Dialektometria, klasszifikáció, hasonlósági súlypontok

Vargha Fruzsina Sára
ELTE Nyelvtudományi Kutatóközpont

Magyar nyelvjárások akusztikai és dialektometriai vizsgálata, NKFIH, FK 138396

Nyelvelmélet és nyelvleírás 2: Nyelvi változatosság, 2025. november 26–27.

Az előadás felépítése

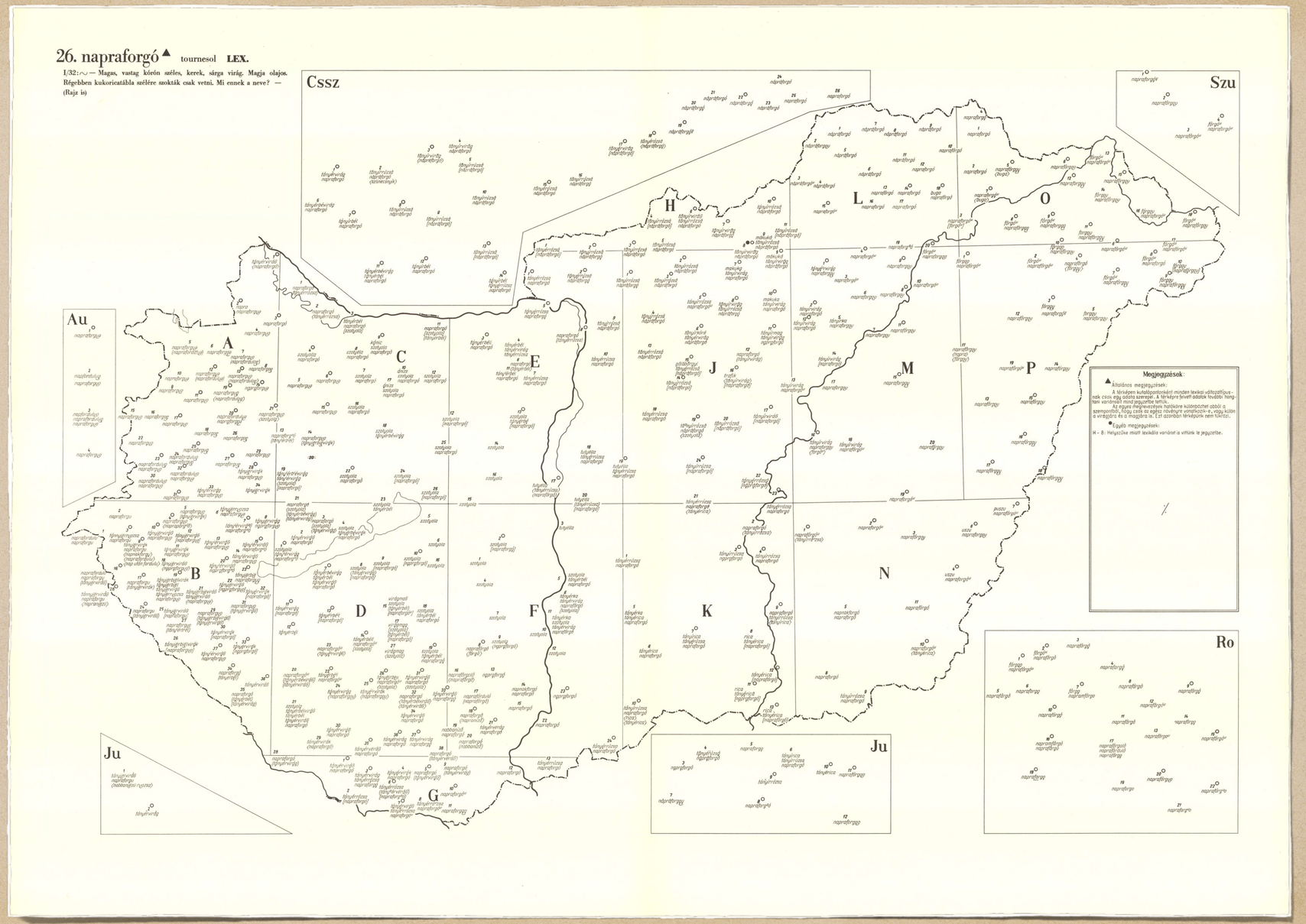
- A dialektometria helye a nyelvjáráskutatásban
- Informatizált nyelvjárási adatok
  - nyelvjárási atlaszok
  - adatok kódolása, konvertálása
  - integrált kutatóponthálózat, integrált atlaszok
- Dialektometria: különböző megközelítések
  - kutatói csoportosítás alapján
  - automatikus adatösszevetés alapján
- Egy kutatópont nyelvi hasonlósági térképe
- Nyelvjárási klasszifikáció és dialektometria (klaszteranalízis, MDS)
- Új módszer: nyelvi hasonlósági súlypontok és valós törésvonalak a nyelvjárási kontinuumban
- Egyetlen nyelvjárási térkép adatainak automatikus csoportosítása
- Konklúzió

A dialektometria helye a nyelvjáráskutatásban

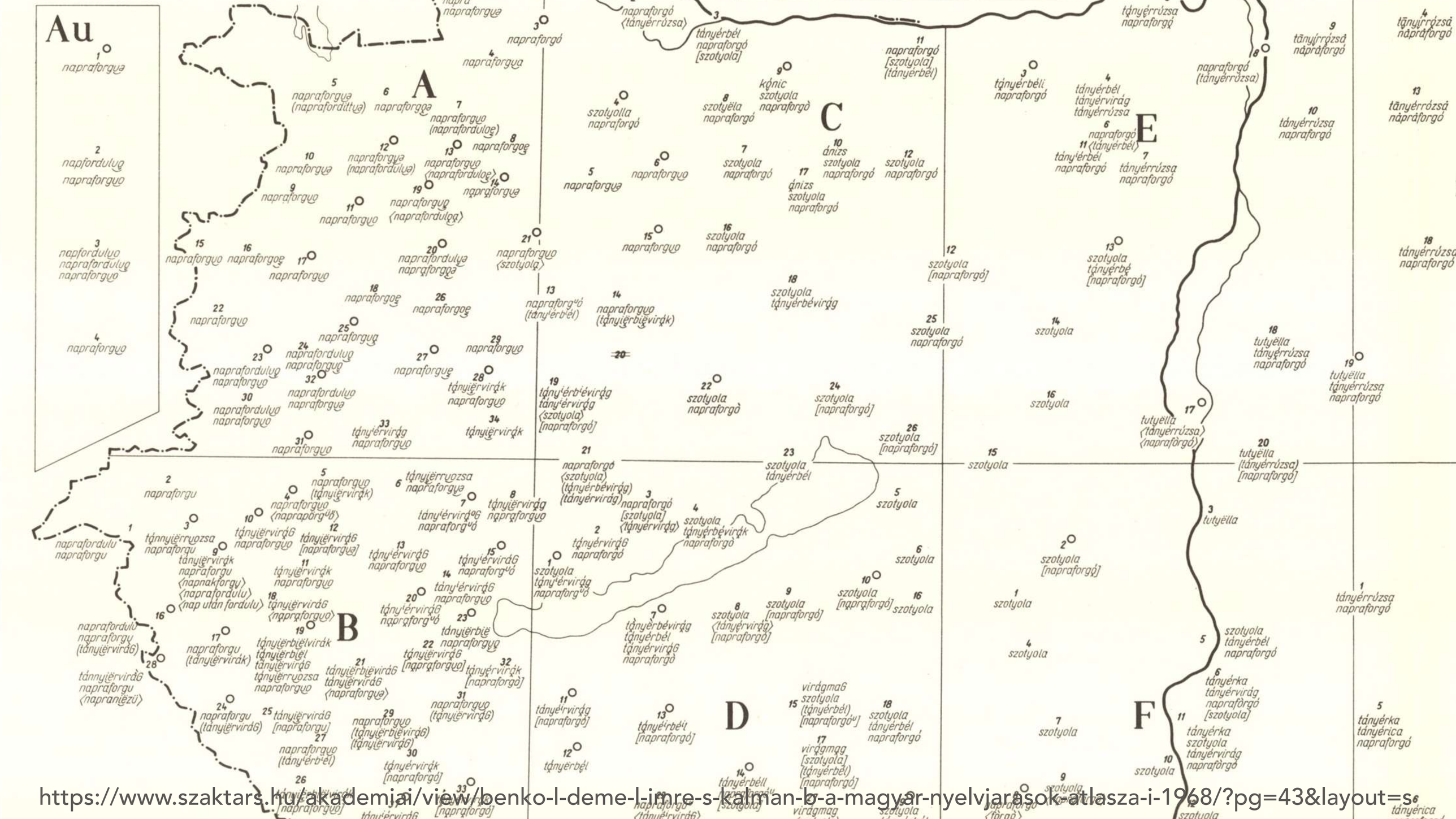
- Kvantitatív nyelvföldrajz
- Feltétel: rendelkezésre álló, elemezhető nyelvi adatok
- Nyelvatlaszok kialakítása, hangfelvételek készítése: a dialektológia a 20. század második felében a magyar nyelvtudomány kiemelt területe (fő cél: a nyelvjárások klasszifikációja)
- Imre Samu 1971. A mai magyar nyelvjárások rendszere: kvantitatív nyelvföldrajz A magyar nyelvjárások atlasza szűkített kutatópont-hálózata alapján
- A kvantitatív dialektológiai vizsgálatok elakadnak, elvétve akad egy-két próbálkozás (pl. Balogh Lajos)
- „A **nyelvjáráskutatás** nyelvtudományon belüli elismertségének megroppanását (a változásról l. Kiss 2015:385) éppen az okozta, ami korábban a legmagasabb presztízst is hozta: a tudományterület belefulladni látszott a fáradságosan felhalmozott adatok holt papírtengerébe. **Nem boldogult megfelelően a feldolgozás és elemzés kihívásaival**, miként a hangzó anyagok kezelésére sem volt megoldása.” (Vargha–Vékás 2024)
- Fordulat: megfelelő módszerek az adatok elemezhetővé tételére – Bihalbocs: a nyelvjáráskutatás hatékony támogatása célzott fejlesztésekkel, térinformatikai megoldásokkal (Vékás 2007)

Mi kell ahhoz, hogy végezhessünk dialektometriai kutatásokat?

- nyelvi adatok: nyelvatlasz (kérdőfüzetek összeállítása, munkamódszerek kidolgozása, az adatok felgyűjtése, rendszerezése, szerkesztés)
- az atlaszadatok informatizálása (adatbevitel és ellenőrzés) és adatbázisba-rendezése (Bihalbocs)
- az adatok összevetése valamilyen módszer, algoritmus alkalmazásával (pl. Levenshtein-algoritmus), hasonlósági mátrix kialakítása (Bihalbocs)
- statisztikai elemzések a mátrix alapján (R)
- az elemzés eredményének színeknek való megfeleltetése és térképezése (Bihalbocs)



https://www.szaktars.hu/akademiai/view/benko-l-deme-l-imre-s-kalman-b-a-magyar-nyelvjarasok-atlasza-i-1968/?pg=43&layout=s


Au
A
B
C
D
E
F
napraforgua
napfordulug
napraforgug
napraforgoe
napraforgoa
napraforgo
napraforgu
napraforgo (tányérrúzsa)
tányérbél napraforgó [szotyola]
napraforgó [szotyola] (tányérbél)
tányérrúzsa napraforgó
napraforgó (tányérrúzsa)
kónic szotyola napraforgó
szotyolla napraforgó
szotyëla napraforgó
tányérbéli napraforgó
tányérbél tányérvirág tányérrúzsa
napraforgó (tányérbél)
tányérbél napraforgó
tányérrúzsa napraforgó
tányérrúzsá náprafórgó
ánizs szotyola napraforgó
szotyola napraforgó
napraforgua (naprafordittua)
napraforgua (napraforduloe)
napraforgua (napraforduloa)
napraforgug (napraforduloa)
napfordulug napraforgug
napfordulug napraforgug napraforgug
napraforgug (szotyola)
napraforgug (tányérbévirák)
napraforgó (tányérbél)
napraforgó (tányérbévirág)
napraforgó [szotyola] (tányérvirág)
napraforgó (szotyola) (tányérbévirág) (tányérvirág)
napraforgó (szotyola) [napraforgó]
napraforgó (tányérvirág)
napraforgó [napraforgó]
napraforgó (tányérvirág) napraforgó
napraforgó (tányérvirág) napraforgó
napraforgó (tányérvirág) napraforgó

Kereshető fonetikai szimbólumok

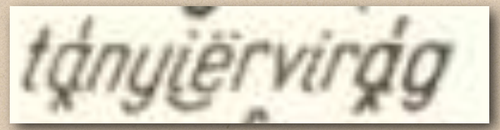
patká̰

á ~ ^ ⇒ á ^ ~ = á̰

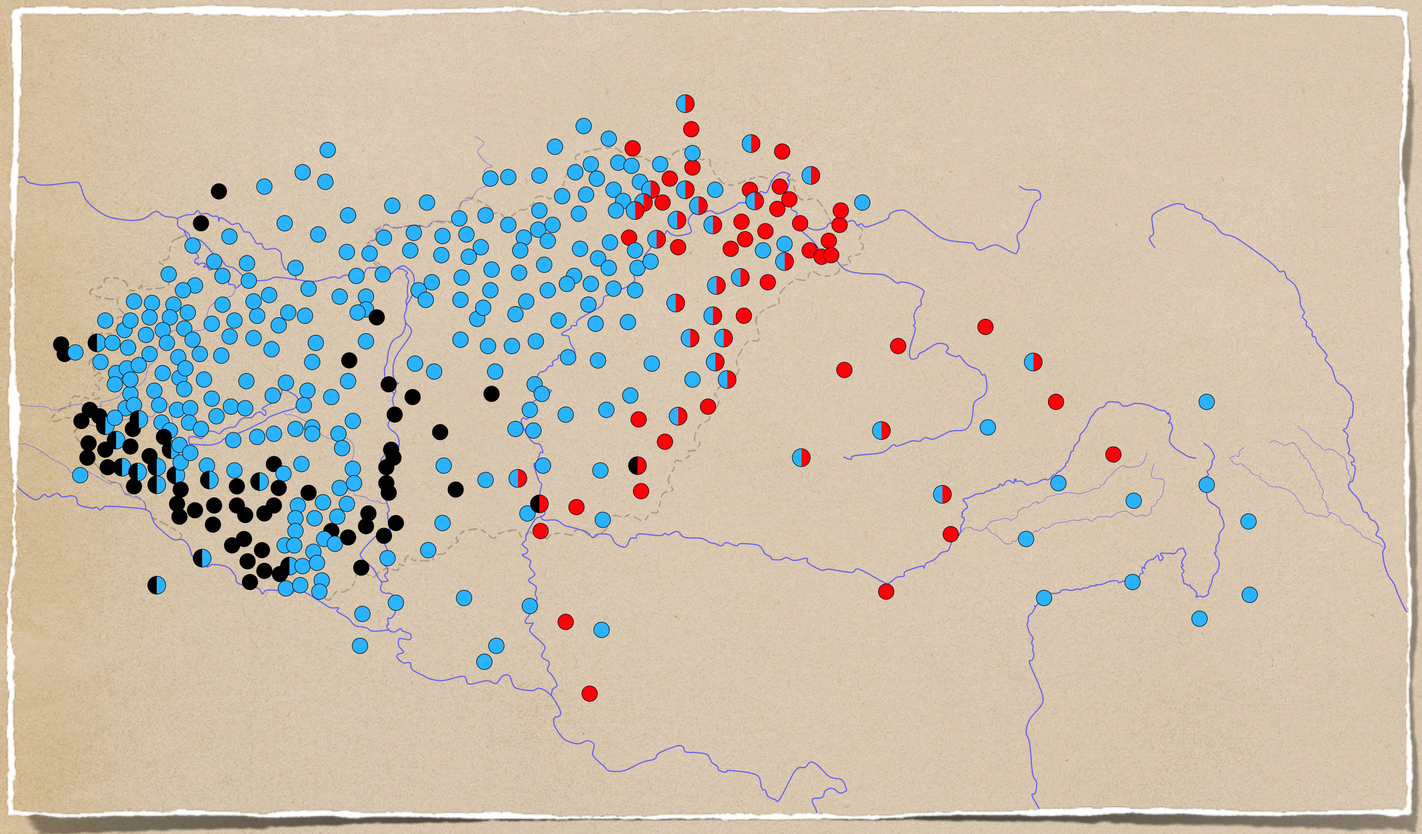
szárnyasägiér szárnyosegëir

Fon.1	·sz·á̰·r·ny·a̰·s·ä·g·i̯é·r· ·sz·á̰·r·ny·o̰·s·e·g·ëi̯·r·
Fon.2	·sz·á·r·ny·a·s·ä·g·ié·r· ·sz·á·r·ny·o·s·e·g·ëi·r·
Fon.3	·sz·â·r·ny·a·s·e·g·ê·r· ·sz·â·r·ny·o·s·e·g·ê·r·
Fon.4	·z·a·l·n·a·zs·a·g·a·l· ·z·a·l·n·a·zs·a·g·a·l·

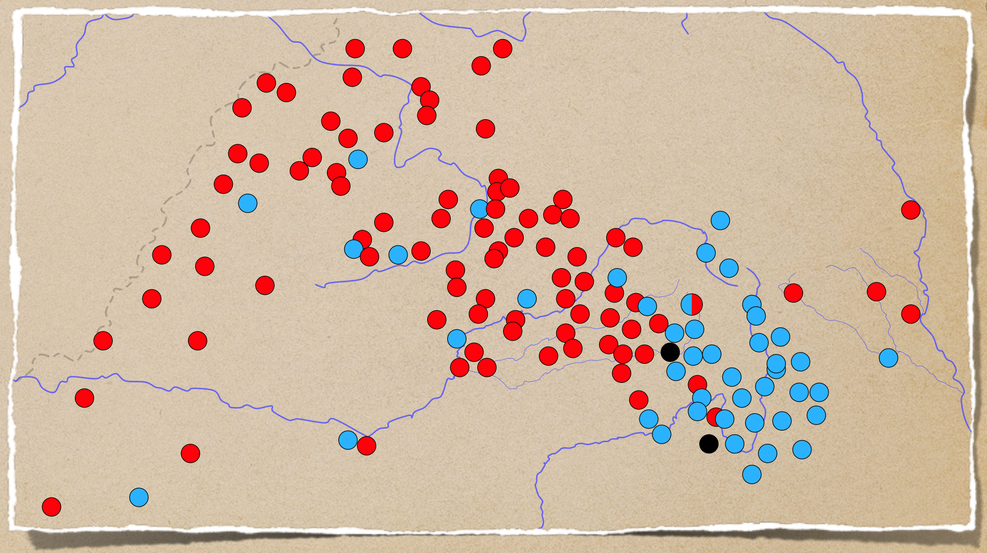
Hogyan lehet kódolni a fonetikai szimbólumokat?

- Eredeti adat a papírtérképen: 
- Balogh Lajosék kódolása: tá↑ny{ië}rvirá↑g / tá(F)ny{ië}rvirá(F)g
- A Bihalbocsban (a szerkeszthető változat): tą̂nyi͡ërvirą̂g
- kódok: 116, 129, 243, 110, 121, 105, 221, 131, 213, 114, 118, 105, 114, 129, 243, 103
- megfelelő algoritmusokkal konvertálható más kódrendszerekbe (UC karakterekkel megjelenítve: tą̂nyi͡ërvirą̂g)

Integrált atlaszok – MNyA., RMNyA.



A magyar nyelvjárások atlasza (MNyA.)



A romániai magyar nyelvjárások atlasza (RMNyA.)

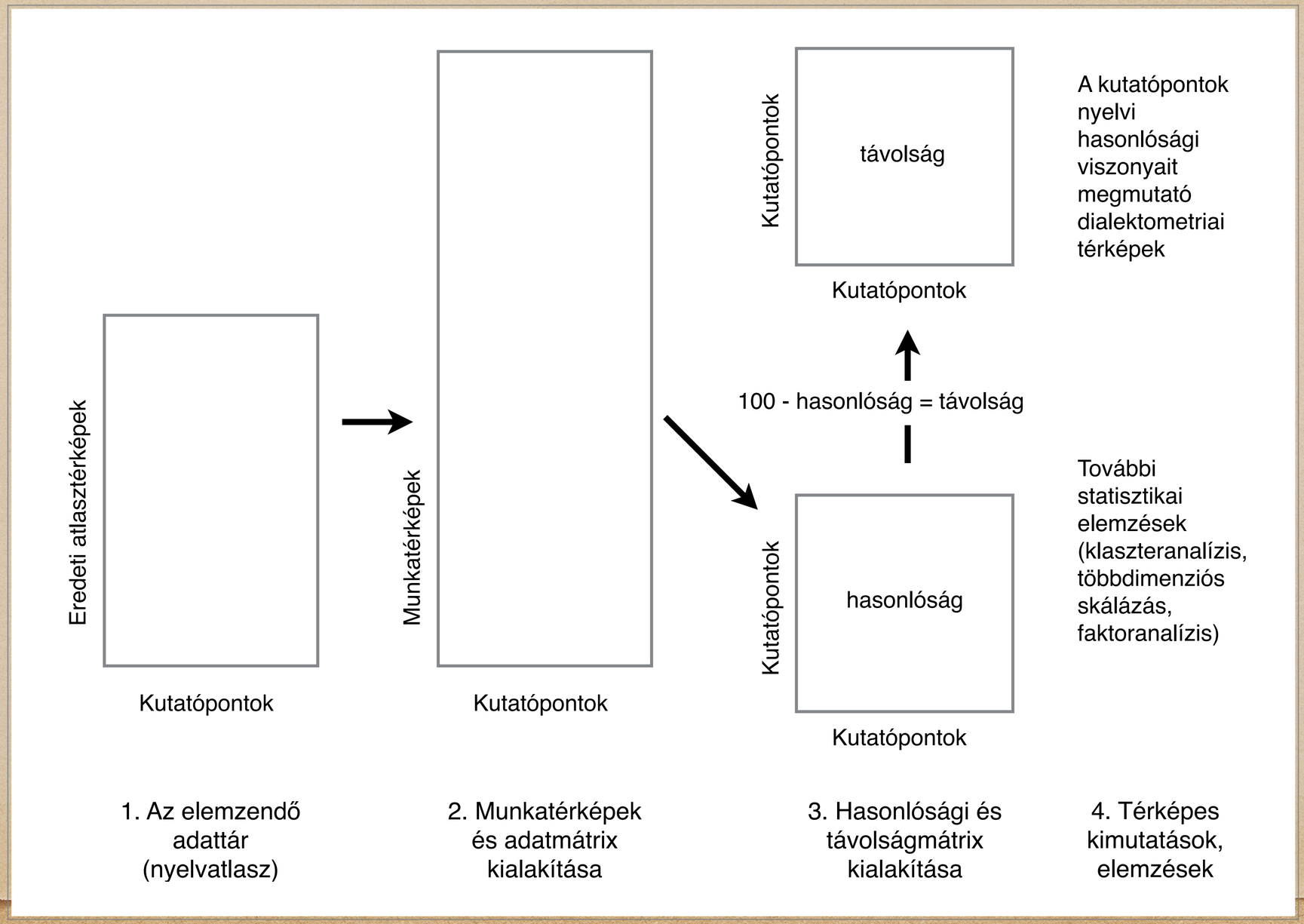
Integrált atlaszok (meggy)

A dialektometria mint kutatási módszer

Dialektometria

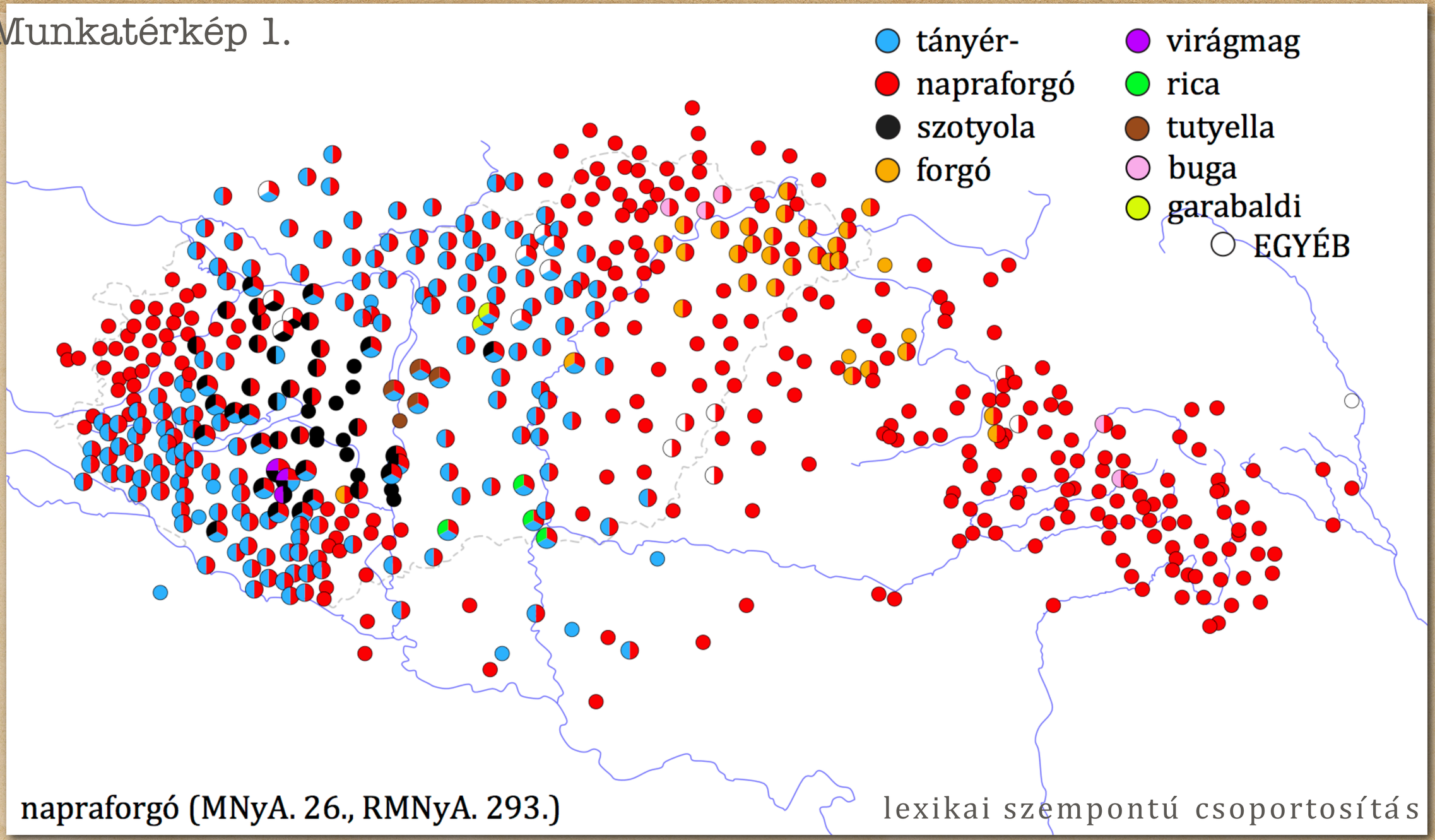
- Hans Goebl, Salzburg, 1971, 1982
- A nyelvjárások közötti hasonlóságot (illetve különbözőséget) vizsgáljuk, kvantitatív módon, statisztikai eszközökkel
- Eredmény: hasonlósági mátrix, térképes szemléltetés
- Hans Goebl és munkátarsainak új honlapja: http://dialektkarten.ch/dmviewer/alf/

A Goebl-féle dialektometria folyamatábrája



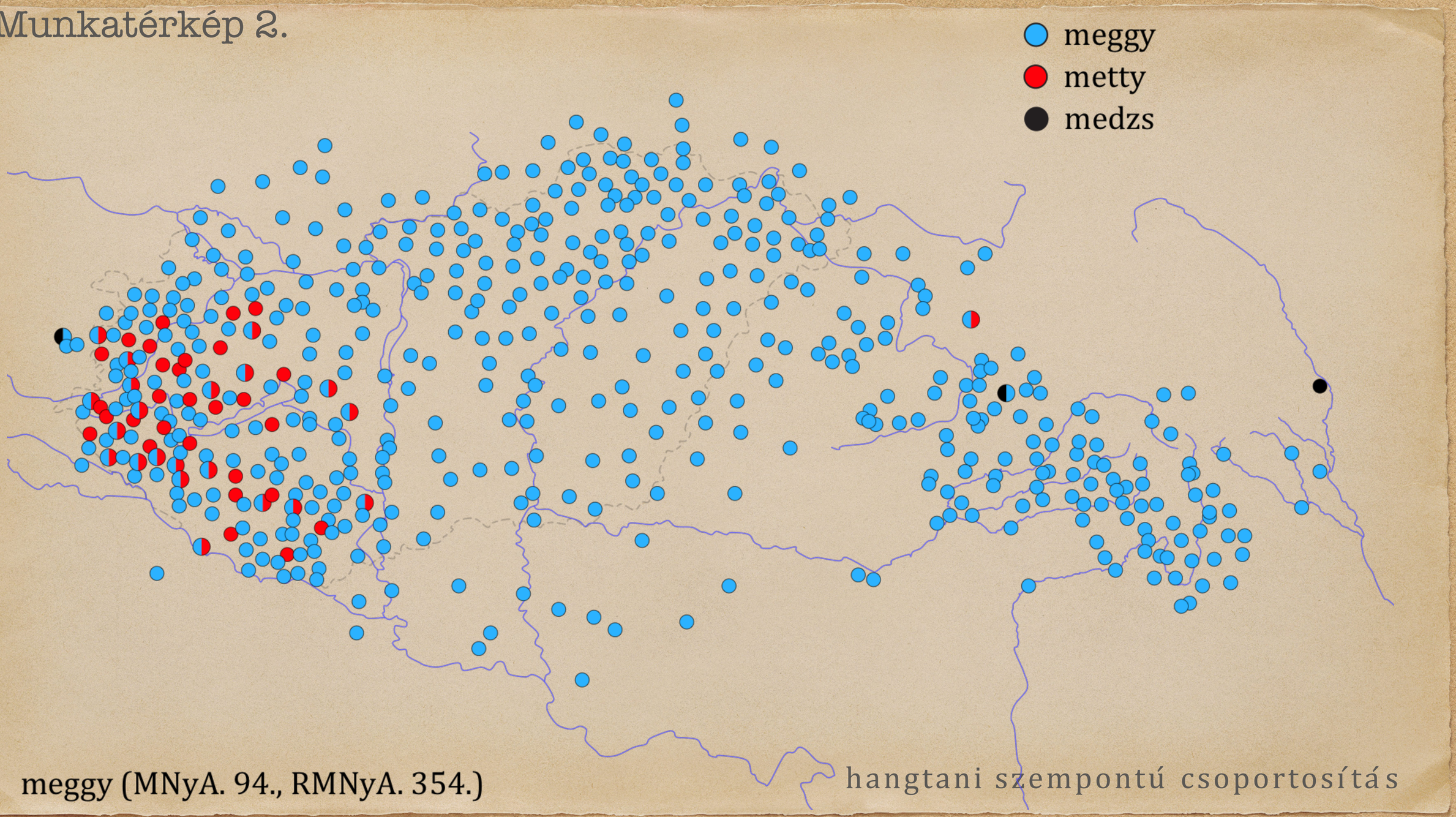
Goebl 2006

Munkatérkép 1.



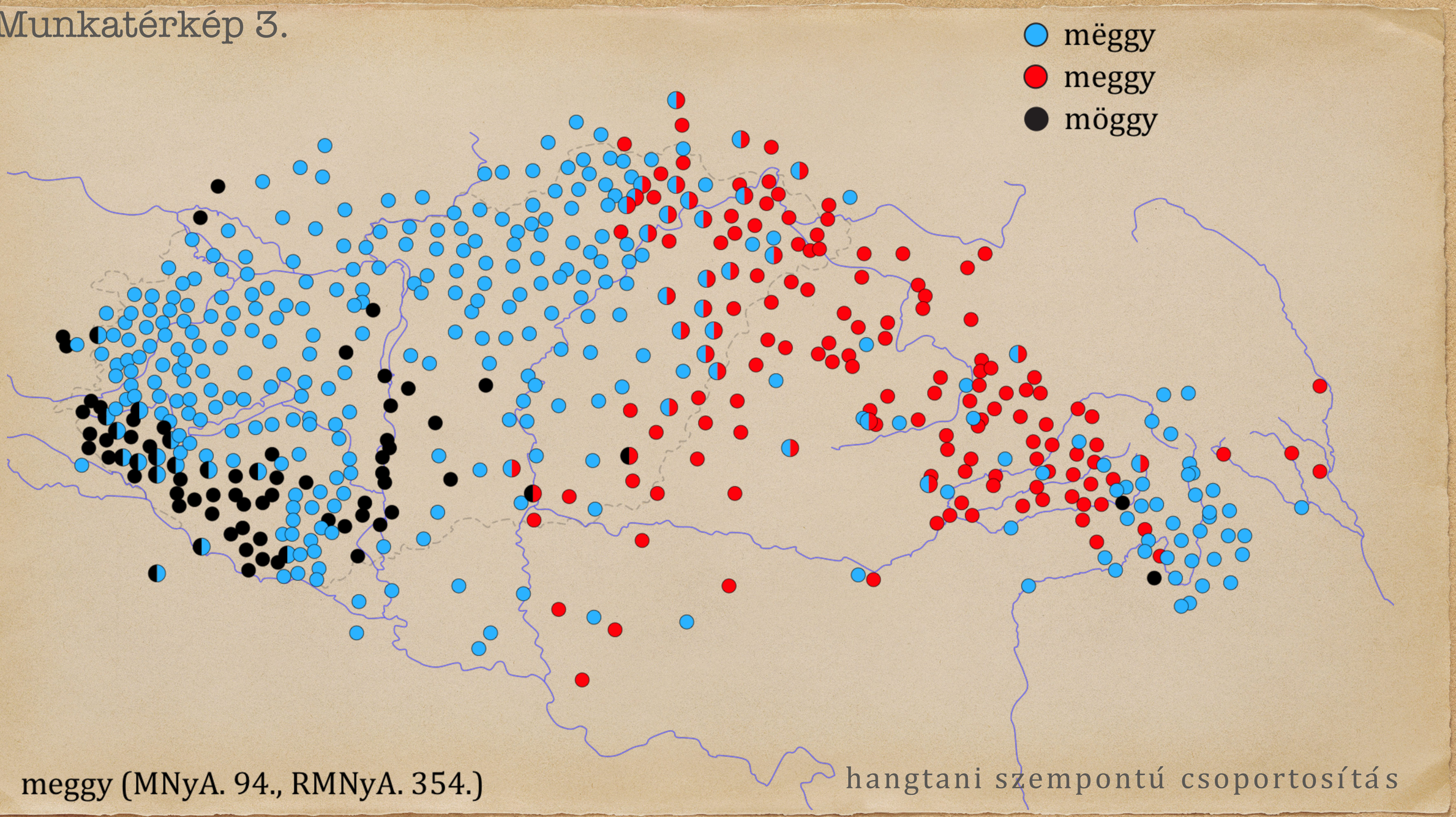
https://www.bihalbocs.hu/mnyarmnya/munkaterk/napraforgo1.html

Munkatérkép 2.



https://www.bihalbocs.hu/mnyarmnya/munkaterk/meggy1.html

Munkatérkép 3.



https://www.bihalbocs.hu/mnyarmnya/munkaterk/meggy2.html

Betűláncok automatikus összevetése

- John Nerbonne és munkatársai (1996), Wilbert Heeringa (2004)
- Automatikus adatösszevetés Levenshtein-algoritmussal
  - betoldás, törlés, csere
    - denevér : benevér (85,8%)
    - denevér : szárnyasegér (30%)
    - denevér : bőrmadár (12,5%)
    - benevér : hajkapó (0%)
    - hajkapó : hajlopó (71,5%)

		h	a	j	k	a	p	ó
	0	1	2	3	4	5	6	7
h	1	0	1	2	3	4	5	6
a	2	1	0	1	2	3	4	5
j	3	2	1	0	1	2	3	4
l	4	3	2	1	1	2	3	4
o	5	4	3	2	2	2	3	4
p	6	5	4	3	3	3	2	3
ó	7	6	5	4	4	4	3	2

		h	a	j	k	a	p	ó
	0	1	2	3	4	5	6	7
b	1	1	2	3	4	5	6	7
e	2	2	2	3	4	5	6	7
n	3	3	3	3	4	5	6	7
e	4	4	4	4	4	5	6	7
v	5	5	5	5	5	5	6	7
é	6	6	6	6	6	6	6	7
r	7	7	7	7	7	7	7	7

Peter Kleiweg, http://www.let.rug.nl/kleiweg/lev/

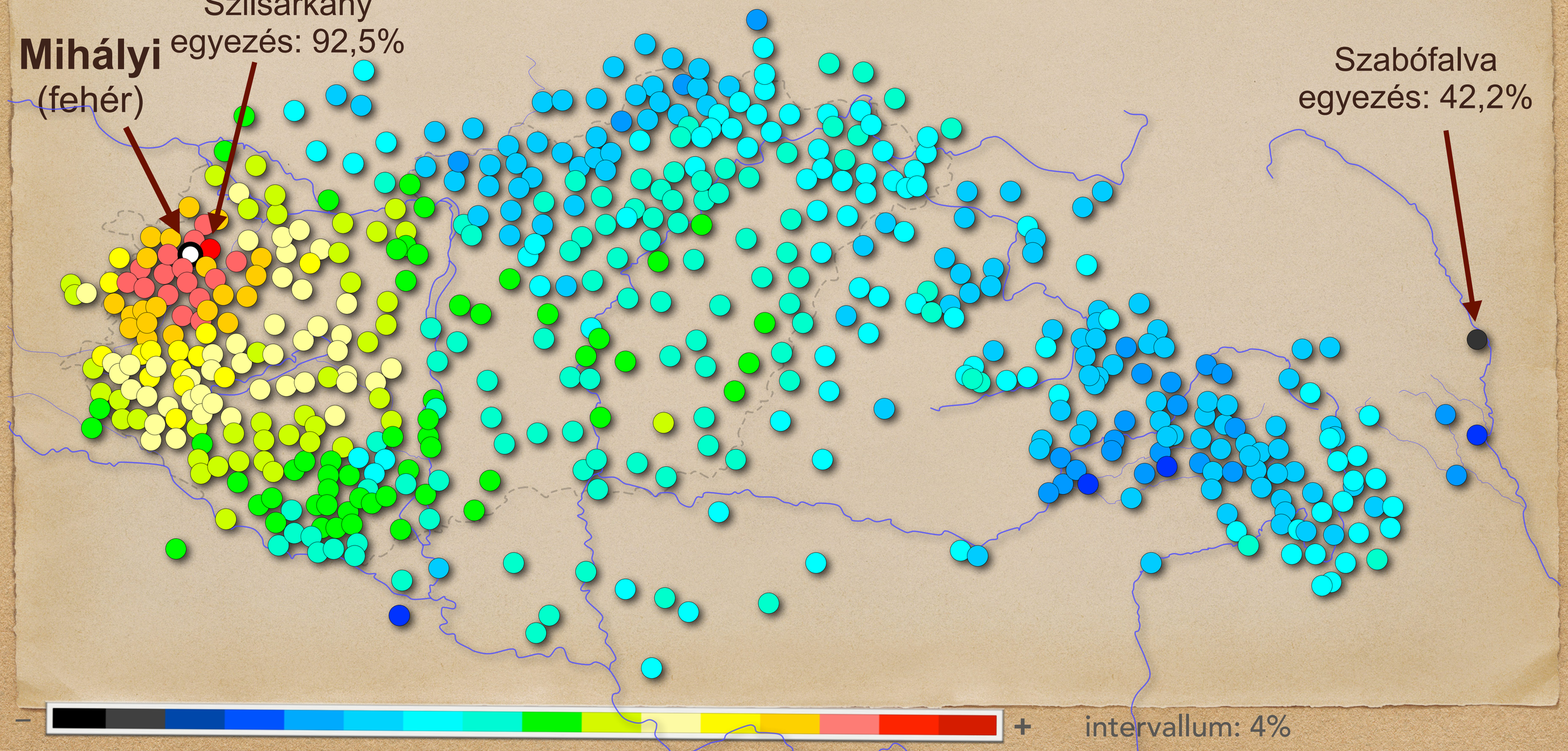
Hasonlósági mátrix

	1. kutatópont	2. kutatópont	3. kutatópont	n. kutatópont
1. kutatópont	100%	95%	78%	65%
2. kutatópont	95%	100%	82%	67%
3. kutatópont	78%	82%	100%	85%
n. kutatópont	65%	67%	85%	100%

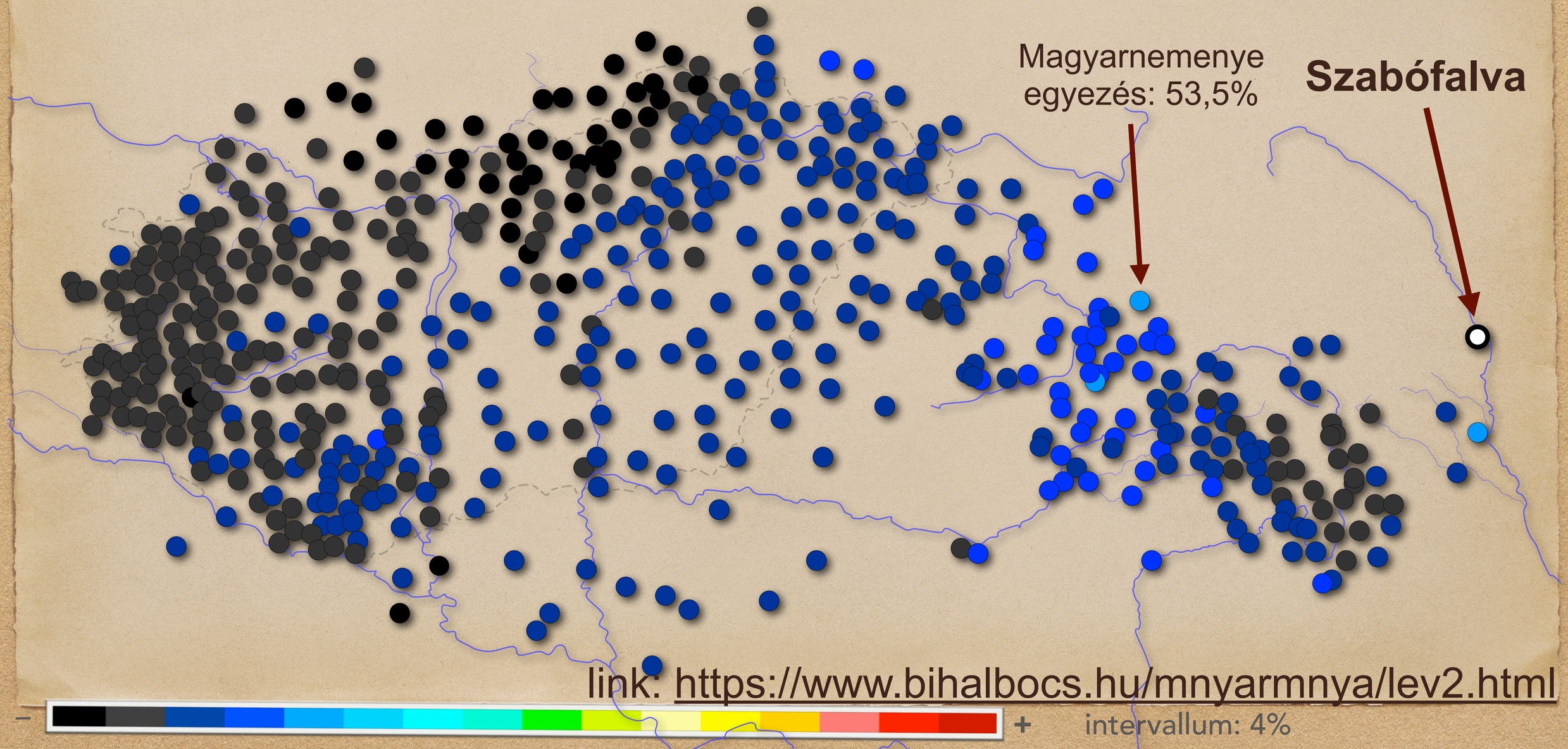
A MNyA. és a RMNyA. integrált dialektometriai elemzése

- 483 **integrált térképlap**
  - A magyar nyelvjárások atlasza sűrű kutatóponthálózatú térképlapjai
  - A romániai magyar nyelvjárások atlasza térképlapjai
- Hangjelölés összefésülése
  - azonos fonetikai tartalmat jelölő különböző szimbólumok egyforma jelölése
  - diftongusok jelölésének egyszerűsítése az RMNyA.-ban (összevethetőség növelése)
- **Levenshtein** algoritmus használatával létrehozott nyelvi hasonlósági mátrix
- „All-word method”: egy térképlap valamennyi adatát felhasználjuk, az elemzés nem korlátozódik a fonetikai különbségekre

Hogyan mutatja meg a dialektometriai elemzés egy kutatópont nyelvi hasonlósági viszonyait?

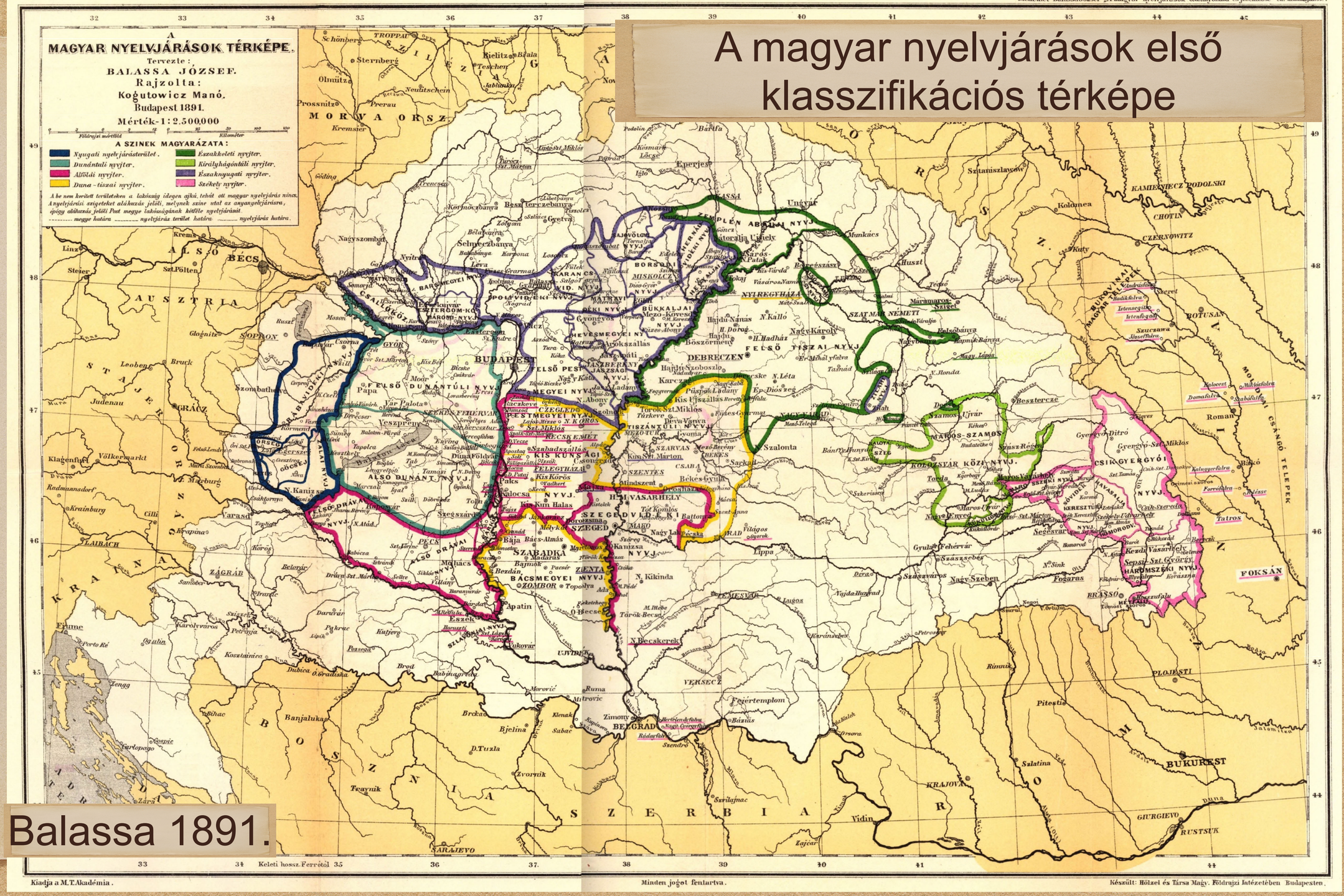


Hogyan mutatja meg a dialektometriai elemzés egy kutatópont nyelvi hasonlósági viszonyait?



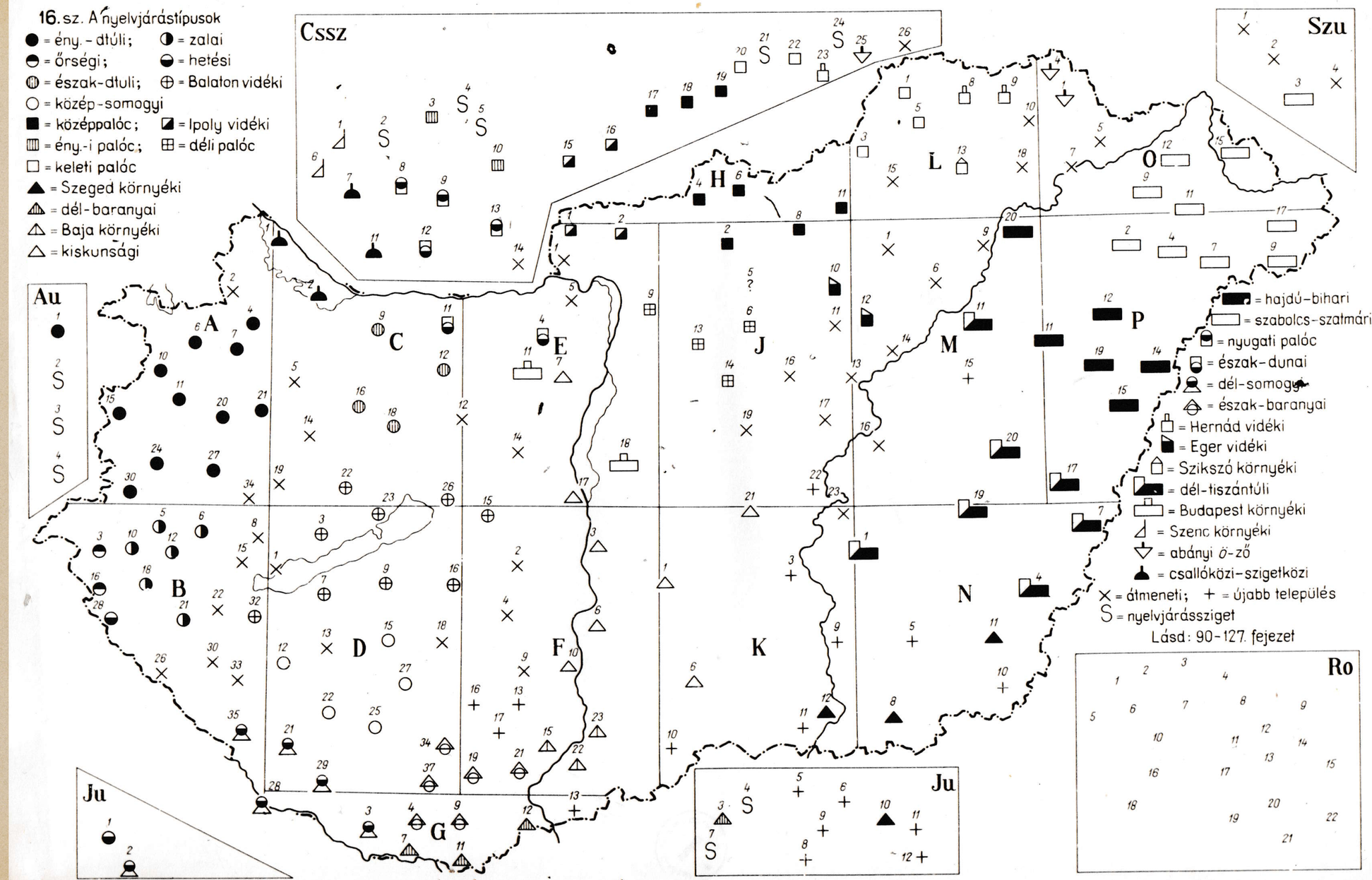
Nyelvjárási klasszifikáció

A magyar nyelvjárások első klasszifikációs térképe

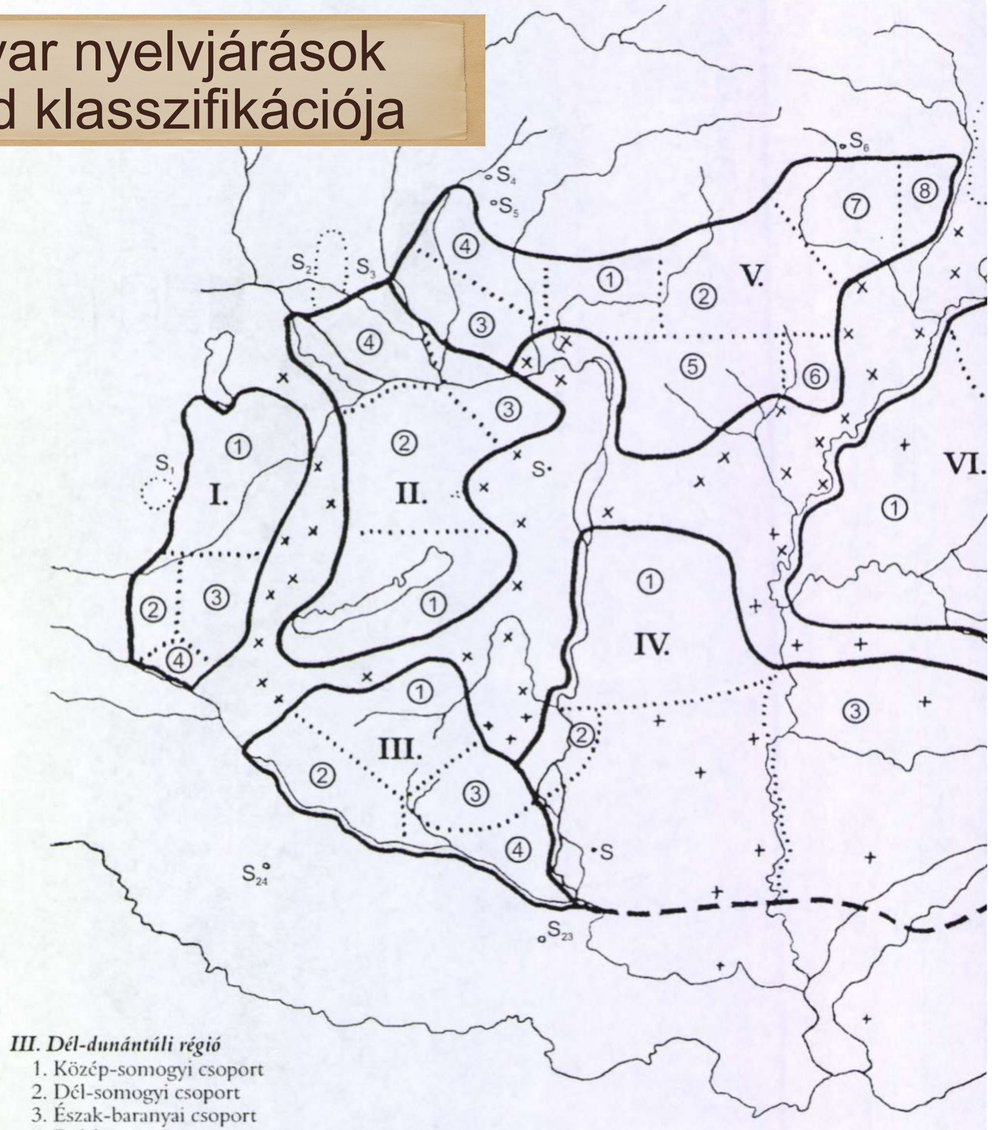


Balassa 1891.

Imre Samu 1971: nyelvjárástípusok



A magyar nyelvjárások standard klasszifikációja



I. Nyugat-dunántúli régió
1. Északnyugat-dunántúli csoport
2. Őrségi csoport
3. Zalai csoport
4. Hetési csoport

II. Közép-dunántúli–kisalföldi régió
1. Balaton vidéki csoport
2. Észak-dunántúli csoport
3. Észak-dunai csoport
4. Csallóközi–szigetközi csoport

III. Dél-dunántúli régió
1. Közép-somogyi csoport
2. Dél-somogyi csoport
3. Észak-baranyai csoport
4. Dél-baranyai csoport

IV. Dél-alföldi régió
1. Kiskunsági csoport
2. Baja környéki csoport
3. Szeged környéki csoport

V. Palóc régió
A) Középső tömb
1. Ipoly vidéki csoport
2. Középpalóc csoport
B) Nyugati tömb
3. Nyugati palóc csoport
4. Északnyugati palóc csop.
C) Déli tömb
5. Déli palóc csoport
6. Eger vidéki csoport
D) Keleti tömb
7. Keleti palóc csoport
8. Hernád vidéki csoport

VI. Tisza–Körös vidéki régió
1. Közép-tiszántúli csoport
2. Hajdú-bihari csoport
3. Kalotaszegi csoport

VII. Északkeleti régió
1. Szabolcs-szatmári csoport
2. Bereg-ugocsai csoport
3. Ungi csoport
4. Észak-szilágyi csoport

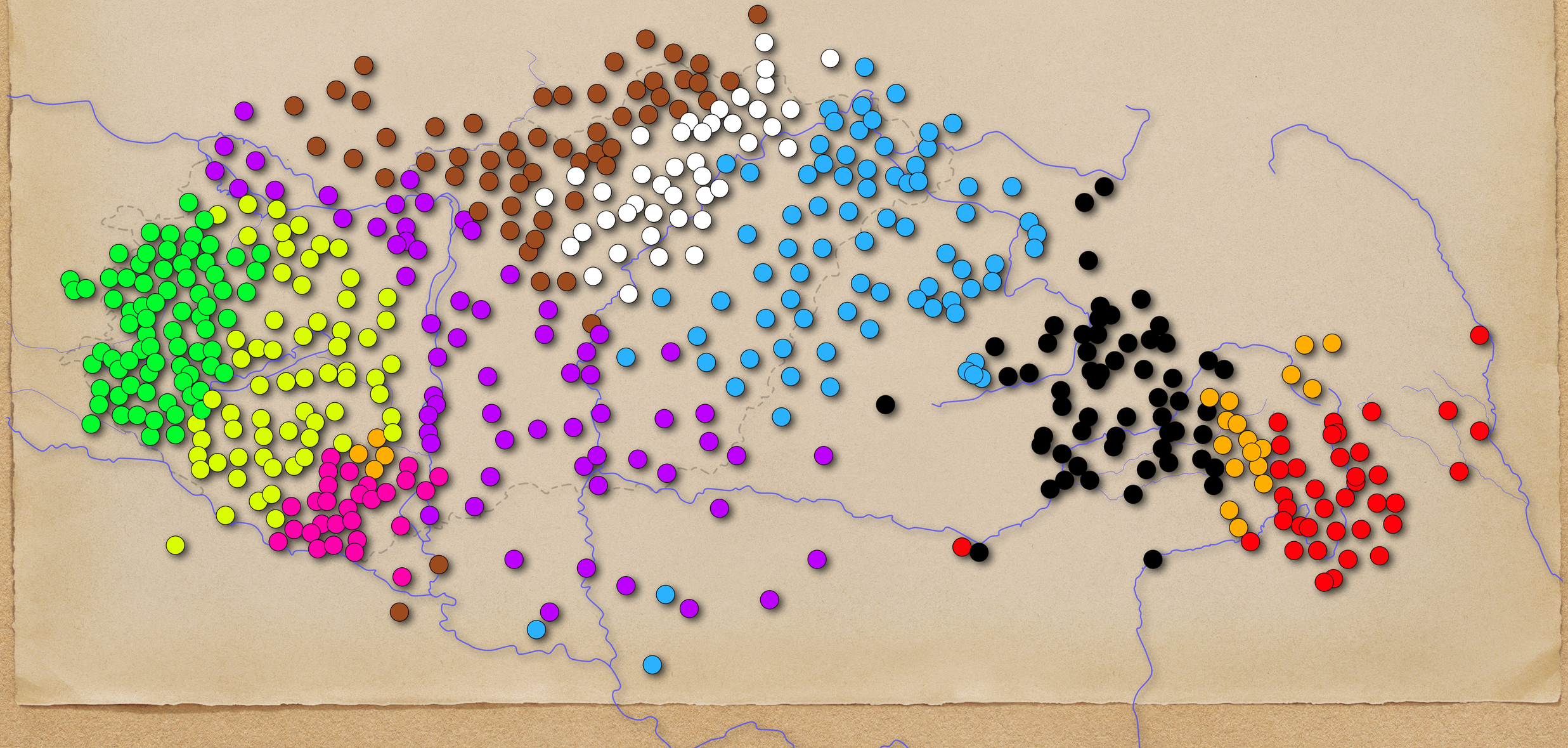
VIII. Mezőségi régió
1. Belső-mezőségi csoport
2. Aranyosvidéki csoport
3. Maros–Küküllő vidéki csop.

IX. Székely régió
1. Udvarhelyszéki csoport
2. Háromszéki csoport
3. Kászoni csoport
4. Alcsíki csoport
5. Felcsíki csoport
6. Gyergyói csoport
7. Kelet-marosszéki csop

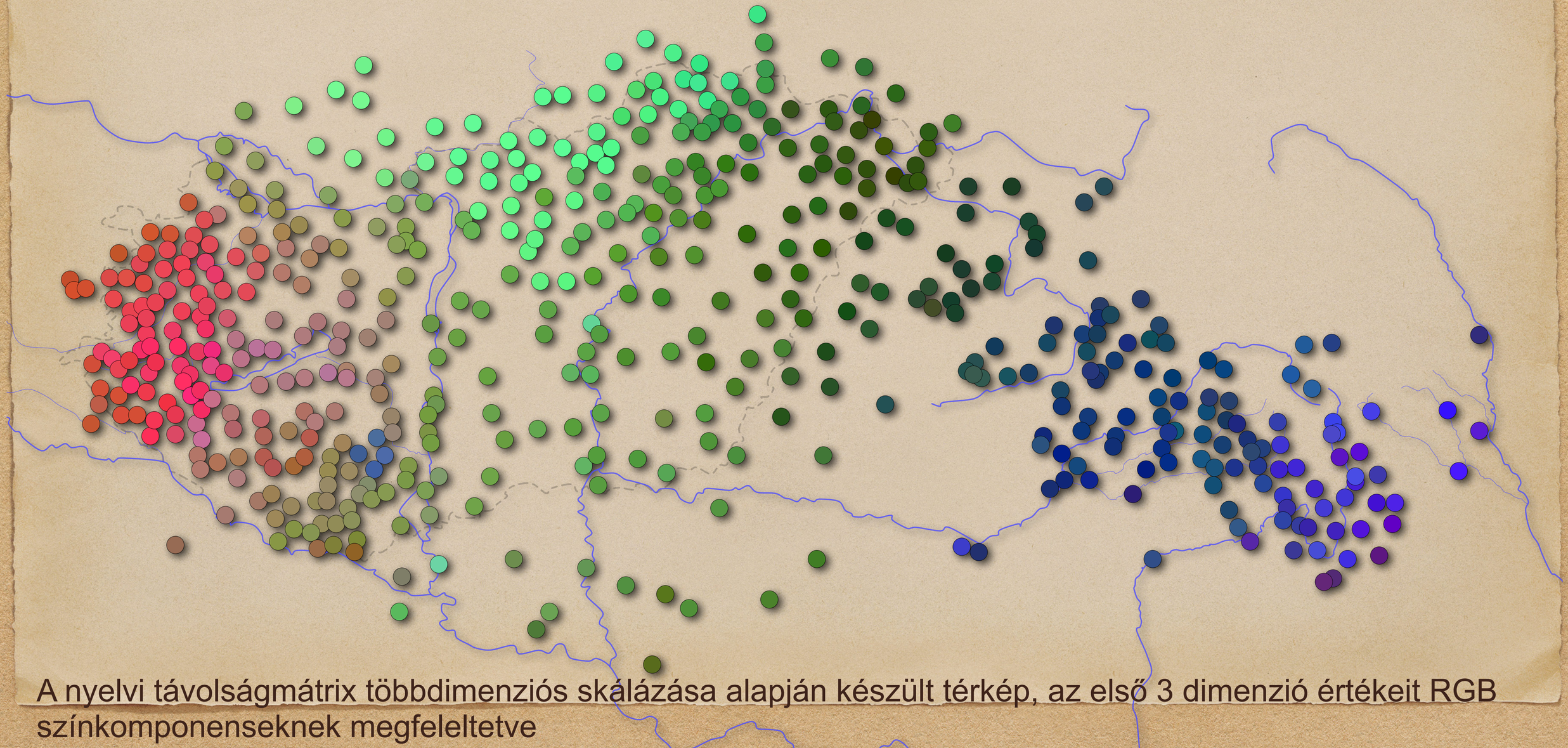
X. Moldvai régió
1. Északi csángó csoport
2. Déli csángó csoport
3. Moldvai székely csop.

Juhász 2001.

Klasztertérkép a nyelvi távolság alapján

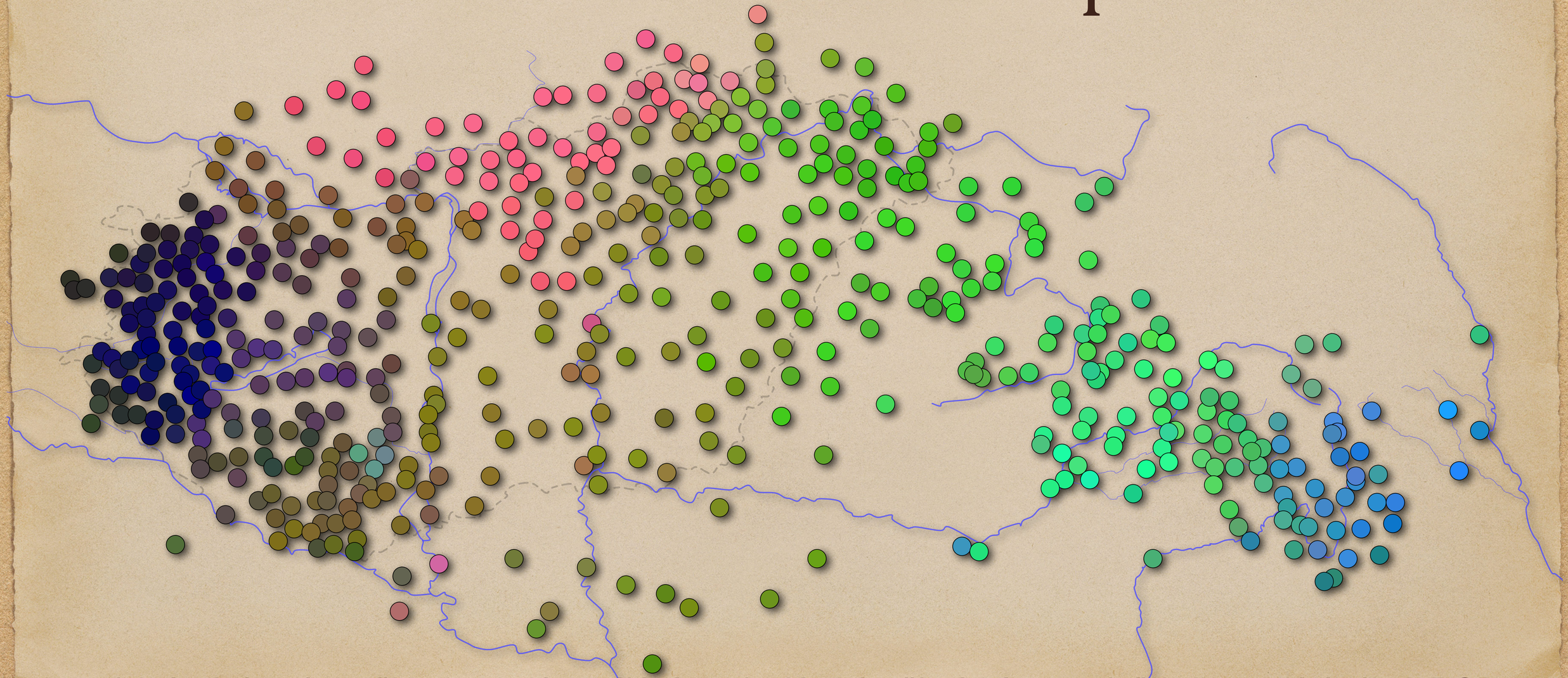


MDS-térkép: nyelvjárási kontinuum



A nyelvi távolságmátrix többdimenziós skálázása alapján készült térkép, az első 3 dimenzió értékeit RGB színkomponenseknek megfeleltetve

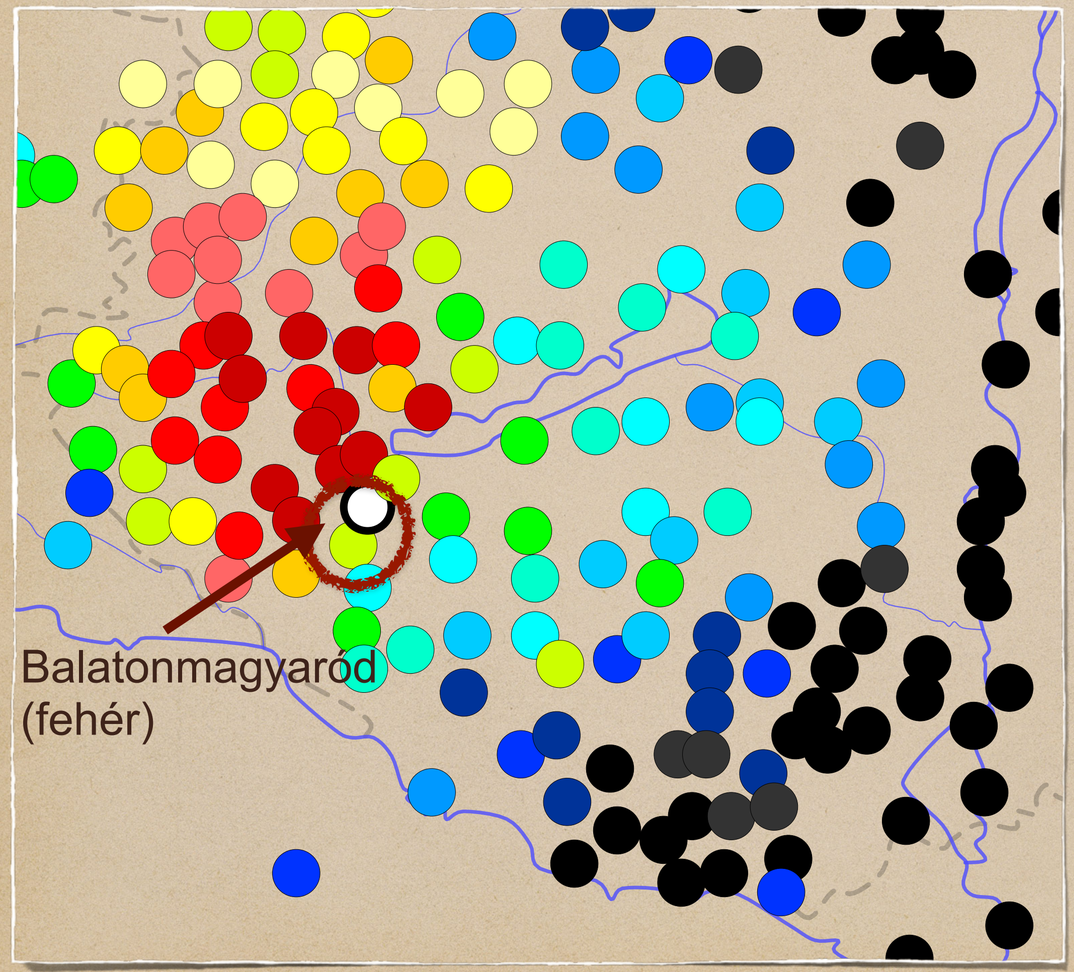
A magánhangzók hasonlósága alapján készült MDS-térkép

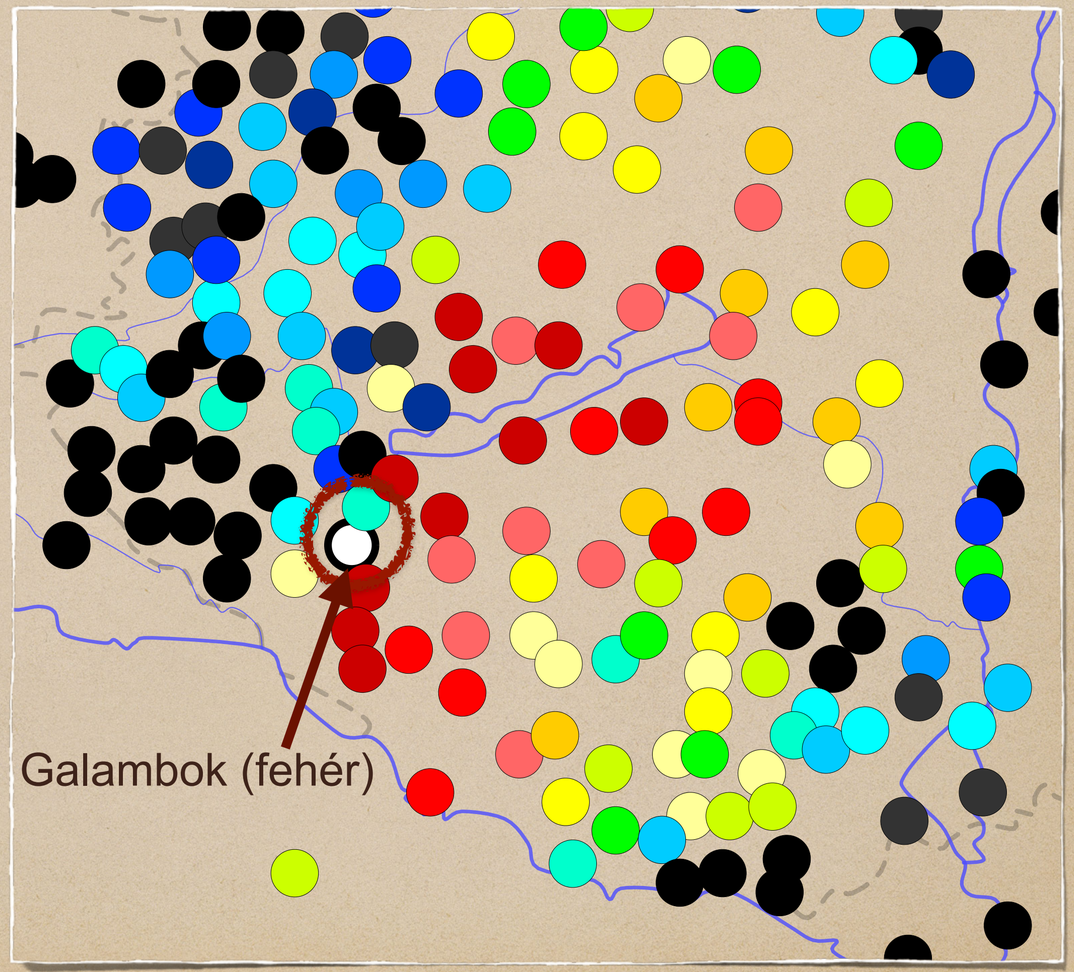


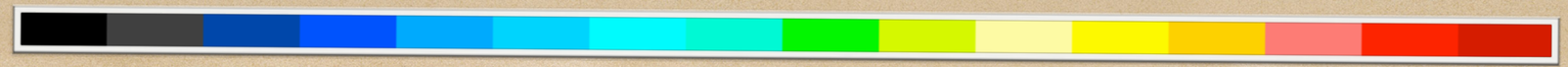
A nyelvi távolságmátrix többdimenziós skálázása alapján készült térkép, az első 3 dimenzió értékeit RGB színkomponenseknek megfeleltetve

Nyelvi hasonlósági súlypontok és valódi törésvonalak a nyelvjárási kontinuumban

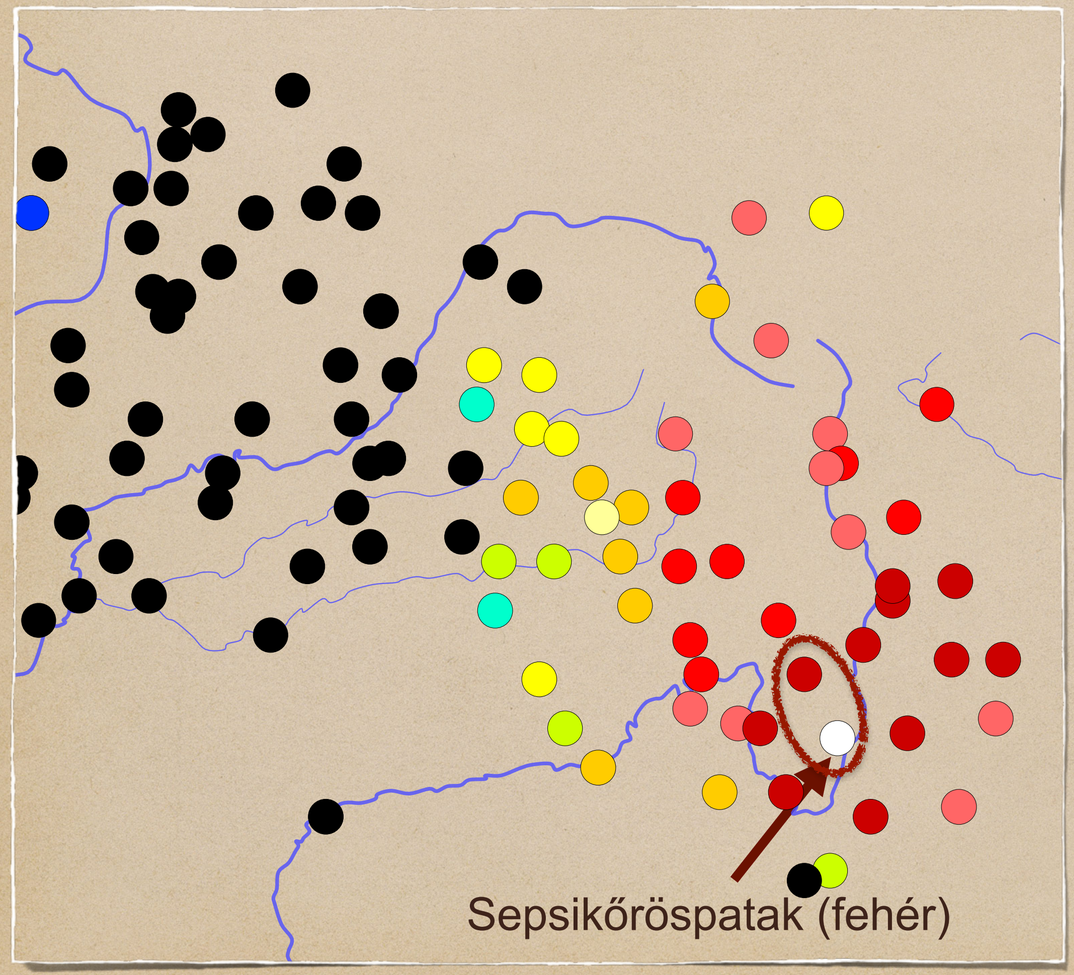
Két szomszédos kutatópont (hasonlóság: 76,5%)
Merőben eltérő nyelv hasonlósági mintázatok

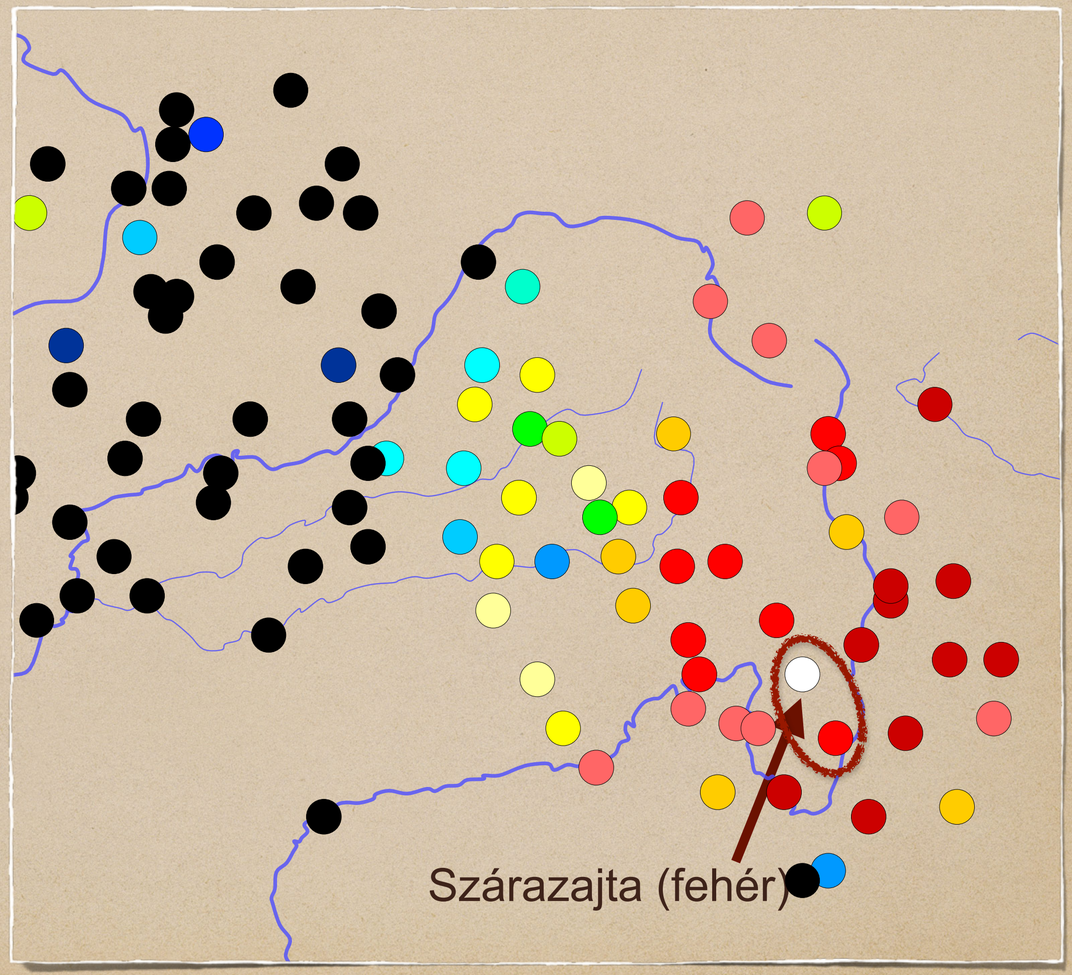


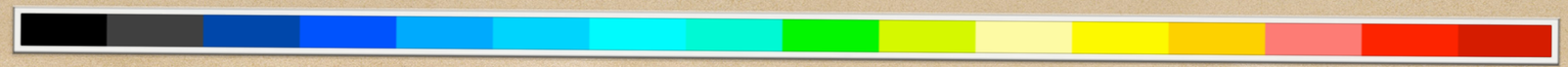




Két szomszédos kutatópont (hasonlóság: 76,1%)
Hasonló mintázatok





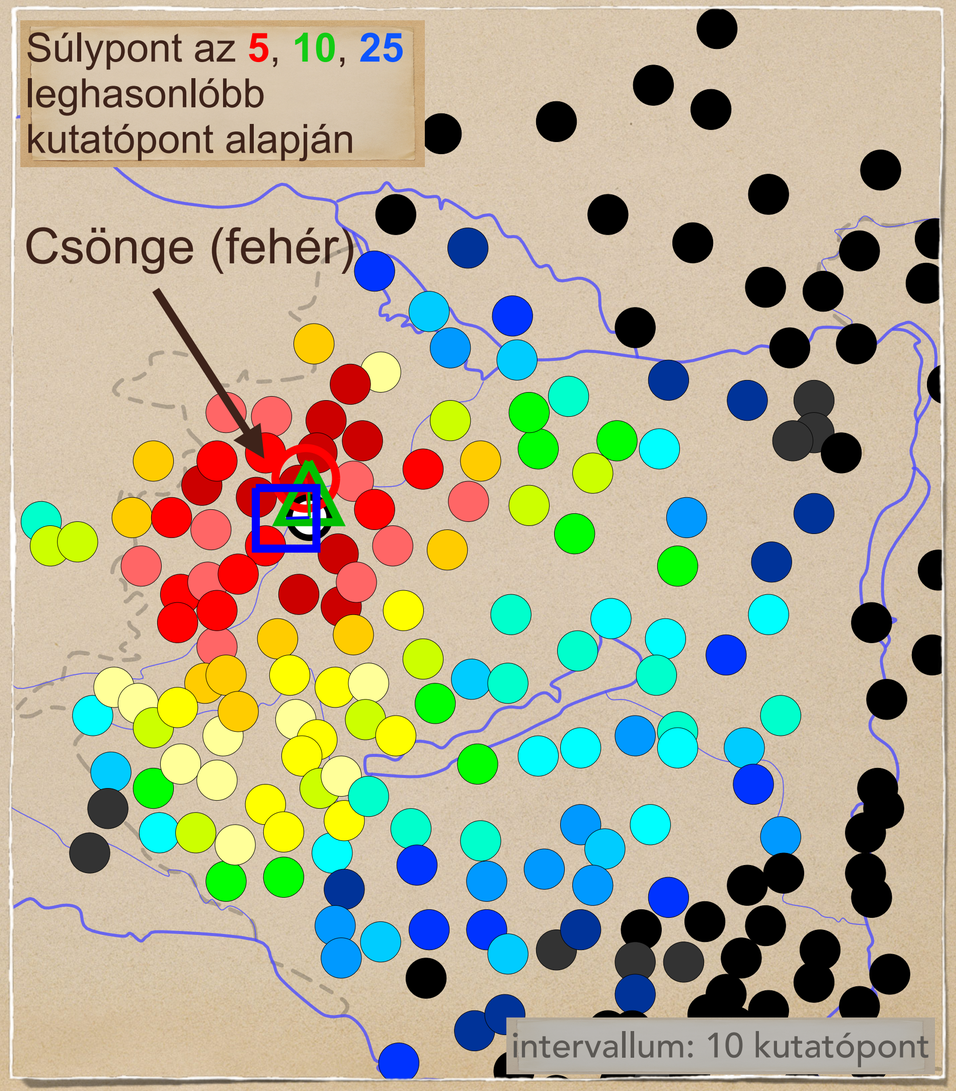


Törésvonal: közeli kutató pontok közti eltérő hasonlósági mintázatok

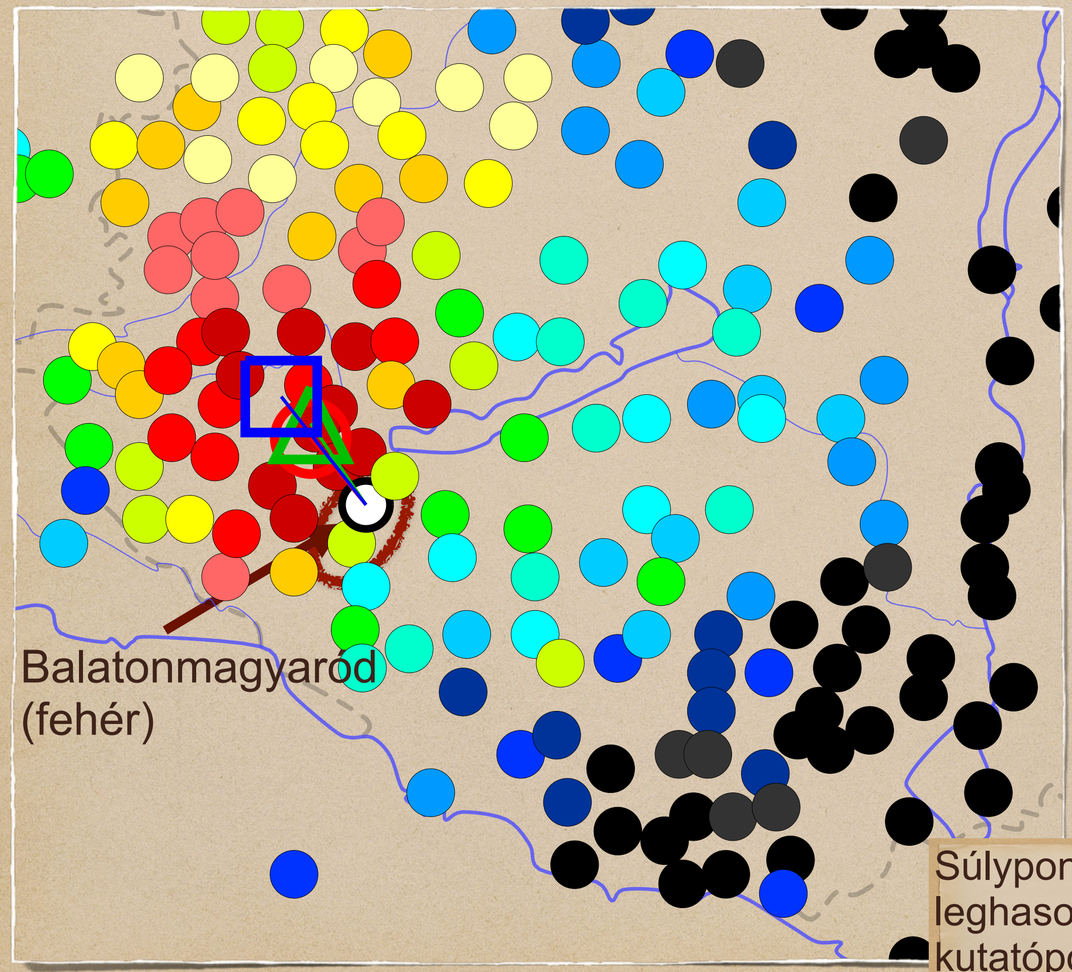
- Nem ott feltételezünk törésvonalat, ahol két kutatópont közötti nyelvi távolság nagyobb.
- Elválasztó vonal feltételezhető ott, **ahol két szomszédos vizsgálati pont nyelvi hasonlósági mintázatai jelentősen eltérnek egymástól**.
- Hogyan lehet egy adott kutatópont nyelvi hasonlósági súlypontját egyetlen pontban lokalizálni?

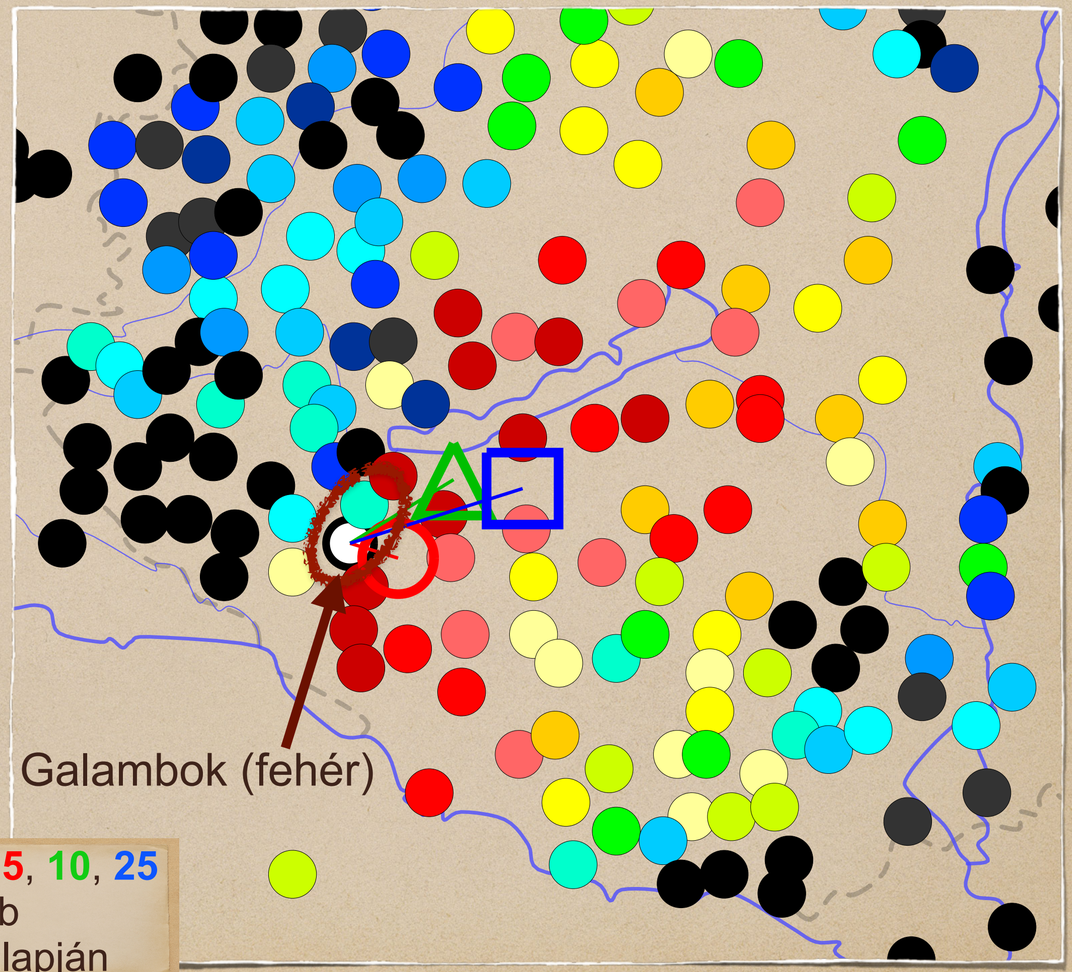
Hogyan határozzuk meg a nyelvi hasonlóság súlypontját?

- A nyelvileg leghasonlóbb kutatópontok földrajzi koordinátáinak a hasonlóság mértéke szerint súlyozott átlaga
- A legtöbb kutatópont nyelvi hasonlósági súlypontja saját földrajzi helyéhez közel esik.

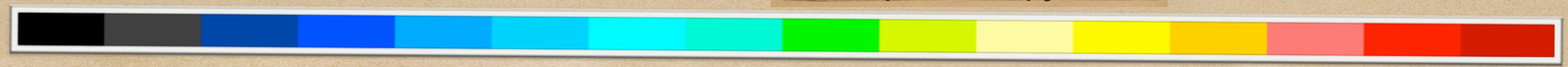


Különböző irányokba mutató nyelvi hasonlósági súlypontok





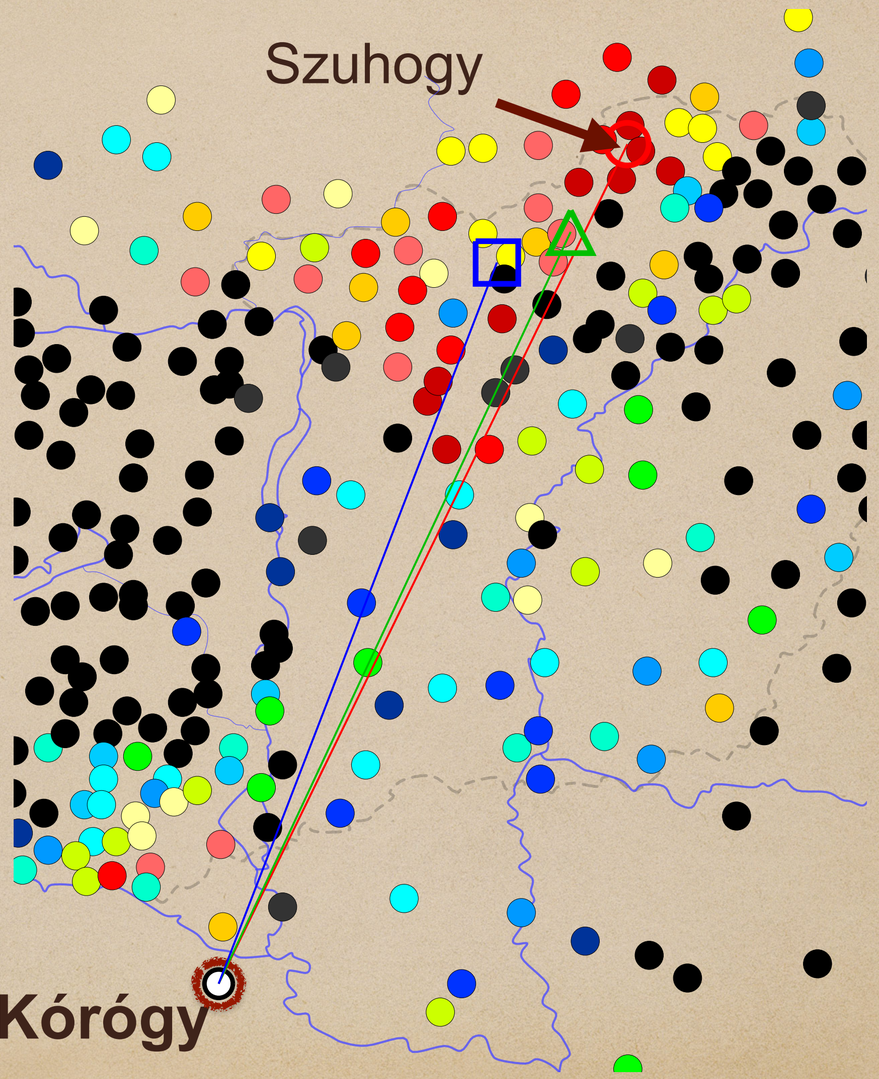
Súlypont az **5**, **10**, **25** leghasonlóbb kutatópont alapján



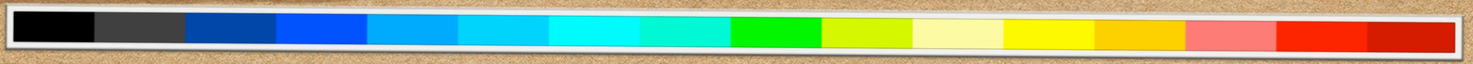
– + intervallum: 10 kutatópont

Nyelvjárásszigetek (Kórógy)

- A nyelvjárásszigeteknek atipikus nyelvi hasonlósági súlypontjuk van
- Kórógy (Vajdaság): a nyelvi hasonlósági súlypont a kibocsátó nyelvjárás szerint Szuhogy közelébe esik

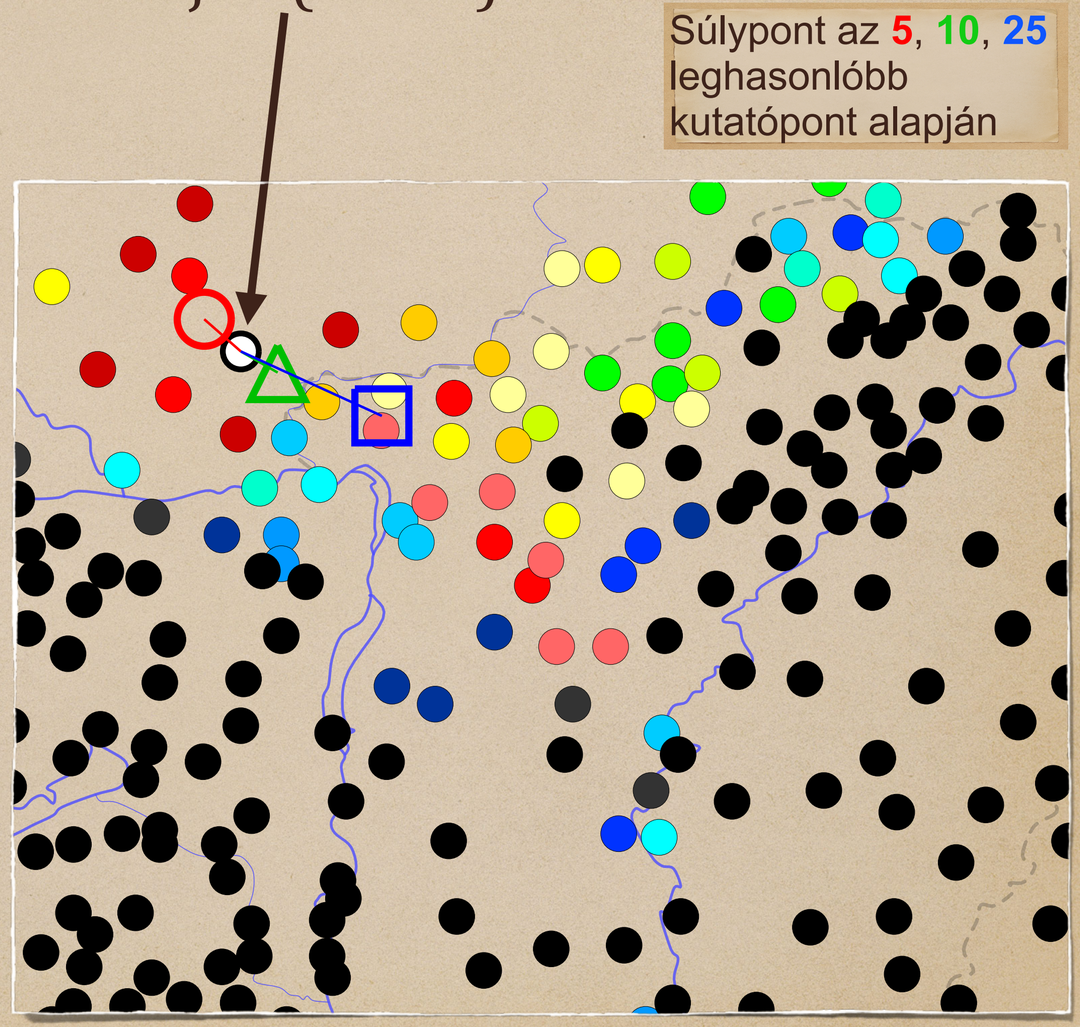


Súlypont az 5, 10, 25 leghasonlóbb kutatópont alapján



A nyelvi hasonlósági súlypont helye függ a figyelembe vett kutatópontok számától: Bajka (white)

- A figyelembe vett hasonló kutatópontok száma alapján eltérő eredményre juthatunk
  - Kevés figyelembe vett kutatópont: a hasonlóság súlypontja egy kirajzolódó kisebb csoportra mutat (nyelvjáráscsoport)
  - Több figyelembe vett kutatópont: a hasonlóság földrajzi súlypontja távolabb eshet, egy nagyobb területre mutat (nyelvjárásterület vagy régió).

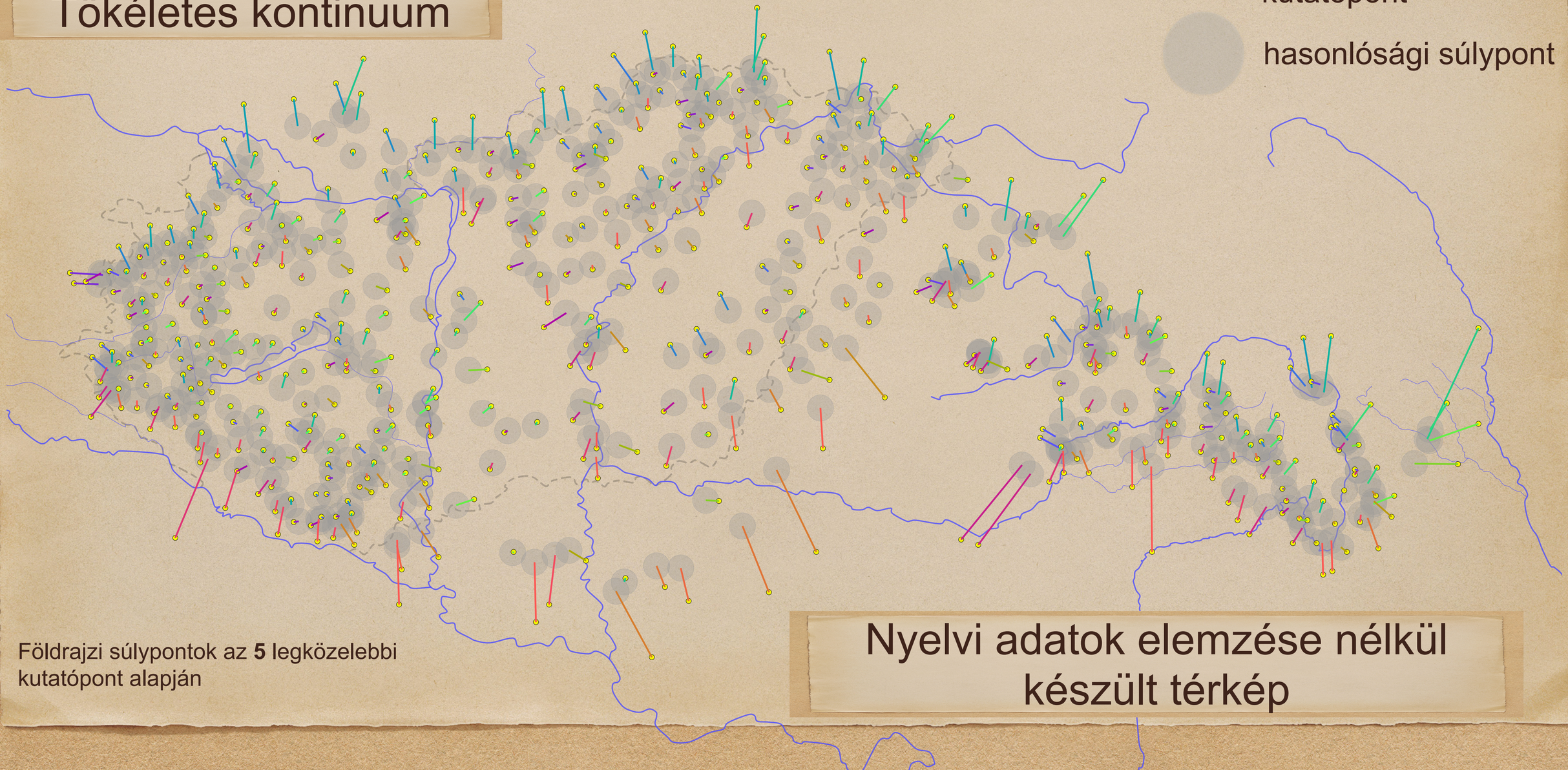


Földrajzi súlypontok (földrajzi távolság) 1.

Tökéletes kontinuum

kutatópont

hasonlósági súlypont

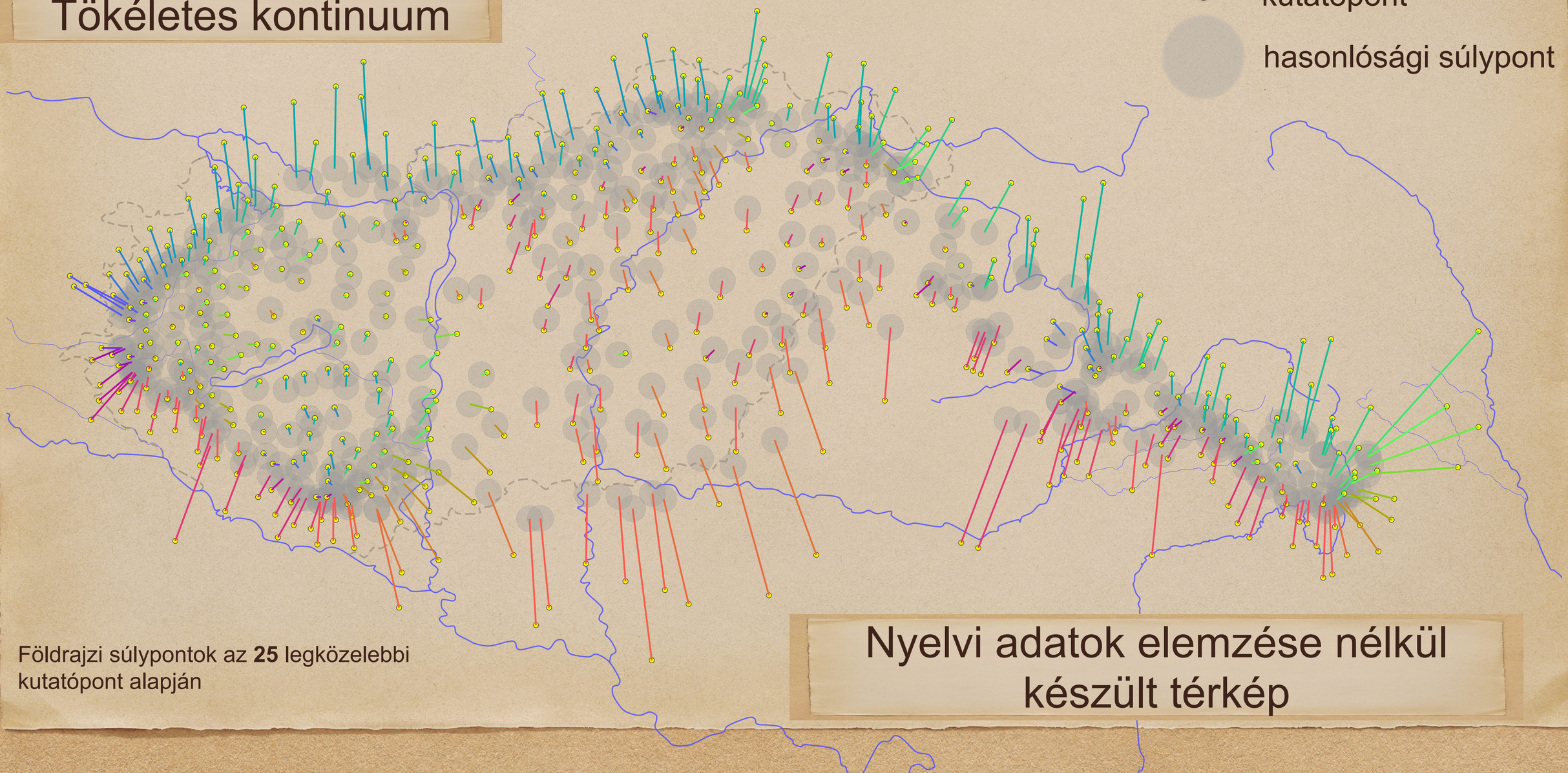


Földrajzi súlypontok az **5** legközelebbi kutatópont alapján

Nyelvi adatok elemzése nélkül készült térkép

Földrajzi súlypontok (földrajzi távolság) 2.

Tökéletes kontinuum



Földrajzi súlypontok az **25** legközelebbi kutatópont alapján

Nyelvi adatok elemzése nélkül készült térkép

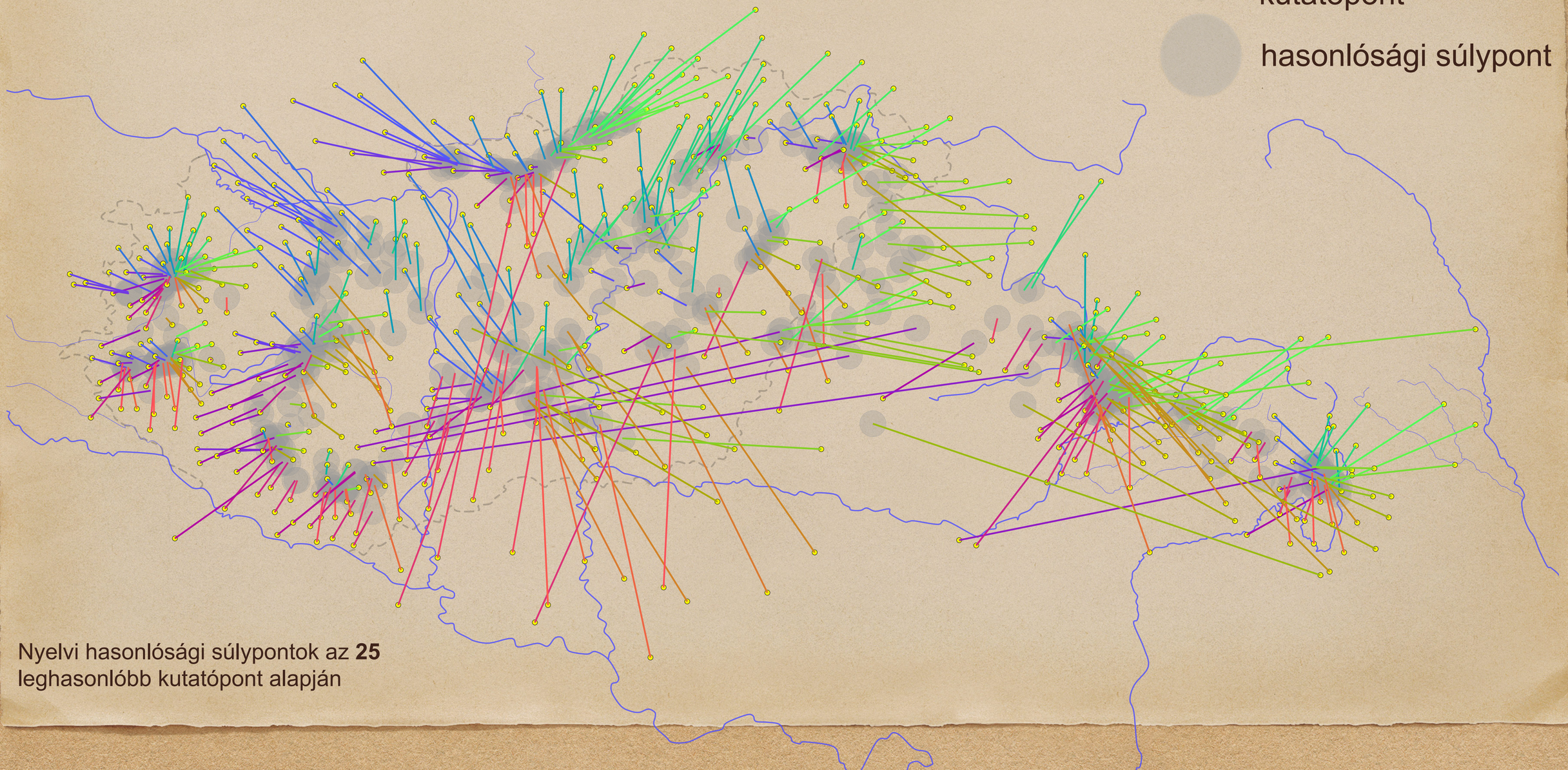
A nyelvi hasonlóság földrajzi súlypontjai 1.

Nyelvi hasonlósági súlypontok az **5** leghasonlóbb kutatópont alapján

A nyelvi hasonlóság földrajzi súlypontjai 2.

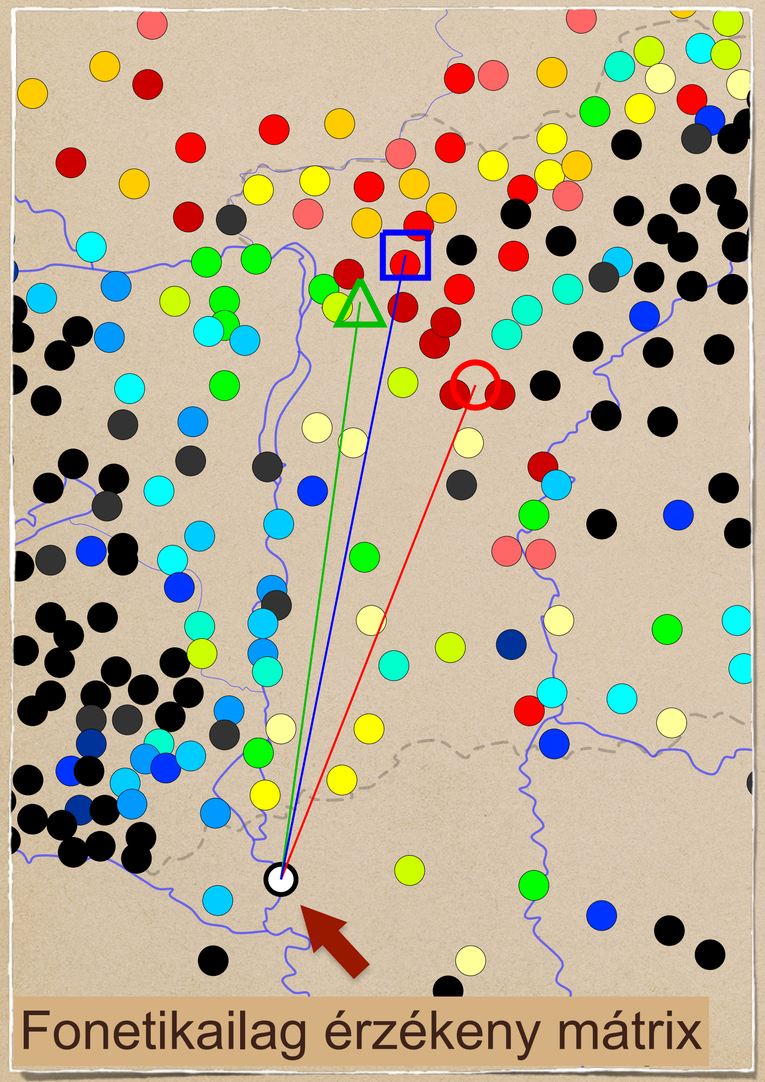
Nyelvi hasonlósági súlypontok az **10** leghasonlóbb kutatópont alapján

A nyelvi hasonlóság földrajzi súlypontjai 3.



Nyelvi hasonlósági súlypontok az **25** leghasonlóbb kutatópont alapján

Nyelvjárásszíget nyelvi hasonlóságának földrajzi súlypontjai a fonetikai részletek figyelembevételének függvényében

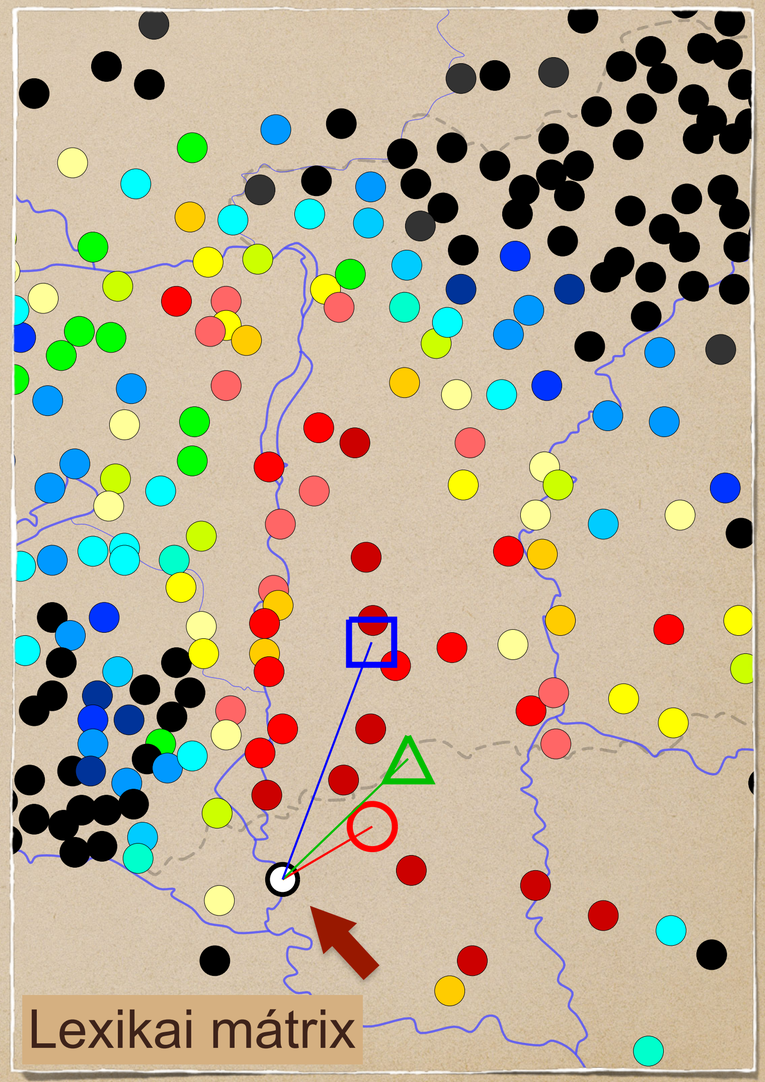


Fonetikailag érzékeny mátrix

Kupuszina (fehér)

- Fonetikailag érzékeny mátrix: a hasonlósági súlypontok a kibocsátó nyelvjárás felé mutatnak (Palóc régió)
- „Lexikai” mátrix: a hasonlósági súlypontok helyét az újabb nyelvi kapcsolatok határozzák meg (közeli kutatópontok)

Súlypont az **5**, **10**, **25** leghasonlóbb kutatópont alapján

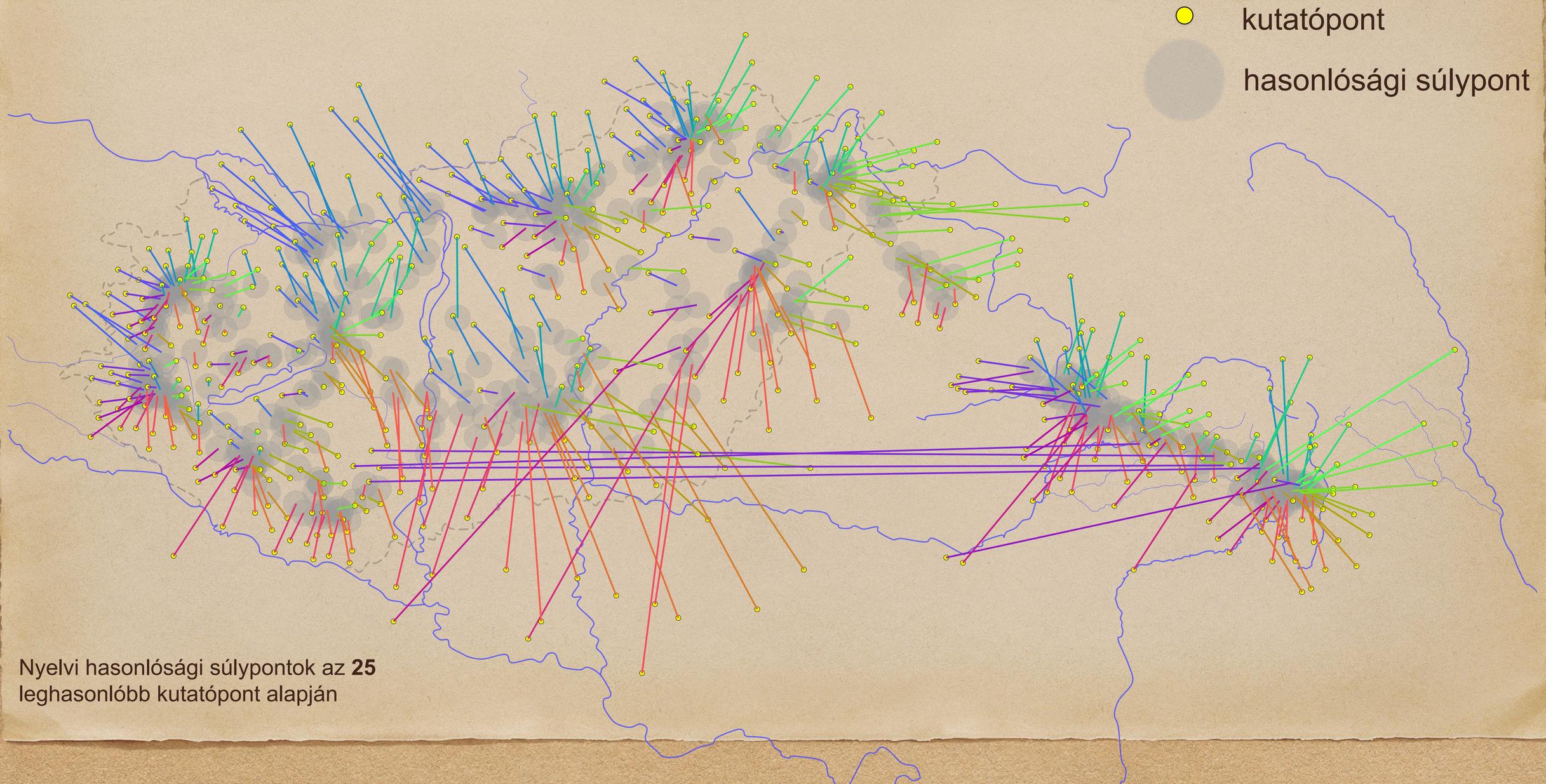


Lexikai mátrix

– +

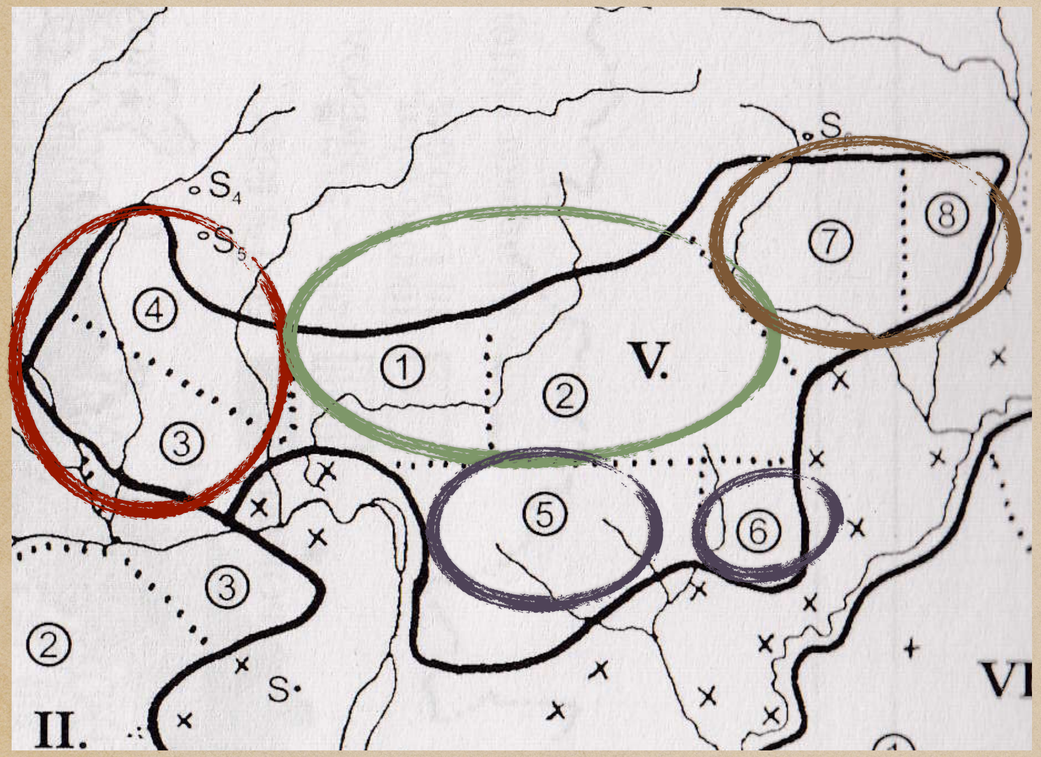
intervallum: 5 kutatópont

Hasonlósági súlypontok („lexikai” mátrix)



Nyelvi hasonlósági súlypontok az **25** leghasonlóbb kutatópont alapján

Palóc nyelvjárások



Juhász 2001.

V. Palóc régió

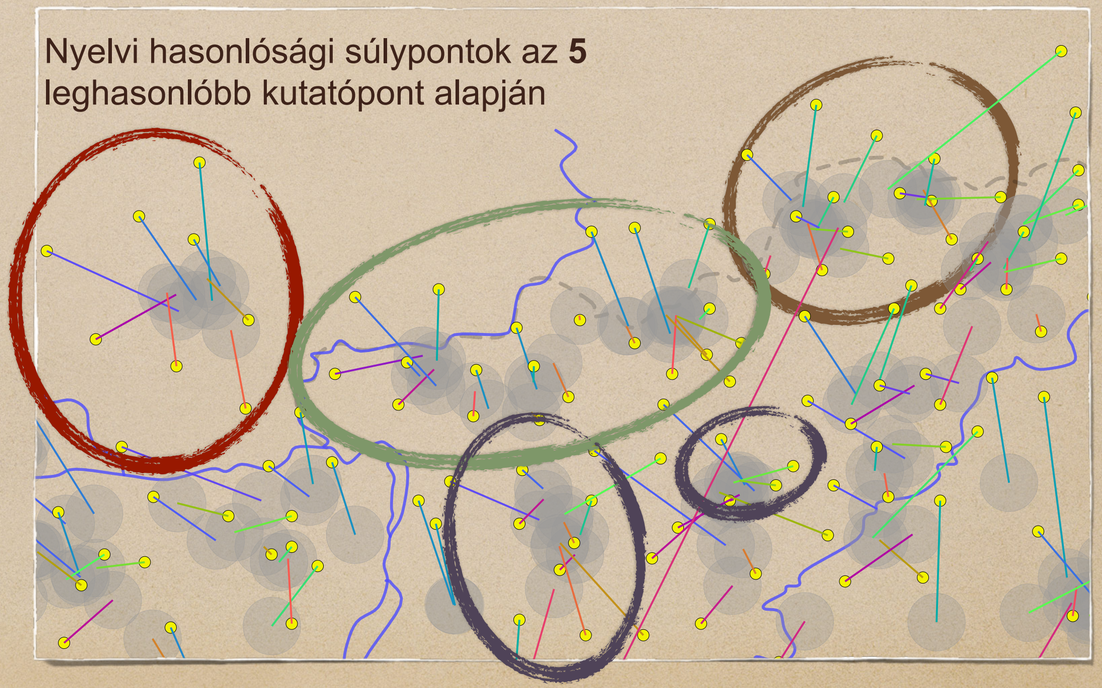
A) Középső tömb
1. Ipoly vidéki csoport
2. Középpalóc csoport

B) Nyugati tömb
3. Nyugati palóc csoport
4. Északnyugati palóc csop.

C) Déli tömb
5. Déli palóc csoport
6. Eger vidéki csoport

D) Keleti tömb
7. Keleti palóc csoport
8. Hernád vidéki csoport

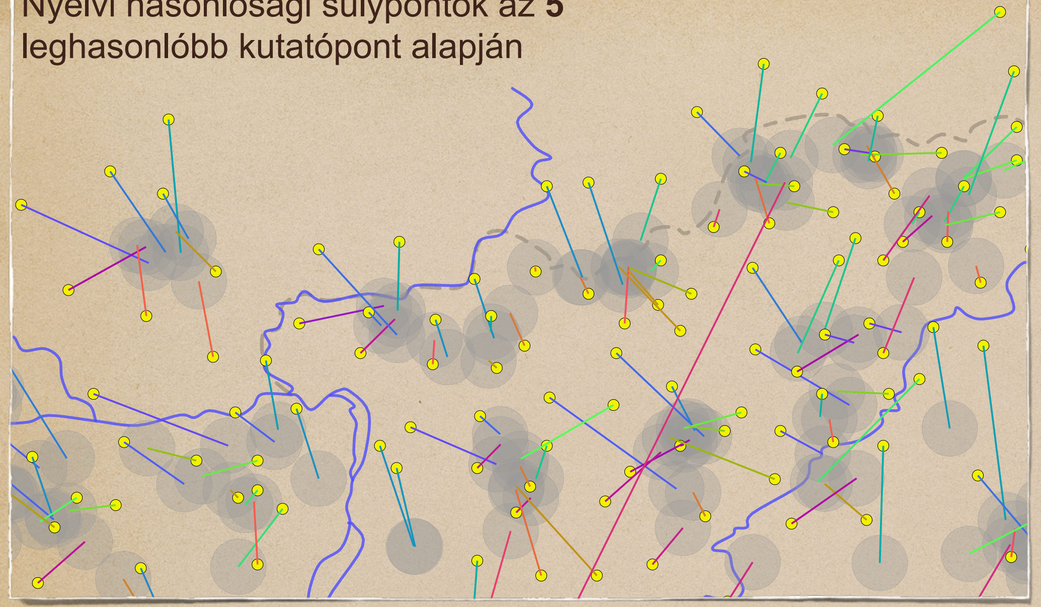
Nyelvi hasonlósági súlypontok az **5** leghasonlóbb kutatópont alapján



Palóc nyelvjárások

- 5 leghasonlóbb kutatópont alapján: kisebb koncentráció, kisebb területek rajzolódnak ki
- 10 leghasonlóbb kutatópont alapján: határozottabb törésvonalak rajzolódnak ki a Palócföld körül
- 25 leghasonlóbb kutatópont alapján: erősebb koncentráció, kirajzolódik a Palóc régió

Nyelvi hasonlósági súlypontok az **5** leghasonlóbb kutatópont alapján

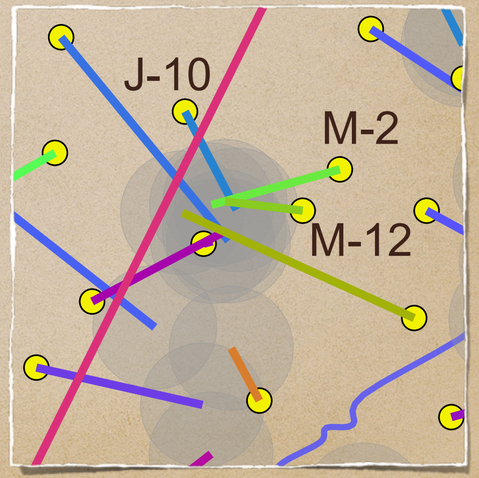


Nyelvi hasonlósági súlypontok a **10** leghasonlóbb kutatópont alapján

Nyelvi hasonlósági súlypontok a **25** leghasonlóbb kutatópont alapján

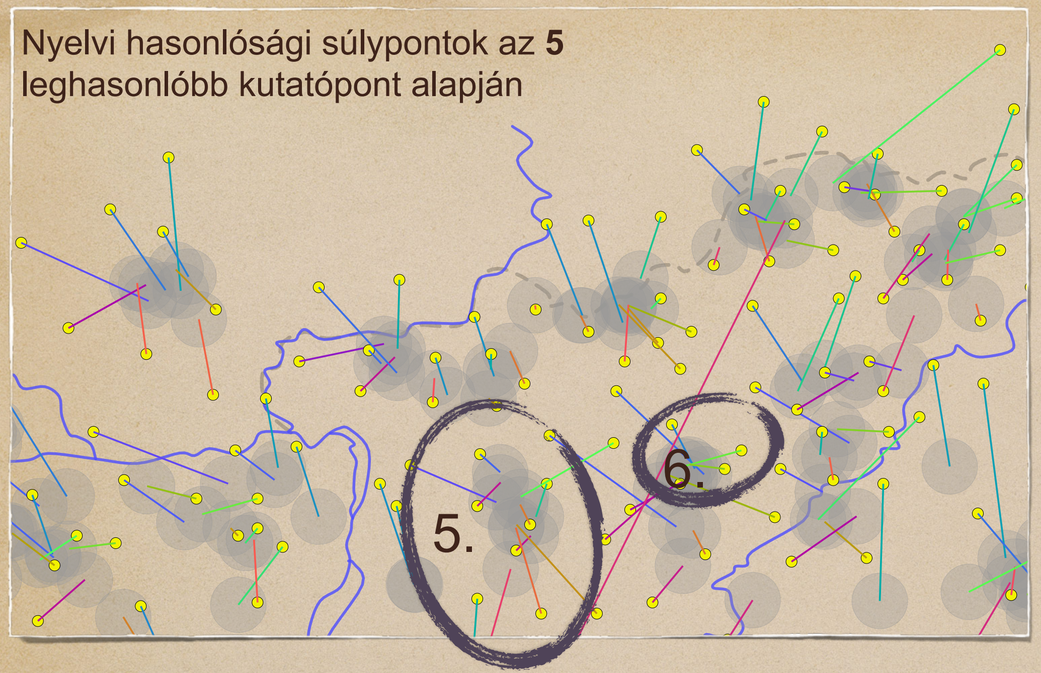
Imre Samu az Eger vidéki 6. csoportról

- „Az Eger környéki nyelvjárástípus. —Az Eger től délre-dny.-ra elterülő kisebb vidék nyelvjárása. Legtöbb vonásában egyezik a középpalóc típussal, esetleg annak keleti változatával. […] A legfontosabb eltérés, hogy az *a* fonéma labiális *(a)*, az *á* illabiális *(á)* megoldású.” (Imre 1971: 359–360)
- Imre elemzésében szereplő kutatópontok:
  - Verpelét (J-10)
  - Dormánd (M-12)
  - Szihalom (M-2)



Palóc nyelvjárások

Nyelvi hasonlósági súlypontok az **5** leghasonlóbb kutatópont alapján

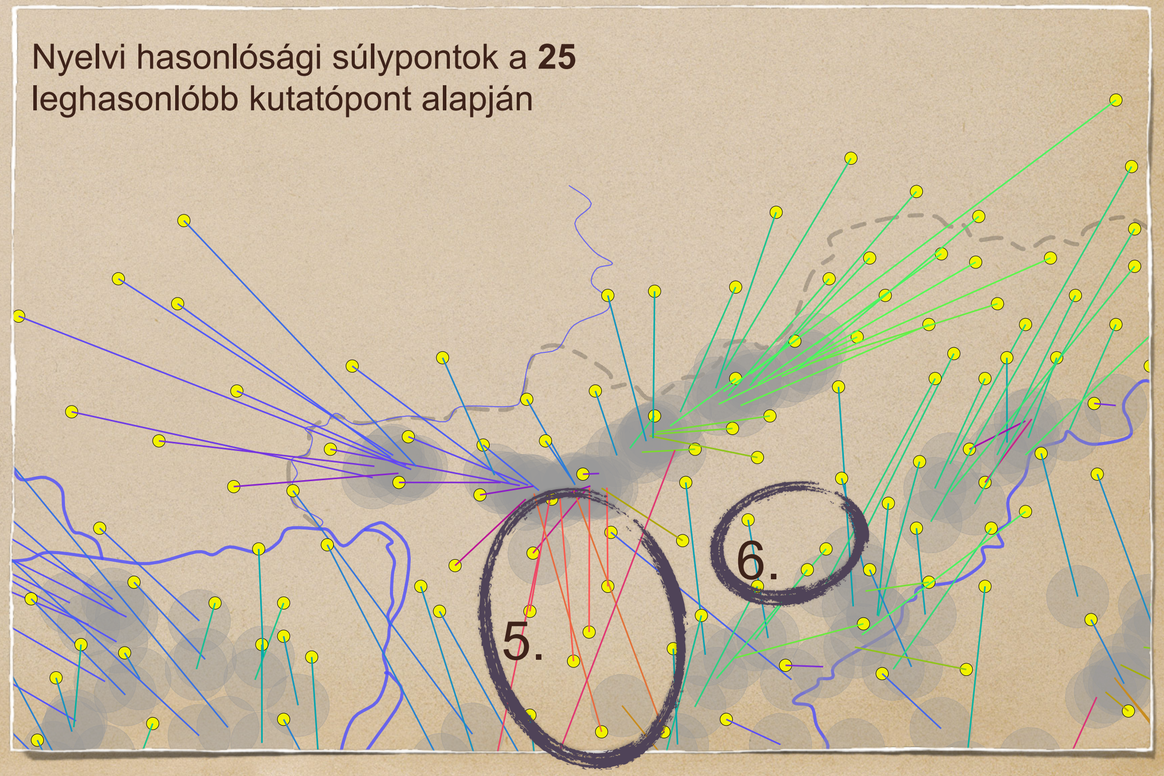


Nyelvi hasonlósági súlypontok a **10** leghasonlóbb kutatópont alapján

Déli tömb

5. Déli palóc csoport
6. Eger vidéki csoport

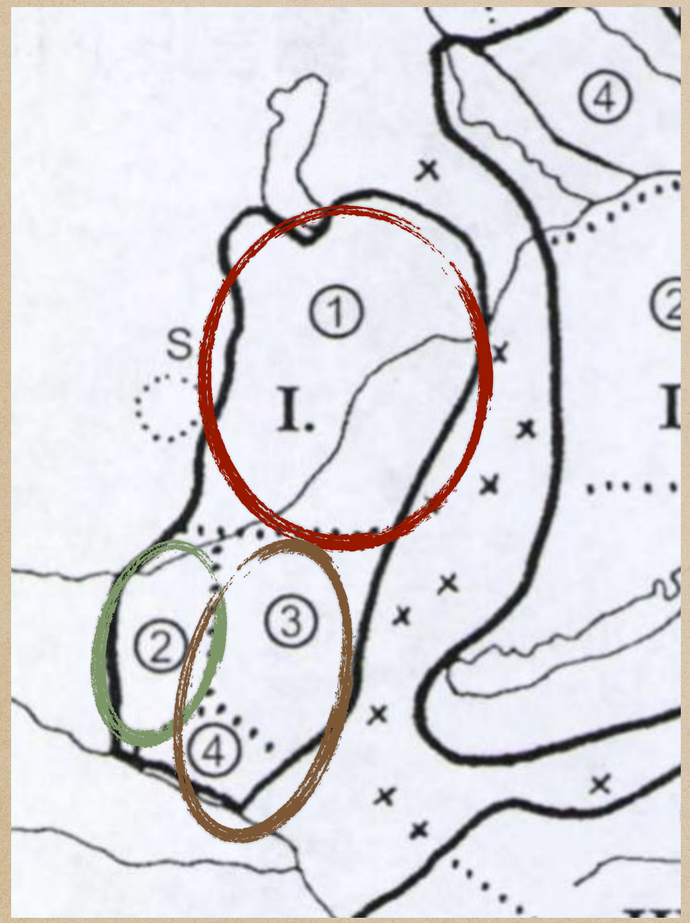
Nyelvi hasonlósági súlypontok a **25** leghasonlóbb kutatópont alapján



Nyugat–dunántúli régió

kutatópont

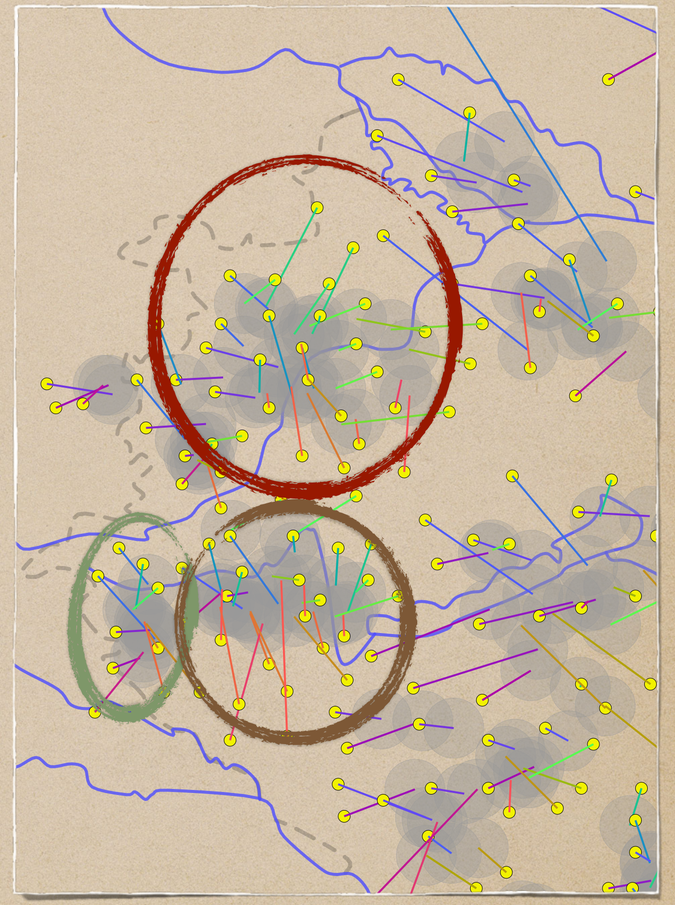
hasonlósági súlypont



Juhász 2001.

I. Nyugat-dunántúli régió

1. Északnyugat-dunántúli csoport
2. Őrségi csoport
3. Zalai csoport
4. Hetési csoport

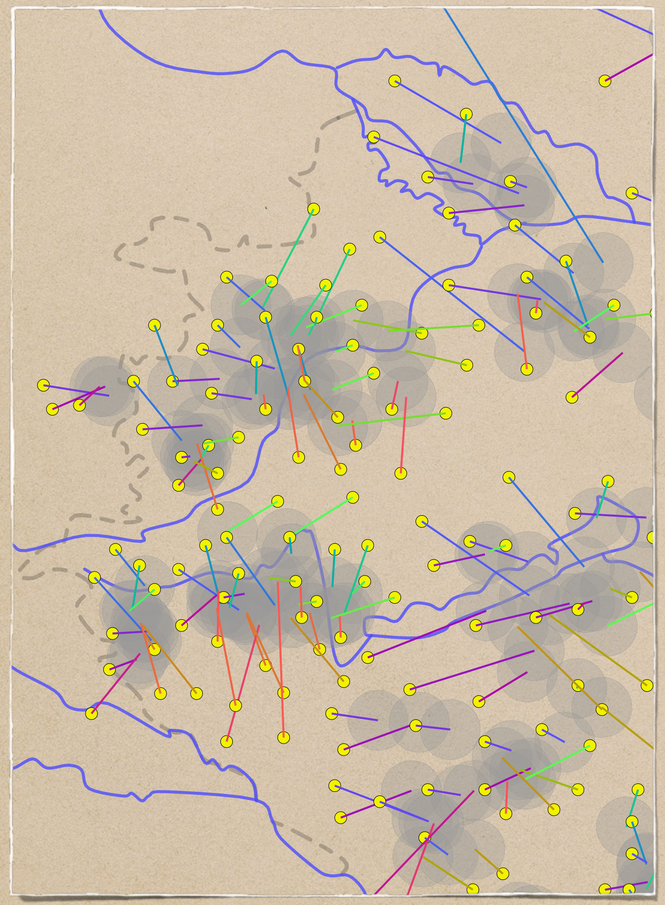


Nyelvi hasonlósági súlypontok az **5** leghasonlóbb kutatópont alapján

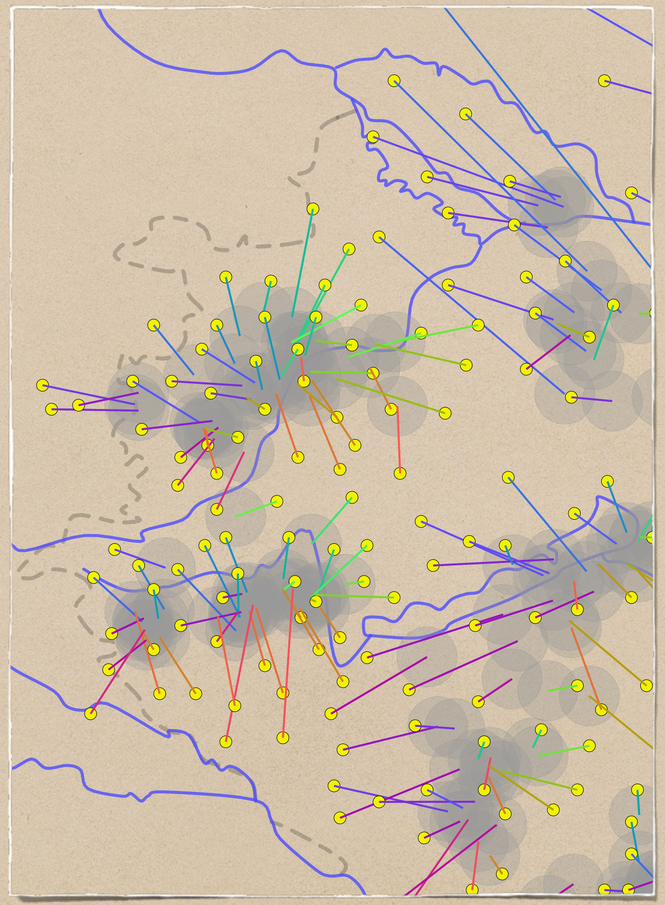
Nyugat-Dunántúl

kutatópont

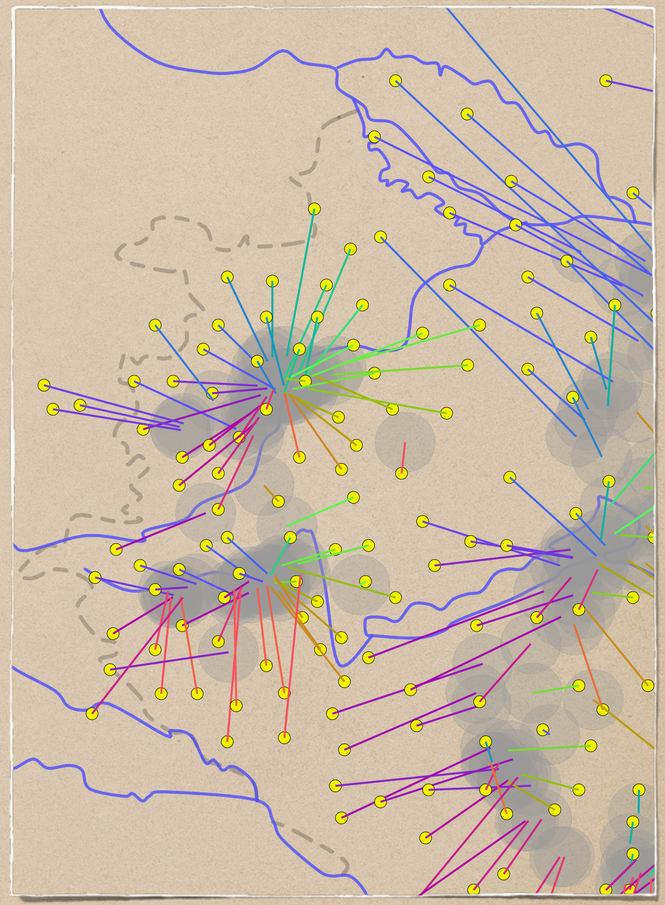
hasonlósági súlypont



Nyelvi hasonlósági súlypontok az **5** leghasonlóbb kutatópont alapján



Nyelvi hasonlósági súlypontok a **10** leghasonlóbb kutatópont alapján

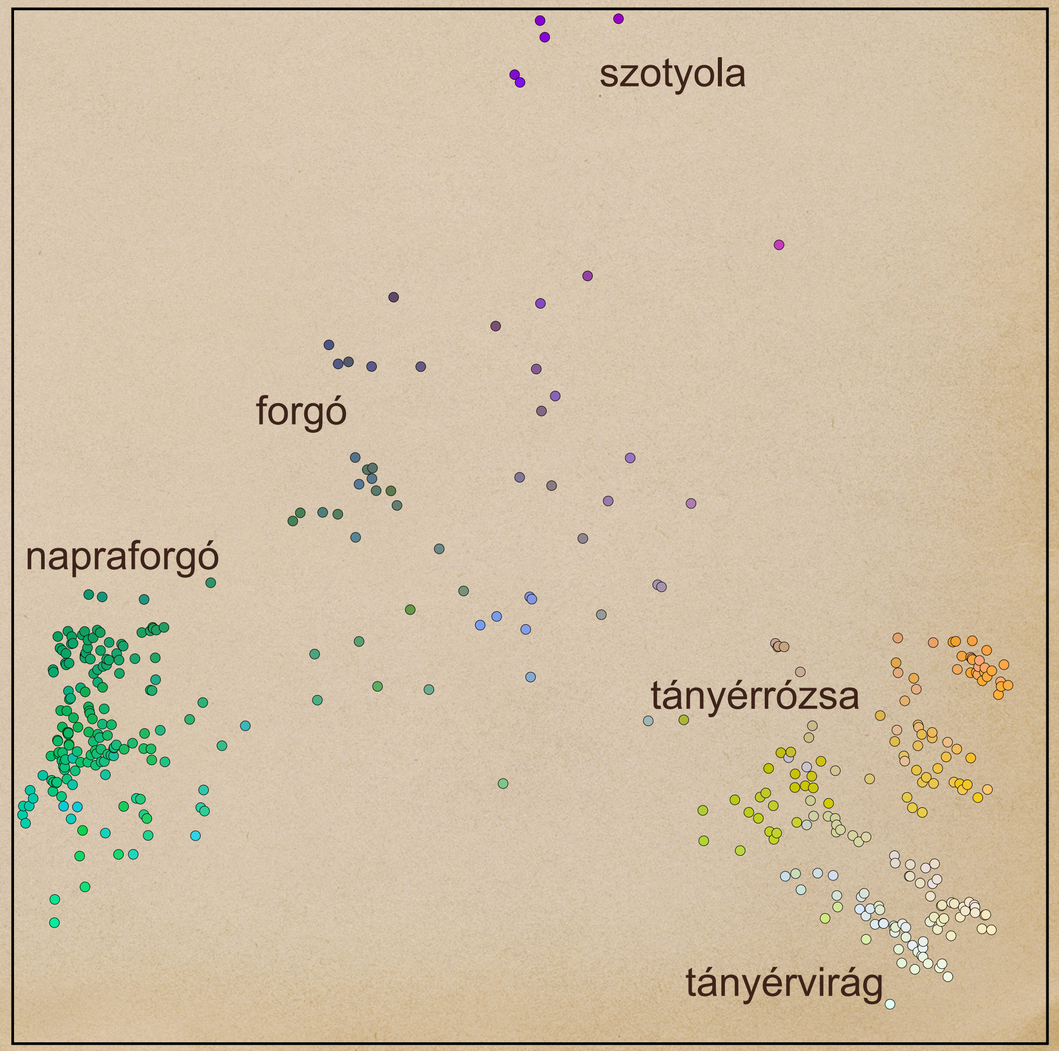


Nyelvi hasonlósági súlypontok a **25** leghasonlóbb kutatópont alapján

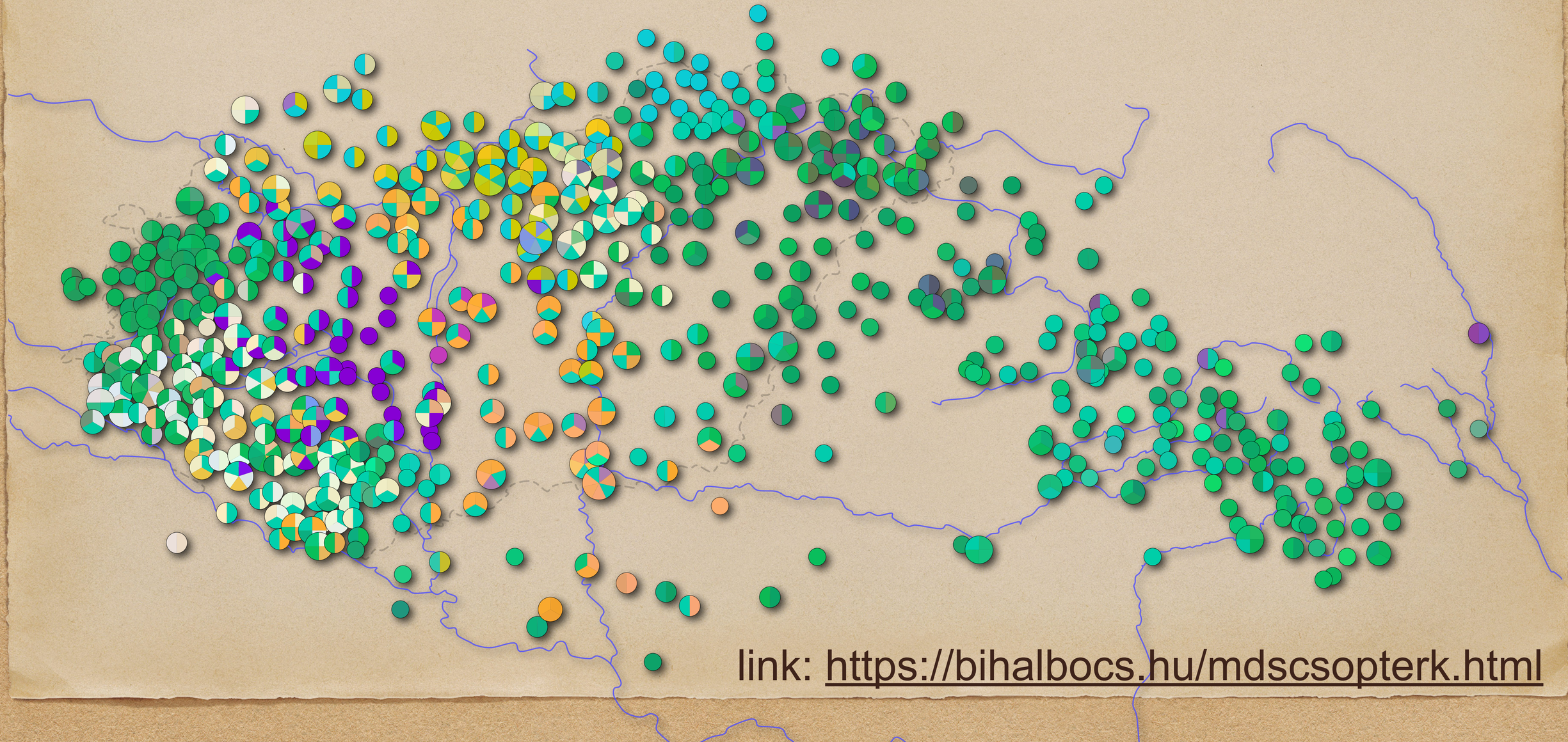
Egy térképlap adatainak csoportosítása többdimenziós skálázással

Adatmátrix és vizualizációja

	napraforgu͡o	naprafòrgo͡u	tányírvirág	tányérrúzsa
napraforgu͡o	100%	78,6%	7,7%	7,7%
naprafòrgo͡u	78,6%	100%	7,2%	7,2%
tányírvirág	7,7%	7,2%	100%	40%
tányérrúzsa	7,7%	7,2%	40%	100%



Napraforgó automatikus csoportosítás



link: https://bihalbocs.hu/mdscsopterk.html

Konklúzió

- A dialektometria mint kvantitatív nyelvföldrajzi módszer felfogható a 20. sz. második felében elindult, de megfelelő adatkezelési módszerek hiányában elakadt nyelvföldrajzi kutatások folytatása, újraélesztése, kiteljesedéseként.
- Elemzési lehetőségek a dialektometriai vizsgálatokban:
  - **egy kutatópont nyelvi hasonlósági mintázatai**nak kiszámítása, térbeli ábrázolása
  - aggregált adatokon alapuló **automatikus csoportosítás (klasszifikáció)** klaszteranalízis segítségével
  - a **nyelvi kontinuum vizualizálása** többdimenziós skálázás segítségével
- A földrajzi információ elemzésbe való bevonásával lehetőségünk van:
  - **nyelvi hasonlóság földrajzi súlypontjainak** azonosítására
  - **nyelvjárási törésvonalak** azonosítására azokon a területeken, ahol szomszédos kutatópontok nyelvi hasonlósági súlypontjai eltérő irányokba mutatnak
  - korábbi klasszifikációk hitelesítése vagy pontosítása
- A nyelvi hasonlóság földrajzi súlypontjainak helyét alapvetően meghatározza
  - a súlypont kiszámításához figyelembe vett leghasonlóbb kutatópontok száma
  - a hasonlósági mátrix hátterében lévő adatok fonetikai pontossága
- Többdimenziós skálázás: **egyetlen térkép adatainak objektív csoportosítása**, egy-egy jelenség nyelvjárási sokszínűségének ábrázolása

Köszönöm a figyelmet!

Az előadás elkészítését az NKFIH FK 138396 számú, *Magyar nyelvjárások akusztikai és dialektometriai vizsgálata* c. projektje támogatta.