

# Névszói kötőhangzók variabilitásának korpuszalapú vizsgálata

Lévai Dániel

ELTE-BTK NyTUD Elméleti nyelvészet tanszék

ELTE TTK AI csoport

# Struktúra

- Bevezetés, problémafelvetés
- Magyar magánhangzóharmónia
- Variáció
- Hipotézis
- Szintaktikai bizonyítás
- Szemantikai bizonyítás
- Konklúzió

# Bevezetés

- Főnév ~ dolog, melléknév ~ tulajdonság
- Alkalmi főnevesülés:
  - Újabb hazai villanyautó-töltők lesznek **fizetősek**.
  - Kirostálták a gyerekeket, a **fizetősök** maradhattak.
  - Amikor az ingyenes helyett „végre” megkapjuk a **fizetőst**.
  - Az ingyenes filmletöltés lenyomja a **fizetőset**.
- Variáció, van de miért, és meg lehet-e jósolni a környezet alapján?

# Magyar magánhangzóharmónia, röviden

- Kerekségi és előlségi harmónia
- Pörkölt + -hoz/-hez/-höz: kerek (o/ö) és előlképzett (e, ö) (magas) magánhangzó kell → pörkölthöz

	front		back	
	unrounded	rounded	unrounded	rounded
high	i <i>, i: <í>	y <ü>, y: <ű>		u <u>, u: <ú>
mid	e: <é>	ø <ö>, ø: <ő>		o <o>, o: <ó>
low	ε <e>		a: <á>	ɒ <a>

# Variáció, 1

- A főnevek a függőleges nyelvállás szerint magasabban képzett toldalékokat kedvelik: fájl-ok, sör-ök (van sok kivétel...)
- A melléknevek nem: piros-ak, hűvös-ek
- Nyitótő (lowering stem): azok a tövek, amelyek az alacsonyabb nyelvállású toldalékokat preferálják (azaz a legtöbb melléknév és néhány főnév)

# Variáció, 2

- Csak a többesszámot (PL) és tárgyesetet (ACC) vizsgáljuk
- Négyféle variáció fordul elő PL és ACC esetekben:
  - 1) a-o (túlsúlyosVk)
  - 2) a-Ø (állóVk)
  - 3) e-ö (szőrösVk)
  - 4) e-Ø (fekvőVk)

# Hipotézisek, 1

- Létezik tipikus főnévi/melléknévi viselkedés
- Szintaktikailag:
  - A tipikus főnév alany- és tárgyesetben esetben NP-fej pozícióban jelenik meg
  - A tipikus melléknév viszont predikatív pozícióban jelenik meg
- Szemantikailag:
  - A tipikus főnév dolgot jelent, a melléknév tulajdonságot
- A nyitási preferencia viszont ezzel interakcióba lép

# Hipotézisek, szintaxis

- **Predikatív pozíció:**
  - A kínából hazajöttek mind pozitívak voltak.
- **NP-fej pozíció, tbsz. :**
  - A pozitívak újra kivizsgálásra kerültek, a negatívak hazamehettek.
- **NP-fej, akkuzatívisz:**
  - A pozitívot nem hittük el a tesztnek.
- **Tőre vonatkoztatva:**  
 $R_1 > R_2 > R_3$

position	plural	accusative
predicative	low : non-low $R_1$	×
NP-head	low : non-low $R_2$	low : non-low $R_3$



# Hipotézisek, szemantika

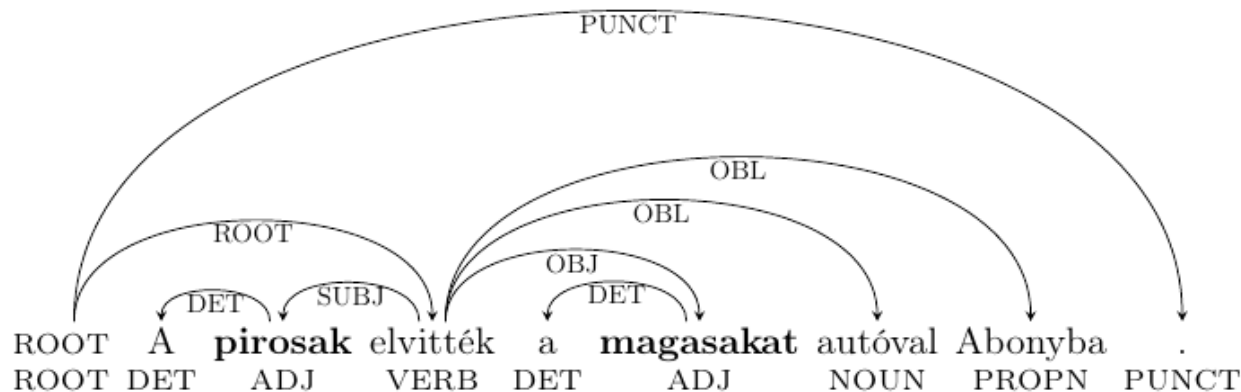
- A főnevek nem nyitnak, a melléknevek nyitnak
- Tehát azt várnánk, hogy a nyitó alakok melléknévszerűek inkább, a nem nyitóak pedig főnévszerűek
- De (szóbeágyazásból) kidisztillálni absztrakt főnévi és melléknévi tulajdonságot nehéz, ezért szomszédsági relációt néztünk
- Egy tőre lebontva azt várjuk, hogy a nyitó alak más nyitóalakokhoz lesz közel, a nem nyitóak pedig nem nyitóakhoz

# Hipotézisek tesztelése

- A korpusz a state-of-the-art, még nem kiadott [Webcorpus 2.0](#)
- Mondatbéli szerepet e-magyarral, azon belül emDeppel tesztelünk
- Szemantikát pedig statikus szóbeágyazásokkal (word2vec)
- Három csoportra bontjuk a töveket: ó (álló), s (fizető), és egyéb tövek (pörkölt).

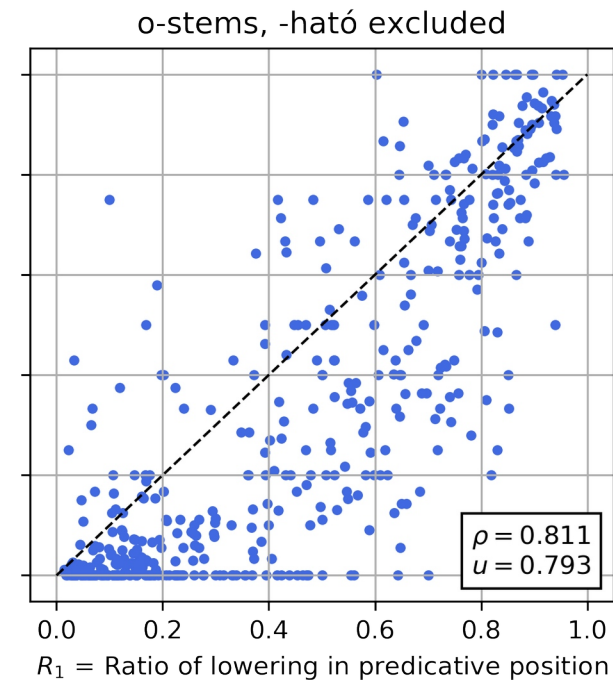
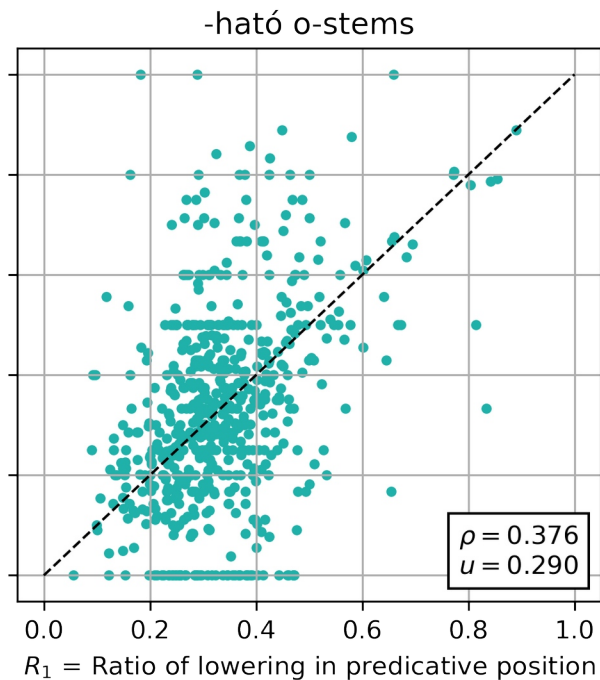
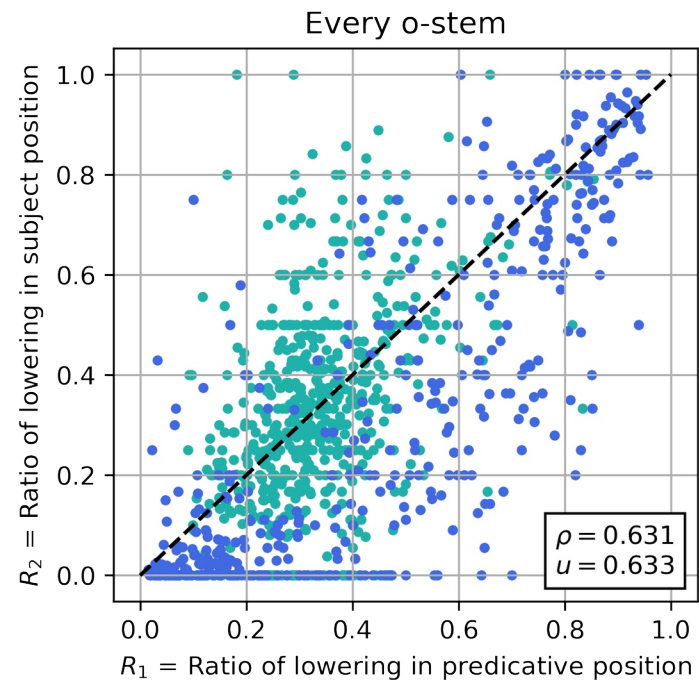
# Szintaxis, 0

- Dependencia-fák ~ iskolában tanult farajzok
- Ennek alapján állapítjuk meg, hogy predikatív vagy NP-fej pozícióban szerepel-e az adott alak
- $O(n^4)$  komp

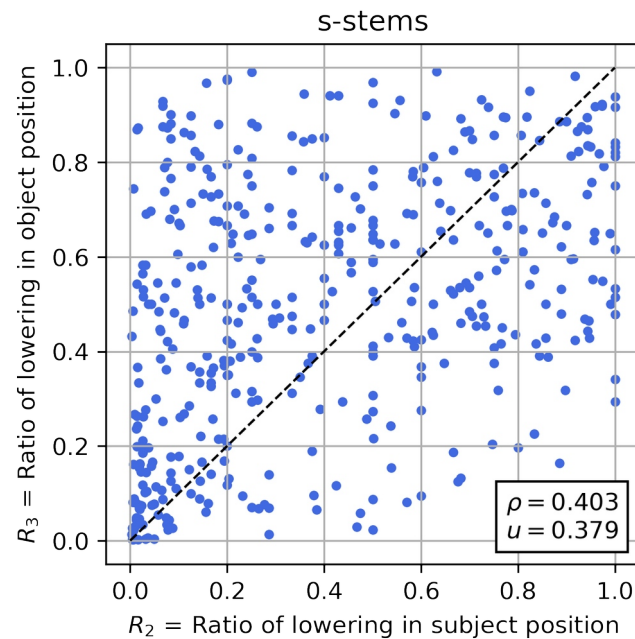
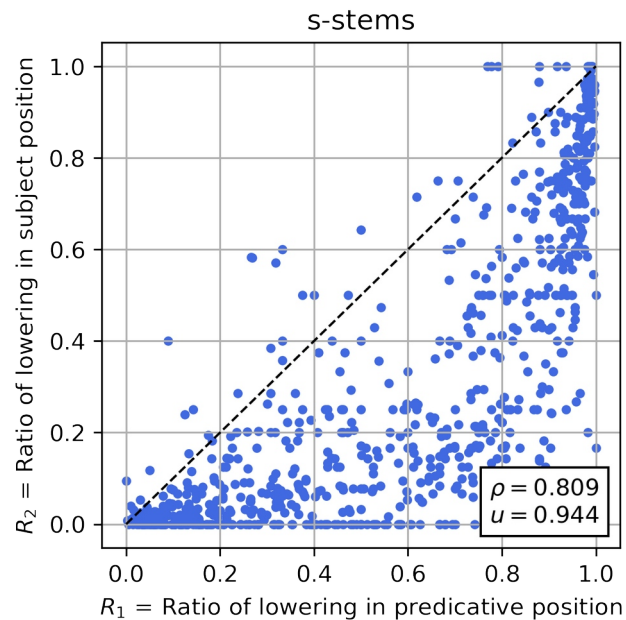


OSSZ

# Szintaxis, 1



# Szintaxis, 2

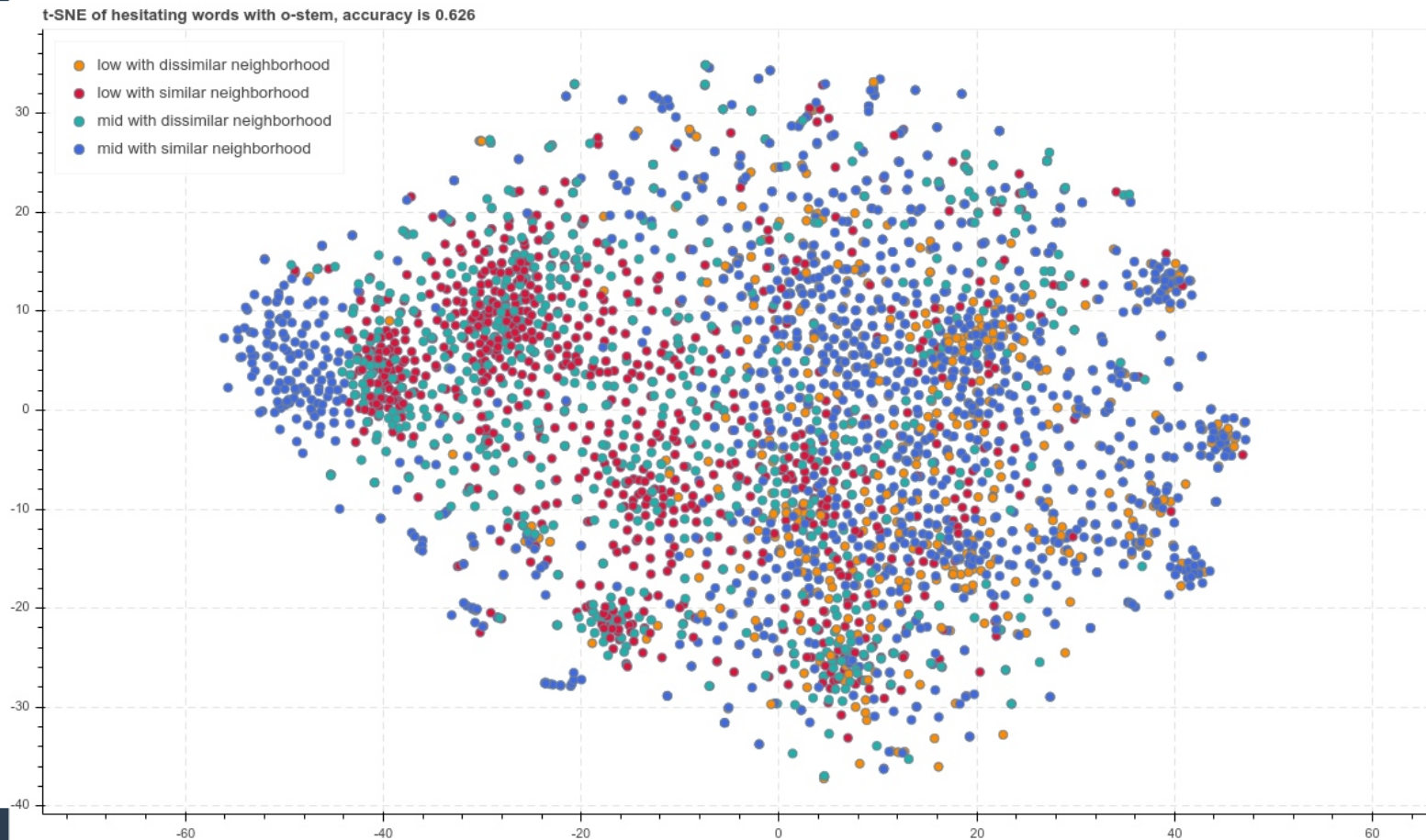


- Az  $R_1 > R_2$  teljesül, az  $R_2 > R_3$  nem.

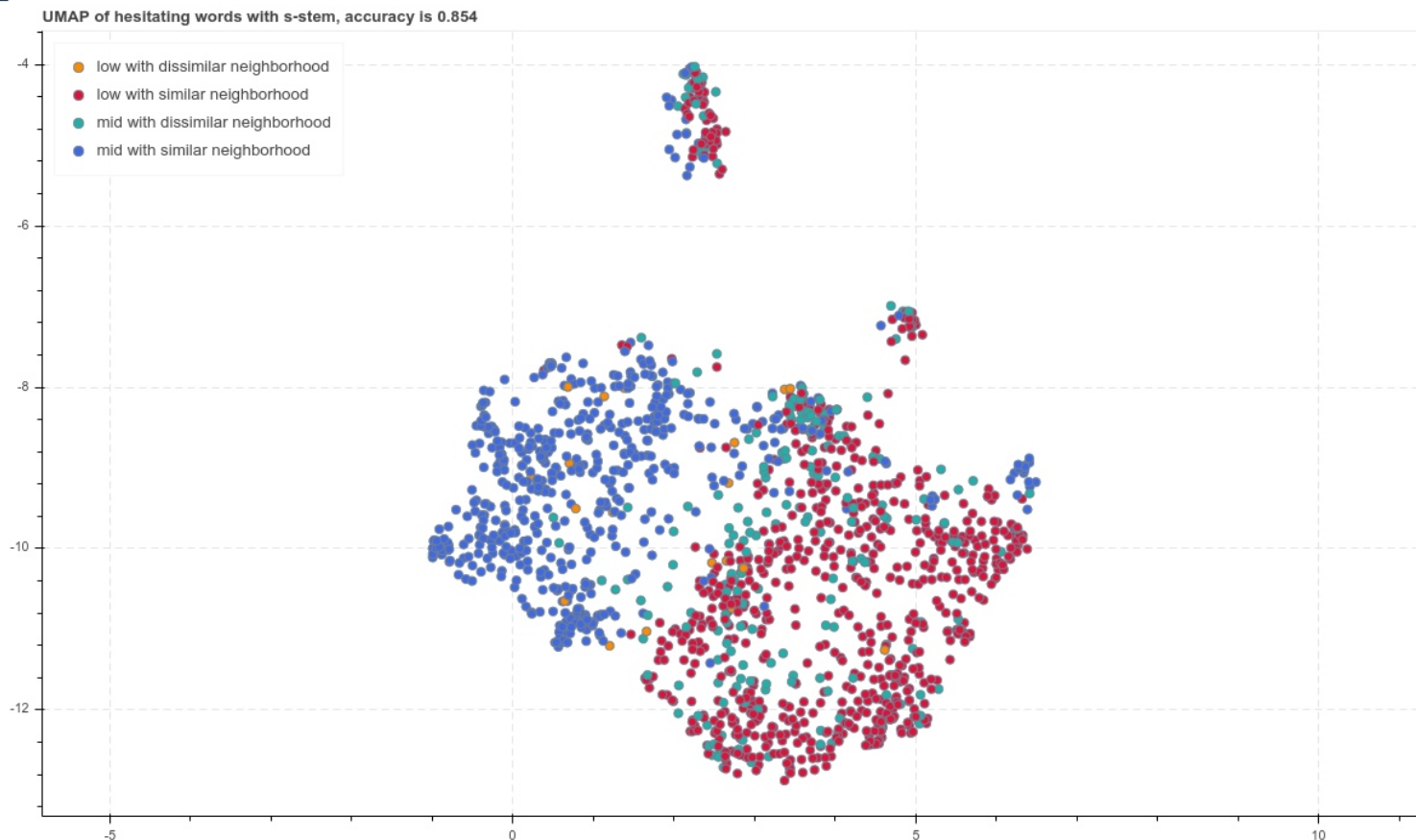
# Szemantika, 0

- Neurális szóbeágyazás
- Szomszédsági mérték: az  $n$  legközelebbi szomszéd közül mekkora az ugyanolyan nyitással rendelkező alakok aránya
- Dimenziócsökkentés (UMAP, t-SNE), hogy interpretálható legyen az eredmény

# Szemantika, 1



# Szemantika, 2





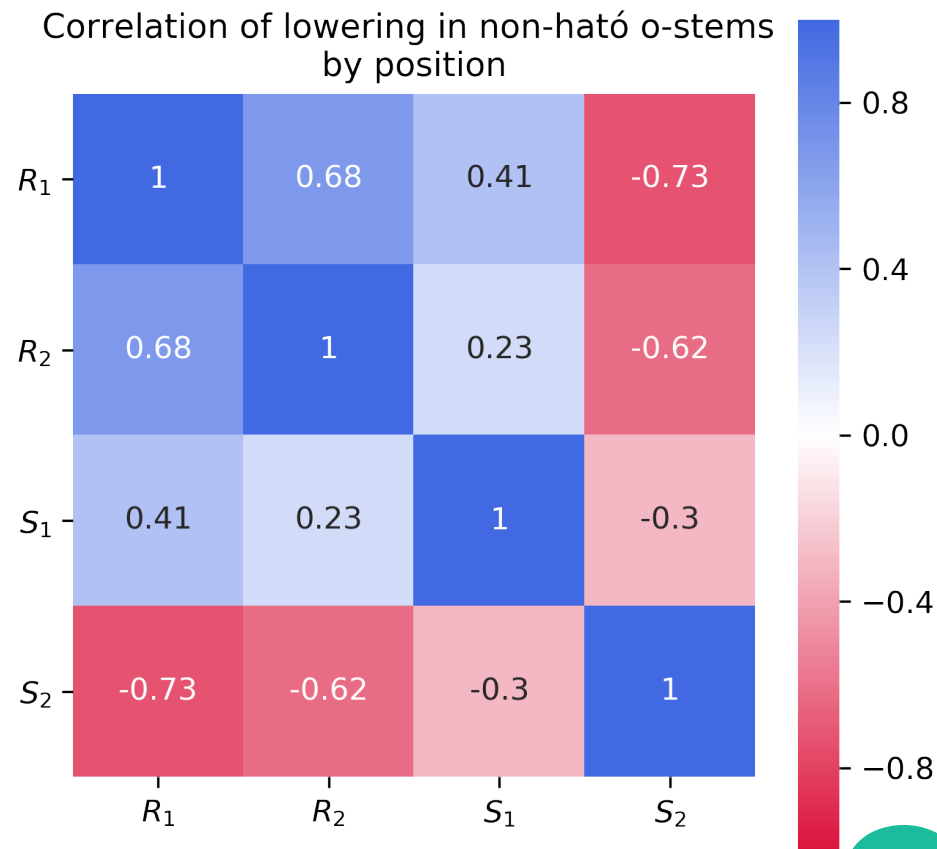
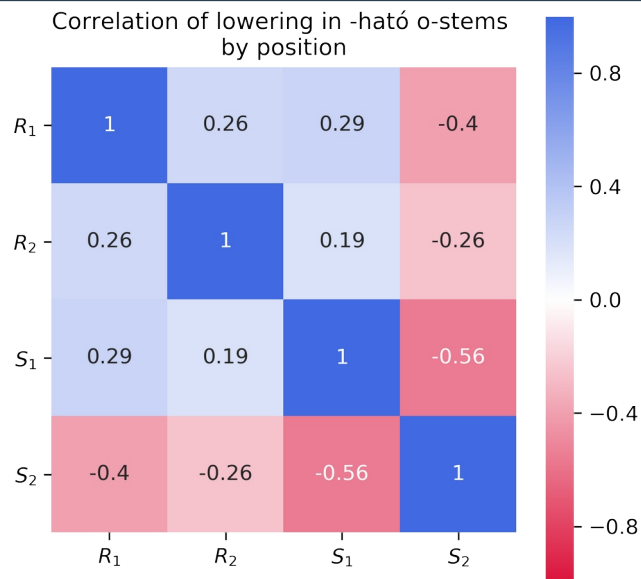
# Szemantika, 3

- Interaktíve megnézhető itt:  
<https://users.renyi.hu/~levaid/public/harmony/>
- Hasonlóak az eredmények, mint a szintaxisban: az ó-sok viszonylag rosszul viselkednek, az s-esek viszonylag jól
- Itt még az akkuzatívsz is szebb

# Hipotézisek összekapcsolása

- Van egy “főnévségi-melléknévségi” skála, ami folytonos, és ahogy haladunk rajta egyik-másik irányba, úgy változik fokozatosan a preferált mondatszerepe és jelentése az alaknak
- Korrelációkat vizsgálunk tövenként aggregálva
- Emlékeztetőül:  $R_i$  – mondatbéli szerepekben a nyitás aránya,  $S_i$  – adott alaknak mennyire egyezik a saját nyitásával a szomszédsága magas dimenzióban

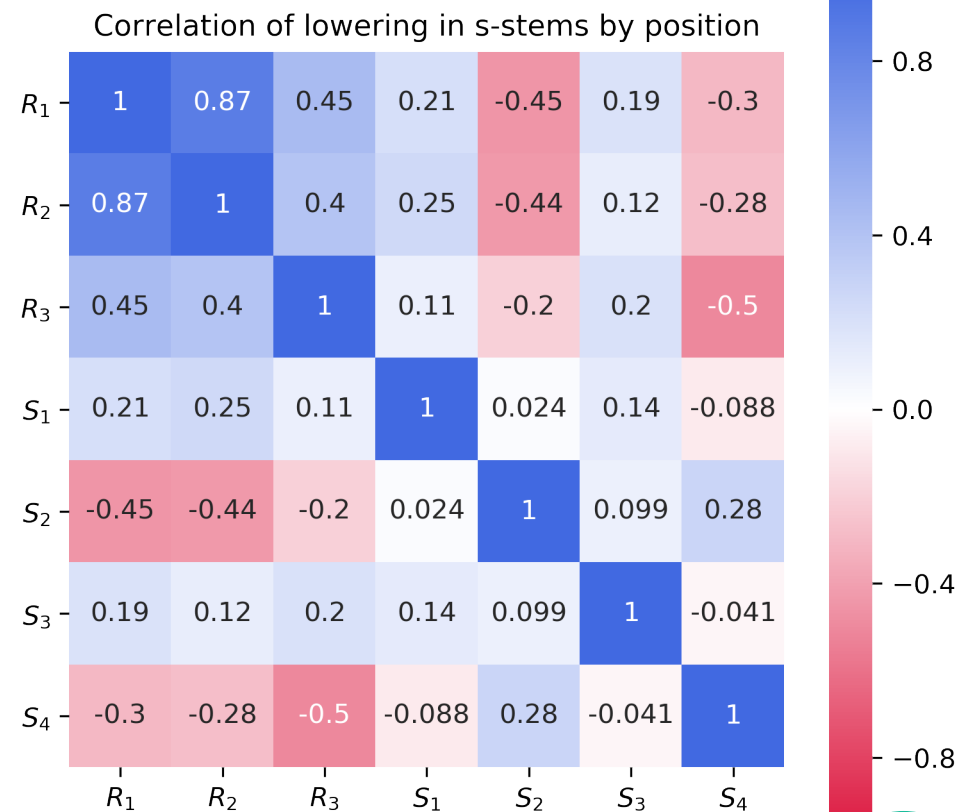
# Hipotézisek összekapcsolása, 2



Pl. predikatív: nemdohányzók (nem nyit) →  
a *nemdohányzóak* saját magához hasonló  
környezetben van

# Hipotézisek összekapcsolása, 3

- Legtöbb információval az R1 rendelkezik
- S-tövéknél alig korreláltak a szemantikai hasonlóságok
- De ó-tövéknél igen
- Van korreláció a szintaktikai és szemantikai jegyek közt



# Konklúzió

- Analogikus alapon közelítettük meg a problémát: ha egy-egy alak felveszi a “tipikus” főnévi és melléknévi toldalékolást, akkor ez a viselkedésén is meglátszik-e
- A magas fokú korrelációk alapján mondhatjuk, hogy igen, a nyitás minősége szemantikai és szintaktikai változást is okoz

**Köszönöm a figyelmet!**