése az MNSZ 2.0.3 korpuszon (Oravecz et al. 2014)

A preverbális igekötő és a finit ige távolsága

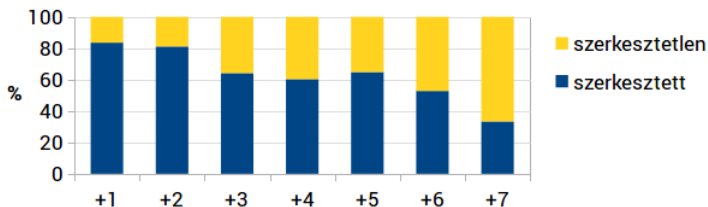
| IK pozíciók | -4 | -3 | -2 | -1 | 0 (FIN) |
|--------------------|---------------|-----------|-------------|--------|------------------|
| | | | meg | is | elmond |
| | | el | sem | nagyon | nyeri |
| | vissza | most | már | nem | jutnánk |
| | | | | | szívhatom |
| találatszám | 6 | 91 | 102768 | | 13715093 |
| százalék | - | - | 0,74 | | 99,26 |

A posztverbális igekötő és a finit ige távolsága

| FIN | +1 | +2 | +3 | +4 | +5 | +6 | +7 |
|------------------|--------------|-------------|-------------|-------------|-------------|-----------|------------|
| jelenik | meg | | | | | | |
| lát | majd | el | | | | | |
| fordulhat | még | egyszer | elő | | | | |
| adok | én | most | itt | elő | | | |
| beszélek | még | mindig | egy | kicsit | haza | | |
| fut | a | HÉV | a | vidám | néppel | ki | |
| csábítson | téged | a | nagy | retyézati | barna | medve | oda |
| | 7527308 | 163993 | 5156 | 1193 | 267 | 101 | 27 |
| | 97,98 | 2,13 | 0,07 | 0,02 | - | - | - |

Mitől függhet az ige – igekötő távolság?

A szöveg szerkesztettsége



- (1) akármilyen is **dobja** valakinek az egyéniségét a hullámhegyre **fel**,
- (2) azért nem **merül** ez a kérdés ilyen sarkallatosan **föl**,
- (3) Azért **mentem** egy kicsit a popzene felé **el**,

Mitől függhet az ige – igekötő távolság?

Fonológia: A növekvő összetevők törvénye

A rövidebb összetevő megelőzi a hosszabbat, ha ezt szintaktikai szabály nem gátolja. (Behaghel 1932)

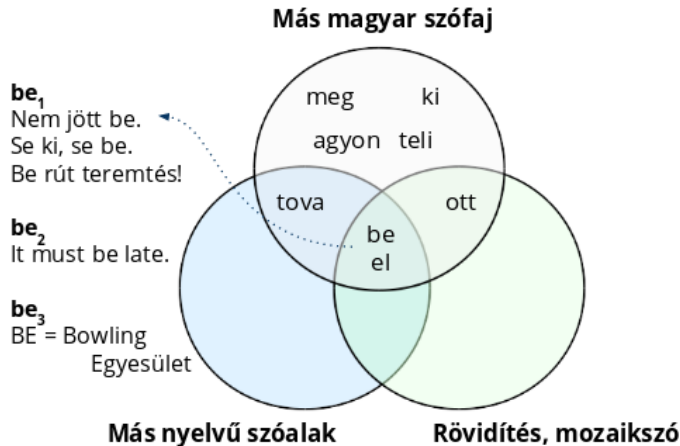
Magyar: az ige utáni mondatzakszakra érvényes ez a tendencia. (É. Kiss 2007)

| Alkorpusz | Átlag szóhossz | | | Gyakoriság |
|-----------|----------------|-----|-----|------------|
| +1-es IK | 2,8 | 4,1 | 4,9 | 97,78% |
| +2-es IK | 3,0 | 3,2 | 4,5 | 2,13% |
| +3-as IK | 3,3 | 5,2 | 4,1 | 0,07% |

II. Az igekötők annotációjának javítása

Kapcsolódó publikáció:
(Kalivoda 2017b)

Homográf szóalakok → hibás annotáció



Az igekötők annotációjának javítása

Szabály-alapú javítás (a kontextus mintázatai alapján)

Kiértékelés: az MNSZ 2.0.4-ből vett 5000 mondaton

Teljesítmény a tesztkorpuszon (%):

- pontosság: 88.2
- fedés: 57.5
- F-mérték: 69.6

Miért csak ennyi?

- nehéz kontextusra támaszkodni, ha rossz a kontextus
- rengeteg az elírás, hibás szóköz

III. Az eredmények felhasználása

Az eredmények felhasználása

Manócska – egységesített igei vonzatkerettár:

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

Kapcsolódó publikációk:

(Vadász – Kalivoda – Indig 2018)

(KaliVoda – Vadász – Indig 2018)

VFrame keresőeljárás:

<https://github.com/ppke-nlpg/vframe>

Kapcsolódó publikáció:

(Vadász – Indig – Kalivoda 2017)



IV. Diakrón vizsgálatok

Hat prototipikus igekötő (*meg, el, fel, ki, be, le*) vizsgálata

| Korpusz | Mondatok | Pozíciók | Preverbális | Posztverbális |
|---------|----------|-----------|-------------|---------------|
| ÓMK | 4873 | [-2; +4] | 83,3% | 16,7% |
| TMK | 26160 | [-3; +7] | 76,2% | 23,8% |
| MTSZ | 364902 | [-3; +7] | 71,0% | 29,0% |
| MNSZ | 21516014 | [-4; +11] | 64,2% | 35,8% |

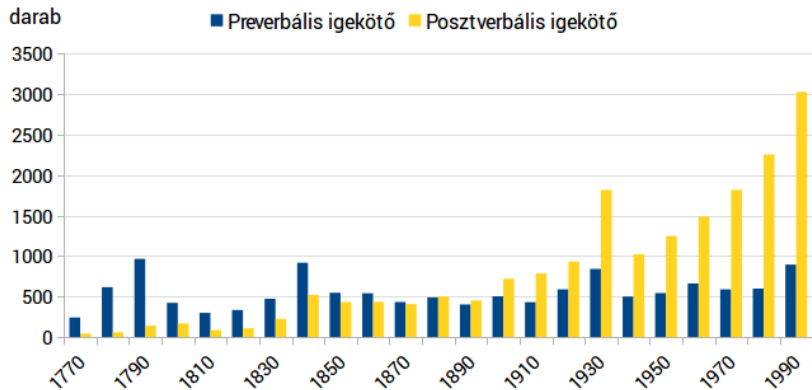
Példák az eltávolodásra:

- (1) **Jelentette** vala az Úristen **meg** asszony Jacobának,
- (2) **küldém** Szokoly Albertet mindjára mintegy posta módon **el**,
- (3) nem **ülepdedhetik** a víz a kávéval annyira **meg** a testben,
- (4) **metszik** a zsidók egy keresztény fiúcska nyakát **el**.

A tagadó szórend változása

Preverbális igekötő: megszakított szórend (*meg nem teszed*)

Posztverbális igekötő: fordított szórend (*nem teszed meg*)



További tervek

A disszertáció váza

- 1 Mik az igekötők?
- 2 Az igekötős szerkezetek szintaxisa
- 3 Az igekötők és az igei vonzatkeret
- 4 Diakrón vizsgálatok

Hivatkozások I.

- Behaghel, Otto (1932). *Deutsche Syntax IV*. Heidelberg, Carl Winters.
- É. Kiss Katalin (2007). *Az ige utáni szabad szórend magyarázata*. *Nyelvtudományi Közlemények* 104. p. 124–152.
- Indig Balázs, Vadász Noémi, Kalivoda Ágnes (2017). *Manócska – integrált igei vonzatkeret adatbázis*. Elérhető: <https://github.com/ppke-nlpg/manocska>
- Kalivoda Ágnes (2017a). *Hungarian particle verbs in a corpus-driven approach*. In: *Computational Linguistics and Intelligent Text Processing – 18th International Conference (CICLing 2017)*, 2017. április 17–23. Budapest. Megjelenés alatt.
- Kalivoda Ágnes (2017). *Az igekötők gépi annotálásának problémái*. In: Ludányi Zsófia (szerk.): *Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből. XI. Alkalmazott Nyelvészeti Doktoranduszkonferencia (Alknyelvdok 2017)*. 2017. február 3. Budapest, MTA Nyelvtudományi Intézet. pp. 100–109.

Hivatkozások II.

- Kalivoda Ágnes (2018). *Az igekötős igék szintaxisa korpuszvezérelt megközelítésben*. In: Scheibl György (szerk.): *Nyelvészdoktoranduszok dolgozatai (LingDok 17)*. 2016. november 30. – december 1. Szeged. Szegedi Tudományegyetem, Nyelvtudományi Doktori Iskola. pp. 159–176.
- Kalivoda Ágnes, Vadász Noémi, Indig Balázs (2018). *Manócska: A Unified Verb Frame Database for Hungarian*. In: 21st International Conference on Text, Speech and Dialogue (TSD 2018), 2018. szeptember 11–14. Brno, Csehország. Megjelenés alatt.
- Oravecz Csaba, Váradi Tamás, Sass Bálint (2014). *The Hungarian Gigaword Corpus*. In: Calzolari, Nicoletta, et al. (szerk.): *Proceedings of the 9th International Conference on Language Resources and Evaluation*. Reykjavík, Izland. ELRA. p. 1719–1723.
- Prószéky Gábor, Indig Balázs, Vadász Noémi (2016). *Performanciaalapú elemző magyar szövegek számítógépes megértéséhez*. In: "Szavad ne feledd!": *Tanulmányok Bánréti Zoltán tiszteletére*. p. 223–232.

Hivatkozások III.

- Vadász Noémi, Indig Balázs, Kalivoda Ágnes (2017). [Ablak által világosan – Vonzatkeret-egyértelműsítés az igekötők és az infinitívuszi vonzatok segítségével](#). In: Vincze Veronika (szerk.): XIII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2017). Szeged, 2017. január 26–27. pp. 3–12.
- Vadász Noémi, Kalivoda Ágnes, Indig Balázs (2018). [Egy egységesített magyar igei vonzatkerettár építése és felhasználása](#). In: Vincze Veronika (szerk.): XIV. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2018). Szeged, 2018. január 18-19. pp. 3–15.