
Az igekötős igék szintaxisa korpuszvezérelt megközelítésben

Kalivoda Ágnes

— Szeged, 2016. nov. 30. – dec. 1 —

PPKE BTK



Az előadás témája

Az igekötők mondatbeli pozícióinak feltérképezése korpuszvezérelt módszerrel. Nyelvtechnológiai motiváció.

Gyakorlati probléma

Az elvált igekötő és az ige összekapcsolása

→ mire készítsük fel a gépi elemzőt?

Elméleti kérdések

- Mennyire érvényesülő tendencia a Behaghel-törvény a magyarban?
- Milyen a magyar igekötő állomány összetétele?

- A korpuszelemzés módszere
 - Technikai nehézségek
- Az igekötők állománya
- A korpuszelemzés eredményei
 - Igekötő a finit igétől balra
 - Igekötő a finit igétől jobbra
 - Komplex esetek
 - Centrum – periféria modell
 - Szemantikai, pragmatikai tényezők
- Összegzés

- **A korpuszelemzés módszere**
 - Technikai nehézségek
- Az igekötők állománya
- A korpuszelemzés eredményei
 - Igekötő a finit igtől balra
 - Igekötő a finit igtől jobbra
 - Komplex esetek
 - Centrum – periféria modell
 - Szemantikai, pragmatikai tényezők
- Összegzés

A korpuszelemzés módszere

- korpuszvezérelt módszer (Sass 2011: 15–16)
- felhasznált korpuszok:
 - Magyar Nemzeti Szövegtár (Oravecz et al. 2014)
 - Pázmány Korpusz (Endrédi 2016)
- igekötő-pozíciók a finit igéhez viszonyítva, automatikusan
 - 0 el+olvas +1 olvasson el +2 olvasson már el
 - *-1 el akart olvasni -2 el sem olvas

Technikai nehézségek

- homonímia
 - **meg:** igekötő vagy kötőszó?
*akkor csak lámpát kell **vennem meg** rácsot*
 - **ki:** igekötő vagy kérdő/vonatkozó névmás?
*az mennyibe fog kerülni és **ki** fogja rá **adni** a pénzt*
- hibás különírás *agyon ver, el intéz*
- ékezetek hiánya *fel tizre mennem kell (fel+megy?)*
- kié az igekötő? *verni el nem mentek (el+ver? el+megy?)*
- **mikor határozószó, mikor igekötő?**

- A korpuszelemzés módszere
 - Technikai nehézségek
- **Az igekötők állománya**
- A korpuszelemzés eredményei
 - Igekötő a finit igétől balra
 - Igekötő a finit igétől jobbra
 - Komplex esetek
 - Centrum – periféria modell
 - Szemantikai, pragmatikai tényezők
- Összegzés

Az igekötők állománya

	HSz8.	HSz10., Rácz–Takács	MMNyR	Temesi–Rónai –Vargha	Szemere	ÉrtSz	Jakab
egybe	+	–	–	+	+	+	+
együtt		–	–		–	+	–
meg	+	+	+	+	+	+	+
oda	+	+	+	+	+	+	+

(Komlósy 1992: 457–459)

Forgács (2005: 11) két közös kritériumot talál a szerzőknél:

- az igekötő produktív, új igék hozhatók létre vele
- ha szemantikailag transzparens → nem igekötő

x haza

...

hazavág, hazabeszél ?

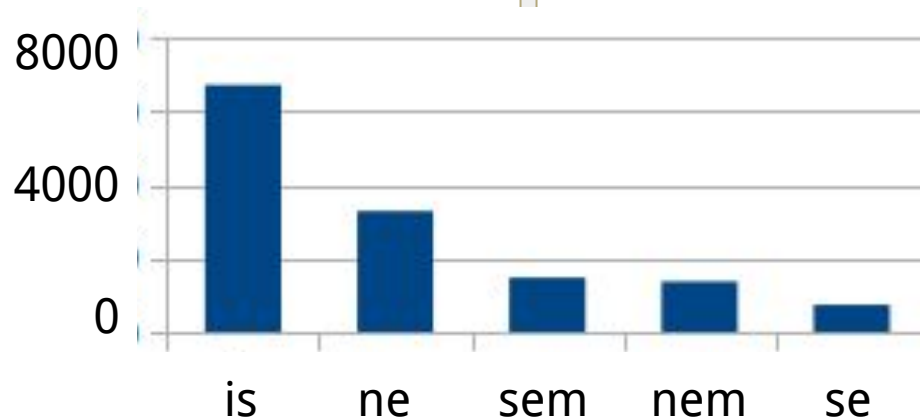
(J. Soltész 1959, Kiefer – Ladányi 2000)

- A korpuszelemzés módszere
 - Technikai nehézségek
- Az igekötők állománya
- **A korpuszelemzés eredményei**
 - Igekötő a finit igétől balra
 - Igekötő a finit igétől jobbra
 - Komplex esetek
 - Centrum – periféria modell
 - Szemantikai, pragmatikai tényezők
- Összegzés

Igekötő a finit igétől balra

IK pozíciók	-4	-3	-2	-1	0 (FIN)
<i>példák</i>					elmond
			meg	is	nyeri
		el	sem	nagyon	jutnánk
	vissza	most	már	nem	szívhatom
találatszám	6	91	102.768		13.715.093

A közbeékelődő szavak
klitikumok:



Igekötő a finit igétől jobbra

FIN	+1	+2	+3	+4	+5	+6	+7	IK poz.
jelenik	meg							<i>példák</i>
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	retyezāti	barna	medve	oda	
	7.527.308	163.993	5.156	1.193	267	101	27	találat -szám

+11 27 gyereket **vitt** egy feltehetően részeg buszsofőr Szentesen még csütörtökön egy sportrendezvény után **vissza** az iskolába.

A Behaghel-törvény – számokban

„Az ige utáni, szabad szórendű mondatszakasz akkor hangzik legjobban, ha eleget tesz a Behaghel (1932) megfogalmazta 'növekvő összetevők törvényének', azaz, ha az ige utáni összetevők növekvő sorrendben követik egymást.” (É. Kiss 2008: 132)

Kvantifikálható:

IK helye	Szavak átlagos hossza				Gyakoriság (%)	FIN és IK közt álló szavak
	+1	+2	+3	+4		
+1	2.8	3.7	4.9	6.2	97.8	∅
+2	3.0	3.2	4.5	5.7	2.1	-e, is, már, csak, még ...
+3	3.3	5.2	4.1	4.8	0.07	egy kicsit, én is, -e majd ...

Komplex esetek

- Igekötő kiemelése fölérendelt tagmondatba

el** kell, hogy **mondjam → több mint 3000 találat

Milyen finit igékkel? *kell* (99.9%), *akar* (4), *szabad* (2), *szeretne* (1)

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

- Igenevek (kié az igekötő? – probléma)

- “Gazdátlan” igekötő infinitívuszos szerkezetben:

***elképzelni** bármit **el** lehet, **becsajozni be** tudnék*

→ több mint 230 találat

Centrum – periféria modell

meg ki be le fel föl át rá el


-3	-2	0	+1	+2	+6
	0.6	58	40	0.9	

össze haza oda ide tönkre vissza

-4	-3	-2	0	+1	+2	+3	+4	+11
	0.2	1.6	64	31	2.5	0.4	0.1	

Szemantikai, pragmatikai tényezők

- Ha a jelentés kompozicionális, jobban eltávolodhat az igekötő?



Ige		Max.	+1 aránya (%)
ki+terjed	N	+3	99
ki+enged	I	+6	88
vissza+él	N	+5	97
vissza+visz	I	+11	83

- Van olyan ige, amelynél az igekötő szinte mindig posztverbális?

Ige	Igekötő	Posztverbális (%)
dolgoz	fel	98.6
tevődik	össze	97.7
áll	hozzá	92.2
ér	el	90.8
helyezkedik	el	90.8

- A korpuszelemzés módszere
 - Technikai nehézségek
- Az igekötők állománya
- A korpuszelemzés eredményei
 - Igekötő a finit igétől balra
 - Igekötő a finit igétől jobbra
 - Komplex esetek
 - Centrum – periféria modell
 - Szemantikai, pragmatikai tényezők
- **Összegzés**

Összegzés

- Az elvált igekötők pozíció szerinti eloszlásaival a Behaghel-törvény kimutatható, kvantifikálható
- Az “igekötőség” centrum-periféria modellben megragadható
- A korpuszelemzés új szerkezetet mutatott ki: *elhinni el tudnám*

Nyelvtechnológiai tanulságok

Az ige beolvasásakor a gépi elemző tudja, milyen igekötőt várhat az ige, milyen valószínűséggel, hány szónyi távolságban.

→ hatékonyabb elemzés

Hivatkozások 1.

- Behaghel, Otto (1932): [Deutsche Syntax IV](#). Heidelberg, Carl Winters.
- Endrédi István (2016): [Nyelvtechnológiai algoritmusok korpuszok automatikus építéséhez és pontosabb feldolgozásukhoz](#). PhD. disszertáció.
- É. Kiss Katalin (2008): [Tagadás vagy egyeztetés? A *senki, semmi* típusú névmások szórendi helye, jelentése és hangsúlyozása](#). In: *Magyar Nyelv* 104. 129–143.
- Forgács Tamás (2005): [Grammatikalizálódás az igekötők körében](#). In: Oszkó Beatrix – Sipos Mária (szerk.): *Uráli grammatizáló*. BUM 4. Budapest, MTA Nyelvtudományi Intézet. 88–116.
- J. Soltész Katalin (1959): [Az ősi magyar igekötők](#). Budapest, Akadémiai Kiadó.

Hivatkozások 2.

- Kiefer Ferenc – Ladányi Mária (2000): [Az igekötők](#). In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan. 3. Morfológia*. Budapest, Akadémiai Kiadó. 453–518.
- Komlósy András (1992): [Régensek és vonzatok](#). In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan. 1. Mondattan*. Budapest, Akadémiai Kiadó. 274–486.
- Oravecz Csaba – Váradi Tamás – Sass Bálint (2014): [The Hungarian Gigaword Corpus](#). In: *Proceedings of LREC 2014*. Reykjavík, 2014. május 26-31. 1719–1723.
- Sass Bálint (2011): [Igei szerkezetek gyakorisági szótára – egy automatikus lexikai kinyerő eljárás és alkalmazása](#). PhD disszertáció.

Köszönöm a figyelmet!

