

# Az igekötők jegyalapú osztályozása

Kalivoda Ágnes

Budapest, 2019. június 28.

# Bevezetés

## Mik az igekötők és hogyan határozható meg az állományuk?

### Javaslat:

- prototípus elméleti megközelítés
- az igekötőség jegyeinek mérése korpuszon, klaszterezés

### Az előadás vázlata:

1. Háttér
2. Vizsgált jegyek
3. Jegyalapú osztályozás
4. Összegzés

# A kutatás háttere

# Elmosódó határok

Az igeekötők az igemódosítók osztályába tartoznak  
→ hasonló szemantikai és szintaktikai viselkedés (Kiefer 2007)

## Igekötő

**El**mentem.

**El** is mentem.

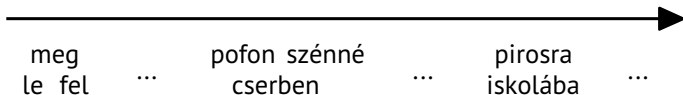
Nem mentem **el**.

## Pusztá névszói igemódosító

**Munkába** mentem.

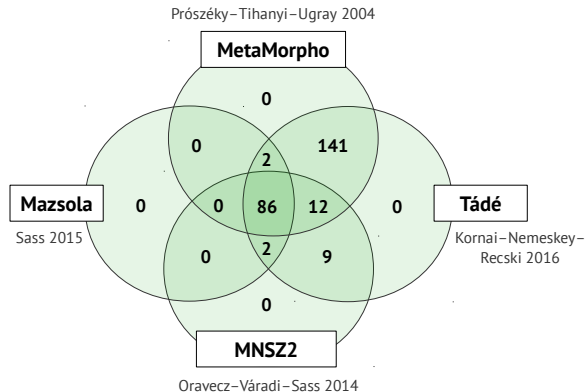
**Munkába** is mentem.

Nem mentem **munkába**.



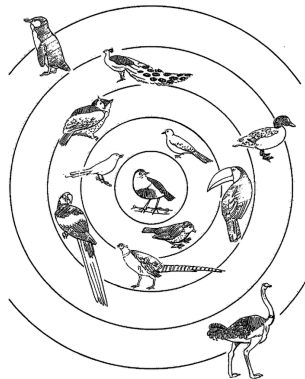
# Az egyetértés hiánya

Lehetséges igekötők a **Manócskában** (Kalivoda–Vadász–Indig 2018):



# Prototípus-elmélet

- holisztikus kognitív szemantika
- a kategóriák prototípusok köré szerveződnek
- a kategória-tagság fokozatos
- **az igekötőkre is alkalmazható**  
(Forgács 2005, Kerekes 2011)



(Aitchison 2004: 54)

# A kutatás módszere

1. igekötő-jelöltek a Manócskából
2. a *mérhető* jegyek meghatározása
3. korpuszmérések: minden igekötő-jelölthöz minden jegy  
MNSZ 2.0.4 (Oravecz–Váradi–Sass 2014)
4. adatelemzés
  - jegypárok közötti korrelációk
  - K-közép klaszterelemzés
  - UMAP vizualizáció

**Köszönet Makrai Mártonnak!**

# Az igekötőség jegyei



# Morfológiai jegyek

- 1. Szerkesztettség:** elemezhető-e többmorfémásnak?
- 2. Duplikálhatóság:**  
pl. **Neki-neki**ütődtek a kocsi falának.
- 3. Fokozhatóság:**  
pl. A három öregúr **összéb**húzódott a padon [...]
- 4. Személyragozhatóság:**  
pl. [...] ráadásul a felnőttek is folyton **belénk** kötnek.
- 5. A -fele/felé szócska csatolhatósága:**  
pl. [...] de most már **oldódik megfele** a dolog.

# Morfológiai jegyek

## 6. Morfológiai produktivitás: kvantitatív vizsgálat (Baayen 2009)

### Megvalósult:

$$P_m = \frac{\text{az affixumot tartalmazó lemmák (típus)}}{\text{a korpusz összes lemmája (típus)}} \quad \textit{el, meg, le}$$

### Terjeszkedő:

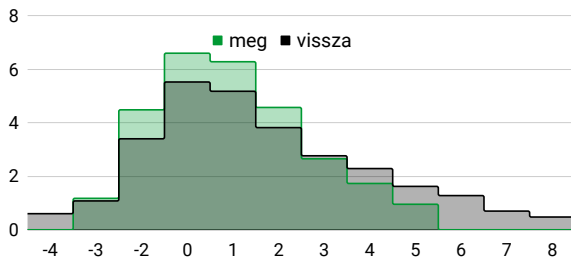
$$P_t = \frac{\text{az affixumot tartalmazó hapaxok}}{\text{a korpusz összes hapaxa}} \quad \textit{el, le, meg}$$

### Lehetséges:

$$P_l = \frac{\text{az affixumot tartalmazó hapaxok}}{\text{az affixumot tartalmazó összes token}} \quad \textit{szarrá, szénné}$$

# Szintaktikai jegyek

## 7. Eloszlás a finit ígérez képest:



pl. [...], **tegyük** ezt most mi is **meg**, [...]

[...] **vitt** egy feltehetően részeg buszsofőr Szentesen még csütörtökön egy sportrendezvény után **vissza** az iskolába.

## Szintaktikai jegyek – Általános jegyek

### 8. Elválás deverbális főnévtől tagadás esetén:

pl. Elsőbbség **meg nem adása** más járműveknek.

Kezdve a szőkítéssel, folytatva a **be sem színezéssel**, [...]

### 9. Szótagszám:

*meg*: 1

*agyon*: 2

*keresztbe*: 3

### 10. Gyakoriság:

*meg*: 4 897 845

*agyon*: 10 196

*keresztbe*: 844

# Az igekötők jegyalapú osztályozása

# Igekötő × Jegy mátrixok

## Igekötők

minden szóalak külön

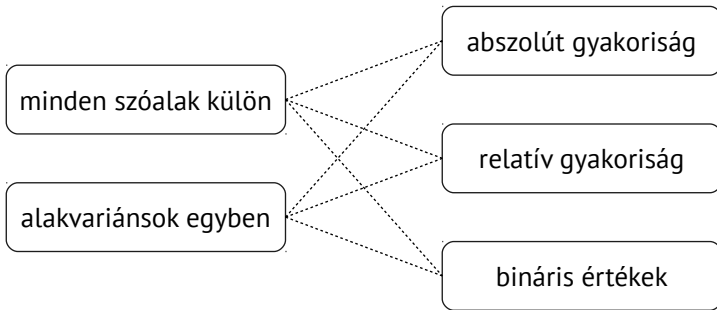
alakvariánsok egyben

## Jegyek értékei

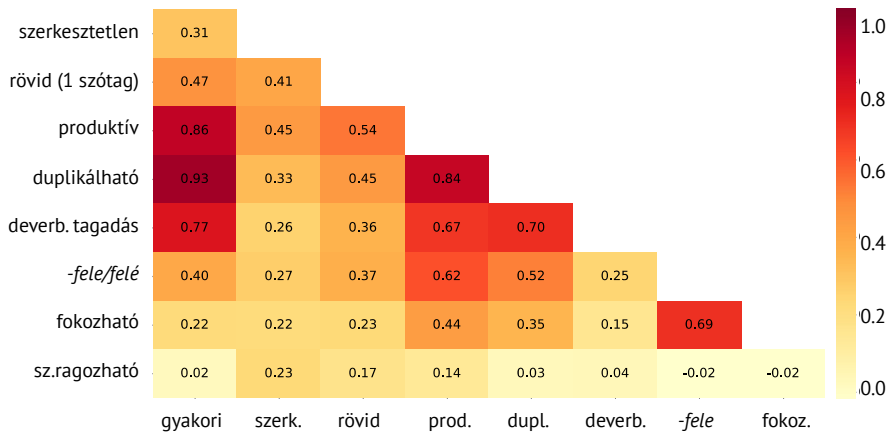
abszolút gyakoriság

relatív gyakoriság

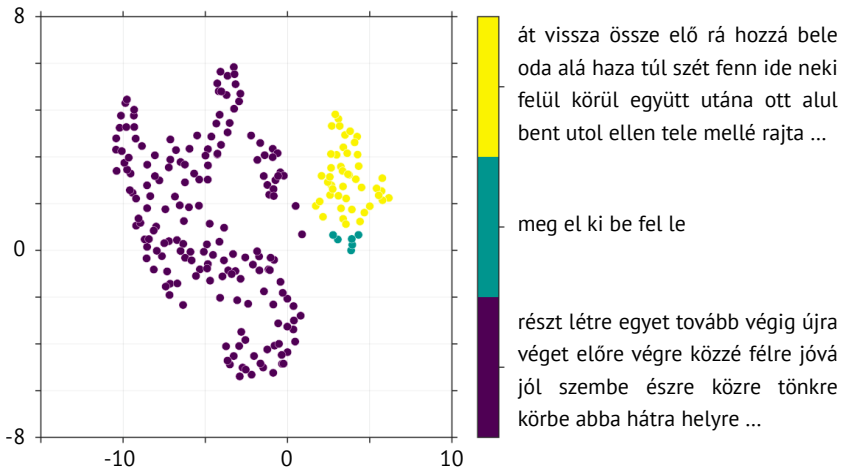
bináris értékek



# Jegyek közötti korrelációk

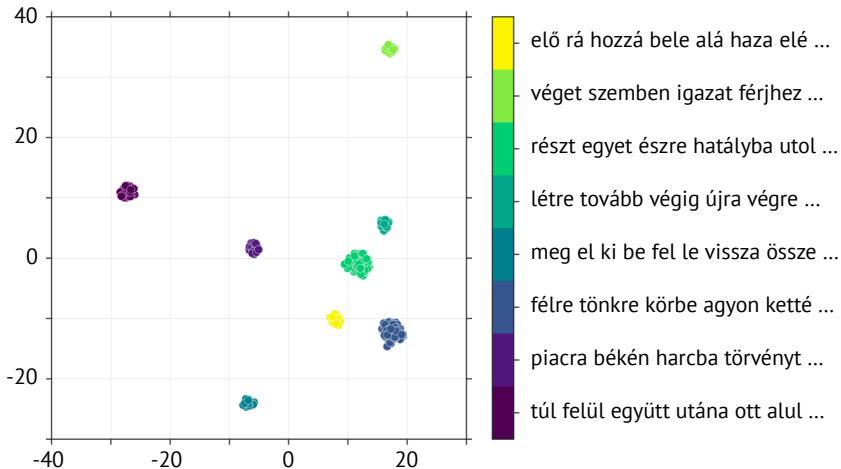


# Klaszterezés – abszolút gyakoriság





# Klaszterezés – bináris értékek



# Az eredmények összefoglalása

## 1. Új, nyílt hozzáférésű erőforrások

- PrevLex táblázat: [github.com/kagnes/prevlex](https://github.com/kagnes/prevlex)
- Igekötő × Jegy mátrixok: [github.com/kagnes/prev\\_features](https://github.com/kagnes/prev_features)

## 2. Átfogó korpuszméréssel vizsgált jegyek

- hasznos más elméleti keretek számára is

## 3. Az igekötő-jelöltek csoportosítása K-közép klaszterezéssel

## 4. Jegyek közti alapvető összefüggések igazolása

# Hivatkozások I.

**Aitchison, J. (2004):** *Words in the Mind: An Introduction to the Mental Lexicon*. 4. kiadás. Oxford, Egyesült Királyság, Basil Blackwell Publishers.

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

**Forgács, T. (2005):** *Grammatikalizálódás az igekötők körében*. In Oszkó, B. – Sipos, M. (szerk.): *Uráli grammatizáló. BUM 4*. Budapest, MTA Nyelvtudományi Intézet.

**Kalivoda, Á. – Vadász, N. – Indig, B. (2018):** *Manócska: A Unified Verb Frame Database for Hungarian*. In Sojka, P., et al. (szerk.): *TSD 2018*, Brno, Csehország, Springer-Verlag.

**Kalivoda, Á. (2019):** *Véges erőforrás végtelen sok igekötős ígére*. In Berend, G., et al. (szerk.): *MSZNY 2019*, Szeged, Szegedi Tudományegyetem Informatikai Tanszékcsoport.

**Kerekes, J. (2011):** *Az igekötők meghatározásának problémái*. In Gécseg, Zs. (szerk.): *Ling-Dok 10. Nyelvészdoktoranduszok dolgozatai*, Szeged, JATEPress.

**Kiefer, F. (2007):** *Jelentéselmélet*. Budapest, Corvina Kiadó.

**Kornai, A. – Nemeskey, D. M. – Recski, G. (2016):** *Detecting Optional Arguments of Verbs*. In Calzolari, N., et al. (szerk.): *LREC 2016*, Portorož, Szlovénia, ELRA.

## Hivatkozások II.

**Oravecz, Cs. – Váradi, T. – Sass, B. (2014):** The Hungarian Gigaword Corpus. In Calzolari, N., et al. (szerk.): LREC 2014, Reykjavik, Izland, ELRA.

**Prószéky, G. – Tihanyi, L. – Ugray, G. (2004):** Moose: a robust high-performance parser and generator. In Hutchins, J. (szerk.): EAMT 2004, La Valletta, Málta, Foundation for International Studies.

**Sass, B. (2015):** 28 millió szintaktikailag elemzett mondat és 500 000 igei szerkezet. In Tanács, A., et al. (szerk.): MSZNY 2015, Szeged, Szegedi Tudományegyetem Informatikai Intézet.