

---

# Ami az informatizált nyelvatlaszokból hiányzik – nyelvjárási hangfelvételek nyelvföldrajzi vizsgálata

Vargha Fruzsina Sára – Vékás Domokos

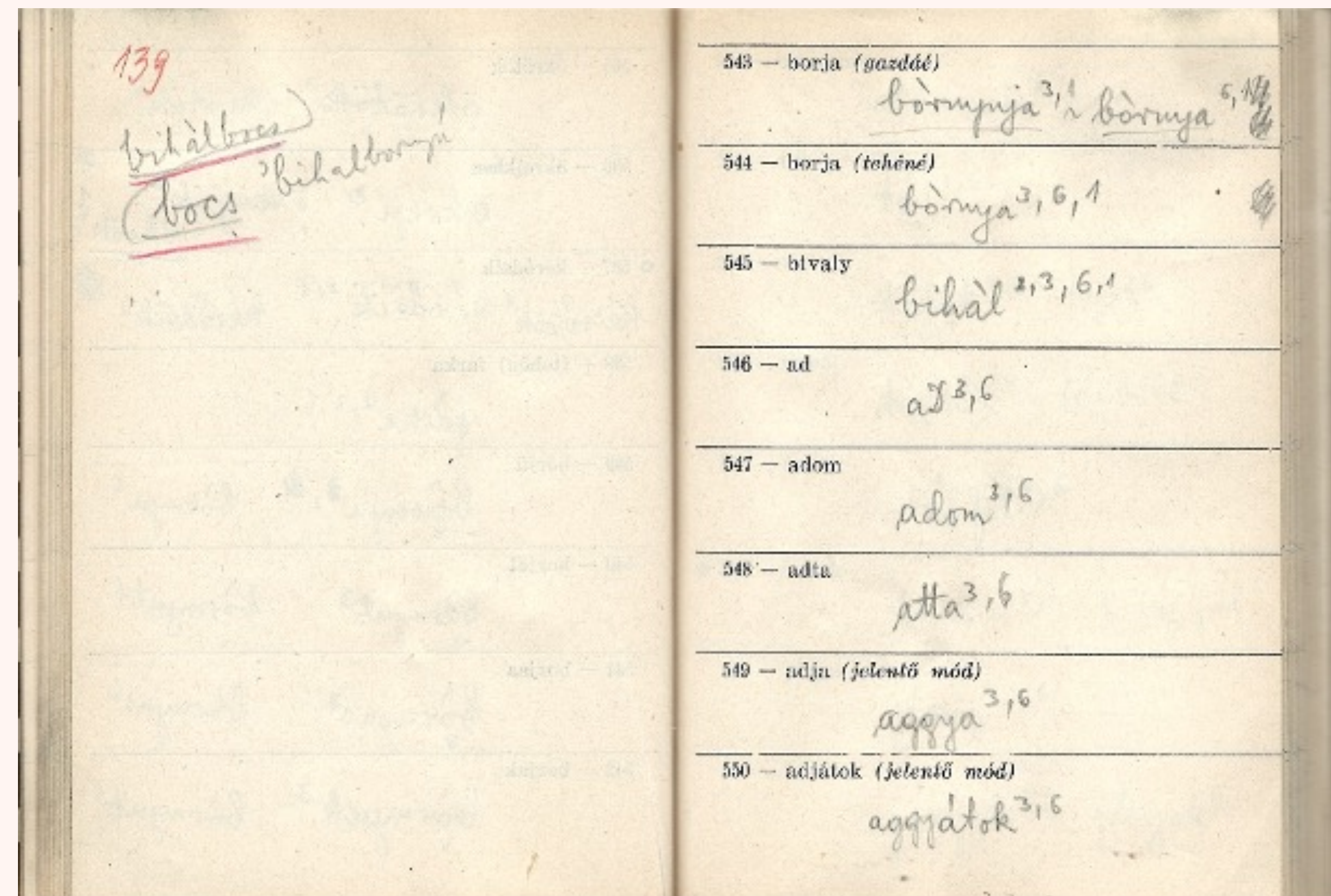
---

# Az előadás menete

- **A (számítógépes) dialektológia helye:**
    - klasszikus, alternatív és standard megközelítések a dialektológiában
    - civil és nem civil erőforrások; eredmények és zsácutcák
  - **Hangzó adattárak mint a klasszikus nyelvatlászok alternatívái**
    - a MNyA. hangfelvételeinek akusztikai vizsgálata
    - Tánczos Vilmos moldvai gyűjtésének feldolgozása
  - **Tanulságok**
-

# Klasszikus dialektológia

- Benkő Loránd Kobátfalván (1957):
  - *bihalbocs*
- Csúcson a dialektológia presztízse
- Készül a nagyatlasz



Kobátfalva, Magyar Autonóm Tartomány,  
1957. június 19-21.

---

# Alternatív kísérlet

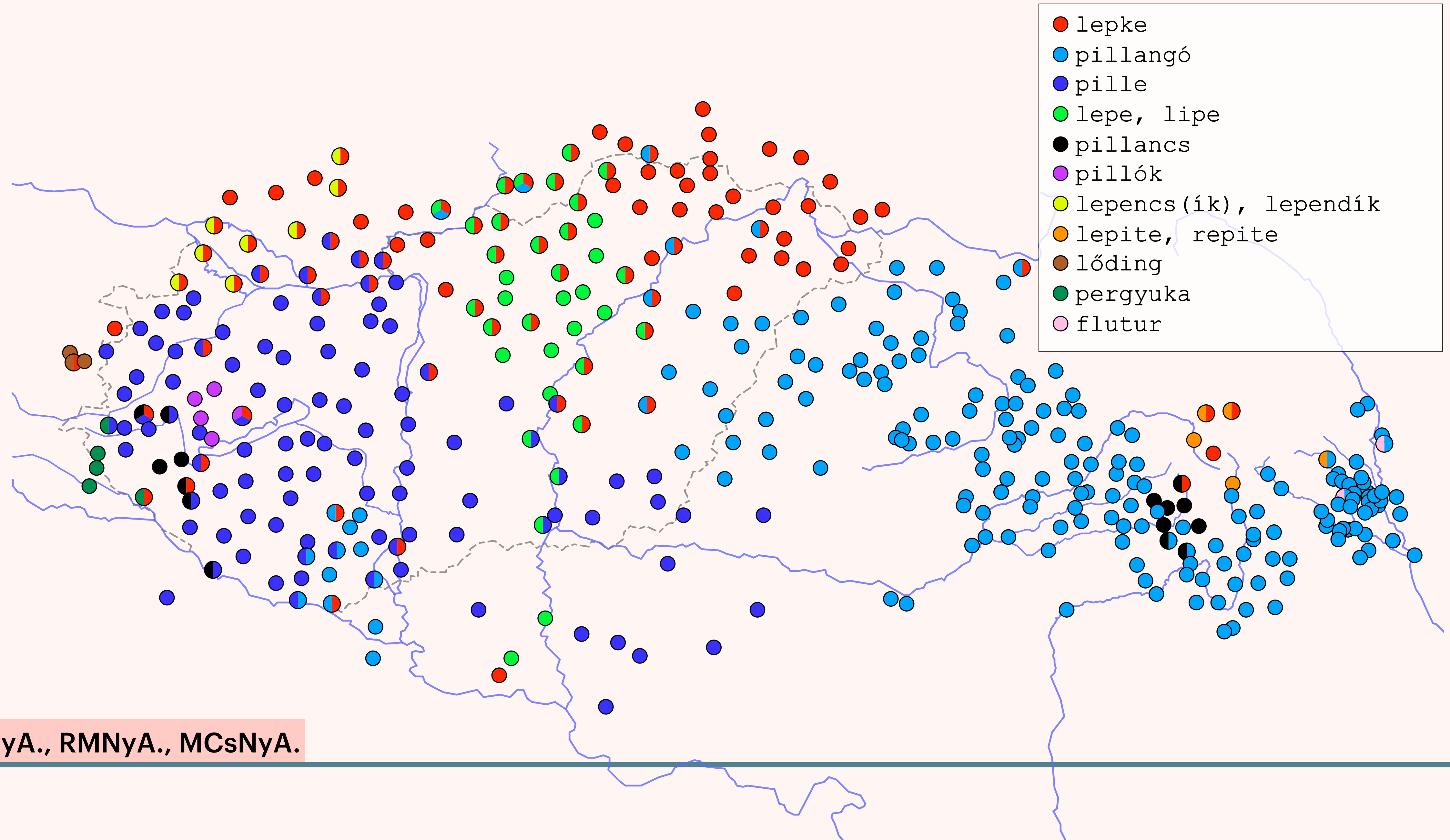
- **Balogh Lajos a nyelvatlaszról: holt anyag (alternatívát kínál)**
  - **Balogh Lajos vs. Benkő Loránd: informatizálás vs. RMNyA papíron:**
    - **többszörös zsácutca (az erőforrások hasznosulása nem megfelelő)**
    - **Tudománytörténet: meddig tart Benkő Loránd előrevivő ereje? (1992)**
-

---

# Alternatív megoldás: Bihalbocs (1996–)

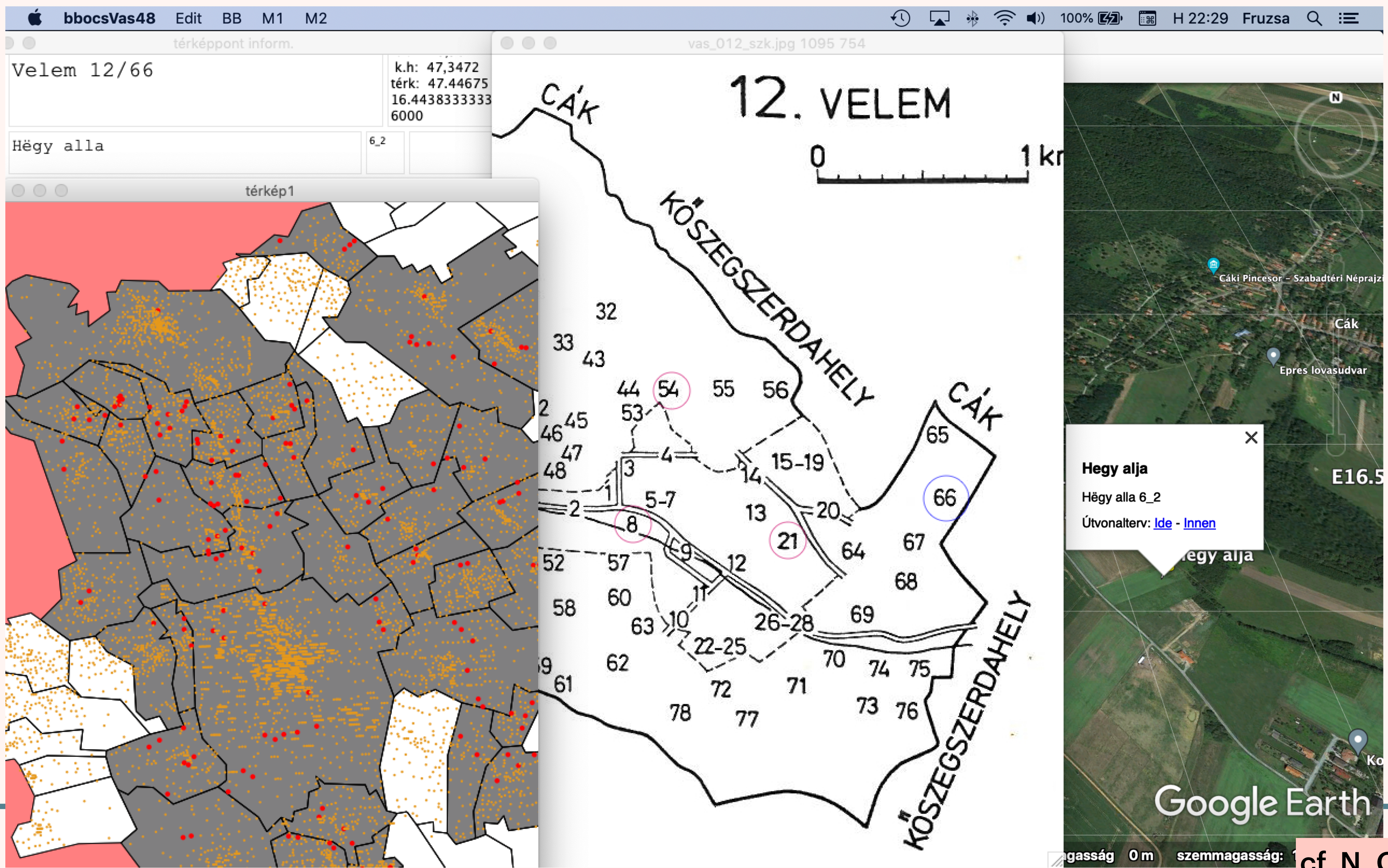
- nem közpénzből! Civil erőforrásokkal, fonetikai alapokon
  - célja: sokoldalúan felhasználható dialektológiai adatok létrehozása:
    - papíratlaszok és hanganyagok használhatóvá tétele (potenciális adat -> informatizált adat)
    - új magyar nyelvjárási atlaszgyűjtések támogatása (hangzó adatok is)
  - erősségeiből:
    - az adatelőállítás hatékonysága (erőforrások megfelelő hasznosulása)
    - az adatok konvertálhatósága, hordozhatósága (Praat, R, stb.)
    - a fejlesztés fenntarthatósága (biztos civil erőforrásokkal, bizonytalan pályázati erőforrásoktól függetlenül)
-

# Bihalbocs, 90-es évek: atlaszintegrálás, Kárpát-medencei (határtalanító) szemlélet



MNyA., RMNyA., MCsNyA.

# Bihalbocs és névtan: mikrotoponímák a Kárpát-medencében



---

# Standard megoldások

- **alternatív, innovatív kísérletként indulnak**
  - **elfogadottá, megkerülhetetlenné válnak:**
    - **nem lehet Bihalbocs nélkül atlaszadatokat informatizálni, amíg egy alternatív megoldást nem hozunk létre**
  - **atlaszinformatizálás 1996–2003. között nagyrészt lezárult, de ma is tart**
  - **Alternatív megoldás ma a Bihalbocs?**
    - **alternatív, mivel a folyamatos innováció és megújulás jellemzi**
-

---

# Civil = erőforrás

- **Ki a civil a (számítógépes) dialektológia szempontjából?**
  - **az, akinek erőforrásai vannak (pénz, idő, tudás, ötlet, helyismeret, motiváció), amelyeket egy klasszikus megközelítést alkalmazó tudományos irányzat nem tud (vagy nem is akar) elérni, hasznosítani (esetleg nem is ismeri fel ezeket az erőforrásokat).**
  - **Lendíthet-e a magyar dialektológiai kutatásokon - civilként - egy nem dialektológus nyelvészprofesszor?**
-

---

# Újabb alternatíva? MI

- Kiss Jenő Szombathelyen (2015) : „Tempora mutantur, [...] a dialektológia már eltűnt a Nyelvtudományi Intézetből.” „S vajon napjainkban hány aktív magyar dialektológus tevékenykedik Erdélyben?”
  - Oszkó, Prószéky, Vargha (2021): Nyelvjáráskutatás, számítógépes módszerek, mesterséges intelligencia
  - a legmodernebb nyelvi technológiák a nyelvjárási vizsgálatok szolgálatában
  - Ki a civil?
-

---

# A dialektológia továbbra is a nyelvészet perspektivikus területe

- **Miért fordul az elméleti nyelvészek figyelme is a dialektológia felé?**
    - a nyelvjárások alkalmasabbak az emberi nyelvi képesség alapelveinek feltárására, mint a kodifikált (tehát kevésbé természetes, normatív manipulációnak jobban kitett) standard nyelvváltozatok (cf. De Vogelaer–Seiler 2012)
    - erősödik a komparatív szemlélet, a variáció és a mikrovariáció vizsgálata, illetve a korpusz alapú megközelítés (adatok és algoritmusok)
  - **Közös projektek? (dialektológia, nyelvtechnológia, elméleti irányok)**
    - „A dialektológusoknak [...] aktívabbnak [kellene] lenniük a dialektológia [...] általános nyelvészeti vonatkozásainak láttatásában, [...] a tudományköziség kínáلتa lehetőségek és a dialektológiai szoftverfejlesztés, a nyelvtechnológia vívmányainak jobb kiaknázásában.” (Kiss 2015)
-

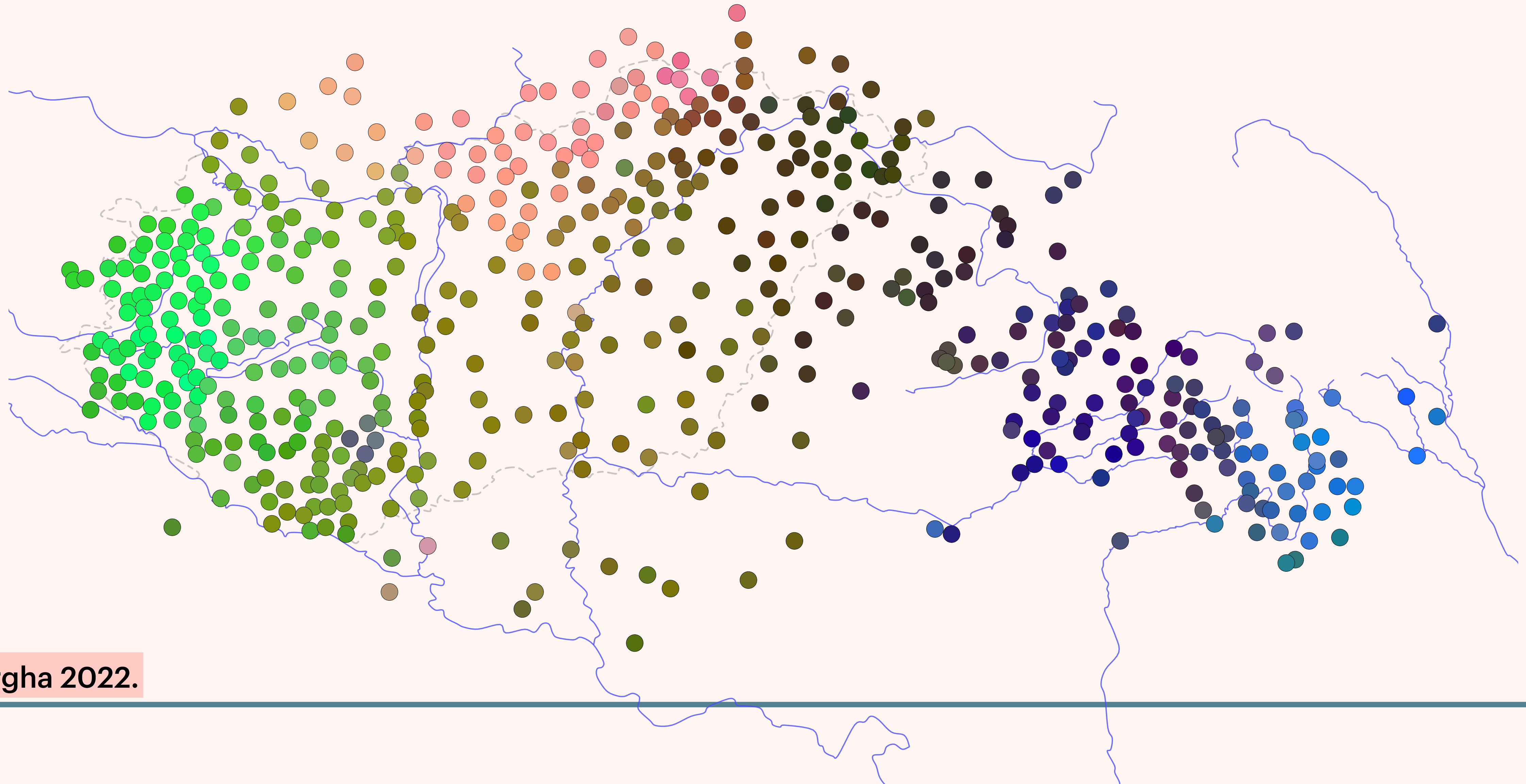
---

# Példák a nyelvjárási atlaszok (nem hangzó) adatainak innovatív vizsgálatára

---

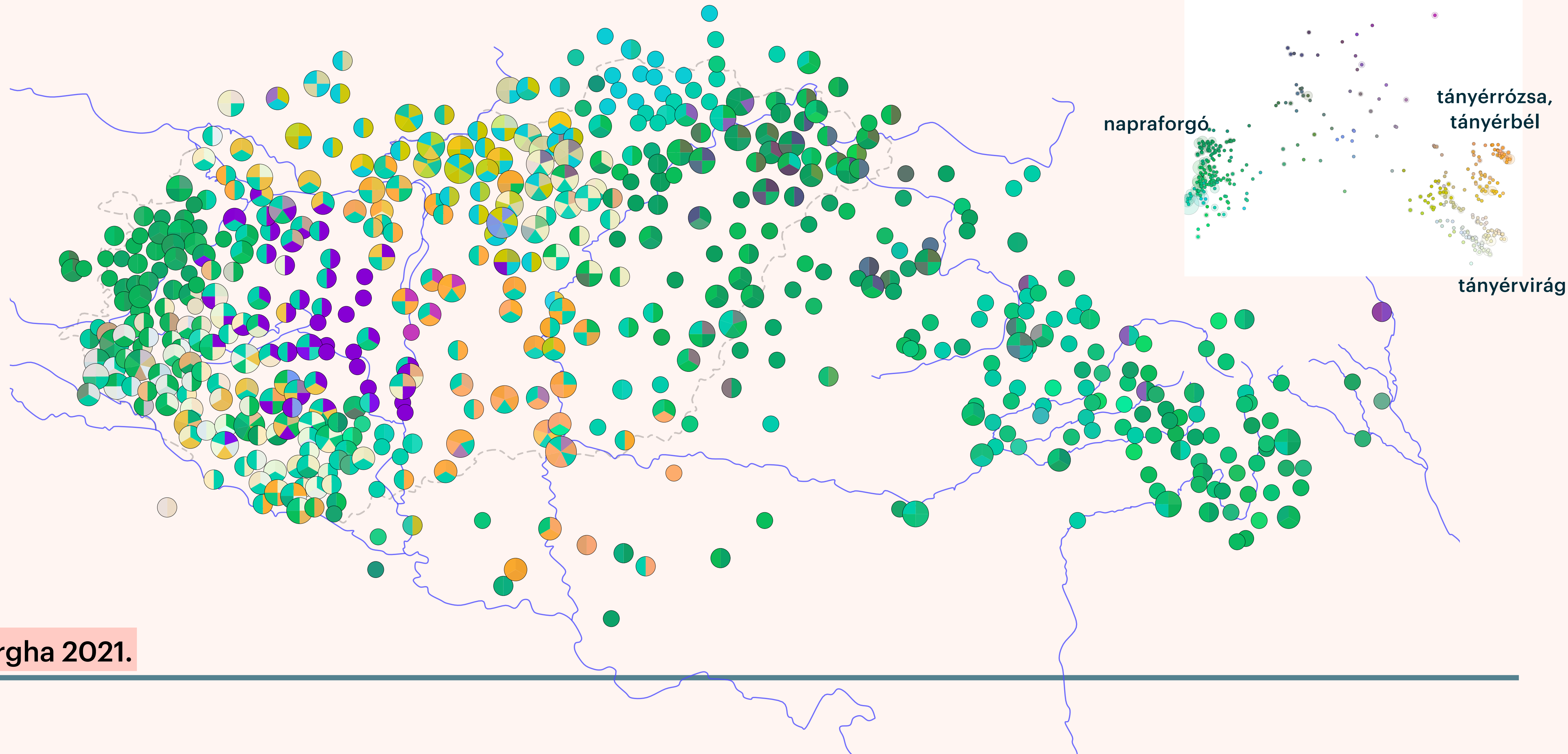
# A vokalizmus dialektometriai elemzése

Többdimenziós skálázás, 483 integrált térképlap adatai alapján a MNyA. és az RMNyA.  
Bihalbocsos adatbázisából



# Automatikus csoportosítás többdimenziós skálázással

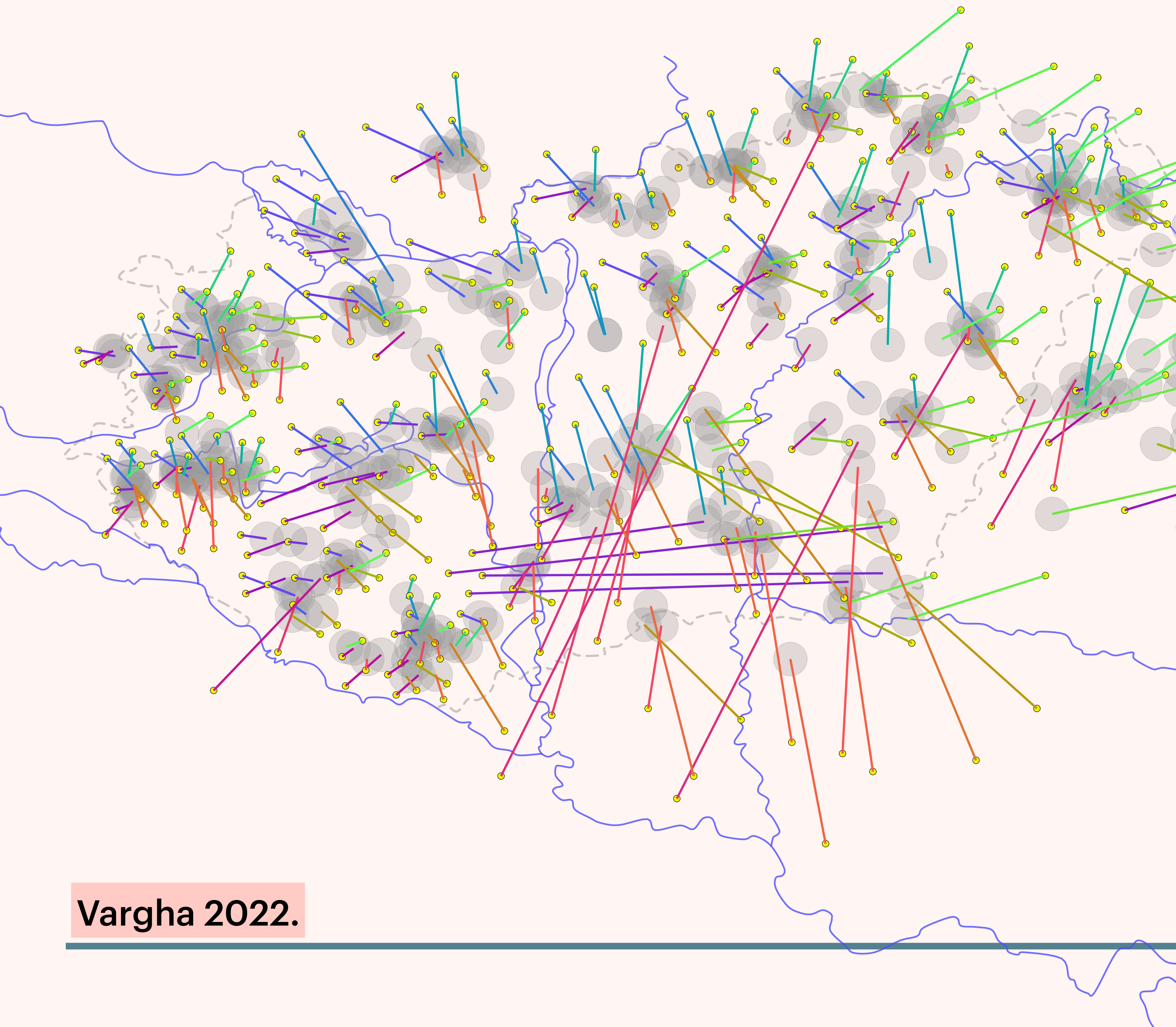
MNyA. 26., RMNyA. 293.



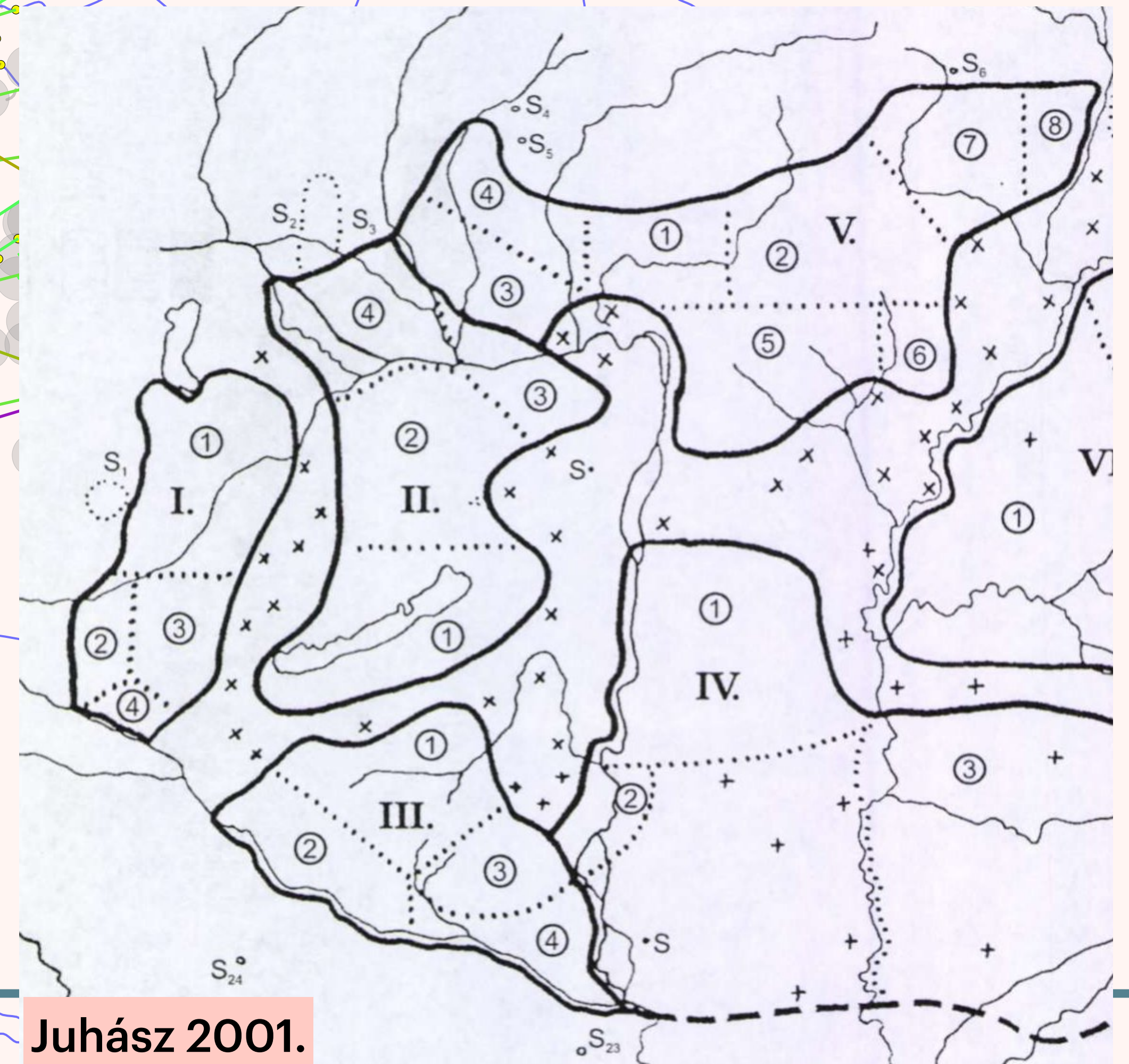
Vargha 2021.

# Kutatópontok csoportosítása földrajzi információ felhasználásával

- nyelvi hasonlóság földrajzi súlypontja
- kutatópont helye



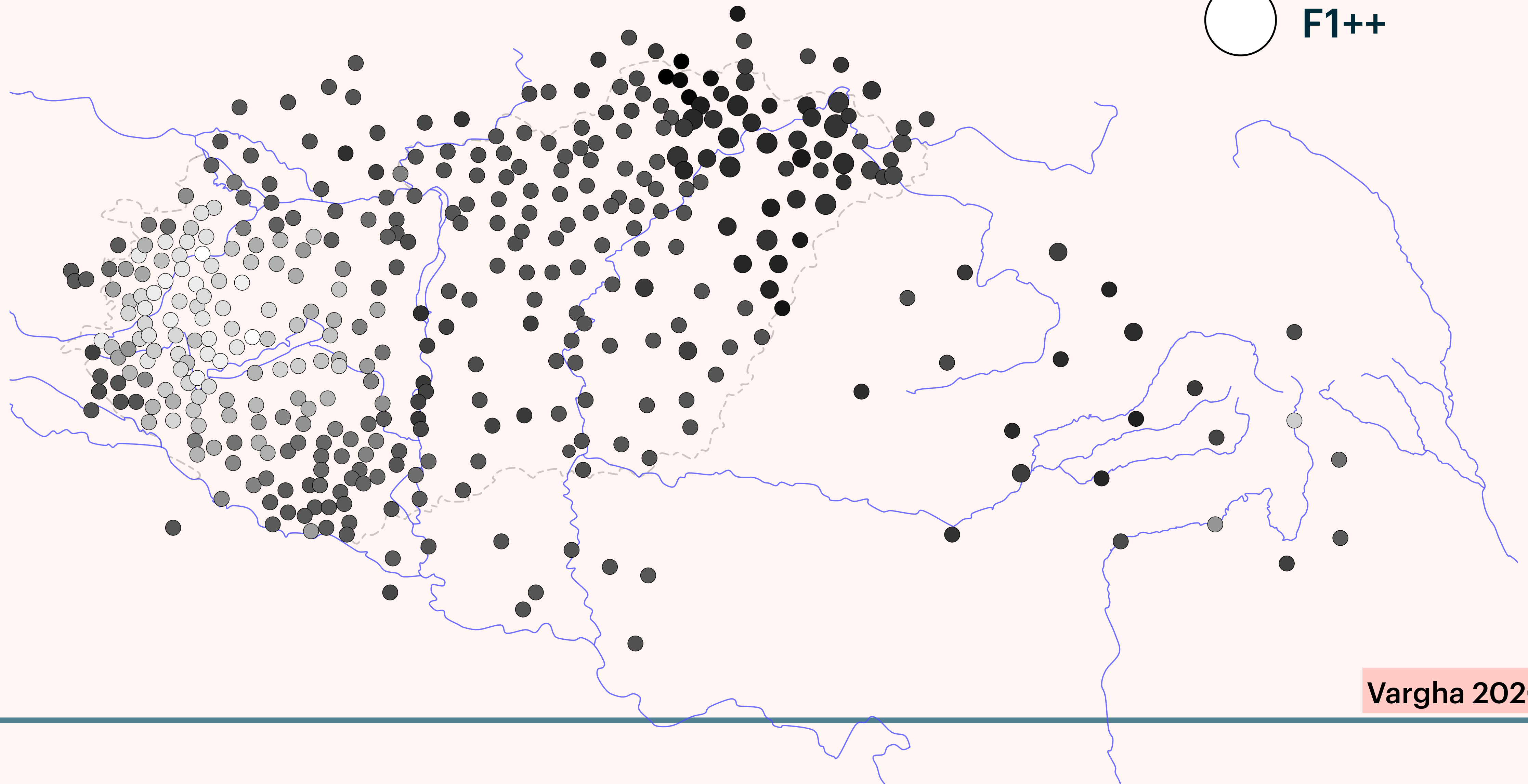
Vargha 2022.



Juhász 2001.

# Magánhangzó-minőségek térképezése pseudo-formánsértékek kiszámításával

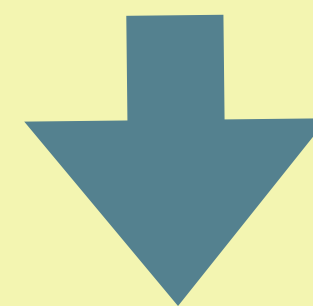
○ F1++



---

# Az atlaszadatok felhasználásának korlátai

- ahol az adatok nem tükrözik a variabilitást: zárt *ë* minősége
- ahol különböző jelek hasonló minőségeket adnak vissza: *á* vs. *ä*
- ahol az adatok eredeti, hangzó változata már nem rekonstruálható, vagyis a lejegyzések pontossága, objektivitása nem ellenőrizhető



**Hangzó adatok, szövegtörzsek jelentősége megnő**

---

---

# Hangzó adatok, szövegkorpuszok

---

---

# Magyar nyelvjárások akusztikai és dialektometriai vizsgálata

(NKFIH FK-138396, Kocsis Zsuzsanna, Szabó Gergely, Vargha Fruzsina Sára)

- **Kutatási cél: A magánhangzók térbeli variabilitásának vizsgálata akusztikai fonetikai módszerekkel**
    - a magánhangzók minőségének minél pontosabb meghatározása
    - a lejegyzésben nem tükröződő ejtésbeli hasonlóságok és különbségek feltárása
    - a magánhangzók közti kölcsönhatások feltárása és elemzése
  - **kutatási módszerek**
    - MNyA. ellenőrző gyűjtésének hangfelvételei, MNyHk. I–XVI.
    - hanggal szinkronizált lejegyzések készítése, ahonnan még nincs (Bihalbocs)
    - F1 és F2 mérése, a mért értékek rögzítése a szövegfájlban (Bihalbocs, Praat)
    - az értékek normalizálása, vizualizálása, összevetése (Bihalbocs)
-

# F1 és F2 mérése

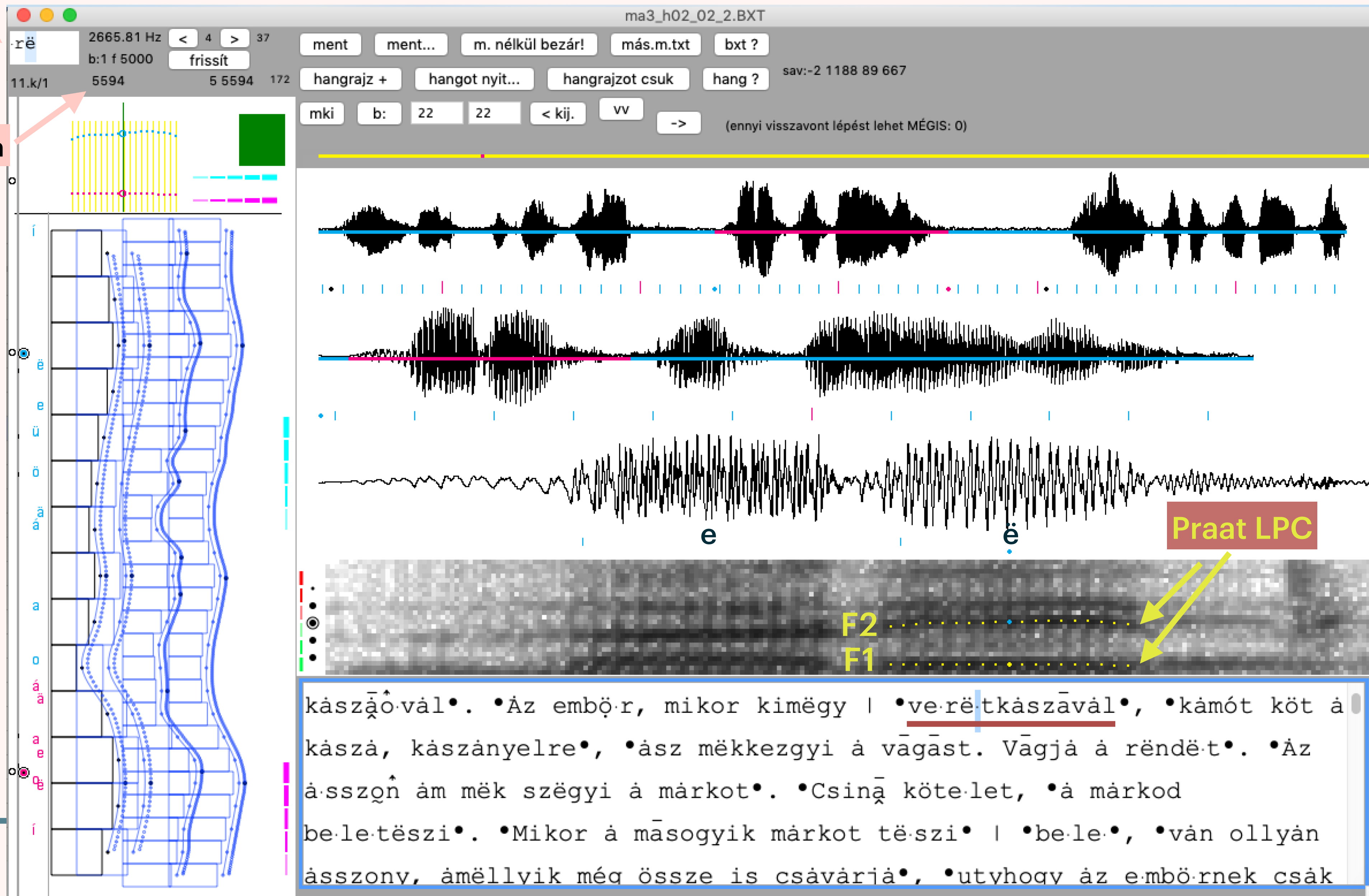
Bihalbocsban, a Praat Burg-algoritmus használatával

mért magánh.

formánsplafon

F2

F1

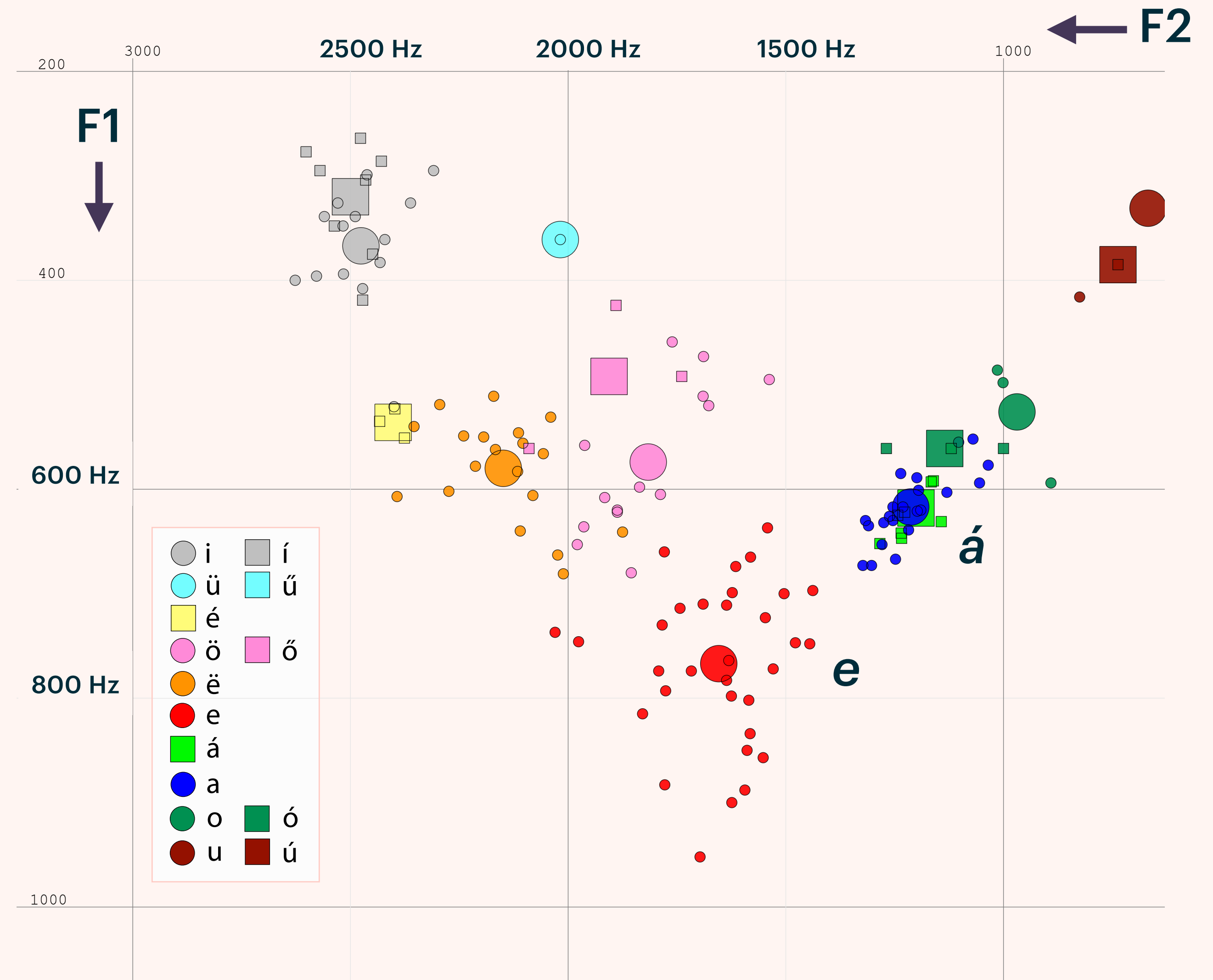
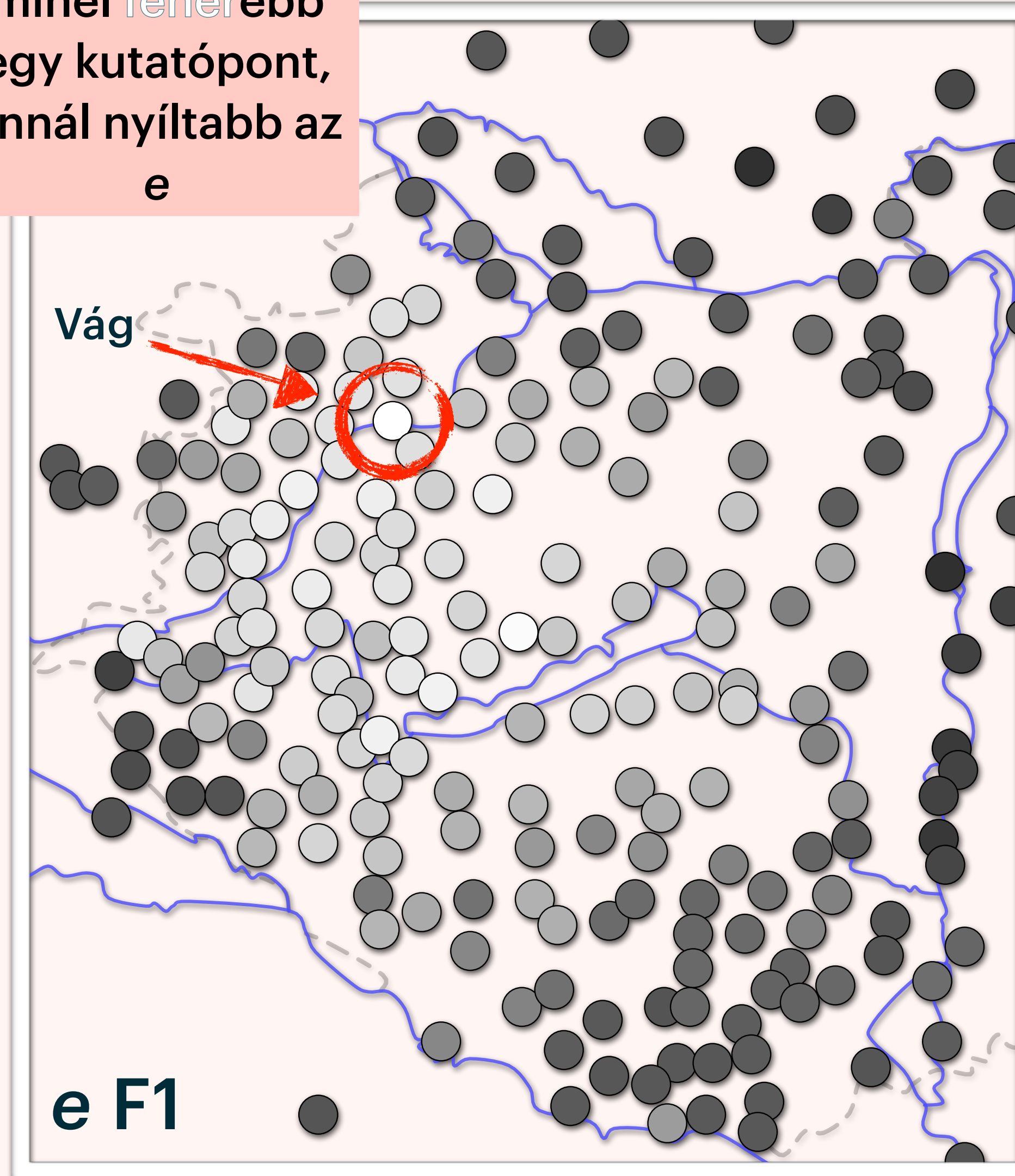


- a hangfãjl egy bizonyos idõpillanatãra mutató marker

- a mært magãnhangzõ után betett marker

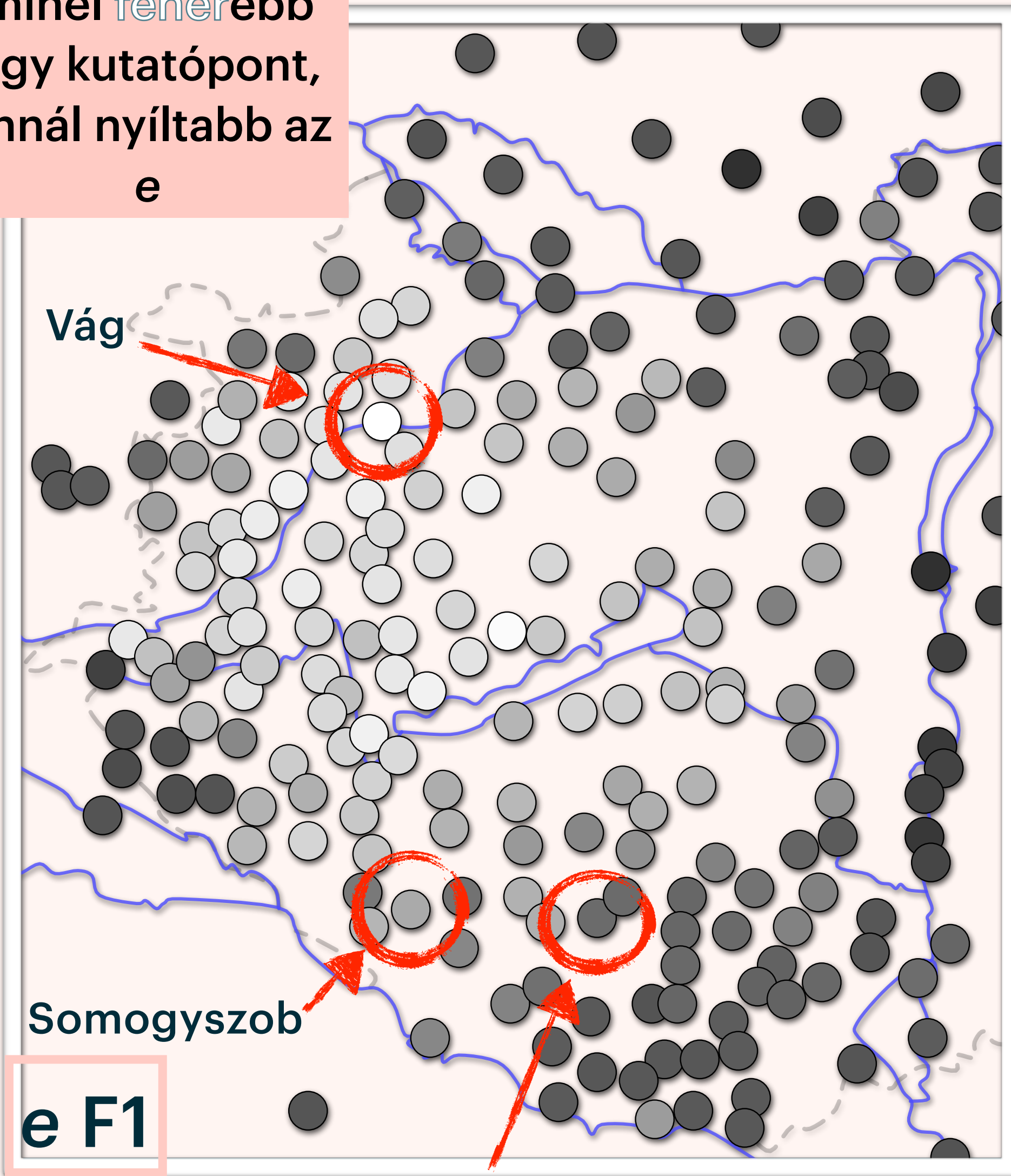
# Vág kutatóponton mért F1 és F2 értékek

minél **fehérebb**  
egy kutatópont,  
annál nyíltabb az  
**e**

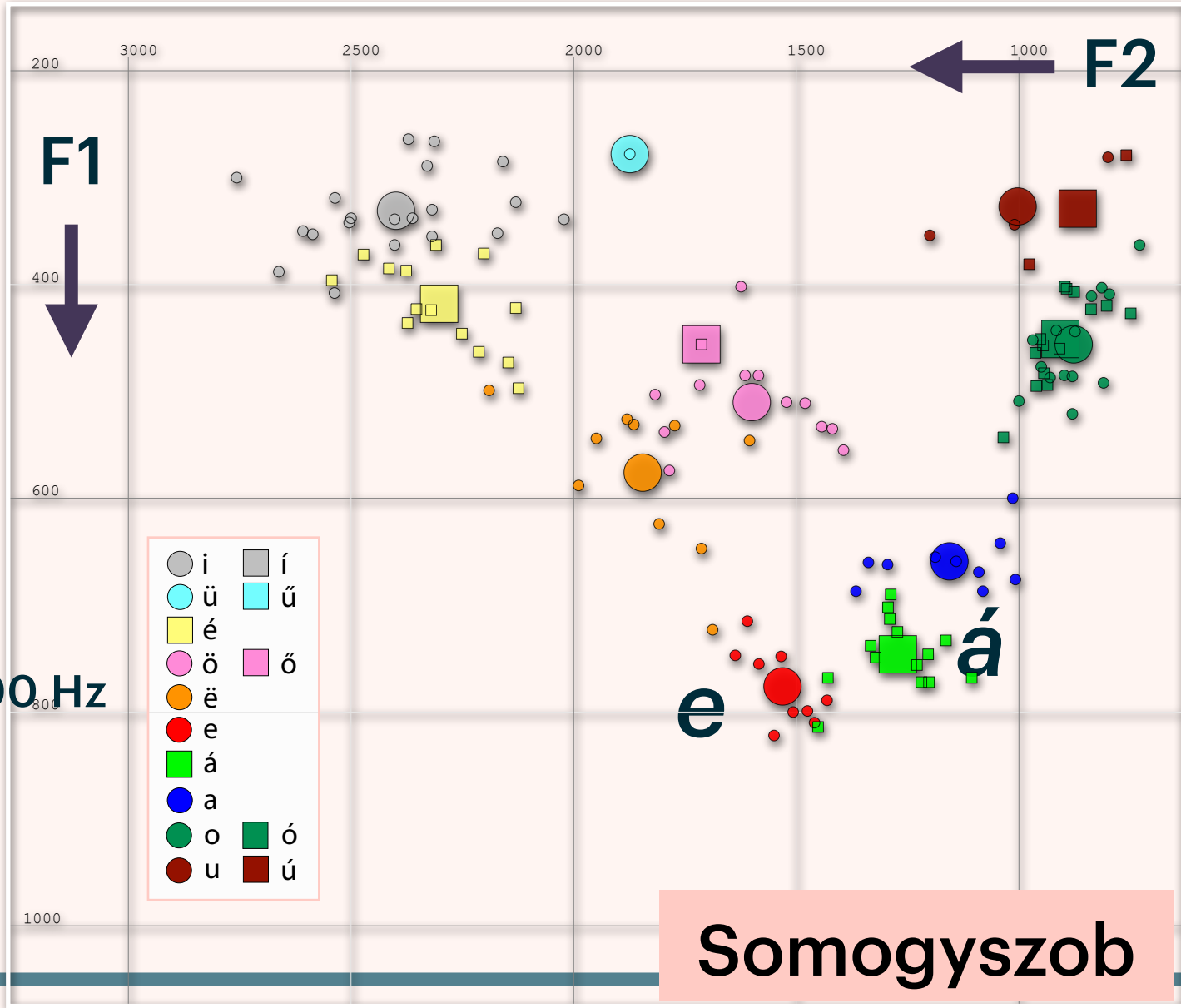
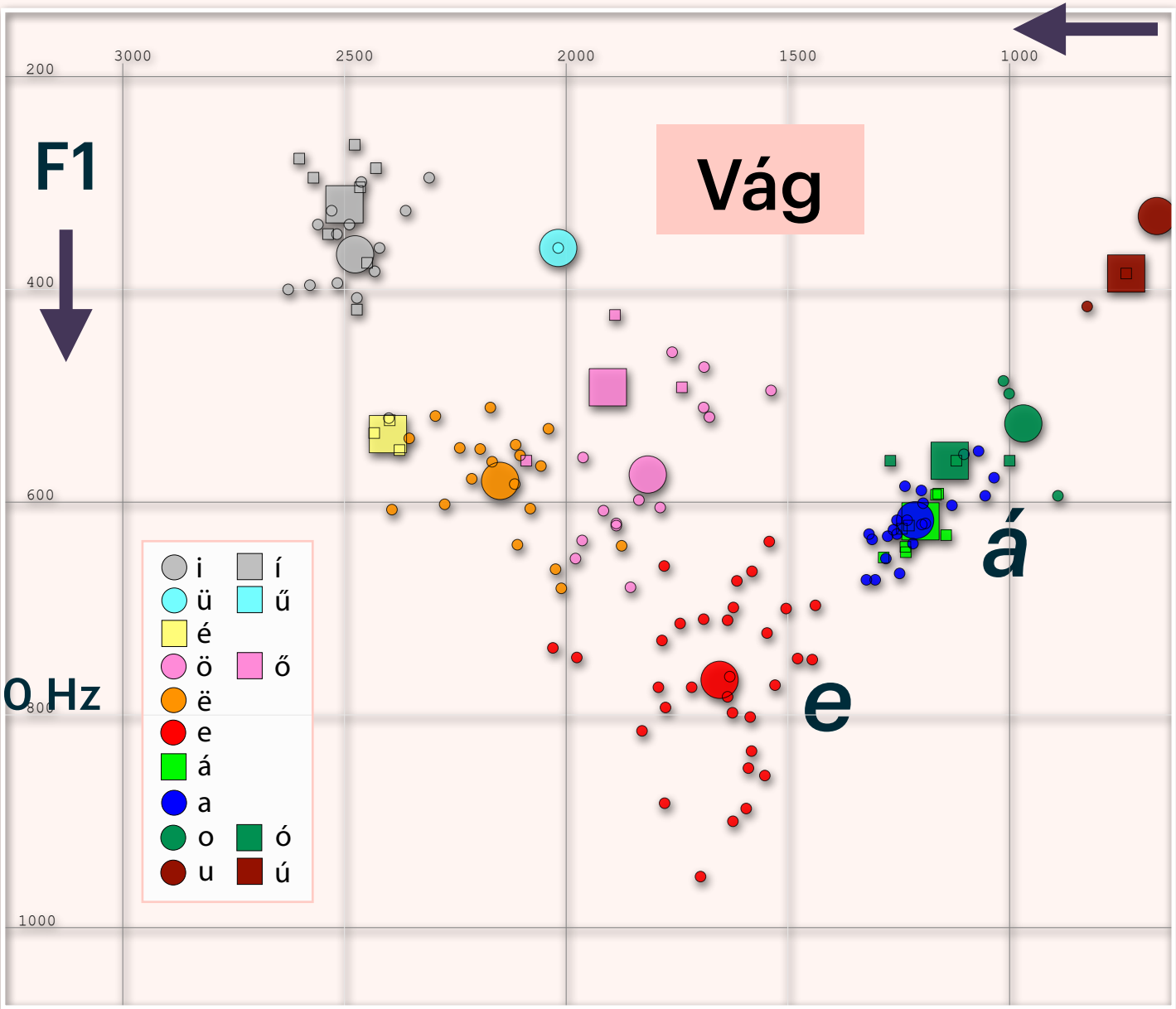


# Az e nyíltsága három dunántúli kutatóponton

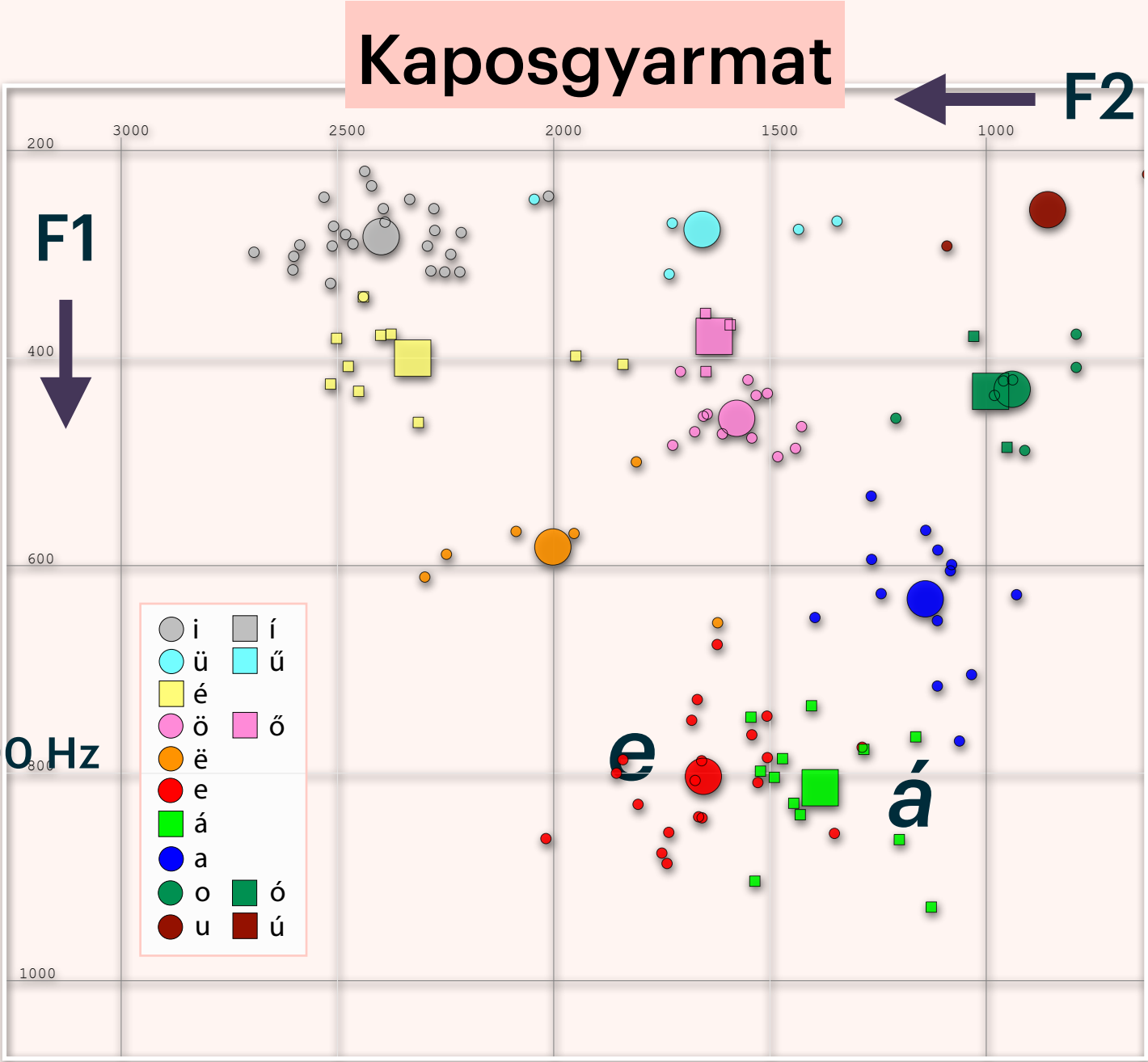
minél **fehérebb**  
egy kutatópont,  
annál nyíltabb az  
**e**



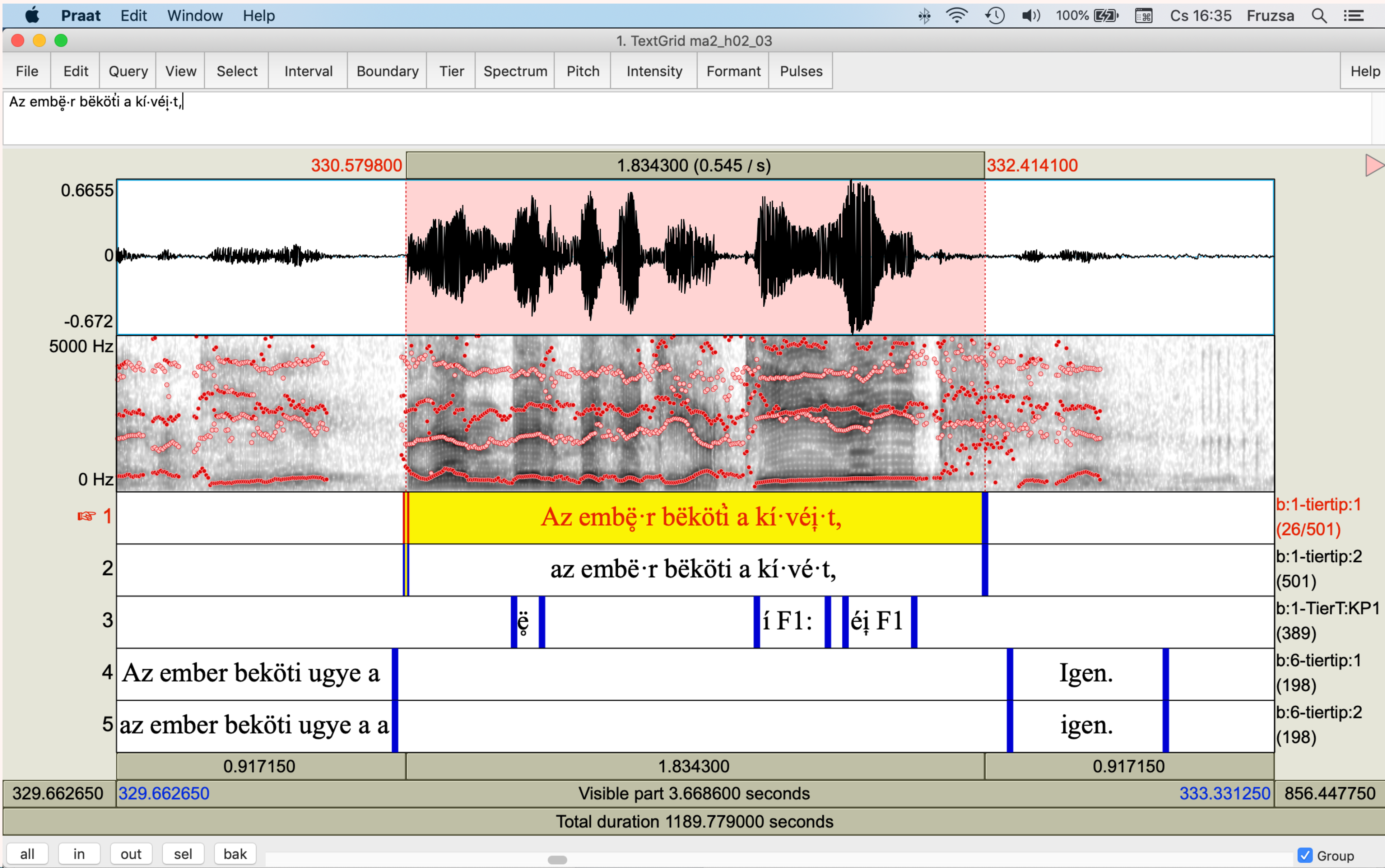
Kaposgyarmat



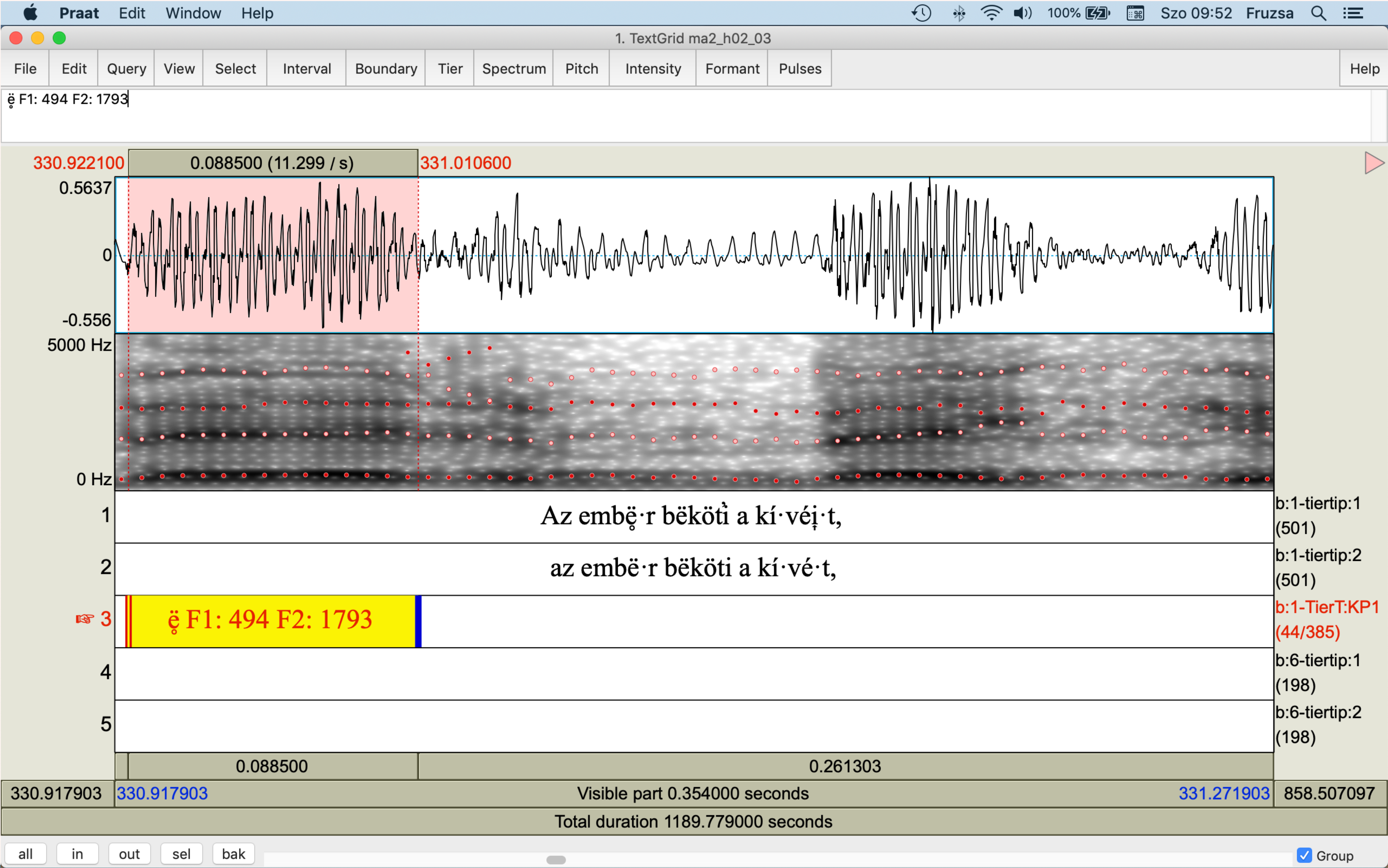
Somogyszob



# A pataki interjú Praat TextGrid változata



# A pataki interjú Praat TextGrid változata



---

# Tánczos Vilmos moldvai gyűjtésének feldolgozása

Veszélyeztetett uráli nyelvek szintaxis-elméleti és magyar nyelvtörténeti tanulságai (vezető kutató: É. Kiss Katalain, Élvonal 129921)

A moldvai magyar nyelv: adatbázisépítés és grammatikai kutatások (vezető kutató: Surányi Balázs, ELKH SA-53/2021)

- **Cél: nyelvészeti elemzésre alkalmas, hanggal összekapcsolt lejegyzésekből korpusz építése**
- **A felvételek jellegzetességei: fókuszban az archaikus népi imádságok**
- **Az adattárépítés folyamata:**
  - hanggal összekapcsolt lejegyzések a magyar egyezményes hangjelölés szerint, Bihalboccsal (Balázs Renáta, Eris Elvira Mária, Vargha Fruzsina Sára)
  - különböző változatok létrehozása a finom fonetikai lejegyzésből, a lejegyzések online közzététele (Vargha Fruzsina Sára)
  - normalizálás és adatbázis kialakítása (Huszár Anna Laura, Vadász Noémi)
  - morfológiai elemzés (Kalivoda Ágnes)
  - morfológiailag annotált, online is elérhető, kereshető változat kialakítása (Sass Bálint)

---

MoMa: egy beszéltnyelvi, nyelvjárási korpusz építésének lépései és érdekességei. Műhelyfoglalkozás a 22. Élőnyelvi Konferencián 2023 szeptemberében (szervező: Vadász Noémi)

# Tánczos Vilmos hangfelvételeinek lejegyzése

Rotár Mária, Szítás, 1998.

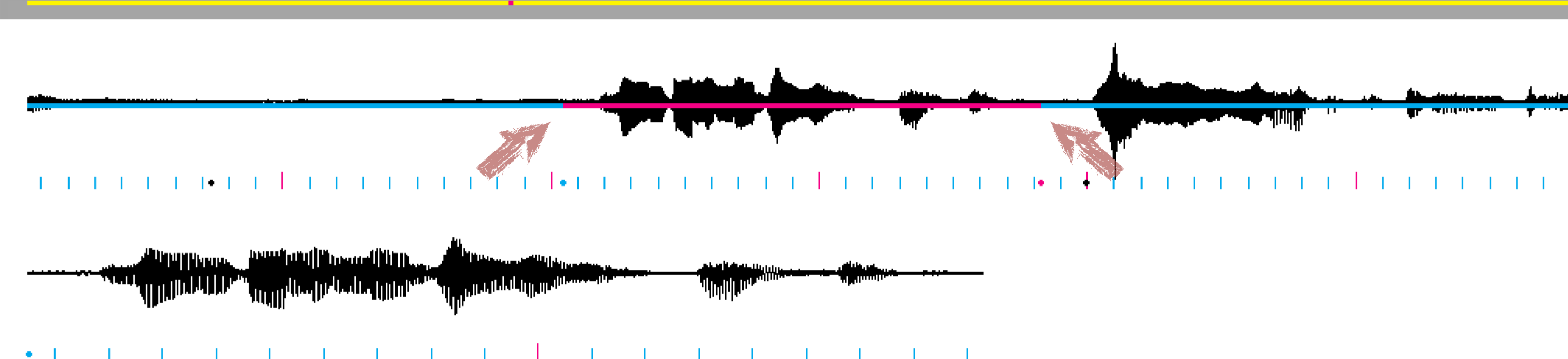
a felvétel a Gyöngyökkel gyökereztél c. dokumentumfilm nyersanyagából való, rendező: Mohi Sándor

m4\_53\_02\_elv.BXT

ment ment... m. nélkül bezár! más.m.txt bxt ? sav:-2 410-9 765

hangrajz + hangot nyit... hangrajzot csuk hang ?

mki b: 22 22 < kij. vv -> (ennyi visszavont lépést lehet MÉGIS: 0)



{-1 •<Magyarul> imádkozik Marika néni?•}

{3 •Magyarul.•}

{-1 •Miért imádkozik magyarul?•}

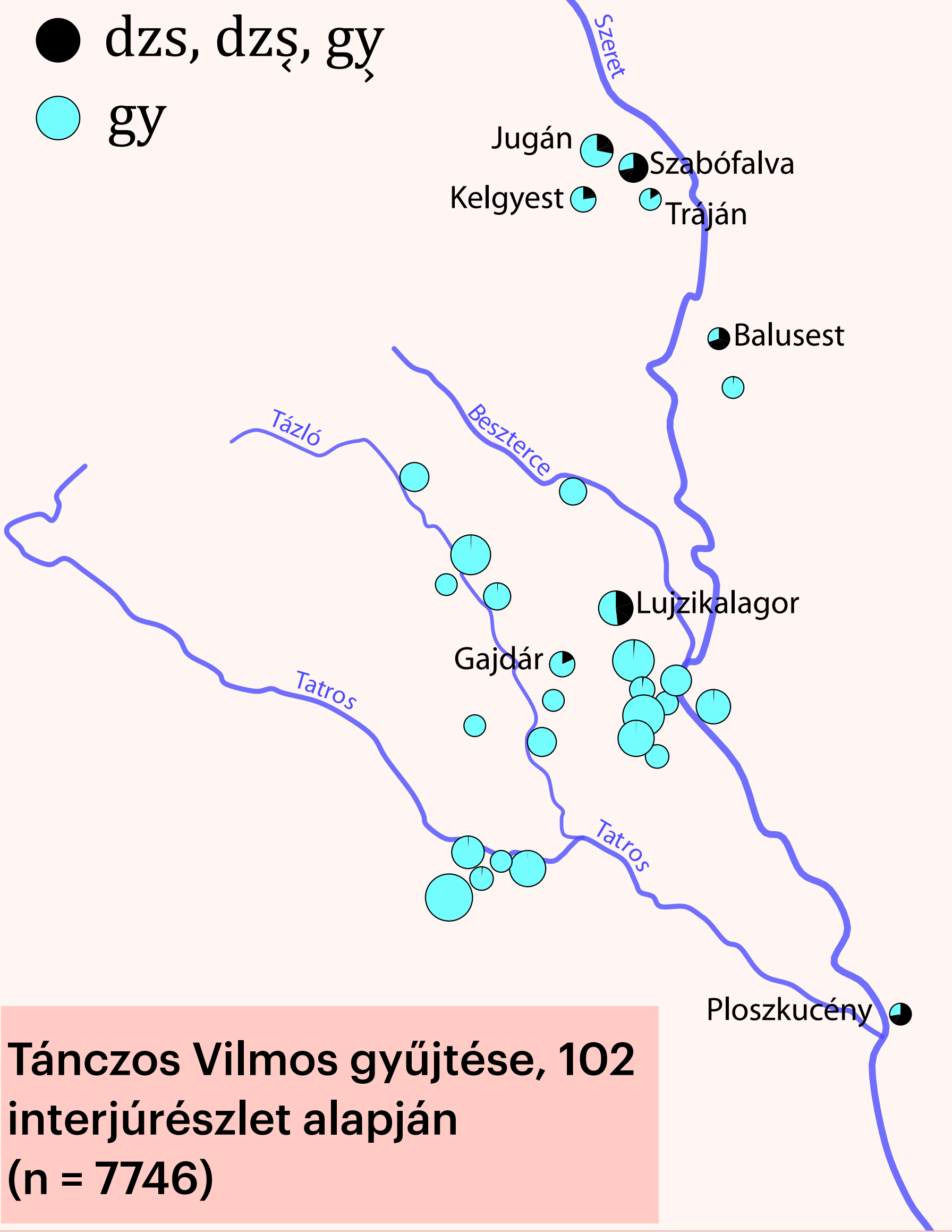
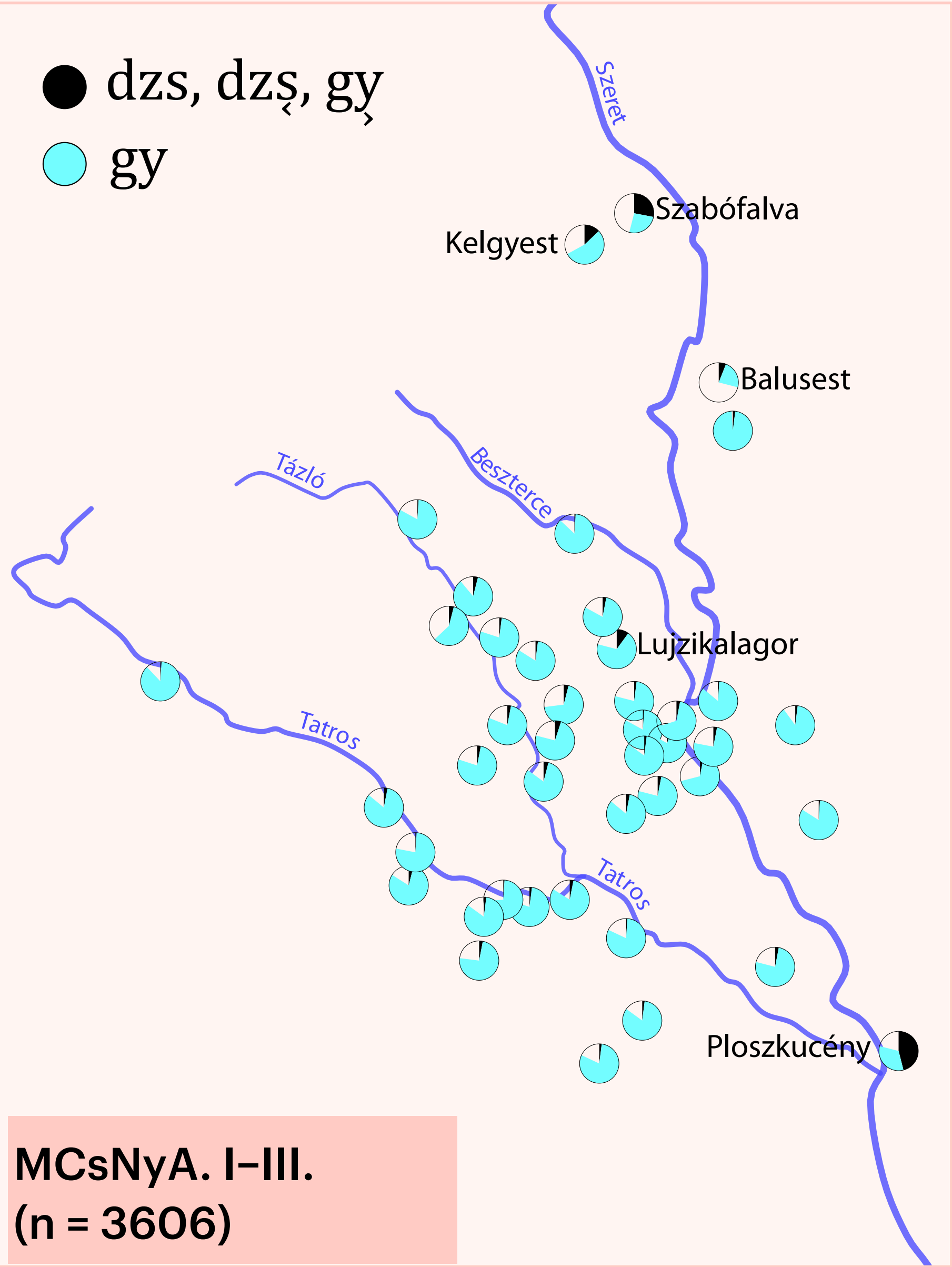
{3 •Hát azét, hogy ə nekem úgy teccik ['tűnik'], hogy mintha könnyebb magyarul imádkozni, mind románul.•}

{-1 •De tud románul is?•}

{3 •Há ê• | •muszáj tudni, tudom románul és az imáccságot, mennyit tudok, azt tudom románul és. • Há. • Asztá• | •mennyit nem tudok magyarul, •végzem románul, s• | •gondolom, hogy a Jòisten tud• | •ijen-ojan nyelvekét, mindénfélét, s mégérti, • [#nevet#]• •mégérti amijenekét• | •imádkozom. • Mèt nem tudok sokat imádkozni | magyarul. • Da áhol tudtam,

# Nyelvjárási térképek generálása a szövegek alapján

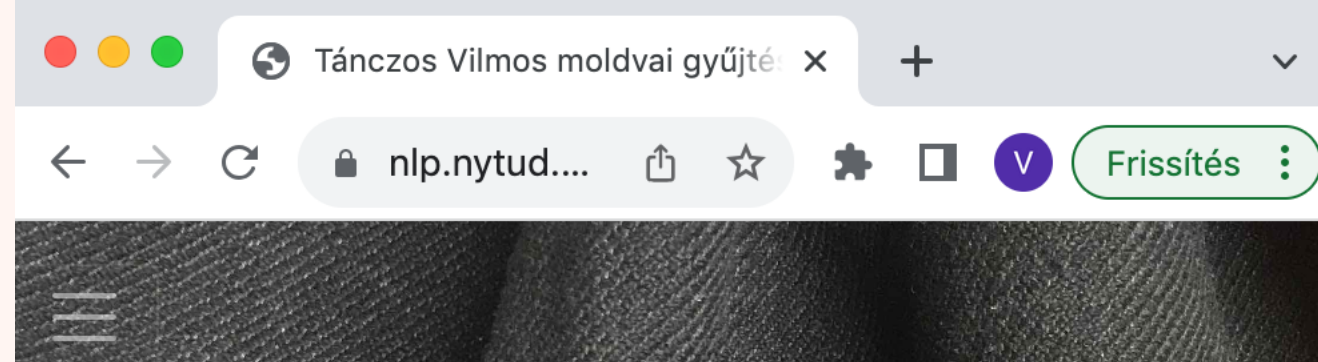
A kutatópontok kitöltöttsége a két kategória együttes előfordulását mutatja az egyes kutatópontokon előforduló adatok számához viszonyítva. A színek mutatják a kétféle kategóriába sorolt találatok gyakoriságát egymáshoz képest.



A kutatópontok színe azt mutatja, hogyan oszlanak meg a dzséző és gyéző adatok egy adott kutatóponton. A kutatópontok mérete a találatok függvényében alakul, összefüggésben azzal is, mennyi lejegyzés van egy adott kutatópontról jelenleg az adattárban.

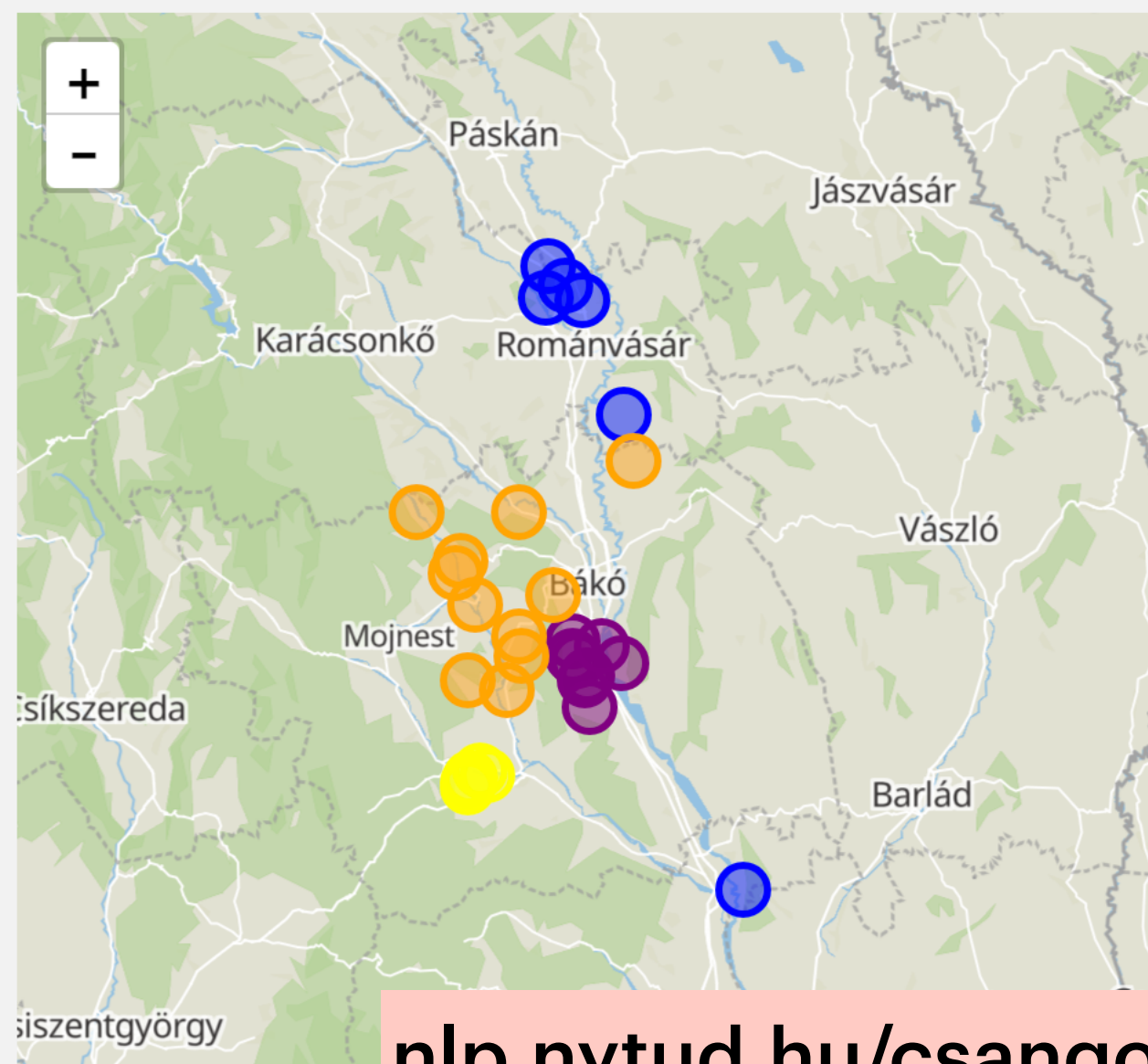
# A Tánczos Vilmos gyűjtéséből készült lejegyzések jelenleg elérhető online változata

30 moldvai településről vannak lejegyzett és ellenőrzött interjúrészleteink. Az adattár folyamatosan bővül.



## Kutatópontok

A térkép azokat a moldvai településeket jeleníti meg, ahonnan már vannak lejegyzéseink. A kutatópontok neve a pontokra kattintva érhető el. A kiválasztott kutatópont oldalának megjelenítéséhez kattints a nevére!



[nlp.nytud.hu/csango](http://nlp.nytud.hu/csango)

Fonetikailag igényes változat

{-1 «Magyarul» imádkozik Marika néni?•}

{3 Magyarul.•}

{-1 Miért imádkozik magyarul?•}

{3 Hát azét, hogy ə nekem úgy teccik ['tűnik'], hogy mintha könnyebb magyarul imádkozni, mind románul.•}

{-1 De tud románul is?•}

{3 Há ə• | muszáj tudni, tudom románul és az imáccságot, mennyit tudok, azt tudom románul és.• Há.• Asztá• | mennyit nem tudok magyarul,• végzem románul, s• | gondolom, hogy a Jòisten tud• | ijen-ojan nyelveket, mindenfélét, s megérti,• [#nevet#]• megérti amijeneket• | imádkozom.• Mèt nem tudok sokat imádkozni | magyarul.• Da əhol tudtam, édesanyám megtanyitot vót, met csak úgy imádkoztunk, de ha,• kifelejtettem.•}

{-1 És ə ugyanez

{3 «À Miatyán»

Gyónom urómot

Mostaniasan, de

imáccságot,• az

{-1 «Most ez a»

én?»•}

[nlp.nytud.hu/csango/m4\\_53/m4\\_53\\_02/index\\_simp.html](http://nlp.nytud.hu/csango/m4_53/m4_53_02/index_simp.html)

Automatikusan egyszerűsített változat

{-1 De tud románul is?•}

{3 Há ə• | muszáj tudni, tudom románul és az imáccságot, mennyit tudok, azt tudom románul és.• Há.• Asztá• | mennyit nem tudok magyarul,• végzem románul, s• | gondolom, hogy a Joisten tud• | ijen-ojan nyelveket, mindenfélét, s megérti,• [#nevet#]• megérti amijeneket• | imádkozom.• Met nem tudok sokat imádkozni | magyarul.• Da ahol tudtam, édesanyám megtanyitot vót, met csak úgy imádkoztunk, de ha,• kifelejtettem.•}

---

# Következtetés, kitekintés

- Nagy, hangzó adattárak építésére van szükség
  - új módszerekkel a korábbinál pontosabb kép a magyar nyelv térbeliségéről
  - a nyelvatlaszokban fellelhető adatok értelmezését is segítő eredmények
- Kulcsszó: együttműködés
  - szoftverek között (pl. Bihalbocs, Praat, R)
  - adattárak között (pl. atlaszok és hangzó adattárak)
  - megközelítések között (pl. elméleti nyelvészet, nyelvtechnológia, fonetika, dialektológia)
  - civilek és kutatók között (pl. VD és VFS)

**Köszönjük a figyelmet! :-)**

---