

A magyar igekötős szerkezetek adatvezérelt kutatása

Kalivoda Ágnes

Nyelvtudományi Kutatóközpont
PPKE BTK Nyelvtudományi Doktori Iskola

2021. május 14.

A főbb kutatási témáim röviden

Bevezetés

A kutatás célja:

- az igekötős szerkezetek minél teljesebb leírása
- adatközpontú szemlélettel, korpuszvezérelt módszerrel

A főbb témák:

1. az igekötők állományának meghatározása
2. az igekötős szerkezetek mintázattípusainak leírása
3. az igekötők produktív kapcsolódási mintáinak feltárása

Az igekötők állománya: elmosódó határok

Az igekötők az igemódosítók közé tartoznak

→ hasonló szemantikai és szintaktikai viselkedés (Kiefer 2007)

el

eltöpreng

nem különíthető **el**

el kell, hogy engedj

elszúrnivaló

el-elszöktem

haza

hazamegy

nem csábíthatók **haza**

haza kell, hogy menjek

hazalógynivalóm

haza-hazajárnak

zsebre

zsebre vág

nem tehető **zsebre**

zsebre kéne, hogy rakjon

zsebrevágnivalót

zsebre-zsebrevágott

Jakab (1976) áttekintő táblázatában az *el* 7, a *haza* 2, a *zsebre* 0 forrás szerint igekötő (összesen 7 forrásból).

Az igekötők állománya: elmosódó határok

Mely lexikai elemeket és mi alapján soroljuk az igekötők közé?

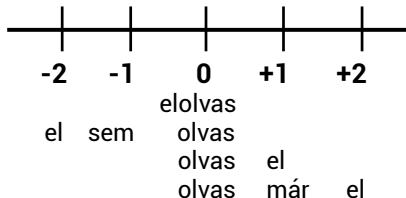
Javaslat: prototípus elméleti megközelítés (Rosch 1973)

Módszer:

1. kiinduló halmaz: 235 igekötő-jelölt
2. automatikusan mérhető jegyek meghatározása
3. korpuszmérések: minden igekötő-jelölthöz minden jegy
4. adatelemzés: jegyek közötti korrelációk
5. kategorizálási kísérlet: viszonyítási pont a *meg* igekötő

Az igekötős szerkezetek mintázattípusai

- Milyen mintázattípusokat mutatnak az igekötős szerkezetek?
- Mikor és mennyire távolodhat el az igekötő az igétől, igenévtől, igei származéktól?
- A módszer lényege:



- Egy nagyobb szinkrón és egy kisebb diakrón korpuszvizsgálat

Az igekötős szerkezetek mintázattípusai

- a szinkrón korpuszvizsgálat eredményeiből adatbázis épül
→ PrevDistro (*Preverb Distributions*)
- 53 szerkezet típus, kb. 34 000 000 korpuszbeli előfordulás
- mutatvány az adatbázisból:

meggondoltan	-(A)n	0	meg	gondol	-	85
felvágósdit	-sdi	0	fel	vág	-	1
vonható csak le	-hAtÓ	+2	le	von	csak	4
dadogván el	-vÁn	+1	el	dadog	-	1
el kell, hogy mondjak	climbing	-3	el	mond	kell, hogy	23
legösszenemillőbb	leg-...-bb	-1	össze	illik	nem	1

Az igekötők produktív kapcsolódási mintái

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

Az igekötők produktív kapcsolódási mintái

Az erre a célra készült adatbázis: PrevCons (*Preverb Constructions*)

- 21 038 igekötős igei hapax vonzatkerettel és meta-adatokkal
- a Mazsola (Sass 2011) mintájára készült
- nyílt hozzáférésű erőforrás
- webes keresőfelület segítségével is böngészhető

Köszönet érte Indig Balázsnak!

PrevCons: Hogyan készült?

A felhasznált korpusz

MNSZ 2.0.4 (Oravecz et al. 2014) → **PrevLex** (Kalivoda 2019)

PrevLex:

- 53 503 igekötős ige és gyakorisága
- UNKNOWN szavak is, pl. *összecuccol*, *szénnécsócsál*
- 235 igekötő(szerűség)

Az itt felhasznált részhalmoz:

hapaxok, amelyeknél az igekötő terjeszkedő produktivitása nagyobb nullánál (Baayen 2009)

→ 122 igekötő(szerűség), 21 038 igekötős igei hapax

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)

Fonológiai és szemantikai információk kinyerése

Kigágogják magukat ezek a gúnárok, aztán abbahagyják...

↓ 1.

ki + gágogják

↓ 2.

gágog → (C)(C)VCOg 4. hangzásséma elemzés, saját script

↓ 3.

a 450-es klaszter tagja

(gágog, brekeg, brummog, cincog, csivog, dübög, kaffog, kurrog, mekeg, nyekereg, nyervog, nyifog, nyihog, rőfög, vakog, vartyog, vijjog)

Az igealkotás produktív esetei

1. Névszó + igeképző

pl. *rokkeresedik, butalibáskodik*

2. Ige + igeképző

pl. *dilemmázgat, aszalódik*

3. Hangzásséma

pl. *slattyog, zsibbaszt*

Névszó + igeképző

Valódi névszói tövekhez járuló, produktív igeképzők:

képző	példák	db
-Vz(ik)	<i>zebracsíkoz, doktoroz, kávézik</i>	2558
-(s)Odik	<i>babásodik, jegesedik, vörösödik</i>	638
-Vl	<i>falfirkál, offtopikol, chatel</i>	525
-(V)(s)ít	<i>bénít, felhőkarcolósít, vidámít</i>	420
-(s)Ul	<i>halkul, szelídül, zombul, szenttehenül</i>	205
-(s)kOdik	<i>flegmáskodik, izmoskodik, mártírkodik</i>	180

Névszó + igeképző

Idegen eredetű névszói tövekhez járuló, produktív igeképzők:

képző	példák	db
-ál	<i>annotál, fókuszál, transzponál</i>	303
-izál	<i>motorizál, szolmizál, tematizál</i>	109
-íroz	<i>lavíroz, morfondíroz, políroz</i>	39
-ifikál	<i>kodifikál, purifikál, identifikál</i>	7

Ritkán magyar szavaknál is, pl. *bolondizál, kántorizál*.

Ige + igeképző

képző	példák	db
-(t)At	<i>kocsikáztat, pötyögtet, oszlat</i>	1412
-Ódik	<i>duplázódik, hekkelődik, szövődik</i>	851
-(O)gAt	<i>hintáztat, fuvintgat, búsulgat</i>	510
-(t)Atik	<i>doboltatik, aranyoztatik, rostáltatik</i>	178

Az -(O)gAt sok alternatívája (nem vagy alig produktívak):
lopkod öldös nézdegél rugdal bukdácsol forrong huzigál

Ezek paradigmába is szerveződnek. (Ladányi 2007)

Hangzásséma

- 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

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

Problémák a vonzatkeret-kinyeréssel

- az adatkinyerés automatikus
→ lehetnek hibák, főleg a vonzatkereteknél
- kötelező vonzat ↔ szabadhatározó
- *magam, magad*, stb. – nem látszik, hogy alany- vagy tárgyeset
- bizonyos esetragok sokszor kiválthatók lennének irány-
hármast jelölő vonzattal, máskor nem, pl:
el {-rA, -bA, ...} → *el* HOVÁ
rá -rA ↗ *rá* HOVÁ

PrevCons: Miből áll?

1-6. oszlopok

	az oszlop tartalma	példa (<i>elcsipszez</i>)
1	ID	7706
2	igekötő	<i>el</i>
3	igelemma	<i>csipszez</i>
4	az igekötős ige szóalakja	<i>elcsipszezem</i>
5	vonzatkeret	-t
6	az igekötő produktivitása	0.0951

7-12. oszlopok

	az oszlop tartalma	példa (<i>elcsipszez</i>)
7	az igekötő hapaxainak száma	2005
8	igealkotási mód	N
9	a képző/hangzásséma típusa	-Vz(ik)
10	CV-séma (hangzásséma esetén)	-
11	stílusregiszter	személyes
12	régió	magyarországi

13-16. oszlopok

	az oszlop tartalma	példa (<i>elcsipszez</i>)
13	word2vec input	chips
14	word2vec klaszter	109
15	tagmondat	<i>meg elcsipszezem ,</i> <i>Én sajnálom a pénzt virágföldre ,</i>
16	teljes mondat	<i><ncl> inkább elsörözöm <ncl> meg</i> <i>elcsipszezem , <ncl> így tőletek kérek</i> <i>segítséget .</i>

PrevCons: Mire használható?

Példa a használatra

Pl. az *el* mellett milyen *névszó* → *ige* képzők és vonzatok jelennek meg legtöbbször?

igekötő	képző	vonzatkeret	db
<i>el</i>	-(s)Odik		196
<i>el</i>	-Vz(ik)	-t	107
<i>el</i>	-(V)(s)ít	-t	72
<i>el</i>	-Vz(ik)		69
<i>el</i>	-(s)Ul		41

Példa a használatra

Milyen igéket jellemez az 'el-N-(s)Odik' szerkezet?

C	el-N-(s)Odik	tagmondat
1	<i>elemdéefesedik</i>	és elemdéefesedett .
2	<i>elbalosodik</i>	ha teljesen elbalosodna ez a fórum .
2	<i>elliberálisodik</i>	ugyanis reménytelenül elliberálisodtak ,
2	<i>elpártosodik</i>	s elpártosodott ,
3	<i>elordasodik</i>	és elordasodik az ő hazája .
3	<i>elsimicskásodik</i>	Mielőtt még teljesen elsimicskásodnánk
6	<i>eltőkésedik</i>	ha a volt elvtársak eltőkésednek ,
7	<i>elmagyarodik</i>	Óhatatlanul elmagyarodott ,

Próbáljuk ki!

<https://aszalo.herokuapp.com>

A további tájékozódáshoz

- a PrevCons elérhetősége:
<https://github.com/kagnes/prevcons>
- Kalivoda Ágnes (2021): Az igekötők produktív kapcsolódási mintái. Argumentum 17. p. 56–82.
<https://doi.org/10.34103/ARGUMENTUM/2021/4>
- Kalivoda Ágnes (2021): Igekötős szerkezetek a magyarban. Doktori értekezés (munkahelyi vitára benyújtott változat).
https://github.com/kagnes/phd_thesis
- kalivoda.agnes@nytud.hu

Hivatkozások I.

Baayen, R. H. (2009): Corpus linguistics in morphology: morphological productivity. In: Lüdeling, A., Kytö, M. (eds.): *Corpus Linguistics. An international handbook, Volume 2*. Berlin: Mouton De Gruyter, 900–919.

Benő, A., Szilágyi N., S. (2015): Hangzásséma és motiváltság a hangutánzó és hangulatfestő igéink körében. In: Kádár, E., Szilágyi N., S. (szerk.): *Motiváltság és nyelvi ikonicitás*. Kolozsvár: Erdélyi Múzeum-Egyesület (EME), 43–57.

Indig, B., Sass, B., Simon, E., Mittelholcz, I., Kundraóth, P., Vadász, N. (2019): emtsv – Egy formátum mind felett. In: Berend, G., Gosztolya, G., Vincze, V. (szerk.): *XV. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2019)*. Szeged: Szegedi Tudományegyetem, TTIK, Informatikai Intézet, 235–247.

Jakab, I. (1976): *A magyar igekötők állományi vizsgálata*. Nyelvtudományi Értekezések 91. Budapest: Akadémiai Kiadó.

Kalivoda, Á. (2019): Véges erőforrás végtelen sok igekötős igére. In: Berend, G., Gosztolya, G., Vincze, V. (szerk.): *XV. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2019)*. Szeged: Szegedi Tudományegyetem, TTIK, Informatikai Intézet, 331–344.

Kiefer, F. (2007): *Jelentéelmélet*. 2. kiadás. Budapest: Corvina.

Hivatkozások II.

Ladányi, M. (2007): Produktivitás és analógia a szóképzésben: elvek és esetek. Budapest: Tinta Könyvkiadó.

Ladányi, M. (2012): Igekötős igék kapcsolódási mintázatai. Vizsgálati lehetőségek. In: Tolcsvai Nagy, G., Tátrai, Sz. (szerk.): *Konstrukció és jelentés. Tanulmányok a magyar nyelv funkcionális kognitív leírására.* Budapest: Eötvös Loránd Tudományegyetem, 71–84.

Novák, A., Rebrus, P., Ludányi, Zs. (2017): Az emMorph morfológiai elemző annotációs formalizmusa. In: Vincze, V. (szerk.): *XIII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2017).* Szeged: Szegedi Tudományegyetem Informatikai Intézet, 70–78.

Oravecz, Cs., Váradi, T., Sass, B. (2014): *The Hungarian Gigaword Corpus.* In Calzolari, N. et al. (szerk.): *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014),* Reykjavík, European Language Resources Association (ELRA).

Rosch, E. H. (1973): Natural categories. *Cognitive Psychology* 4/3. 328–350.

Sass, B. (2011): Igei szerkezetek gyakorisági szótára – egy automatikus lexikai kinyerő eljárás és alkalmazása. Budapest, Pázmány Péter Katolikus Egyetem.

Siklósi, B., Novák, A. (2016): Beágyazási modellek alkalmazása lexikai kategorizációs feladatokra. In: Tanács, A. et al. (szerk.): *XII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2016).* Szeged: Szegedi Tudományegyetem, TTIK, Informatikai Intézet, 3–14.