



Vokalizmus és konsonantizmus a magyar nyelvjárások hasonlósági mintázataiban¹

Vargha Fruzsina Sára

Nyelvtudományi Kutatóközpont

vargha.fruzsina@nytud.hu

Kivonat: A hangtani jelenségek döntő szerepet játszanak a magyar nyelvjárások közötti kapcsolatrendszerek feltárásában, kérdés azonban, hogy e jelenségek mely kategóriái a leginkább meghatározók. Dolgozatomban a vokalizmus és a konsonantizmus hatását vizsgálom a nyelvjárások közti hasonlóság alakulására, illetve a nyelvjárások egymás közti hasonlóságán alapuló csoportosításokra. A kutatásomat informatizált nyelvjárási atlaszok adatain végzem, a korábbi, más nyelveken végzett korrelációs dialektometriai elemzésekből inspirálódva, ám az adatok előzetes manuális osztályozása nélkül, automatizált elemzési módszerek kidolgozásával, tehát nagyobb hatékonyságra és objektivitásra törekedve. Az elemzések eredményét térképeken szemléltetem, megjelenítve a konsonantizmus és a vokalizmus alapján meghatározott nyelvi hasonlóság térbeli mintázatait. Amellett érvelek, hogy elsősorban a magánhangzó-minőségek bizonyulnak meghatározónak a nyelvjárások térbeli mintázatainak alakulásában.

Kulcsszavak: kvantitatív nyelvföldrajz; dialektometria; nyelvjárások klasszifikációja; vokalizmus; konsonantizmus

1. Bevezetés

Nyelvatlaszaink a 20. században azzal a céllal jöttek létre, hogy reprezentatív forrásanyagot biztosítsanak a magyar nyelvjárások tanulmányozásához, a magyar nyelv térbeli jellegzetességeinek feltárásához, továbbá – a jellegzetességek alapján – nyelvjárási alakulatok elkülönítéséhez. Felhasználásuk a dialektológiai munkákban azonban jobbra megállt egy-egy jelenség szemléltetésénél, vagy néhány „jól kiválasztott” („kimazsolázott”) térképlap felhasználásánál valamely hipotézis alátámasztására. Az egyetlen nagy ívű, kvantitatív szemléletű munka, amely A magyar nyelvjárások atlasza (Deme–Imre 1968–1977, a továbbiakban MNyA.) alapján vállalkozott elsősorban egyes magánhangzók ejtésében megmutatózó térbeli mintázatok megragadására, Imre Samu monográfiája (1971).

¹ A dolgozat az FK-138396 számú NKFI-projekt keretében készült. Köszönöm a két anonim lektornak és a szerkesztőknek a dolgozat korábbi változatához fűzött értékes tanácsaikat, észrevételeiket.



A számítógépes dialektológia elindulása által vált lehetővé, hogy a sok évtizedes munkával összegyűjtött nyelvatlaszadatokat megfelelően rögzítsük, és az ezeket kezelő nyelvészeti technológiák létrehozásával és használatával számításintenzív műveleteket is végezhesünk rajtuk. A dialektometria mint kutatási módszer révén olyan elemzéseket is készíthetünk, amelyekben az egyes jelenségek tekintetében megmutatkozó hasonlóság mértéke összeadódik, s így egyetlen, aggregált adatként mutatja meg a kutatópontpárok közti általános nyelvi hasonlóságot, illetve különbözőséget (Goebel 2005; 2010; Nerbonne et al. 1996; Vargha 2017). A MNyA. dialektometriai elemzése abbahozó irányba mutatnak, hogy a nyelvi hasonlóság térbeli mintázatainak kialakításában a hangtannak, az adatok lejegyzésében is tükröződő fonetikai tényeknek van kulcsszerepe (Vargha 2015; 2017, 49–66).

Jelen dolgozatban arra vállalkozom, hogy a MNyA., A romániai magyar nyelvjárások atlasza (Murádin–Juhász 1995–2010, a továbbiakban RMNyA.) és A moldvai csángó nyelvjárás atlasza (Gálffy et al. 1991, a továbbiakban MCsNyA.) adatai alapján megnézzem, a fonetikai jellegzetességeken belül a magánhangzók, illetve a mássalhangzók hogyan és milyen mértékben határozzák meg a nyelvjárások közti hasonlóság földrajzi mintázatait. Előzetes elgondolásom szerint a magánhangzók játszanak nagyobb szerepet a nyelvjárások térbeli mintázatainak alakításában, mivel a magánhangzó-minőségek rendszerén belül erős a kölcsönhatás a rendszerlemek között, továbbá nagyobb az ejtésbeli variabilitás, mint a mássalhangzók esetében. Vizsgálataimat az informatizált (számítógépen megfelelően rögzített, l. Vékás 2007) nyelvjárási adatok automatikus elemzési módszereinek megteremtésével, előzetes kutatói csoportosítás nélkül, az adatok automatikus összevetésével végzem.

1.1. Atlaszadatok fonetikus lejegyzése és kvantitatív nyelvföldrajz

A dialektológiában a kezdetektől jelen van a törekvés a nyelvi adatok fonetikaikailag minél pontosabb lejegyzésére, hiszen az egyes nyelvjárások közötti eltérések gyakran az egyes hangárnyalatok közötti akár apró, ám következetesen érzékelt különbségekben jelentkeznek. A magyar nyelvatlaszok nyelvi adatait magyar egyezményes hangjelöléssel rögzítették, és ezt a jelölő rendszert használhatjuk nyelvjárási szövegek lejegyzéséhez, de akár újabb nyelvatlaszok vagy más nyelvjárási adattárak létrehozásához is. A lejegyzett adatok fonetikai szempontú vizsgálatakor fontos figyelembe venni, hogy milyen jellegű ez





a fajta hangjelölés, mennyire koherens a lejegyzés gyakorlata, és mennyire tekinthetők megbízhatónak az ebben a rendszerben rögzített adatok egyenként nézve, illetve kvantitatív keretben feldolgozva. Az alábbiakban mindenekelőtt ezt a kérdéskört járjuk körbe röviden, a teljesség igénye nélkül, néhány példát is bemutatva, ugyanis a nyelvi adatok értékét növeli, felhasználhatóságukat erősíti, ha tisztában vagyunk a lejegyzés főbb erősségeivel és gyenge pontjaival egyaránt.

Az egyezményes hangjelölés fonetikus jellegű: nemcsak alapjel kiválasztásával ragadhatunk meg egy szóban forgó hangárnyalatot, a fonetikai részleteket a képzés helyét, módját vagy időtartamát módosító mellékjelek használatával pontosíthatjuk. A kifejezett fonetikus jelleget mutatja, hogy még azonos fonémához rendelhető hangárnyalatokat is jelölhetünk különböző alapjellel, még azonos nyelvjáráson belül is. Így például az *e* magánhangzó különböző árnyalatainak megragadására alkalmasak – többek között – a következő fonetikus szimbólumok: *ɛ*, *ä*, *ä*.

Az egyezményes hangjelölés az igen különböző, gyakran térbeliséget is mutató hangárnyalatok lejegyzésére alkalmas, tehát az akár igen apró, ám jellegzetesnek gondolt hangárnyalatok megkülönböztetésének igénye hozta létre a jelölés rendszerét. Az egyes hangminőségek jelölésmódja következetes, ugyanis a szimbólumok (akár alapjel-mellékjel kombinációk) használatát a lejegyzés kialakult gyakorlata pontosítja.

Az alapjelekkel (például: *a*, *á*, *é*, *ty*) kifejezett viszonyítási minőségek viszonylag szilárdak, a 20. század közepén a nyelvatlászokon dolgozó nyelvészek körében köznyelvi ejtésűnek ítélt hangzóminőségek alkotják a hangjelölés gerincét. Így például az *a* szimbólum egyszerűen és megbízhatóan utal az *a* magánhangzó köznyelvinek tartott ejtés módjára (például az *alma* szóban), a mellékjellel nyíltabbnak (*a*) vagy zártabbnak (*ɑ*) jelölt változat esetében is egészen egyértelmű, hogy milyen viszonyítási hangzóminőséghez képest van apró eltérés a nyelvtálasztásban. (Ha valamilyen megfontolásból az IPA szimbólumaival akarnánk leírni valamelyik *a*-féleséget, sokkal bizonytalanabb lenne, milyen hangértékre utalhatunk például az [ɔ] szimbólummal, különösen azért, mert az IPA esetében a magyar nyelvváltozatok vonatkozásában nincs kikristályosodott lejegyzői gyakorlat, amely eligazító hatású lenne.)

Nem szükséges elméleti síkon is vitatható fonetikai definíciót tulajdonítanunk a jelölt hangoknak: a *gyár* szó első mássalhangzóját a *gy* fonetikus szimbólummal jelöljük, de ebből nem következik, hogy azt zárhangnak vagy affrikátának tekintjük. Hasonlóképpen a *dzsem* szó első mássalhangzójának fonetikus jele a *dzs*, és a hangjelölés gazdag gyakorlata a *gy* és *dzs* között két köztes hangár-





nyalat jelölési módját is meghatározza (*gy*, *dzs*), miként például a *gy* és *j* közti átmenetek jelölését is (*gy*, *j*). Ezek a szimbólumok lényegében ugyanazokra a hangárnyalatokra utalnak minden képzett dialektológus számára, függetlenül attól, hogy az illető a *j* hangot approximánsnak vagy netán zöngés réshangnak tekinti. A hangjelölés lehetőségét biztosít például *l* és *r* közötti átmenet megragadására is, méghozzá azért, mert az ilyen hangárnyalatok egy-egy területre jellemzők lehetnek. A magyar dialektológia igényeire választ adó egyezményes hangjelölés tehát számos lehetőséget biztosít, némelyiket nehézkes vagy erőltetett volna egy olyan jelölő rendszerben megoldani, amelyet nem kifejezetten a magyar nyelvváltozatok lejegyzésére alkottak meg.

A köznyelvszerű *á*-nál kissé hátrább, kissé labiálisabban és kissé magasabb nyelvháttal ejtett magánhangzó-árnyalatot (amelynek, ha akusztikailag vizsgáljuk, az első és különösen a második formánsát is alacsonyabbnak találjuk) a már évtizedekkel ezelőtt kiforrott gyakorlat szerint egyetlen, látszólag a nyelvállást módosító mellékjellel szokás jelölni (*ǎ*). Korábban az enyhe labializáltságra utaló mellékjel használata is előfordult, különösen az erdélyi hagyományban, ám a domináns gyakorlat ezt teljesen kiszorította, így a nyelvállásra utaló mellékjelek váltak a legszokásosabbá a magánhangzó-árnyalatok jelölésekor. Az elméletileg rendelkezésre álló alapjel-mellékjel kombinációknak tehát csak egy részét használhatjuk, ha meg akarunk felelni a hangjelölés kialakult gyakorlatának.

A fentiek értelmében elmondhatjuk, hogy az egyezményes lejegyzésnek – különösen az apró fonetikai különbségek megragadásában – komoly erősségei vannak, amelyek elsősorban a viszonyítási hangminőségek megnyugtató egyértelműségéből és a lejegyzésnek a gyakorlatban kikristályosodott normáiból következnek. Mindez azt is jelenti, hogy elviekben nagy a koherencia az egyezményes hangjelölés használatában, legalábbis világos az erre való törekvés. Szükséges ugyanakkor egy-két példa erejéig kitérni következetlenségekre is, amelyek a lejegyzési gyakorlat gyenge pontjaira világítanak rá.

Noha a magánhangzók lejegyzésénél a MNyA. készítői nagyfokú fonetikai pontosságra törekedtek, az *ĕ* (a dialektológiában standard elnevezéssel: „zárt *ĕ*”) hang jelölése nem mutat változatosságot az atlaszban. Annak ellenére sem, hogy hangfelvételeket vizsgálva az *ĕ* nyíltabb realizációi a Dunántúl nyugati részén egészen általánosnak tűnnek (vö. Vargha 2013; 2018), helyenként nyíltságban a köznyelvszerű *e*-vel vetekednek – miközben az *ĕ*, egyezményes fonetikai minősége szerint, csak egy árnyalattal lenne nyíltabb, mint az *é*. Az atlasz lejegyzési gyakorlatának erre a furcsaságára már Imre Samu is felfigyelt (Imre 1971, 272), hiszen a zárt *ĕ* esetében nem tudott sem kvantitatív elemzést, sem térképet készíteni azokról az ejtésbeli különbségekről, amelyekről a gyűjtőmunka során



tapasztalatokat szerzett. A zárt *ĕ* hangok előfordulásai a MNyA. lejegyzéseiben tehát inkább csak arra utalnak, hogy egy adott kutatóponton megkülönböztetnek kétféle *e*-szerű hangot, de nem alkalmasak arra, hogy következtetni tudjunk a zárt *ĕ* pontos hangminőségére a szóban forgó településen. Ez a gyakorlat tehát egy fonológiai szempont érvényesítésére szorítkozik, így szembemegy az egyezményes lejegyzés hangsúlyosan fonetikus jellegével.

Az egyezményes hangjelölésnek az atlaszmunkák gyakorlatában jelentkező másik gyenge pontja a megcélzott fonetikai pontossági szinttel függ össze. A RMNyA. és a MCsNyA. hangjelölési gyakorlata, bár a MNyA.-ével közös alapelvekre épül (a hangjelölés gyakorlatáról a MNyA. esetében részletes leírás is készült, Benkő 1975), néhány apróságban mégis eltér, a legfontosabb különbség a diftongusok lejegyzését érinti. Az erdélyi dialektológusok által készített adattárakban a diftongusok jelölése igen változatos, a nyomatékeloszlás jelölésére három fokozatot is megkülönböztettek. A MNyA. terepmunkásai e tekintetben tudatosan bizonyos mértékű egyszerűsítésre törekedtek, hogy minél jobban kiküszöbölhető legyenek az egyéni lejegyzői gyakorlatokból, illetve a gyűjtők nyelvjárási háttéréből adódó különbségek, következtelenségek (Benkő 1975, 142–143).²

A teljes magyar nyelvterületet a MNyA. önmagában nem fedi le, ezért, ha átfogó nyelvföldrajzi vizsgálatokat akarunk végezni, mindenképp szükséges két nagyatlaszunk, a MNyA. és a RMNyA. integrálása. A lejegyzői, gyűjtői gyakorlatból adódó akár apró, ám következetes eltérések a kvantitatív, sok száz térképlapot automatikusan elemző eljárások használatakor fölerősödhetnek, összeadódnak, ezért számolnunk kell az adattárhatással: a hangjelölési gyakorlatban mutatkozó bizonyos különbség könnyen nyelvi különbség látszatát eredményezi. A lejegyzés fonetikai információtartalmának csökkentésével mérsekkelhető ugyan az adattárhatás (Kocsis–Vargha 2016; Vargha 2017, 67–89), ám ez a megoldás nem alkalmazható akkor, ha kifejezetten a finom fonetikai különbségek mentén vizsgáljuk a térbeli mintázatok alakulását.

² „...tehát: a jelölési rendszert úgy kellett átalakítani, hogy a kettőshangzókban csak azokat a fonetikai tényeket tüntessük föl, melyeket a hallásban megközelítőleg biztosan észlelni tudunk, s elveszünk mindazon sajátságok jelölési kényszerét, amelyekben a pontosabb apperpciálás bizonytalan vagy éppen lehetetlen. Magától értetődik, hogy a változások az időtartam- és a nyomatékviszonyokat érintették. A részletekről majd alább lesz szó. Összefoglalóul csak annyit, hogy a kettőshangzók időtartam- és nyomatékjelölésében végbevitt módosítások egyszerűbbé, kevésbé színessé tették jelölési rendszerünknek ezt a részlegét, s így látszólag az élő nyelv bonyolultságát elmosták, elfedték; a valóságban azonban hangjelölésünk e vonatkozásban föltétlenül megbízhatóbb, igazabb lett, mert nem vállalkozik arra, amit nem tud hitelesen megoldani.” (Benkő 1975, 142–143)



Az egyezményes hangjelöléssel lejegyzett adatok nagy, és jelenleg is folyamatosan növekvő száma óhatatlanul magában hordozza különböző anomáliák veszélyét. Ilyen például hehezetesnek jelölt mássalhangzók előfordulása a MNyA. kórógyi kutatópontján. Magyar beszélőközösség esetén eleve nem is számíthatunk hehezetes mássalhangzókra, Kórógy földrajzi környezetében nincsenek is olyan nem magyar nyelvváltozatok, amelyek ilyen mintát jelentenének, ráadásul nem tudunk olyan hangfelvételtől, amelyen nyoma lenne a hehezetes ejtésnek. Nehéz magyarázatot találni e jelölésre, talán egy idiolektális jelenséget értelmezett hehezetként a terepmunkás.

Ha a fenti példákban illusztrált következtelenségektől és furcsaságoktól el is tekintünk, az adatok lejegyzése akkor is, még megfelelő jelölési rendszerrel és kiforrott gyakorlattal is igen összetett feladat.

A nyelvjárási atlaszok általában impresszionisztikusan lejegyzett nyelvi adatokat adnak közre. Noha ezek az adatok részletekbe menő fonetikai lejegyzésben kerülnek elének, esetlegességeket tartalmaznak, és szükségszerűen tükrözik a lejegyző előzetes elvárásait, vagyis alapvetően az adott nyelvjárást már ismert jellegzetességeit igyekeznek visszaadni. A hangjelölési rendszer önkényessége (gondoljunk például a viszonyítási hangzóminőségek között a lejegyző által választható hangárnyalatok kötött számára) elméletileg sem teszi lehetővé az ideális lejegyzést, a konkrét realizációkban meglévő bizonyos mértékű ingadozás, változatosság csak növeli a lejegyzői benyomások esetlegességeit. Nem véletlen, hogy a fonetikus lejegyzések megbízhatóságával kapcsolatosan az utóbbi évtizedekben felerősödött a kritika.

A fenti észrevételekkel szembenézve minden egyes lejegyzett nyelvi adatot – egyenként véve – esetlegesnek, bizonytalan értékűnek kell tekintenünk, hasonlóan ahhoz, ahogy egy magánhangzó akár manuálisan mért, akár valamilyen beállítások szerint automatikusan megállapított formánsértékeit is – önmagukban – pontatlannak, megbízhatatlannak gondoljuk. Ugyanakkor azt sem felejtethetjük el, hogy minden lejegyzés tükröz fonetikai tényeket, és alighanem több fonetikai információt tartalmaz, mint esetlegességet. Ha lejegyzésekkel dolgozunk, az a feladatunk, hogy csökkentjük a lejegyzésben rejlő szubjektivitást.

Ahhoz, hogy egy kutatópontban egy jellegzetes hangminőséget megragadjunk, olyan módszerre van szükségünk, amely egyaránt képes fölülemelkedni a hangjelölési rendszer önkényességén, a lejegyzői szubjektivitáson és a konkrét realizációkban meglévő ingadozáson. A megoldást a kvantitatív megközelítés jelenti, lehetővé téve a lejegyzett adatokban rejlő szubjektivitás bizonyos mértékű kiküszöbölését: az egyes adatok lejegyzésében lévő bizonytalanságok, eset-

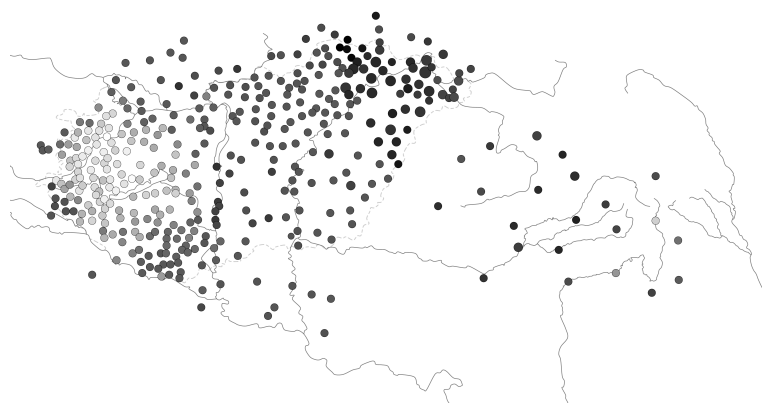




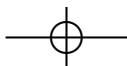
legességek (vagyis az ideális lejegyzéshez képest a különböző irányú eltérések) több száz adat együttes áttekintésekor kiegyenlítik egymást.

Kvantitatív megközelítésben még olyan nyelvföldrajzi elemzésre is használhatjuk az informatizált atlaszadatokat, amely alkalmas az egyes magánhangzó-minőségek folyamatos és fokozatos térbeli változást mutató mintázatának megjelenítésére is. Az egyes szimbólumokhoz, a kialakult lejegyzési szokások ismeretében, hanggal összekapcsolt lejegyzések alapján becsült formánsértékeket rendelhetünk, amelyek magánhangzó-minőséget fejeznek ki. Az 1. ábrán látható térkép elkészítéséhez kidolgozott elemzési módszer alapvetése, hogy – bármely magánhangzó esetében – az egyes hangszín-realizációk gyakoriságát figyelembe véve, a különböző hangszínrnyalatokat jelző szimbólumokhoz rendelt jellemző formánsértékek súlyozott átlaga szerint megállapítható a kutatópontot jellemző magánhangzó-minőség (részletesebben l. Vargha 2020).

Az 1. ábrán az *e* lejegyzett hangtani variánsai alapján készített, az *e* kiszámított első formánsa tipikus értékeinek térbeliségét mutató térképet láthatjuk. Noha az elemzésnél figyelembe vett minden egyes magánhangzó-szimbólumot bizonytalannak és esetlegesnek tekintünk (a lejegyző néha inkább az egyik, máskor inkább egy másik fonetikai szimbólum elképzelt tartalmához közeleink érezte az elhangzott változatot, döntése mindig szubjektív és esetleges volt), a kvantitatív módszer minden településre egy tipikusnak tekinthető, számszerűen kifejezett minőséget (nyíltságbeli fokozatot) határozott meg, olyan ideális pontosságot célozva, amelynek kifejezésére maga a hangjelölés nem is alkalmas.



1. ábra: Az *e* lejegyzés alapján kiszámított F1 értékeinek térképe. Minél világosabb egy kutatópont, annál nagyobb az F1 kiszámított értéke (Vargha 2020, 454).





1.2. Dialektometria

A számítógépes dialektológia lényeges vonatkozása, hogy az adatok felhasználásának lehetősége, így használati értéke megnő. Az informatizált adattárak egymással integrálhatók, vagyis a különböző adattárakból származó adatok könnyen összekapcsolhatók, együtt is elemezhetők, térképezhetők. Megfelelő mennyiségű informatizált adat előállítás után akár igen bonyolult kvantitatív elemzéseket végezhetünk, más tudományágakban már használatos numerikus eljárásokat és térinformatikai módszereket is alkalmazhatunk az adatelemzés során. A nyelvjárási adatok páronkénti összevetésén alapuló dialektometria például kvantitatív módon, statisztikai eszközökkel vizsgálja a nyelvjárások közötti hasonlóság, illetve különbözőség mértékét, az elemzések eredményeit többnyire térképen szemléltetve (Vargha 2017 11–36).

A dialektometria módszertanának kidolgozása Hans Goebel salzburgi romanista nyelvész nevéhez fűződik. Módszere az adatok térképenkénti osztályozásán, kutatói csoportosításán alapul. Az adatokat közlő eredeti térképek alapján a kutatók úgynevezett munkatérképeket hoznak létre, mindig egyetlen változó különböző értékei szerint csoportosítva az adatokat. Az adatok munkatérképeken kialakított csoportosítását táblázatos formában összegzik, majd ebből az úgynevezett adatmátrixból hozzák létre a kutatópontpárok nyelvi hasonlósági értékeit megmutató hasonlósági mátrixot (a módszerről összefoglalóan l. Goebel 2010; 2011; illetve: <http://dialektkarten.ch/dmviewer/index.en.html>).

A jelen kutatásban azonban a nyelvi hasonlóság kiszámítása nem előzetes kutatói csoportosítás alapján, hanem az adatok automatikus összevetésével történik. A módszer alapja Levenshtein betűláncok összevetésére kidolgozott algoritmus (Levenshtein 1966). Az algoritmust Kessler (1995) használta elsőként dialektometriai kutatásban, az ír nyelvjárások elemzésére. A módszer lényege, hogy az adatok közti különbségek megállapítása nem előzetes kutatói klasszifikáció alapján, hanem a lejegyzett adatok elemzésével, automatikusan történik. A módszer további tesztelése és általános használatának megalapozása a dialektometriai kutatásokban a groningeni egyetem kutatói, elsősorban John Nerbonne és Wilbert Heeringa nevéhez fűződik (Heeringa 2004; Heeringa–Nerbonne 2001; 2013; Nerbonne et al. 1996).

Az algoritmus két, fonetikus lejegyzett és karakterláncnak felfogott adat egymáshoz képesti távolságát méri, kiszámítva, hogyan lehet legkönnyebben (legkevesebb lépéssel, legkisebb transzformációs „költséggel”) „átforgatni” az egyik karakterláncot a másikba, azaz a köztük lévő különbözőség mértékét fejezi ki számszerűsített formában. A lejegyzés minden fonetikus alapjele (például



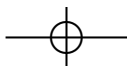


dzs) és mellékjele külön karakternek számít. Az *epër* és az *epör* például egyetlen cserével forgatható át egymásba, a köztük lévő távolság százalékban megadva 25%, vagyis a két adat 75%-ban azonos. Az *epër* és a *szëdër* között már három lépésre van szükség, hogy az egyiket a másikba átalakítsuk, így a hosszabb adat, a *szëdër* hosszát alapul véve (összesen öt karakter) a két adat 60%-ban különböző, vagyis 40%-ban azonos. A térképenként kiszámított kutatópontok közti nyelvi távolságok alapján számítjuk ki a kutatópontok közti átlagos nyelvi távolságot. Valamennyi kutatópont minden más kutatóponttal való nyelvi hasonlóságának számszerűsített mértékét (illetve a köztük lévő nyelvi távolságot) egy mátrixban ábrázolhatjuk. (A módszerről bővebben l. Vargha 2017, 21–24.)

Akár manuális klasszifikáció eredményeként, akár automatikus adatösszevetések révén jön létre a hasonlósági mátrix, ugyanúgy alkalmas arra, hogy további műveleteket végezzünk rajta. A kutatópontok automatikus csoportosítására leggyakrabban alkalmazott módszer a mátrixon végzett klaszteranalízis, amelynek során a kutatópontok közti távolságok mentén alakítunk ki klasztereket oly módon, hogy az egymáshoz nyelvileg közelebb lévő kutatópontok kerüljenek azonos csoportba. Az elemzés eredményét klaszterterképen is ábrázolhatjuk. A 2. ábrán erre látunk példát: 12 csoportot jelenítettünk meg térképesen. A megjelenített csoportok száma a kutató döntésén múlik, valójában kevesebb, de akár sokkal több csoportot is megkülönböztethetnénk. A korábbi, klasszikus (nem kvantitatív alapú) nyelvjárás klasszifikációk sem azonosan határozták meg a nyelvjárás alakulatok számát, Balassa József (1891) például nyolc nyelvjárás-területet, Juhász Dezső (2001) tíz nyelvjárás régiót különböztet meg. Valójában ekkora terület esetén akkor informatívabb, amit látunk, ha több csoportot különítünk el, legalábbis ebbe az irányba mutatnak az újabb vizsgálatok eredményei (vö. Vargha 2021), és 12 csoport még viszonylag áttekinthetően megjeleníthető. Borsos Balázs (2011) például a Magyar néprajzi atlasz hasonló módszerű elemzésekor többféle klaszterterképet is készít egy-egy mátrix alapján, a csoportok számának változtatásával, akár 103 csoportot megkülönböztetve: így jut el felosztásban a néprajzi kiségiók szintjéig.

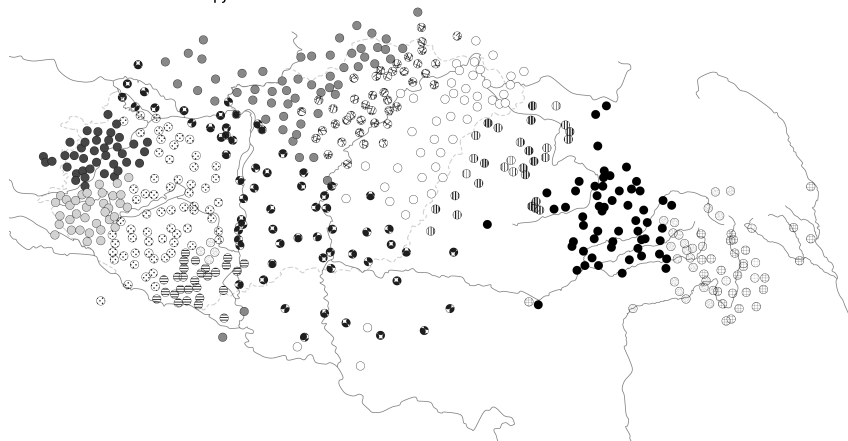
A nyelv térbeli változatossága alapvetően kontinuum jellegű. Ez azt jelenti, hogy a nyelvjárások közti különbség a földrajzi távolsággal arányosan növekszik (Chambers–Trudgill 1998, 5–7). Ahogyan Balassa József fogalmaz a magyar nyelvjárások első átfogó leírásában: „rendes körülmények közt annál nagyobb a különbség [sic!] az egyes nyelvjárások közt, minél messzebb esnek egymástól” (1891, 5).

Bár alapvetően minden összefüggő nyelvterületre jellemző a nyelvjárás kontinuum-jelleg, vannak olyan tényezők, amelyek ezt az összefüggést valamelyest





Klaszterkép a MNyA. és a RMNyA. 483 integrált térképlapjának dialektometriai elemzése alapján



2. ábra: A magyar nyelvjárások automatikus csoportosítása Ward-féle klaszteranalízissel a MNyA. és a RMNyA. alapján 12 csoport megkülönböztetésekor

meg is törik. Így az egyes nyelvi jelenségek térbeli változatosságában megfigyelhető, hogy vannak gócpontok, ahol az adott jelenség általánosabban jellemző, máshol viszont kevésbé. A különböző jelenségek mentén alakulhatnak ki olyan területek, amelyek nyelvi szempontból is hasonlóknak mutatkoznak. A nehezen átjárható politikai határok – mint amilyen például a megyehatár volt néhány száz évvel ezelőtt Somogy és Zala megye között (l. Király 2005, 30) –, vagy az olyan természetes határok, mint egy hegy vagy átjárhatatlan mocsárvidék, szintén befolyásolhatják a madártávlatban közeli települések közti kapcsolatokat, így a köztük lévő nyelvi hasonlóságot is. A magyar nyelvjárások esetében fokozottan számolnunk kell a nyelvterületen belüli népmozgásokkal is, így például a török háborúk következtében elnéptelenedett déli vidékek újratelepülésével, a székelyek nyugatról keletre költözésével vagy erdélyiek Moldvába településével.

A földrajzi és a nyelvi távolság összefüggését vizsgáló, holland nyelvjárások alapján készült tanulmányukban John Nerbonne és Wilbert Heeringa úgy fogalmaznak, hogy a földrajzi távolság valóban nagymértékben magyarázza a nyelvjárások közti eltéréseket, de nem teljes mértékben. A nyelvjárás kontinuum tehát nem egyenletes, így van létjogosultsága annak, hogy nyelvjárás alakulatokat különítsünk el. A csoportok között azonban, éppen a kontinuum jellegéből adódóan, általában nem lesznek éles határok (Nerbonne–Heeringa 2001). A nyelvi tényezők közül azok lesznek számunkra a leginkább meghatározók

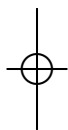


a nyelvjárások közti rendszerszerű összefüggések feltárásában, amelyek olyan nyelvi hasonlósági mintázatokat alakítanak ki, amelyek kevésbé járnak együtt a földrajzi távolsággal.

A dialektometria indulása óta fontos kérdés, hogy a nyelvi jelenségek különböző fajtái milyen módon befolyásolják a kutatópontok közötti nyelvi hasonlóságot; ennek a kérdésnek a megválaszolására magyar nyelvjárási adatok alapján is van példa. A magyar nyelvatlász adatok fonetikai információtartalmának automatikus csökkentése révén különböző nyelvi szinteknek megfelelő hasonlósági mátrixok hozhatók létre. A különböző mátrixok esetén a földrajzi távolsággal való korreláció vizsgálatából, illetve a szigethelyzetű kutatópontok különböző mátrixok használatakor kirajzolódó hasonlósági mintázatainak összevetéséből arra következtethetünk, hogy a fonetikai részleteknek, vagyis a hangtannak van meghatározó szerepe a nyelvjárások közötti kapcsolatrendszerek alakulásában és feltárásában (Vargha 2015; 2017, 49–66).

További kérdésként fölmerül, és a jelen dolgozat is erre keresi a választ, hogy a fonetikán belül a magánhangzók és a mássalhangzók hasonlóan határozzák meg a nyelvjárások közti távolságot. Hans Goebel nyelvatlász adatok manuális csoportosításával létrehozott munkatérképeiből kialakított, vokalizmust és konsonantizmust reprezentáló részkorpuszok alapján végzett korrelációs dialektometriai elemzéseket (2005, 349–350, 365, 367). A mátrixok közti, kutatópontonkénti korrelációs elemzés térképe alapján kirajzolódnak azok a területek (Franciaország déli részén), ahol legkevésbé korrelálnak a kutatópontok hasonlósági mintázatai, vagyis az ilyen kutatópontok a magánhangzókat érintő fonetikai jelenségek alapján más kutatópontokkal mutatnak nagyobb hasonlóságot, mint a mássalhangzók szerint.

Konsonantizmus alatt itt azokat a jelenségeket érthetjük, amelyek a mássalhangzóban mutatnak variabilitást. Ezek többfélék lehetnek, érinthetik a hangok képzési helyét és módját (pl. *borza* ~ *boddza* ~ *bozza* ~ *bojza*), de akár a sorrendjét (pl. *kanál* ~ *kalán*) vagy az időtartamát is (*alugyon* ~ *aluggyon*). A vokalizmus körébe azokat a jelenségeket sorolhatjuk, amelyek a magánhangzók minőségében vagy időtartamában mutatnak különbséget (pl. *buza* ~ *búza* ~ *búza*; *vödör* ~ *vödör* ~ *vüdör*). Az egyes jelenségek, tehát a mássalhangzók, illetve magánhangzók használatában jelentkező azonosságok és különbségek, atlaszbeli előfordulási gyakoriságuk függvényében határozzák meg a kimutatásokat. Számos ritka, lexikailag kötött jelenség is óhatatlanul befolyásolja az eredményeket, de csak előfordulási arányuk mértékében.





2. Anyag és módszer

Az informatizált magyar nyelvjárási adattárak közül jelen vizsgálathoz a MNyA., a RMNyA. és a MCsNyA. adatain végzek elemzéseket. A válogatásból adódó önkényességet kizárva, kivétel nélkül valamennyi olyan térképlapot bevettem az elemzendő korpuszba, amelyeket a gyűjtés során a MNyA. teljes kutatópont-hálózatán kikérdeztek, és megvan a párjuk a RMNyA. térképlapjai között, tehát páronként integrálhatók. Ez összesen 483 MNyA. és ugyanennyi RMNyA. térképlapot jelent. A MCsNyA. esetében a teljes adattárral dolgoztam, ez a megjelent I. és II., valamint a kéziratban maradt III. kötet teljes anyagát jelenti (Gálffy et al. 1991).³ A három kötet összesen 1049 címszót tartalmaz.

Jelen kutatásban nem az adatok előzetes csoportosítása (munkatérképekből álló korpuszok kialakítása) révén valósítom meg az adattárak vokalizmus, illetve konzonantizmus alapú dialektometriai elemzését (ahogyan korábban Goebel), hanem az adatok automatikus, Levenshtein-algoritmussal történő összevetésével. Ahhoz, hogy külön tudjam elemezni a magánhangzókat, illetve mássalhangzókat a páronkénti adatösszevetés folyamatában, első lépésként úgy módosítom az adatokat, hogy azokban csak a magánhangzókat (illetve fordított esetben csak a mássalhangzókat) hagyom meg, és csak azokon végzem el az összevetéseket.

A *meggy* címszavú MNyA. és RMNyA. térképek esetében például ez azt jelenti, hogy a szóvégi mássalhangzó térbeli variabilitásának (*meggy* ~ *metty* ~ *meddzs* stb.) a konzonantizmus alapján létrehozott hasonlósági mátrix, a magánhangzó minőségét érintő különbségeknek (*meggy* ~ *mëggy* ~ *möggy* stb.) pedig a vokalizmus alapján létrehozott mátrix kialakításában lesz szerepe.⁴

Vannak azonban olyan hangtani jelenségek, amelyek mind a magánhangzók, mind a mássalhangzók lejegyzésbeni variabilitásában jelentkeznek. Ilyen az *l*-kiesés és a nazalitás: az *l* redukálódása együtt járhat a megelőző magánhangzó megnyúlásával (*nélkül* ~ *nélkü^l* ~ *nékű*), ugyanígy az *n* ejtése redukálódhat, miközben a megelőző magánhangzó nazalizációja erősödik, amit mellékjellel jelölünk (*patkán* ~ *patkáⁿ* ~ *patká*). E jelenségek figyelemre méltó térbeli mintázatokat mutatnak, a magánhangzó és mássalhangzó közti koartikuláció területenként változó, tehát nyelvváltozat-specifikus jellegével (az interszegmentális kohézió mértékével) függnek össze. A jelen dolgozatban azonban e jelenségek

³ A III., kéziratban maradt kötet fennmaradásának és megtalálásának körülményeiről, az adatok informatizálásáról Bodó Csanáddal közös tanulmányunkban számoltunk be 2008-ban.

⁴ A *meggy* ejtésbeli változatai a magánhangzó és a szóvégi mássalhangzó variabilitása szerint csoportosítva elérhetők egy korábbi kutatás (Kocsis–Vargha 2016) honlapján: <http://www.bihalbocs.hu/mnyarmnya/munkaterk/meggy1.html>





hatása, mint jeleztem, egyaránt jelentkezik a vokalizmushoz, illetve a konzonantizmushoz tartozó kimutatásokban.

Az elemzés során tehát két mátrix jön létre: az egyik a magánhangzók, a másik a mássalhangzók használatában és ejtésében lévő nyelvjárási hasonlóság mértékét mutatja meg a kutatópontpárok között. A mátrixok alapján korrelációs elemzés és korrelációs térképek formájában összevetem a nyelvi távolságot a földrajzi távolsággal, illetve a konzonantizmusból adódó nyelvi távolságot a vokalizmusból adódóval. A mátrixok klaszteranalízise alapján csoportosítom a kutatópontokat, a csoportosítás eredményét térképre vetítem és összevetem egymással a konzonantizmus és a vokalizmus alapján kirajzolódó felosztásokat, illetve mindkét felosztást összevetem az eredeti (magánhangzókat és mássalhangzókat egyaránt tartalmazó) adatokból készült mátrix klaszterterképével. Néhány kutatópont hasonlósági térképeinek összevetésével bemutatom, jellegükben hogyan térnek el egymástól a vokalizmus és konzonantizmus által kirajzolt nyelvi hasonlósági mintázatok. A MCsNyA. dialektometriai elemzése alapján külön megvizsgálom, hogyan alakul a kutatópontok csoportosítása Moldvában, ahol nemcsak a magánhangzók, hanem a mássalhangzók ejtése is közismerten jellegzetes területi különbségeket mutat (Juhász 2001, 307–314).

A nyelvatlazadatok összevetéséhez, vagyis a nyelvi hasonlósági mátrixok kialakításához a Levenshtein-algoritmust használom (Levenshtein 1966), az adatösszevetéseket a Bihalbocs⁵ magyar dialektológiai alkalmazással valósítom meg. A földrajzi távolság meghatározásához a Haversine-távolságot alkalmazom, szintén a Bihalbocsban, a kutatópontpárok GPS-koordinátái alapján, és ugyanezen GPS-koordináták alapján történik a térképezés is. A statisztikai elemzéseket az R statisztikai programcsomaggal végzem. A mátrixok közötti korreláció meghatározására a távolságmátrixok esetében a tudományterületen is szokásosan alkalmazott Mantel-tesztet használom, a kutatópontonkénti korrelációs együtthatókat Pearson-korrelációval számítom ki az ade4 csomag használatával (Dray–Dufour 2007). A korrelációs együttható értéke mindkét eljárás esetében -1 és 1 közé esik, 1 közelében beszélhetünk erős pozitív korrelációról. A kuta-

⁵ Vékás Domokos még a 90-es években Peter Ladefoged, a Nemzetközi Fonetikai Társaság akkori elnöke példája és személyes iránymutatása nyomán vágott bele azoknak a nyelvészeti technológiáknak a fejlesztésébe, amelyek később Bihalbocs néven váltak ismertté és a magyar dialektológiában lehetővé tették adattárak informatizálását, kvantitatív feldolgozásukat, térképezésüket, hanganyagok és lejegyzésük összekapcsolását, biztosítva többek között a hullámforma megjelenítését és spektrális elemzéseket is (Vékás 2007). A Bihalbocs olyan keretrendszernek is felfogható, amelybe a felhasználók a forráskód módosításával beépíthetik a kutatási igényeiknek megfelelő elemzési vagy adatmegjelenítési funkciókat.



tópontok csoportosításához a Ward-féle klaszterezési eljárást alkalmazom, az R statisztikai programcsomagba beépített *ward.D2* algoritmus használatával. Az R-ben kiszámított értékeket minden esetben Bihalboccsal vetítem térképre.

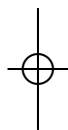
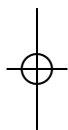
3. Eredmények

3.1. Nyelvi hasonlóság és földrajzi távolság

A nyelvi jelenségek közül várhatóan azok lesznek meghatározóbbak a nyelvjárássok közti viszonyrendszerek alakításában, amelyek kevésbé mutatnak együttjárást a földrajzi távolsággal, vagyis kisebb-nagyobb mértékben megtörik a nyelvjárási kontinuumot. A mássalhangzók dialektometriai elemzésekor kialakított mátrixot összevetve a földrajzi távolságokon alapulóval azt látjuk, hogy Mantel-teszttel kiszámítva a két mátrix közötti korrelációt 1-hez közeli értéket kapunk ($r = 0,802$), ami azt jelenti, hogy a konzonantizmusból adódó nyelvi hasonlóság leginkább a földrajzi távolság függvényében alakul. Térképre vetítve a Pearson korrelációs együttható értékét az egyes kutatópontok esetében (3. ábra) azt látjuk, hogy a kutatópontok túlnyomó többsége 0,9 körüli értéket mutat, a nyelvterület középső részén valamivel gyengébb (0,8 körüli) értékeket találunk, összességében véve igen erősnek mutatkozik az együttjárás a földrajzi, illetve a mássalhangzók összevetése alapján kiszámított nyelvi távolság között. Kifejezetten alacsony értékeket csak olyan kutatópontok esetében találunk, amelyek (a mássalhangzók összevetése alapján is) leginkább térben távoli kutatópontokkal mutatnak nagyobb hasonlóságot: a délvidéken Piros és Kórógy, illetve a 20. században Bukovinából áttelepültek esetében, a Dunántúl déli részén, Tolna és Baranya határán.

A földrajzi távolság és a magánhangzók alapján számított nyelvi távolság korrelációja más képet mutat, mint a mássalhangzók esetében (4. ábra). A magánhangzók alapján kirajzolódó nyelvi hasonlósági mintázatok már csak közepes együttjárást mutatnak a földrajzi távolság szerintivel ($r = 0,564$). Ebből az következik, hogy alapvetően a magánhangzók felelősek a nyelvjárássok közötti olyan összefüggésekért, amelyeket a földrajzi távolság nem magyaráz (vö. Heeringa-Nerbonne 2001).

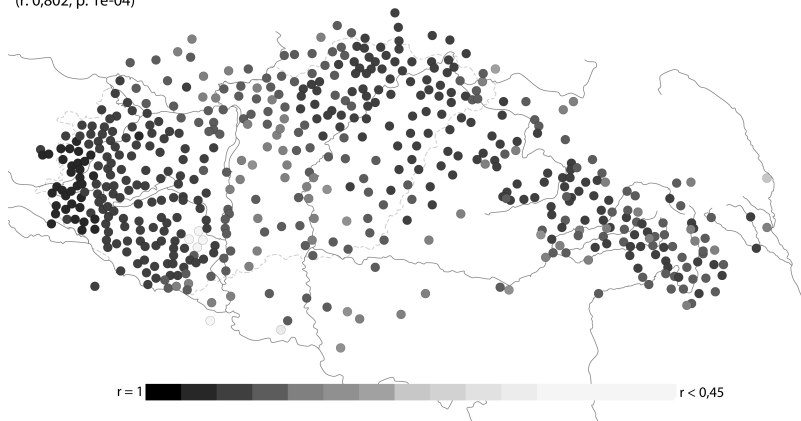
A vokalizmus és konzonantizmus nyelvi hasonlósági mátrixának korrelációja alapján kiszűrhetjük azokat a területeket, ahol a legnagyobb az eltérés a különböző mátrixok alapján kirajzolódó hasonlósági mintázatok között (5. ábra). A két mátrix közti korreláció közepes, az r értéke 0,663. A kutatópontonkénti





Vokalizmus és konzonantizmus a magyar nyelvjárásokban

Korreláció a konzonantizmus és a földrajzi távolság között
($r: 0,802$; $p: 1e-04$)

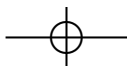


3. ábra: Korreláció a konzonantizmus és a földrajzi távolság között. A két mátrix közötti korreláció Mantel-tesztel kiszámított értéke 0,802. A kutatópontok színe az egyes kutatópontok esetében kiszámított Pearson-korrelációs együttható értéke szerint alakul. Minél inkább a földrajzi távolság függvényében alakul a nyelvi hasonlósági mintázat egy kutatópont esetében, annál sötétebb a kutatópont.

Korreláció a vokalizmus és a földrajzi távolság között
($r: 0,564$; $p: 1e-04$)



4. ábra: Korreláció a vokalizmus és a földrajzi távolság között. A két mátrix közötti korreláció Mantel-tesztel kiszámított értéke 0,565. A kutatópontok színe az egyes kutatópontok esetében kiszámított Pearson-korrelációs együttható értéke szerint alakul. Minél inkább a földrajzi távolság függvényében alakul a nyelvi hasonlósági mintázat egy kutatópont esetében, annál sötétebb a kutatópont.

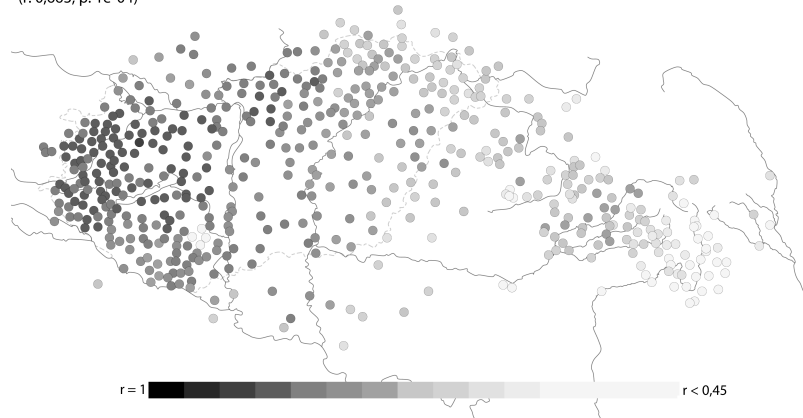




összevetéseket térképezve azt látjuk, hogy – hasonlóan a 4. ábrához, ahol a vokalizmust vetettük össze a földrajzi távolsággal – keleten világosabbak a szürke árnyalatai, a keleti székely kutatópontokon kifejezetten halványak, csakúgy, mint a még keletebbre fekvő moldvai Bogdánfalva esetében. Néhány kutatóponton az r értéke még a 0,3-at sem éri el. Ahhoz, hogy lássuk, milyen jellegű a különbség a Székelyföld keleti részén a vokalizmus és a konsonantizmus által meghatározott mintázatok között, érdemes megnéznünk néhány kutatópont hasonlósági térképét.

A vokalizmus és konsonantizmus nyelvi hasonlósági mátrixának korrelációja alapján kiszűrhetjük azokat a területeket, ahol a legnagyobb az eltérés a különböző mátrixok alapján kirajzolódó hasonlósági mintázatok között (5. ábra). A két mátrix közti korreláció közepes, az r értéke 0,663. A kutatópontenkénti összevetéseket térképezve azt látjuk, hogy – hasonlóan a 4. ábrához, ahol a vokalizmust vetettük össze a földrajzi távolsággal – keleten világosabbak a szürke árnyalatai, a keleti székely kutatópontokon kifejezetten halványak, csakúgy, mint a még keletebbre fekvő moldvai Bogdánfalva esetében. Néhány kutatóponton az r értéke még a 0,3-at sem éri el. Ahhoz, hogy lássuk, milyen jellegű a különbség a Székelyföld keleti részén a vokalizmus és a konsonantizmus által meghatározott mintázatok között, érdemes megnéznünk néhány kutatópont hasonlósági térképét.

Korreláció a vokalizmus és a konsonantizmus között
(r : 0,663; p : $1e-04$)

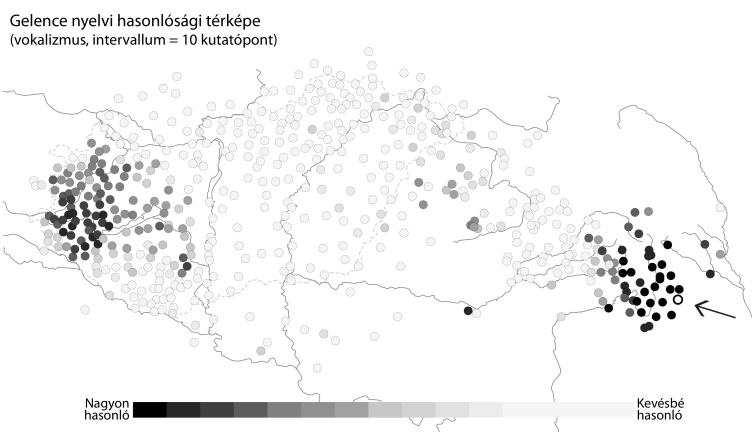


5. ábra: Korreláció a vokalizmus és a konsonantizmus között. A két mátrix közötti korreláció Mantel-teszttel kiszámított értéke 0,663. A kutatópontok színe az egyes kutatópontok esetében kiszámított Pearson-korrelációs együttható értéke szerint alakul. Minél inkább hasonlóak a különböző mátrixok szerint kirajzolódó nyelvi hasonlósági mintázatok egy kutatópont esetében, annál sötétebb a kutatópont.

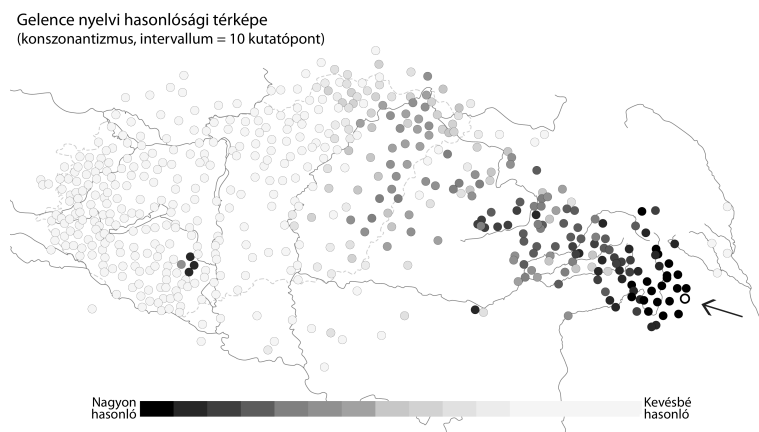




A 6. ábrán Gelence kutatópont nyelvi hasonlósági mintázatai rajzolódnak ki a magánhangzók összevetésével kialakított mátrix alapján. A korábbi dialektometriai elemzések is abba az irányba mutattak, hogy a hangtani részleteket is figyelembe vevő dialektometriai vizsgálatok révén tudunk rátalálni a település-történeti okokkal magyarázható rendszerszerű összefüggésekre.



6. ábra: Gelence kutatópont nyelvi hasonlósági térképe a vokalizmus dialektometriai elemzése alapján. A kijelölt kutatópontot fekete körvonalú, fehér színű kör, a nagyobb nyelvi hasonlóságot a fekete és a mélyszürke árnyalatok jelölik.



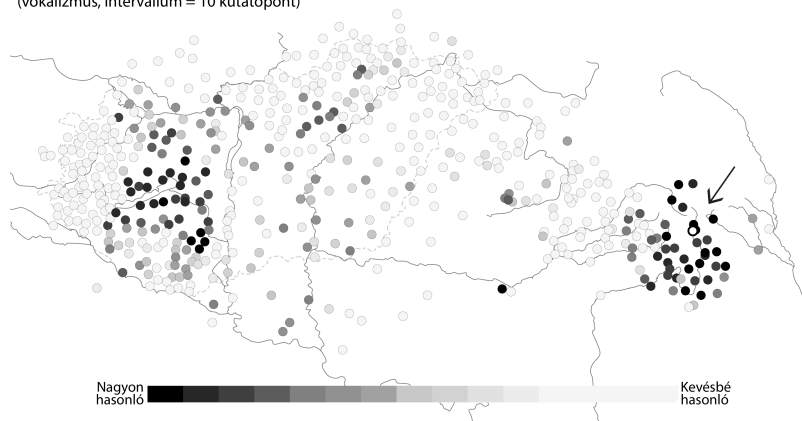
7. ábra: Gelence kutatópont nyelvi hasonlósági térképe a konsonantizmus dialektometriai elemzése alapján. A kijelölt kutatópontot fekete körvonalú, fehér színű kör, a nagyobb nyelvi hasonlóságot a fekete és a mélyszürke árnyalatok jelölik.



A magánhangzó-minőségekre alapozott dialektometriai elemzés a korábbi elemzéseknél látványosabban mutatja a minden bizonnyal a magánhangzórendszerek közti párhuzamokból eredő hasonlóságot a Háromszék keleti részén fekvő Gelence, és a MNyA. nyugat-dunántúli kutatópontjai között. A nyelvi hasonlóságot a mássalhangzók alapján vizsgálva azonban teljesen megváltozik a térkép (7. ábra), a nyelvi hasonlóság a földrajzi közelséggel mutat együttjárást. A Gelence két hasonlósági térképe közt szemmel láthatóan is jelentős különbség tükröződik a mátrixok összevetése alapján készült korrelációs térképen is (5. ábra).

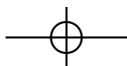
Hasonló megállapításokat tehetünk egy másik keleti-székely kutatópont, Csíkrákos esetében is. Annyi különbséggel, hogy Csíkrákos nem a nyugat-dunántúli, hanem inkább Balaton környéki kutatópontokkal mutat nagyobb nyelvi hasonlóságot a vokalizmus alapján (8. ábra, Csíkrákos konsonantizmus alapján kirajzolódó nyelvi hasonlósági mintázatát l. a 9. ábrán).

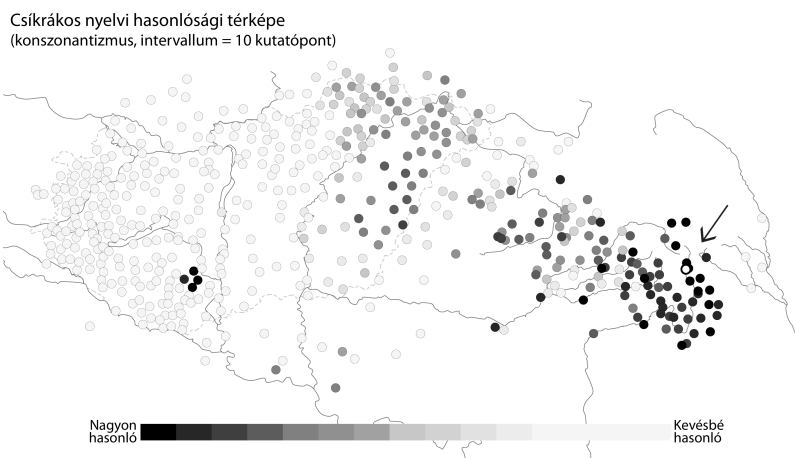
Csíkrákos nyelvi hasonlósági térképe
(vokalizmus, intervallum = 10 kutatópont)



8. ábra: Csíkrákos nyelvi hasonlósági térképe a vokalizmus dialektometriai elemzése alapján. A kijelölt kutatópontot fekete körvonalú, fehér színű kör, a nagyobb nyelvi hasonlóságot a fekete és a mélyszürke árnyalatok jelölik.

A székely nyelvjárások vokalizmusukban jobban hasonlítanak a dunántúli, mint a földrajzilag közelebbi mezőségi, valamint észak-keleti nyelvjárásokhoz: ez magyarázza, hogy a földrajzi távolság és a vokalizmus korrelációs térképén (4. ábra) keleten egyre alacsonyabb korrelációs értékekkel találkozunk. Mivel a konsonantizmus a keleti székely kutatópontokon egészen más mintázatokat mutat, amelyek a földrajzi távolság függvényében alakulnak (7. és 9. ábra), nem meglepő, hogy a vokalizmus és a konsonantizmus korrelációs térképén is rendre alacsonyabb értékekkel találkozunk kelet felé haladva.





9. ábra: Csíkrákos nyelvi hasonlósági térképe a konzonantizmus dialektometriai elemzése alapján. A kijelölt kutatópontot fekete körvonalú, fehér színű kör, a nagyobb nyelvi hasonlóságot a fekete és a mélyszürke árnyalatok jelölik.

A továbbiakban azt nézzük meg, milyen nyelvjárási alakulatok alakíthatók ki automatikusan a vokalizmus és a konzonantizmus hasonlósági mátrixa alapján, illetve azt, miért változhat egyes kutatópontok csoportosítása.

3.2. Magánhangzók, mássalhangzók és nyelvjárás-tipológia

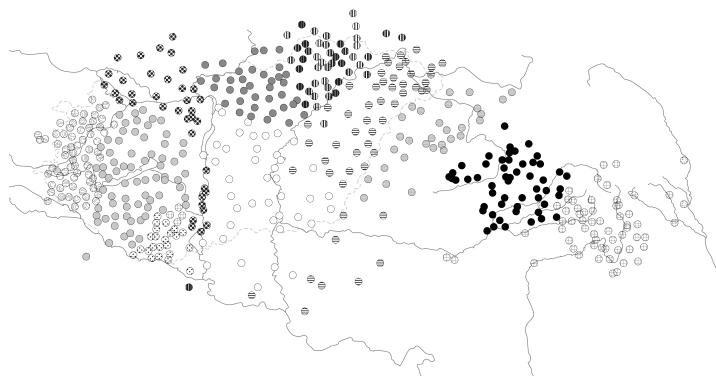
A dialektometriai kutatásokban a nyelvjárások csoportosítása, nyelvjárási alakulatok kialakítása legtöbbször a hasonlósági mátrixok klaszteranalízisével történik, amint ezt már a 2. ábrán is láttuk, ahol 12 csoportot különböztettünk meg. A konzonantizmus és vokalizmus alapján kialakított csoportokat a 10. és a 11. ábra szemlélteti (szintén 12-es felosztással, biztosítandó az összevethetőséget).

A két klaszterkép között több jelentős különbség is van. A legszembeötlőbb ezek közül a Palócföldet érinti: ez a terület markánsan elkülönül a 11., a vokalizmus alapján készített csoportosítás szerint, összhangban a korábbi dialektometriai elemzések eredményeivel (vö. Vargha 2015; 2017), és a Juhász Dezső által készített régiós felosztással (2001). A 10. ábrán azonban, amely a mássalhangzók összevetésén alapul, a palóc régió egyáltalán nem rajzolódik ki. A két legkeletibb moldvai kutatópont egy külön csoportot alkot, akár azokon a klaszterképeken, amelyeket korábbi vizsgálatokban a fonetikai információ kivonásával hoztam létre (l. Vargha 2017, 119).



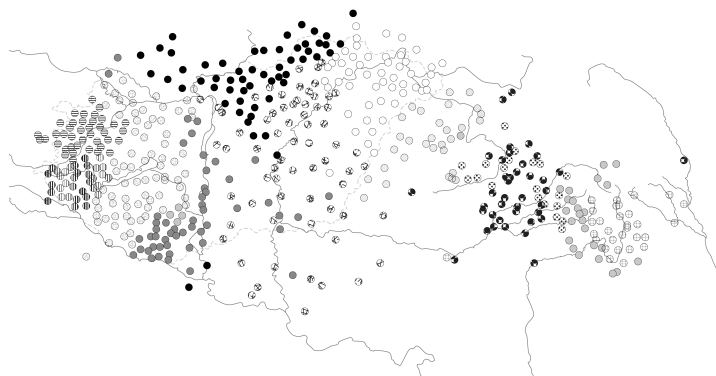


Klaszterterkép a konzonantizmus alapján



10. ábra: Nyelvjárások csoportosítása Ward-féle klaszteranalízissel a konzonantizmus alapján

Klaszterterkép a vokalizmus alapján



11. ábra: Nyelvjárások csoportosítása Ward-féle klaszteranalízissel a vokalizmus alapján

A vokalizmus alapján egyértelműen kirajzolódik a palóc régió, akár a 2. ábrán. Kirajzolódik továbbá egy nagyobb területen jelenlévő, sötétszürke körökkel megjelölt kutatópontokból álló csoport, amelynek legtöbb pontja a Dunántúl dél-keleti részén található, egyértelmű összefüggést mutatva az *ő*-zés térbeliségével (erre jó példa a MNyA. és a RMNyA. integrált csoportosított térképei közül a *szem* magánhangzójának variabilitását bemutató térképlap: <http://www.bihalbocs.hu/mnyarmnya/munkaterk/szem1.html>). E csoport mégsem foglalja azonban magában a Dél-Dunántúl nyugatabbra lévő, *ő*-ző kutatópontjait, nyilván más magánhangzókat érintő, gyakoribb jelenségek miatt (ilyen lehet például a diftongálás vagy a felső nyelvállású magánhangzók időtartama), hiszen

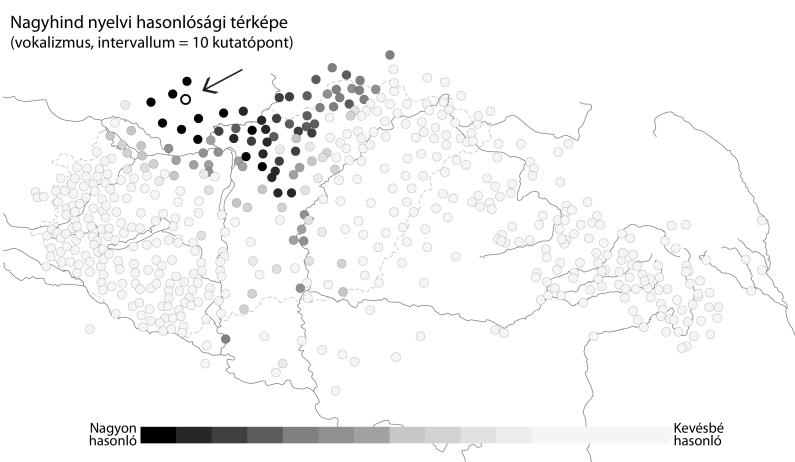


a dialektometriai elemzés kimenetelét minden jellegzetesség együttesen, az atlaszbeli előfordulás arányában határozza meg.

Megemlíthető még a 11. ábrával kapcsolatban, hogy a Nyugat-Dunántúl kétbomlik egy északi és egy déli csoportra, akár az eredeti mátrix alapján készült 2. ábrán. Ami a négy moldvai kutatópontot illeti, közülük három (Pusztina, Diószeg és Bogdánfalva) a keleti székely, az északi Szabófalva pedig mezősegi kutatópontokkal tart, vagyis a vokalizmus alapján nézve moldvai régió nem rajzolódik ki.

Azt, hogy milyen mintázatok állnak az eltérő csoportosítás háttérében, érdemes konkrétan is megnéznünk néhány kiválasztott kutatópont különböző mátrixok alapján készült hasonlósági térképe alapján. Olyan kutatópontokat érdemes kiválasztanunk, amelyek más csoportba kerülnek a konsonantizmus, illetve a vokalizmus alapján.

Nagyhind például a Palócföld nyugati részén van a magánhangzók alapján készített klaszterterképen (11. ábra), viszont a konsonantizmus alapján egy kirajzolódó észak-nyugati területhez tartozik (10. ábra). A 12. ábra Nagyhind nyelvi hasonlósági térképét mutatja a vokalizmus alapján.



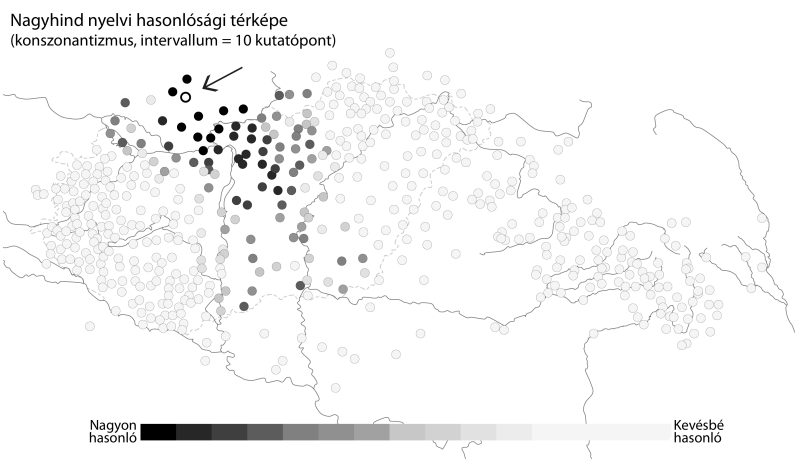
12. ábra: Nagyhind nyelvi hasonlósági térképe a vokalizmus dialektometriai elemzése alapján. A kijelölt kutatópontot fekete körvonallú, fehér színű kör, a nagyobb nyelvi hasonlóságot a fekete és a mélyszürke árnyalatok jelölik.

A nyelvileg hasonló kutatópontok sötétebb színárnyalatokban látszanak, és jól kivehetően megjelenítik a vokalizmus klaszterterképén (11. ábra) is kirajzolódó palóc csoportot.





Nagyhind másik térképén (13. ábra), amely a konzonantizmus alapján készült, más terület bontakozik ki a nagyobb nyelvi hasonlóságra utaló sötétebb színárnyalatokkal. Csak halványabb szürke színt kapnak a palóc régió keleti kutatópontjai, amelyek földrajzilag is távolabb vannak, a Nagyhindtől dél-nyugatra fekvő kutatópontok viszont sötétebb árnyalatúak. Nagyhind dialektometriai térképei is azt támasztják alá, hogy távolabbi kutatópontok között általában csak a magánhangzók alapján van esélyünk nagyobb mértékű hasonlóságot találni.

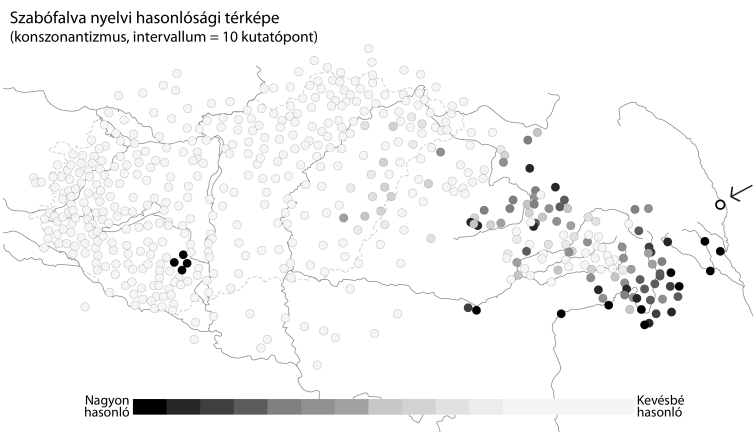


13. ábra: Nagyhind nyelvi hasonlósági térképe a konzonantizmus dialektometriai elemzése alapján. A kijelölt kutatópontot fekete körvonalú, fehér színű kör, a nagyobb nyelvi hasonlóságot a fekete és a mélyszürke árnyalatok jelölik.

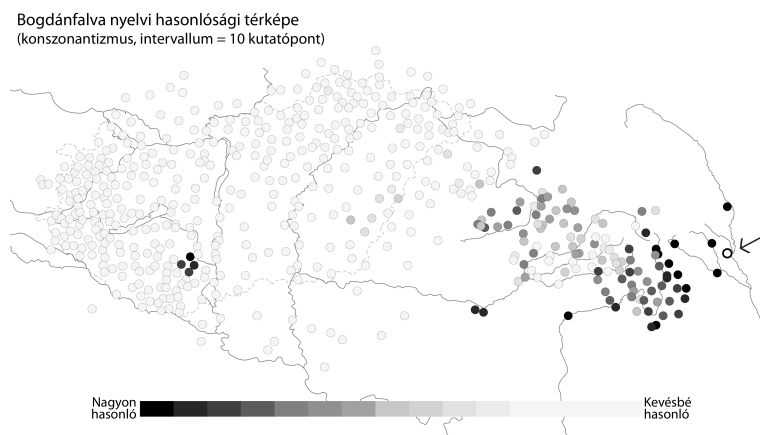
Érdeemes megnéznünk azt a két moldvai kutatópontot, Bogdánfalvát és Szabófalvát, amelyek a konzonantizmus szerint elkülönülő csoportot alkotnak, a vokalizmus alapján azonban különválnak, és keleti székely, illetve mezőszéki kutatópontokkal kerülnek egy csoportba. A konzonantizmus alapján Szabófalva és Bogdánfalva egymással, illetve a másik két moldvai kutatóponttal mutat leginkább hasonlóságot (14. és 15. ábra), a vokalizmus azonban teljesen eltérő mintázatokat alakít ki (16. és 17. ábra). Bogdánfalva leginkább a két székelyes nyelvjárású moldvai kutatóponttal mutat hasonlóságot, de közel azonos mértékben egyeznek adatai a keleti székely kutatópontokkal is. A vokalizmust tekintve Szabófalva egyértelműen észak-mezőszéki kutatópontokkal tart. A további két moldvai kutatópont, Pusztina és Diószeg ugyan jóval nagyobb mértékben mutat hasonlóságot Bogdánfalvával a magánhangzók, mint a mássalhangzók alapján, a



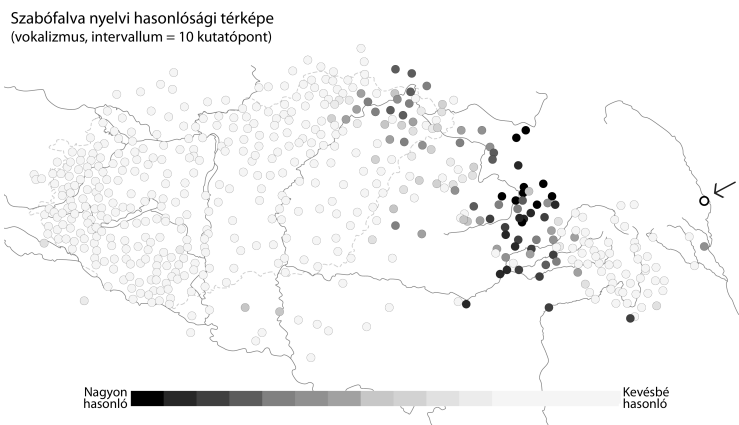
mátrixokban szereplő (itt nem térképezett) százalékos értékek szerint leginkább a keleti székely nyelvjárásokra hasonlítanak.



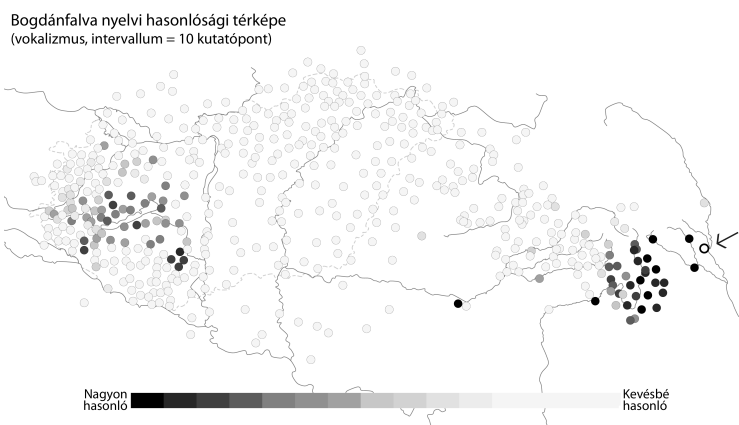
14. ábra: Szabófalva nyelvi hasonlósági térképe a konsonantizmus dialektometriai elemzése alapján. A kijelölt kutatópontot fekete körvonalú, fehér színű kör, a nagyobb nyelvi hasonlóságot a fekete és a mélyszürke árnyalatok jelölik.



15. ábra: Bogdánfalva nyelvi hasonlósági térképe a konsonantizmus dialektometriai elemzése alapján. A kijelölt kutatópontot fekete körvonalú, fehér színű kör, a nagyobb nyelvi hasonlóságot a fekete és a mélyszürke árnyalatok jelölik.



16. ábra: Szabófalva nyelvi hasonlósági térképe a vokalizmus dialektometriai elemzése alapján. A kijelölt kutatópontot fekete körvonalú, fehér színű kör, a nagyobb nyelvi hasonlóságot a fekete és a mélyszürke árnyalatok jelölik.



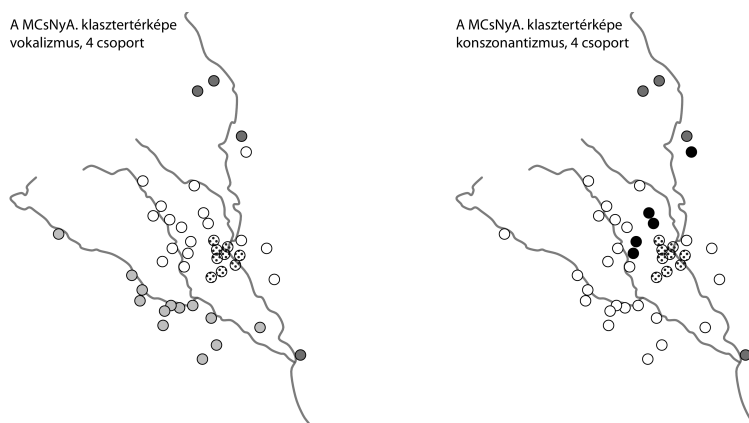
17. ábra: Bogdánfalva nyelvi hasonlósági térképe a vokalizmus dialektometriai elemzése alapján. A kijelölt kutatópontot fekete körvonalú, fehér színű kör, a nagyobb nyelvi hasonlóságot a fekete és a mélyszürke árnyalatok jelölik.

A fentiek értelmében feltételezhetjük, hogy Moldván belül másképpen alakulnak a nyelvi hasonlósági viszonyok a konsonantizmus, mint a vokalizmus alapján, de ezt mindössze négy kutatópont alapján kevésbé tudjuk vizsgálni. A moldvai regionális atlasz, a MCsNyA. adatainak elemzése révén azonban össze tudjuk vetni a vokalizmus és a konsonantizmus mintázatait Moldván belül (18. ábra). Négy csoport megkülönböztetése esetén a Tatros menti, illetve a Tázló menti és a Szerettől keletre található székeleyes nyelvjárású kutatópontok (a 18. ábra bal



Vokalizmus és konzonantizmus a magyar nyelvjárásokban

357



18. ábra: A MCsNyA. kutatópontjainak négy csoportba sorolása klaszteranalízissel a vokalizmus és a konzonantizmus alapján

oldali térképén világosszürke, illetve fehér színnel) a konzonantizmus alapján ugyanabba a csoportba sorolódnak (a 18. ábra jobboldali térképén fehér színnel), kivéve 5 kutatópontot (fekete színnel, északról dél felé haladva: Kalugarény, Szakatura, Lujzikalagor, Gajdár, Esztufuj), amelyek külön csoportot alkotnak. Elkülönülésüket feltehetőleg olyan, a mássalhangzók ejtését érintő jelenségek magyarázhatják, amelyek a székelységben általában nem jellemzők. Ilyen lehet a *dzs*-zés (a más nyelvváltozatokban ejtett *gy* hang helyén *dzs* áll, pl. *dzséng*, különösen Lujzikalagorban) vagy a székelyes változatok *cs*, esetleg *ty* hangja helyén ejtett palatalizált *s'* (pl. *kus'a*, *kuls'*, leginkább Szakatura, Lujzikalagor és Gajdár kutatópontokon) és különösen a palatális laterális *ly* megmaradása a mássalhangzórendszerben (mind az öt említett kutatóponton).

Nincs azonban különbség a két mátrix között az északi (ide tartozik a délre települt Ploszkucény is) és a Bákó környéki csoport elkülönülését illetően. Az északi csoport markánsan elvág az összes többi kutatóponttól, és a Bákó környéki csoport is egyértelműen eltér a székelyes csoportoktól mind a konzonantizmus, mind a vokalizmus alapján (a MCsNyA. kutatópontjainak korábbi automatikus csoportosításairól összefoglalóan l. Vargha 2017, 99–107). Fontos leszögeznünk, hogy a főbb csoportok kialakítása tekintetében nincs különbség a két moldvai mátrix között, vagyis három csoport megkülönböztetése esetén a vokalizmus és a konzonantizmus klaszterterképe teljesen azonos volna: a magánhangzók, illetve a mássalhangzók lejegyzésében rögzített fonetikai információk alapján azonos nyelvjárási alakulatokat jeleníthetünk meg. A különbség négy csoport kialakításakor jelentkezik, és kizárólag a székelyes kutatópontok belső elkülönülését érinti.





4. Következtetések

Felvetésünk, hogy a magyar nyelvjárások tipizálását leginkább a magánhangzók változatosságára alapozhatjuk, annyiban nem új, hogy már a legelső nyelvjárási osztályozások is a magánhangzók alapján próbálták csoportosítani a magyar nyelvjárásokat (l. pl. Simonyi 1889). Lényeges különbség azonban, hogy a korábbi csoportosítások (beleértve ebbe még Imre Samu 1971-es, módszerében kvantitatív monográfiáját is) az egyes jelenségeket külön vizsgálva történtek, a klasszifikációt néhány előre meghatározott nyelvi változó térbeli variabilitására alapozva végezték el. A lejegyzésben rejlő fonetikai információk automatikus összevetésén alapuló dialektometriai elemzés során azonban semmilyen kutatói előfeltevés vagy elvárás sem befolyásolja a jelenségek osztályozását, hiszen a kutatópontok hasonlósági mátrixa kizárólag a lejegyzett adatok alapján jön létre. Így tehát minden, a magánhangzók minőségét befolyásoló és a lejegyzésben is tükröződő tényező érvényesül az elemzésben. Az egyes jelenségek gyakoriságuk függvényében vannak hatással a kutatópontok közti hasonlóság értékére. A magánhangzók dialektometriai elemzésén alapuló statisztikai módszerekkel megvalósított csoportosítás tehát úgy tér vissza a nyelvjárások osztályozásának a 19. századi legelső munkákban már megalapozott hagyományaihoz, hogy közben kiiktatja a folyamatból a hangtani jelenségek önkényes válogatásában rejlő szubjektivitást.

Elemzéseink alapján a konsonantizmus inkább mutat együttjárást a földrajzi távolsággal, mint a vokalizmus, legalábbis a nyelvterület egészét szemlélve. A térben távolabbi kutatópontok közti nyelvi hasonlóságot általában csak a vokalizmus alapján tudjuk kimutatni. Az eredmények arra engednek következtetni, hogy elsősorban a magánhangzó-minőségek bizonyulnak meghatározónak a nyelvjárások térbeli mintázatainak kialakításában és a településtörténeti kapcsolatok feltárásában egyaránt.

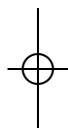
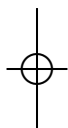
Irodalom

- Balassa József 1891. A magyar nyelvjárások osztályozása és jellemzése. Budapest: MTA.
- Benkő Loránd 1975. A magyar nyelvjárások atlaszának hangjelölési rendszere és gyakorlata. In: Deme László – Imre Samu (szerk.): A magyar nyelvjárások atlaszának elméleti–módszertani kérdései. Budapest: Akadémiai Kiadó. 123–165.
- Bodó Csanád – Vargha Fruzsina Sára 2008. Régi nyelvatlások – új módszerek: Nyelvtechnológiai eljárások a nyelvföldrajzban. Magyar Nyelv 104: 377–392.
- Borsos Balázs 2011. A magyar népi kultúra regionális struktúrája a Magyar Néprajzi Atlasz számítógépes feldolgozása fényében I–II. Budapest: MTA Néprajzi Kutatóintézet.





- Chambers, J. K. – Peter Trudgill 1998. *Dialectology*. 2nd edition. Cambridge: Cambridge University Press.
- Deme László – Imre Samu (szerk.) 1968–1977. *A magyar nyelvjárások atlasza 1–6*. Budapest: Akadémiai Kiadó.
- Dray, Stéphane – Anne-Béatrice Dufour 2007. The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software* 22: 1–20.
- Gálffy Mózses – Márton Gyula – Szabó T. Attila (szerk.) 1991. *A moldvai csángó nyelvjárás atlasza 1–2*. (A kiadás előkészítői: Murádin László – Péntek János. MNyTK. 193). Budapest: Magyar Nyelvtudományi Társaság.
- Goebel, Hans 2005. La dialectométrie corrélative. Un nouvel outil pour l'étude de l'aménagement dialectal de l'espace par l'homme. *Revue de linguistique romane* 69: 321–367.
- Goebel, Hans 2010. Dialectometry and quantitative mapping. In: Alfred Lameli – Roland Kehrein – Stefan Rabanus (szerk.): *Language and space*. Vol. 2: *Language mapping*. Berlin: De Gruyter Mouton. 433–457.
- Goebel, Hans 2011. Introduction aux problèmes et méthodes de l'«École dialectométrique de Salzbourg» (avec des exemples gallo-, italo- et ibéroromans). In: Afonso Álvarez Pérez – Ernestina Carrilho – Catarina Magro (szerk.): *Proceedings of the International Symposium on Limits and Areas in Dialectology (LimiAr)*. Lisboa: Centro de Linguística da Universidade de Lisboa. 117–166.
- Heeringa, Wilbert 2004. *Measuring dialect pronunciation differences using Levenshtein distance (Groningen Dissertations in Linguistics)*. Groningen: University of Groningen.
- Heeringa, Wilbert – John Nerbonne 2001. *Dialectareas and dialect continua*. *Language Variation and Change* 13: 375–400.
- Heeringa, Wilbert – John Nerbonne 2013. *Dialectometry*. In: Frans Hinskens – Johan Taeldeman (szerk.): *Language and space. An international handbook of linguistic variation, Volume III: Dutch*. (Handbook of Linguistics and Communication Science (HSK) 30/3). Berlin – New York: Walter de Gruyter. 624–646.
- Imre Samu 1971. *A magyar nyelvjárások rendszere*. Budapest: Akadémiai Kiadó.
- Juhász Dezső 2001. *A magyar nyelvjárások területi egységei*. In: Kiss Jenő (szerk.): *Magyar dialektológia*. Budapest: Osiris Kiadó. 262–324.
- Kessler, Brett 1995. *Computational dialectology in Irish Gaelic*. In: *Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics*. Dublin: EACL. 60–67.
- Király Lajos 2005. *Somogy–zalai nyelvatlasz*. Budapest: Magyar Nyelvtudományi Társaság.
- Kocsis Zsuzsanna – Vargha Fruzsina Sára 2016. *A magyar nyelvjárások atlasza és A romániai magyar nyelvjárások atlasza integrált dialektometriai elemzése*. In: Czetter Ibolya – Hajba Renáta – Tóth Péter (szerk.): *VI. Dialektológiai Szimpozium*. Szombathely – Nyitra: NYME Magyar Nyelvtudományi Intézeti Tanszék, UKF Közép-európai Tanulmányok Kara, Szlovákiai Magyar Akadémiai Tanács. 193–207.
- Levenshtein, Vladimir I. 1966. Binary codes capable of correcting deletions, insertions and reversals. *Cybernetics and Control Theory* 10: 707–710.
- Murádin László (gyűjt.) – Juhász Dezső (szerk.) 1995–2010. *A romániai magyar nyelvjárások atlasza*. I–XI. kötet. Budapest: Magyar Nyelvtudományi Társaság.
- Nerbonne, John – Wilbert Heeringa 2001. *Computational comparison and classification of dialects*. *Dialectologia et Geolinguistica* 9: 69–83.
- Nerbonne, John – Wilbert Heeringa – Eric van den Hout – Peter van de Kooij – Simone Otten – Willem van de Vis 1996. *Phonetic distance between Dutch dialects*. In: Gert Durieux – Walter





- Daelemans – Steven Gillis (szerk.): CLIN VI: Papers from the Sixth CLIN Meeting. Antwerp: Centre for Dutch Language és Speech (UIA). 185–202.
- R Development Core Team 2020. R: A language and environment for statistical computing. Vienna: Foundation for Statistical Computing. <http://www.R-project.org>.
- Simonyi Zsigmond 1889. A magyar nyelv I. A magyar nyelv élete. Budapest: MTA. 187–234.
- Vargha Fruzsina Sára 2013. A hangzó adat szerepe a magyar dialektológiában. In: Szoták Szilvia – Vargha Fruzsina Sára (szerk.): Változó nyelv, nyelvváltozatok, területiség: A VII. Hungarológiai Kongresszus nyelvészeti tanulmányai. Kolozsvár: Egyetemi Műhely Kiadó – Bolyai Társaság. 94–204.
- Vargha Fruzsina Sára 2015. Lexikai, fonológiai, fonetikai stabilitás (és relevancia) a magyar nyelvjárásokban. In: É. Kiss Katalin – Hegedűs Attila – Pintér Lilla (szerk.): Nyelvelmélet és dialektológia 3. Budapest: Szent István Társulat. 246–261.
- Vargha Fruzsina Sára 2017. A nyelvi hasonlóság földrajzi mintázatai. Magyar nyelvjárások dialektometriai elemzése. Budapest: Magyar Nyelvtudományi Társaság.
- Vargha Fruzsina Sára 2018. A zárt *ē* fonetikai minősége a magyar nyelvjárásokban. Előadás a 20. Élőnyelvi Konferencián. Budapest, 2018. augusztus 30.
- Vargha Fruzsina Sára 2020. Magánhangzó-minőségek meghatározása nyelvatlaszadatok alapján. Magyar Nyelv 116: 442–461.
- Vargha Fruzsina Sára 2021. Vélthatárok és valós törésvonalak a nyelvjárás kontinuumában. Plenárius előadás a 21. Élőnyelvi Konferencián. Beregszász, 2021. november 5.
- Vékás Domokos 2007. Számítógépes dialektológia. In: Guttmann Miklós – Molnár Zoltán (szerk.): V. Dialektológiai Szimpozion. Szombathely: Berzsenyi Dániel Főiskola. 289–293.

Vocalism and consonantism in the similarity patterns of Hungarian dialects

Abstract: Phonetic phenomena play a crucial role in the exploration of the relationships between Hungarian dialects, but the question is which categories of these phenomena are the most decisive. In this paper, I examine the impact of vocalism and consonantism on the similarity patterns between dialects and on groupings based on similarity between dialects. I conduct my research on computerised Hungarian atlas data, inspired by previous correlational dialectometric analyses performed on other languages, but without prior manual data classification, instead developing automated analysis methods, thus aiming at greater efficiency and objectivity. I visualize the results on maps, showing the spatial patterns of linguistic similarity as determined by consonantism and vocalism. I argue that mainly the vowel qualities prove to be determinant in the spatial patterns of dialects.

Keywords: quantitative linguistic geography; dialectometry; dialect classification; vocalism; consonantism

