

# Az ige helyhatározói bővítményeinek megkülönböztetése és az argumentumszerkezeti variánsok korpusz alapú szétválasztása

Szécsényi Tibor, Virág Nándor  
SZTE Általános Nyelvészeti Tanszék

MSZNY 2022. január 27. Szeged

# Vizsgálati témák

- Hogyan lehet az igék argumentumszerkezeti jellemzőit korpuszadatokból meghatározni?
  - modell: argumentumszerkezeti vektor
- Kilenc helyhatározóragos bővítménytípus vizsgálata
  - *bízik/lakik/énekel a ParlamentBEN vagy nevet kínjáBAN*
- Milyen viszonyban lehetnek az ige mellett megjelenő bővítménytípusok?
  - argumentumszerkezeti variánsok meghatározása

## Elméleti háttér

Az argumentumszerkezet vektor  
alapú modellje

## Helyhatározói esetragok

Tematikus vonzat, valódi vonzat,  
szabad bővítmény

## Az adatok feldolgozása

A korpuszadatokról  
a gyakorisági értékekig

## Variánsok elkülönítése

A bővítménytípusok együtt-  
előfordulási értékei

## Összegzés, Irodalom

A tanulmány, a prezentáció és az adatok elérhetők:

<https://github.com/szecsényi/MSZNY2022-Szecsényi-Virag>

# Elméleti háttér

- Mi a baj a hagyományos argumentumszerkezeti modellel?
- Argumentumszerkezeti vektor
  - az ige mellett bővítményként megjelenő bővítménytípusok megjelenési gyakorisága
- Fogalmak magyarázata
  - tagmondat
  - bővítmény, maximális XP
  - bővítménytípusok
  - gyakorisági értékek
- Összehasonlítás más korpusz alapú argumentumszerkezeti modellekkel

# Vonzat-adjunktum megkülönböztetés

Hagyományos argumentumszerkezet-értelmezés: milyen (és mennyi) vonzata van az igének

- A vonzatok nem mindig jelennek meg az ige mellett: ellipszis, pro-drop, hiányos mondat
- Nem csak vonzatok jelennek meg az ige mellett: nem kötelező argumentum, szabad bővítmény, adjunktum
- Nem mindig egyértelmű, hogy valami vonzat vagy adjunktum: *képeslapot küldött **Debrecenből***
- Olyan összetevő is megjelenhet, amivel semmilyen szemantikai kapcsolata sincs az igének: ***El** kell, hogy engedj!* (Demjén F.)

# Argumentumszerkezeti vektor

- Vonzat–adjunktum helyett bővítmény
  - az ige tagmondatában az ige mellett megjelenő maximális összetevők
  - [*Marival*] *adtam* [*oda*] [*Péternek*] [*a könyvet*]
- A bővítményeket legjellemzőbb tulajdonságukkal reprezentáljuk: bővítménytípus
  - általában esettel
  - oda+ad [pv, acc, dat, VAL]
- Nem az egyedi ige+bővítménytípusok adatokkal jellemezzük az igét, hanem a bővítménytípusok ige melletti megjelenési gyakoriságával ( $\approx$ valószínűség)
  - korpuszadatokból származik

	pv	Nom	Acc	Dat	VAL
oda+ad	0,05	0,19	0,68	0,47	0,05

# Tagmondat, maximális összetevő

- A szintaktikai jellemzők függőségi elemzésből származnak
- **megszakítatlan kifejezés:** olyan szólánc, amelyben (legfeljebb) egy szóláncon kívüli függőségi címke található – **fej**
- **maximális megszakítatlan kifejezés:** olyan szólánc (valamilyen korlátok között), amely semelyik irányban nem bővíthető tovább megszakítatlan kifejezéssé
- **bővítmény:** egy maximális megszakítatlan kifejezésben (tagmondatban) a fejet (igét) megelőző vagy azt követő rész maximális megszakítatlan összetevője
- **tagmondat:** igei fejű maximális megszakítatlan kifejezés
- A nem projektív (keresztelő függőség!) kifejezések nem alkotnak egy bővítményt

# Bővítménytípusok

- 32 bővítménytípus
  - Az igék argumentumszerkezete 32 dimenziós valószínűségi vektorral vannak reprezentálva
  - A bővítmény fejének szófaja, morfoszintaktikai jegyei alapján meghatározva
- A funkcionális kifejezéseket nem tekintettük bővítménynek (*nem; sem, is, csak, -e, már, még* stb.)
  - **PV** – igekötő (elváló, vagy beférkőzéssel „megszerzett”)
  - **CP\_cnd; CP\_imp; CP\_ind; HKM; inf** – mellékmondati bővítmény
  - **nom; acc; dat; BAN; ON; RA; VAL; UL; BA; RÓL; HOZ; BÓL; TÓL; NÁL; VÁ; IG; ÉRT; KÉNT; KOR; SZOR; NKÉNT** – eset, főnévi csoport, ragozott névmás, ragozott melléknév
  - **ADP** – névutós kifejezés
  - **ADV** – határozói kifejezés
  - **FROM; IN; TO** – az irányhármasság osztályaiba beilleszthető kifejezések metakategóriája; névutók, névmások, határozók



# Összevetés a duplakocka modellel

## Duplakocka

Sass (2018; 2020)

- az igéhez rendszeresen kapcsolódó bővítménytípusokat és lexémákat határozza meg
- véges háló reprezentáció:  $n$  dimenziós duplakocka
- vonzat–szabad bővítmény megkülönböztetés
- kollokációs szótár, a szótári egységek közötti kapcsolat bemutatásával

## Argumentumszerkezeti vektor

(Szécsényi 2019)

- az igéhez kapcsolódó bővítménytípusokat határozza meg
- $n$  dimenziós folytonos vektor reprezentáció: minden irányban 0 és 1 közötti érték
- nincs vonzat–szabad bővítmény–egyéb bővítmény megkülönböztetés
- az argumentumszerkezetet befolyásoló grammatikai jelenségek vektorműveletként értelmezhetők

# Az adatok feldolgozása

- Az adatkiválasztás szempontjai
- Adatforrások
- Adatfeldolgozás, az adatok szűrése
  - duplikátumok, mondattöredékek
  - magyarlánc függőségi elemző
  - elváló igekötők
  - tagmondatok és bővítményszerkezet
  - gyakorisági táblázatok
- Kontrollkorpusz: Szeged Dependenciakorpusz
  - összesített bővítménymegjelenési gyakoriság

# Az adatkiválasztás szempontjai

- Nem a összes (32) bővítménytípussal dolgoztunk:
  - BÓL, BAN, BA, RÓL, ON, RA, TÓL, NÁL, HOZ – helyhatározói esetragok (3x3)
  - az argumentumszerkezet-variánsokra jellemző bővítménytípusok
- Néhány ige kiválasztása
  - legyen elég sok előfordulás
  - ne legyen túl sok variáns (pl. *megy, tesz*)
  - az igekötő befolyásolja az argumentumszerkezetet: ne legyenek igekötős igék (Gyulai 2019)
  - reprezentálják a várt igitípusokat
- Teljes mondatok
  - duplikátumok szűrése
  - nem nagybetűvel kezdődő adatok szűrése
  - a vizsgált igék ne legyenek főnévi igeneves alakban, és ne legyen főnévi igenévi bővítményük
- Igénként 1000-2000 mondattal indultunk
  - a szűrések után is maradjon elegendő

# Adatforrások

A két feladathoz különböző forrást használtunk

- Helyhatározók: MNSZ2 (Oravecz és mtsai 2014)
  - sok és sokféle adat
  - olyan mondatokat kértünk le, amelyek tartalmazzák a ragozott igekötőtlen igét
  - az elváló igekötős mondatokat utólag kellett szűrni
- Argumentumszerkezeti variánsok: Mazsola (Sass 2009)
  - az elváló igekötőket is számon tartja
  - sokkal több mondattöredék
  - bizonyos bővítménytípusokra nincs példa, pl. *készül* + inf

# Adatfeldolgozás 1. Előszűrés

- A lekért adatokat puszta szöveggé alakítottuk
  - egy sor egy mondat
- A többször szereplő sorok közül csak egy előfordulás maradt
- A kisbetűvel / nem nagy betűvel kezdődő sorokat töröltük
  - lehetőleg csak teljes mondatok maradjanak
- A sorokat (mondatokat) ábécésorrendbe rendeztük

# Adatfeldolgozás 2. Magyarlánc

- A szöveges fájlokat a magyarlanc elemzővel feldolgoztuk depparse módban (Zsibrita és mtsai 2013)
- A kapott .tsv fájlokat .xml fájlá alakítottuk
  - a mondatok és a szavak (tokenek) egyedi azonosítót kaptak
  - könnyebb a hozzáférés a szavak egyes morfoszintaktikai jegyeihez
- A magyarlanc elemzése nem tökéletes, de tegyünk úgy, mintha az lenne

# Adatfeldolgozás 3a. Igekötő-kezelés

A feldolgozott korpuszban megkülönböztetjük az igék igekötőtlen és igekötős lemmáját

- Tapadó igekötők: a magyarlanc nem jelöl minden tapadó igekötőt
  - a Szeged korpuszból származó igelista alapján az ismert igekötős igék szétválasztása
  - a Szeged korpuszban nem szereplő alakok esetében szétválasztás ismert igekötőre és ismert igekötő nélküli igére
  - probléma pl.: *kirándul* =?= *ki+rándul*

# Adatfeldolgozás 3b. Igekötő-kezelés

- Elváló igekötők igéjükhöz kapcsolása
  - nem a függőségi elemzés megváltoztatása!
  - kézi ellenőrzés (csak a határozóragos korpusznál)
    - ha a vizsgált ige igekötős: törölni a mondatot
    - ha nem igekötő lett igekötőnek elemezve: javítani
    - ha nem megfelelő igéhez lett kapcsolva: javítani
    - egyébként az igekötőt az igéhez kapcsolni
- Az igék kibővített jellemzése:

<code>&lt;form&gt;lelőtték&lt;/form&gt;</code>	szóalak
<code>&lt;lemma&gt;lelő&lt;/lemma&gt;</code>	a magyarlánc által meghatározott lemma
<code>&lt;lemma2&gt;lő&lt;/lemma2&gt;</code>	az igekötő nélküli lemma
<code>&lt;pv pvid="0"&gt;le&lt;/pv&gt;</code>	az igekötő (0 ha igekötő nélküli) és az igekötő azonosítója (0 ha tapadó)



# Adatfeldolgozás 4. Tagmondatok és bővítmények

- A függőségi elemzés alapján meghatároztuk a tagmondatokat és a tagmondati bővítményeket
  - külön .xml fájlban felsorolva

*...Közben kilencvenforintos órabérért bejön, és mondja...*

```
<cp>      <start>s115w2</start>
          <head pv="be" pvid="0" lemma="jön" mood="ind">s115w5</head>
          <end>s115w13</end>
          <xp>      <start>s115w2</start>
                    <head lemma="közben" f="ADV">s115w2</head>
                    <end>s115w2</end></xp>
          <xp>      <start>s115w3</start>
                    <head lemma="órabér" f="ÉRT">s115w4</head>
                    <end>s115w4</end></xp>
</cp>
```

# Adatfeldolgozás 5. Variánsok

- A vizsgált igék argumentumszerkezeti variánsainak elkülönítése
  - különböző jelentésű variánsok
  - anyanyelvi intuíció alapján
  - kézi annotálás
- Csak az argumentumszerkezeti variánsok elkülönítésénél!

# Adatfeldolgozás 6. Előfordulási gyakoriságok

- A tagmondatokban szereplő bővítménytípusok táblázatba foglalása
- Első oszlopok: a tagmondat fejének (=ige) tulajdonságai
- További oszlopok: előfordul-e az adott bővítménytípus a tagmondatban?

id	pv	lemma	pvv	mood		PV	nom	acc		ÉRT		ADV	
s115w5	be	jön	be+jön	ind	...	0	0	0	...	1	...	1	...
...													

- A táblázatból igénkénti összesítéssel megkapható az ige korpuszban való előfordulási számai, illetve az összes bővítménytípus ige melletti előfordulási számai
- Ebből származtatható a bővítménytípusok ige melletti előfordulási gyakorisága: argumentumszerkezeti vektor
- → **Adatok elemzése**

# Kontrollkorpusz: Szeged Dependency Treebank

- Az argumentumszerkezeti vektorokat a Szeged Dependency Treebank (Vincze és mtsai 2010) összesített argumentumszerkezeti vektorával vetettük össze
  - mert kellően sokféle adatot tartalmaz
  - mert kézzel annotált (gold standard)
    - de eltér a magyarlánc annotációs sémájától
  - korábbi vizsgálatokból rendelkezésre állt
- Összesített vektor: milyen gyakorisággal fordul elő egy bővítménytípus egy tagmondatban
  - nem az igénkénti gyakoriságot vettük
  - csak a kijelentő módú ígés tagmondatokat vettük figyelembe

# Az SZDT összesített argumentumszerkezeti vektora

argType	freq
nom	0,559251
acc	0,326353
dat	0,055825
BAN	0,115328
ON	0,099435
RA	0,085415
VAL	0,083331
UL	0,062121
BA	0,033539
RÓL	0,024550
HOZ	0,023189

argType	freq
BÓL	0,022399
TÓL	0,016555
NÁL	0,011734
VÁ	0,008620
IG	0,010440
ÉRT	0,005942
KÉNT	0,006258
KOR	0,005671
SZOR	0,003663
NKÉNT	0,001271

argType	freq
PV	0,117021
CP_cnd	0,012245
CP_imp	0,004107
CP_ind	0,186934
HKM	0,094095
inf	0,104655
ADP	0,105482
ADV	0,324841
FROM	0,065994
IN	0,261352
TO	0,161969

# Helyhatározói esetrágok eloszlási mintázatai

- Kilenc helyhatározórágos bővítménytípus vizsgálata
  - BÓL, BAN, BA, RÓL, ON, RA, TÓL, NÁL, HOZ
  - + 3 irányhármasság: FROM, IN, TO
- Miért van az, hogy egyes ígék mellett váltakoznak, mások mellett nem
  - *DebrecenBEN/SzegedEN lakik vagy egyetemRE/iskoláBA jár*
  - *DebrecenBEN/\*SzegedEN bízik vagy panaszkodik az egyetemRE/\*iskoláBA*
- Hogyan lehet korpuszvizsgálattal meghatározni a helyhatározói bővítmények funkcióját?
- 13 ige vizsgálata: *ad, beszél, fél, hisz, indul, javasol, jön, kap, lát, nevet, rak, terem, úszik*

# Helyhatározói esetrágok eloszlási mintázatai

- Hipotézis: a helyhatározói alapjelentésű rágos kifejezéseket a főigékükhöz való viszonyuk alapján korpuszalapú módszerekkel 4 csoportra lehet osztani
- Csoportok:

	Vonzat	Szabad bővítmény
Nem kompozicionális	1. Valódi vonzat pl. <i>bíz</i> ik PéterBEN	3. Egyéb szabad határozó pl. <i>kín</i> jaBAN nevet
Kompozicionális	2. Tematikus vonzat pl. <i>Debrecen</i> BEN/SzegedEN lakik	4. Szabad helyhatározó pl. <i>énekel</i> az erdőBEN

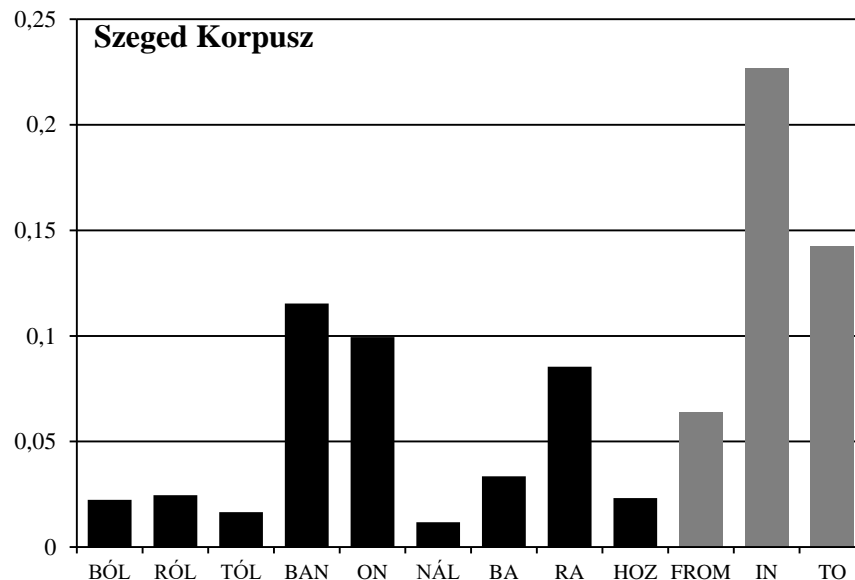
# Helyhatározói esetrágok eloszlási mintázatai

- A vizsgált igék kiválasztási szempontjai:
  - az adott ige megfelelően nagy számban forduljon elő
  - előreláthatólag minden, a hipotézis szerint létező kategória reprezentatív mennyiségben jelenjen meg
  - ne legyen túl sok variánsa



# A Szeged Korpusz adatainak eloszlási mintázatai

- Következtetések a Szeged Korpusz mintázataiból:
  - az adatok megadják a vizsgált ragok általános eloszlását
  - a -bAn, -On és -rA a leggyakoribb, a többi viszonylag ritka
  - a legritkább terc a közelítő viszonyt kifejező

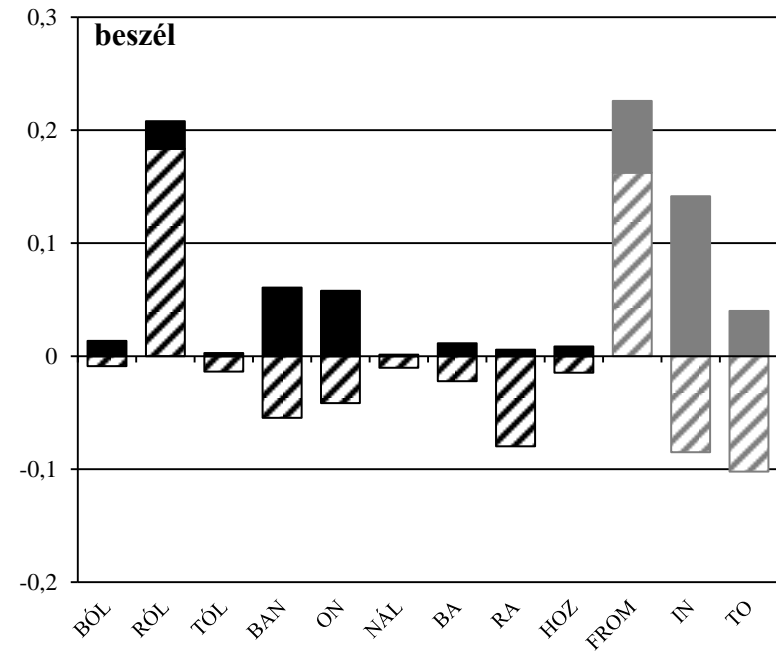


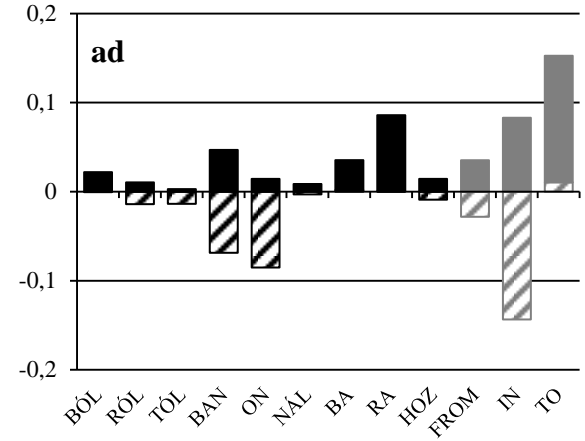
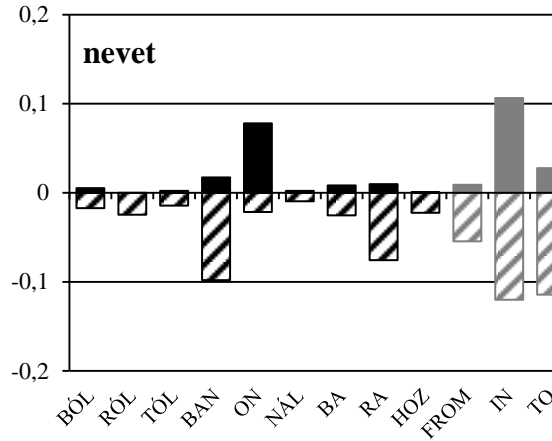
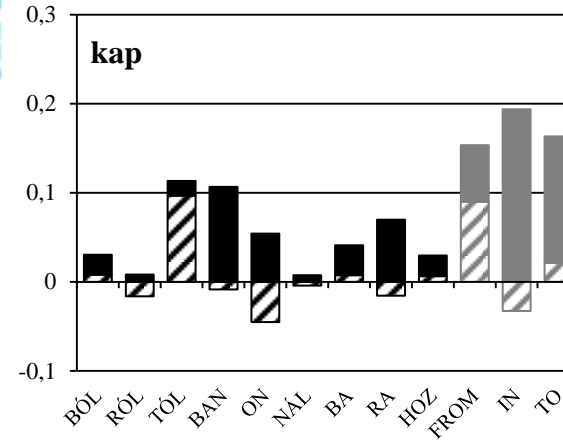
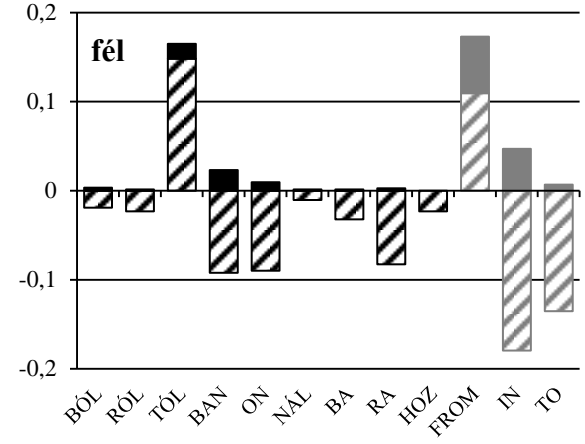
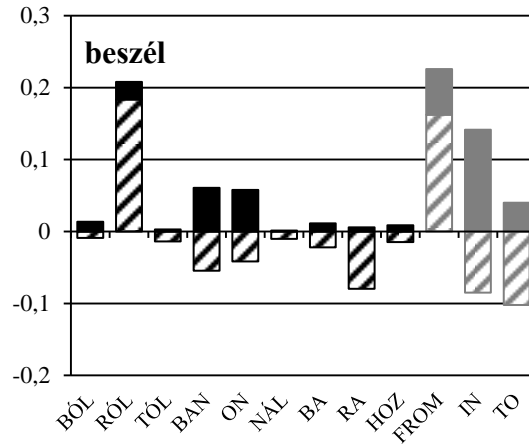
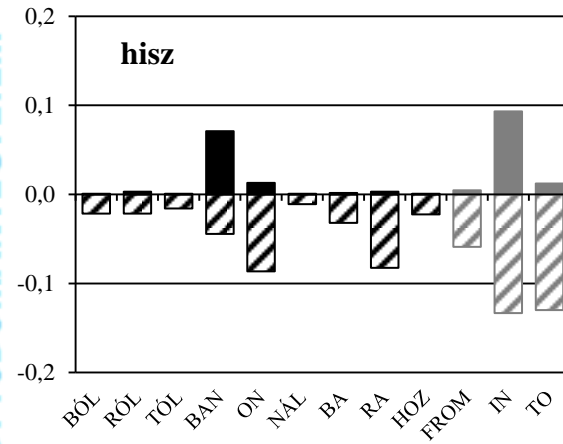
# Valódi vonzatok

- A csoport jellemzői:
  - az ilyen bővítménnyel álló igék egy konkrét ragos kifejezést vonzanak
  - a vonzatként álló ragos kifejezések nagyon magas gyakorisággal fordulnak elő az általános eloszláshoz képest, míg a többi vizsgált rag előfordulása alacsony
  - jelentésük nem kompozicionális, valamilyen absztrakt viszonyt fejeznek ki
  - a vizsgált igék közül ebbe a kategóriába tartozik: *beszél, kap, hisz, nevet, fél, ad*

# Valódi vonzatok

- A *beszél* jól mutatja a csoport jellemzőit:
  - vonzata (RÓL) jelentősen megnövekedett előfordulással áll az általános eloszláshoz képest
  - minden más vizsgált ragos kifejezése csökkenést mutat
- X tengely fölötti oszlopmagasság: előfordulási gyakoriság a vizsgált igénél
- Becsített rész: változás a SZDT adataihoz képest (növekedés, csökkenés)



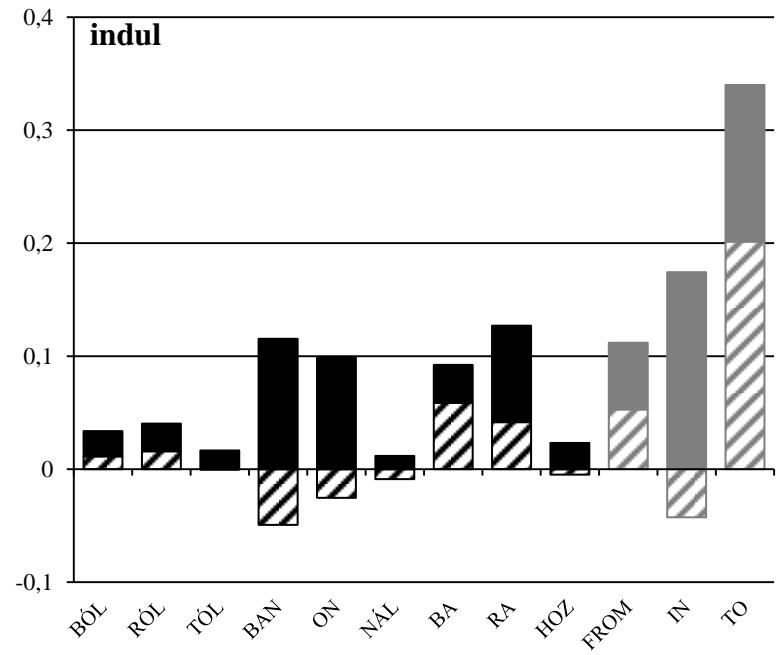


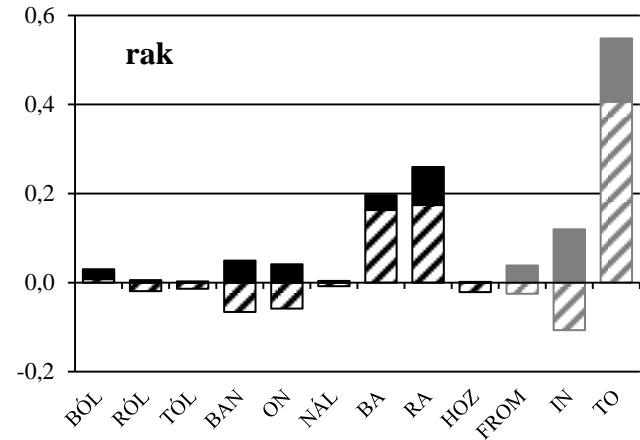
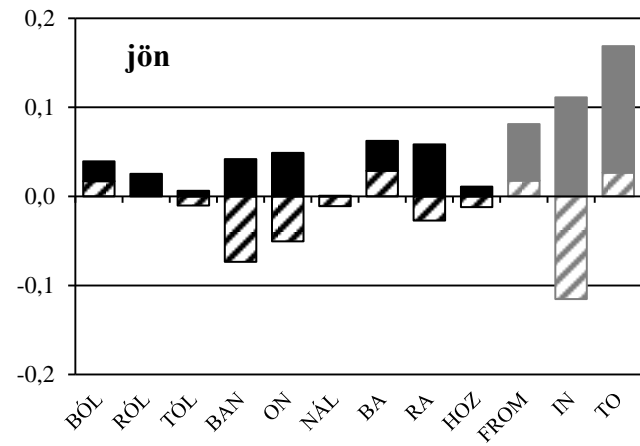
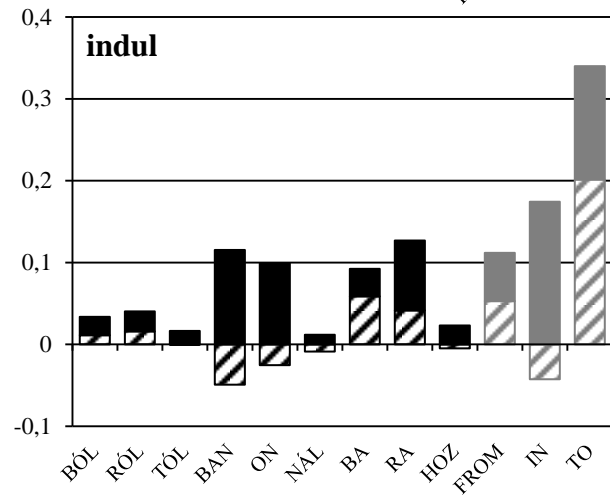
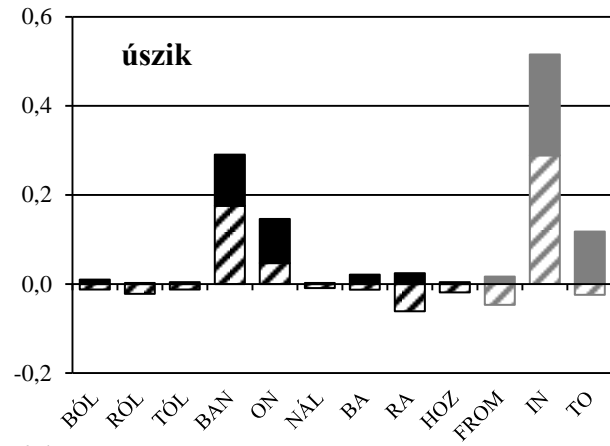
# Tematikus vonzatok

- A csoport jellemzői:
  - az igei alaptag több különböző, de irányhármasság szerint azonos csoportba tartozó ragos kifejezést vonz
  - a csoport tagjai egymással paradigmaticus viszonyban vannak, úgy viselkednek, mintha egyazon vonzat különböző alakváltozatai lennének
  - jelentésük kompozicionális, helyviszonyt fejeznek ki
  - a vizsgált igék közül ebbe a kategóriába tartozik: *úszik, rak, indul, jön*

# Tematikus vonzatok

- Az *indul* jól mutatja a csoport jellemzőit:
  - látható, hogy azokban a csoportokban, amelyek tematikus vonzatként állnak, emelkedett előfordulás látható, míg ahol nem, ott csökkenés
  - ha külön-külön nem is, a vonzatként álló ragos kifejezések együttesen jelentősen megemelik a gyakorisági mutatókat





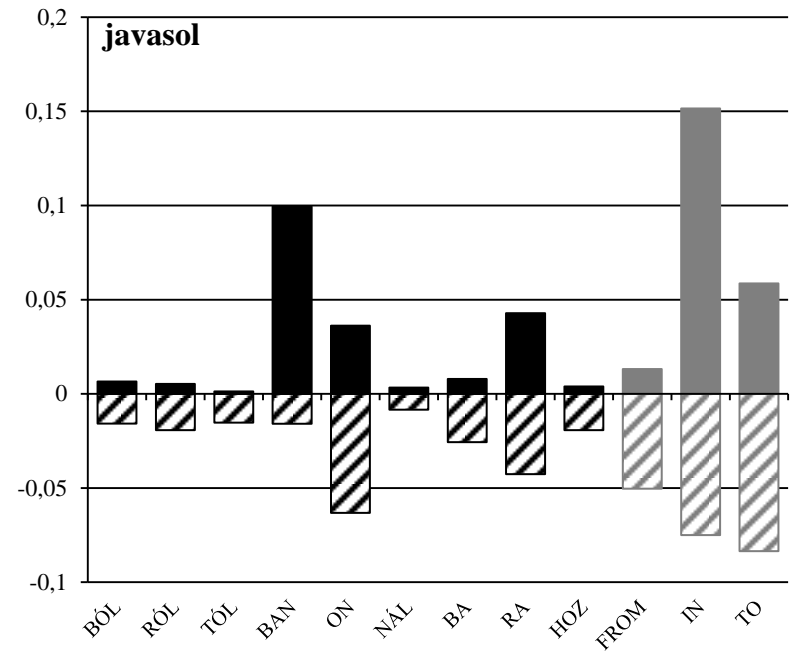
# Szabad bővítmények

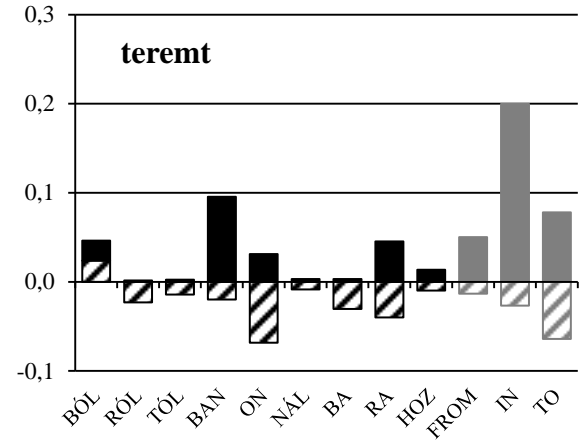
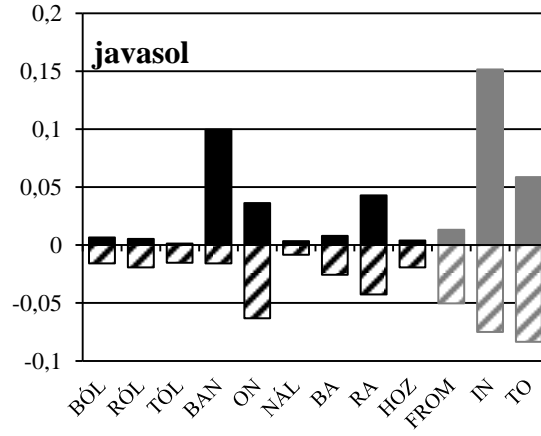
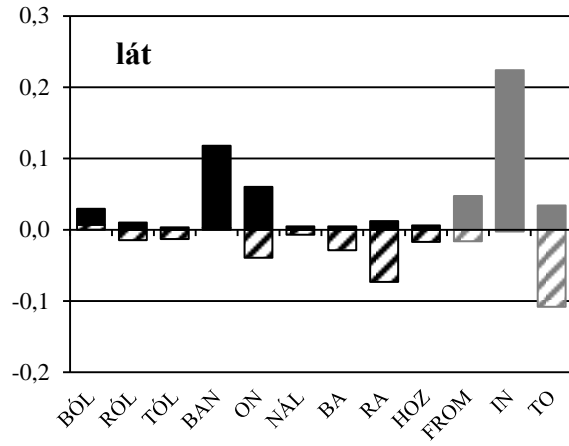
- A csoport jellemzői:
  - két különböző csoport tagjaiból áll:
    - egyéb szabad határozók – nem kompozicionális jelentés, esetleges előfordulás
    - szabad helyhatározók – kompozicionális jelentés, esetleges előfordulás
  - a csoportokat csak jelentés alapján lehet egymástól elkülöníteni, ezt azonban jelen tanulmány nem vizsgálja
  - a vizsgált igék közül ebbe a kategóriába tartozik: *javasol*, *lát*, *teremt*



# Szabad bővítmények

- A *javasol* jól mutatja a csoport jellemzőit:
  - a Szeged Korpuszhoz hasonló mintázatot mutat
  - minden ragnál csökkenés látható – valószínűleg azért, mert a Szeged Korpuszban lévő igék között van olyan, amely ezeket a ragos kifejezéseket vonzza





# Helyhatározói esetragok eloszlási mintázatai

- Összegzés:
  - a megfigyelt adatok alapján igazolódni látszik a hipotézis
  - felmerülő kérdés: hogy lehet megkülönböztetni a tematikus vonzatokat az egymástól független vonzatszerkezetektől?

# Argumentumszerkezeti variánsok elkülönítése

- Milyen kapcsolat van két bővítménytípus között, ha egy ige mellett mindkettő gyakran megjelenik (vonzat)?
- A bővítménytípusok együtt-előfordulási gyakoriságát vizsgáljuk
  - várt gyakoriság vs. megfigyelt gyakoriság
- Ha a megfigyelt együtt-előfordulási gyakoriság töredéke a vártnak, akkor több argumentumszerkezet-variánst feltételezhetünk

# Lehetséges kapcsolatok

- Egyszerre jelennek meg
  - *mesélt a kalandjaiRÓL a közönségNEK*
- Váltják egymást, de összetartoznak
  - *elindult az egyetemRE/iskoláBA*
  - „tematikus vonzat”, egy argumentumszerkezeti variáns
- Váltják egymást, de nem tartoznak össze
  - *a katona ólomBÓL/hadgyakorlatRA készült*
  - két argumentumszerkezeti variáns
- Egyéb kapcsolat
  - *Mari (kávéT) főz* implicit argumentum
  - *Mari (bajkeverőNEK) hívta a rendőrségeT*

# Vizsgálati módszer

- Az egyedi mondatok vizsgálata nem elegendő
  - csak értelmezéssel tudjuk megállapítani az argumentumszerkezet hiányzó elemeit, ahhoz viszont ismerni kell az argumentumszerkezetet
- A bővítménytípusok gyakorisági értékei nem mutatják meg az együtt-előfordulási hajlandóságukat
- A bővítménytípusok egyedi előfordulási gyakoriságát és a bővítménytípusok együtt-előfordulási gyakoriságait kell összevetni

# Kiinduló feltételezések

- Egy ige ( $v$ ) melletti két bővítménytípus ( $A$  és  $B$ ) kapcsolatát vizsgáljuk
  - $A$  és  $B$  előfordulási gyakorisága  $p_v(A)$  és  $p_v(B)$
- Feltételezzük, hogy egy variánsnál a bővítménytípusok előfordulásai nem befolyásolják egymást
  - $p_v(A) * p_v(B) = p_v(A \& B)$
- Az ige két argumentumszerkezeti variánsa ( $v_1$  és  $v_2$ ) esetében
  - mindkét variánusra teljesül az előző állítás:  $p_{v_1}(A) * p_{v_1}(B) = p_{v_1}(A \& B)$
  - a bővítménytípusok előfordulási gyakorisága különbözhet:  $p_{v_1}(A) \neq p_{v_2}(A)$ 
    - az adjunktumok esetében megegyezhet
  - az ige összesített, mindkét variánst számba vevő előfordulási gyakoriságainál viszont nem:  $p_v(A) * p_v(B) \neq p_v(A \& B)$ 
    - $p_v(A) = p(v_1) * p_{v_1}(A) + p(v_2) * p_{v_2}(A)$
    - ahol  $p(v_i)$  az  $i$ . variáns relatív előfordulási gyakorisága:  $p(v_1) + p(v_2) = 1$

# Hipotézisek

- Ha egy ige összesített előfordulási gyakoriságánál  $p_v(A) * p_v(B) = p_v(A \& B)$ , akkor egyetlen argumentumszerkezeti variáns van
- Ha A és B különböző argumentumszerkezeti variánsra jellemzők, akkor  $p_v(A \& B) = 0$ 
  - kiegészítő disztribúció
  - ha  $p_v(A \& B) = 0$ , akkor legalább két argumentumszerkezeti variáns van
  - persze A és B lehet adjunktum is, ezért inkább:  $p_v(A) * p_v(B) \gg p_v(A \& B)$
  - *karácsonyRA ólomBÓL készülnek a katonák (, egyébként másból)*



# Hipotézisellenőrzés

- Négy ige korpuszadatai – várhatóan különböző eloszlások, reprezentánsok
  - $0+készül_1 + BÓL; 0+készül_2 + RA$ 
    - két variáns
  - $be+von_1 + acc + VAL; be+von_2 + acc + BA$ 
    - két variáns, van közös argumentum
  - $0+hív_1 + acc; 0+hív_2 + acc + dat$ 
    - két variáns, van közös argumentum, az egyik variánsban van másik argumentum is
  - $0+mesél + acc + dat + RÓL$ 
    - egy variáns, több argumentum
- A korpuszból nyert adatokat manuálisan soroltuk variánsokba
- **Eredmények** (részletesen a köv. dián)
  - $p_{készül}(BÓL) * p_{készül}(RA) = 0,076 \gg p_{készül}(BÓL \& RA) = 0,004$  két variáns!
  - $p_{bevon}(VAL) * p_{bevon}(BA) = 0,108 \gg p_{bevon}(VAL \& BA) = 0,021$  két variáns!
  - $p_{hív}(acc) * p_{hív}(dat) = 0,309 > p_{hív}(acc \& dat) = 0,213$  két variáns?
  - $p_{mesél}(RÓL) * p_{mesél}(dat) = 0,057 > p_{mesél}(RÓL \& dat) = 0,023$  egy variáns?

	bevon <sub>1+2</sub>		bevon <sub>1</sub>		bevon <sub>2</sub>	
ige	289		251		38	
acc	258	.893	222	.884	36	.947
VAL	41	.142	5	.020	36	.947
BA	219	.758	218	.869	1	.003
p(acc)*p(VAL)		.127		.018		.897
acc&VAL	38	.131	4	.016	34	.894
p(acc)*p(BA)		.676		.768		.025
acc&BA	195	.674	194	.773	1	.026
p(VAL)*p(BA)	(31)	<b>.108</b>		.017		.025
VAL&BA	6	<b>.021</b>	5	.020	1	.026

	készül <sub>1+2</sub>		készül <sub>1</sub>		készül <sub>2</sub>	
ige	475		46		429	
BÓL	42	.088	36	.782	3	.007
RA	409	.861	0	.000	409	.953
p(BÓL)*p(RA)	(36)	<b>.076</b>		.000		.007
BÓL&RA	2	<b>.004</b>	0	.000	2	.005

	mesél	
ige	302	
acc	77	.255
RÓL	157	.520
dat	33	.109
p(acc)*p(RÓL)		.133
acc&RÓL	32	.106
p(acc)*p(dat)		.028
acc&dat	9	.030
p(RÓL)*p(dat)	(17)	<b>.057</b>
RÓL&dat	7	<b>.023</b>

	hív <sub>1+2</sub>		hív <sub>1</sub>		hív <sub>2</sub>	
ige	681		287		394	
acc	377	.554	224	.780	153	.388
dat	380	.558	4	.014	376	.954
p(acc)*p(dat)	(208)	<b>.309</b>		.011		.370
acc&dat	145	<b>.213</b>	3	.010	142	.360

# Értékelés

- Ha a bővítmények várt együtt-előfordulási gyakorisága ( $p_v(A) * p_v(B)$ ) sokszorosa a megfigyeltnek ( $p_v(A \& B)$ ), akkor két argumentumszerkezeti variánsba tartoznak
- Ha megegyeznek, akkor ugyanabba a variánsba tartoznak
- Ha eltérnek, de nem nagyon, akkor
  - *mesél*: egy variánsba tartoznak, de nem elsőrendű argumentumok (a kommunikáció során nem akarunk minden körülményt megnevezni, csak a fontosakat), vagy
  - *hív*: két variáns van, amelyeknél az egyik bővítménytípus mindkettőben jelen van, a másik viszont csak az egyikben (?)
- Mi a helyzet a tematikus vonzatokkal?

# Összegzés

1. A SZDT összes igéjének valószínűségi vektorához viszonyítottunk
  - csak az egyik bővítménytípus megjelenési gyakorisága növekedik: valódi vonzat
  - több bővítménytípus gyakorisága is megnő: tematikus vonzat
  - egyetlen bővítménytípus gyakorisága sem növekedik: szabad bővítmény
2. Bővítménytípus-párok várt ( $p_v(A) * p_v(B)$ ) és megfigyelt ( $p_v(A \& B)$ ) együtt-előfordulási gyakoriságát vizsgáltuk
  - ha a várt és megfigyelt gyakoriság megegyezik: egy argumentumszerkezeti variánsba tartoznak
  - ha a várt gyakoriság sokszorososa a megfigyeltnek: két argumentumszerkezeti variánsba tartoznak

# Hivatkozások

- Gyulai, L.: Nem kompozicionális igekötős igék argumentumszerkezetének korpuszalapú vizsgálata. In: Ludányi, Zs., Grácsi, T. E. (szerk.) Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből 2019. XIII. Alkalmazott Nyelvészeti Doktoranduszkonferencia. pp. 44–58. MTA Nyelvtudományi Intézet, Budapest (2019) doi:10.18135/Alknyelvdok.2019.13.4
- Oravecz, Cs., Váradi, T., Sass, B.: The Hungarian Gigaword Corpus. In: Calzolari, C., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., Piperidis, S. (szerk.) Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14). pp. 1719–1723. European Language Resources Association (ELRA), Reykjavik (2014)
- Sass, B.: „Mazsola” - eszköz a magyar igék bővítményszerkezetének vizsgálatára. In: Váradi Tamás (szerk.) Válogatás az I. Alkalmazott Nyelvészeti Doktorandusz Konferencia előadásaiból. pp. 117–129. MTA Nyelvtudományi Intézet, Budapest (2009)
- Sass, B.: Az igei szerkezetek algebrai struktúrája, avagy a duplakocka modell. In: Argumentum 14. pp. 12–44. (2018)
- Sass, B.: A duplakocka modell és az igei szerkezeteket kinyerő „ugrik és marad” módszer nyelvfüggetlensége, valamint néhány megjegyzés az UD annotáció univerzalitásáról. In: XVI. Magyar Számítógépes Nyelvészeti Konferencia. pp. 399–407. Szegedi Tudományegyetem, Informatikai Intézet, Szeged (2020)
- Szécsényi, T.: Argumentumszerkezet-variánsok korpusz alapú meghatározása. In: Berend, G., Gosztolya, G., Vincze, V. (szerk.) XV. Magyar Számítógépes Nyelvészeti Konferencia. pp. 315–329. Szegedi Tudományegyetem, Informatikai Intézet, Szeged (2019)
- Vincze, V., Szauter, D., Almási, A., Móra, Gy., Alexin, Z., Csirik, J.: Hungarian Dependency Treebank. In: Proceedings of the Seventh Conference on International Language Resources and Evaluation. pp. 1855–1862. European Language Resources Association, Valletta, Málta (2010)
- Zsibrita, J., Vincze, V., Farkas, R.: magyarlanc: A Toolkit for Morphological and Dependency Parsing of Hungarian. In: Proceedings of RANLP 2013. pp. 763–771 (2013)

# KÖSZÖNJÜK A FIGYELMET!