

VARGHA FRUZSINA SÁRA



Budapest, Geolingvisztikai Műhely

fruzsa@gmail.com

## Nyelvatlaszok a szekrényben

2009 tavaszán kollégáimmal Kolozsváron jártunk, bemutatót tartottunk a Geolingvisztikai Műhely nyelvföldrajzi és helynévföldrajzi kutatásairól a magyar szakos hallgatóknak. Az oktatók közül többen is megtiszteltek bennünket jelenlétükkel. Informatikai alapú módszereink későbbi továbbfejlesztéséhez bátorítást jelentett, hogy a nyelvészet és az új technológiák kapcsolódásához különösen értő Szilágyi N. Sándor – kritikus észrevételei mellett is – kedvezően reagált, úgy éreztük, a látottakra.

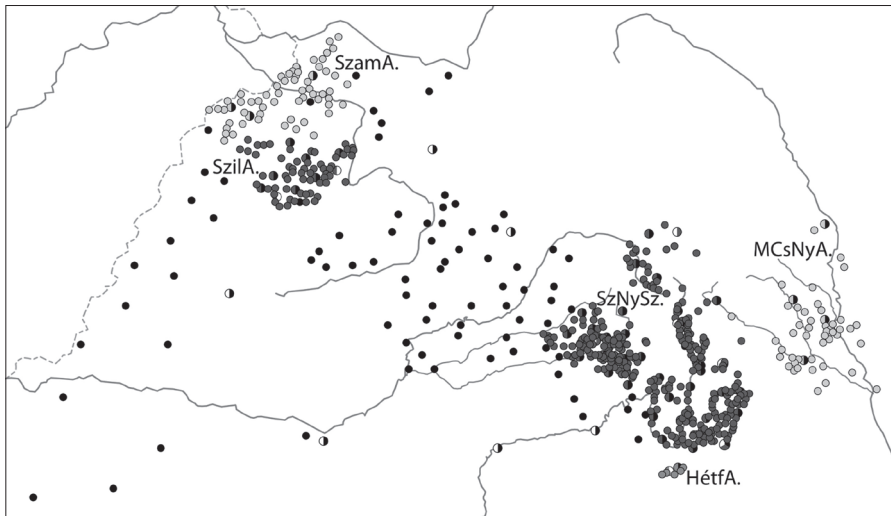
Az előadások után a tanszék vendégeiként bekukkanthattunk a fiókokba és szekrényekbe, s az ott rejtgetett kincsekbe, vagyis a cédulákba és kérdőfüzetekbe, amelyeket a kolozsvári nyelvészeti iskola dialektológusai gyűjtöttek a XX. század közepétől. Az egyik ilyen fiókban rábukkantunk a Szamosháti atlasz céduláira.

Fénymásolatot készítettünk az egyik kérdőfüzetről és egy kötegni céduláról, azzal a szándékkal, hogy megkezdjük az atlasz informatizálásának előkészítését. A munkálatokba Hochbauer Mária, később Gál Noémi is bekapcsolódott, végül ő vállalta az oroszlánrészt: a megtalált kérdőfüzetek adatainak rögzítését. Első lépésként azonban szükség volt a kérdőfüzetek felkutatására, rendszerezésére. 2009 nyarán sor került a tanszéki szekrények átrendezésére, az addig dobozokban és fiókok alján, szekrény tetején rejtőző kérdőfüzetek katalogizálására. A rendszerezést Hochbauer Mária végezte el, Szilágyi N. Sándor segítségével, aki nemcsak nyelvészként, hanem asztalosként is hozzájárult a gyűjtemény összesítéséhez,<sup>1</sup> újfent bizonyítva a tudományterületünkre jellemző interdiszciplináris megközelítés eredményességét. Alig egy év alatt az informatizálásba bevont kolozsvári atlaszok száma eggyel bővült. Jelenleg 22 kutatópont kérdőfüzeteit sikerült megtalálni és számítógépen rögzíteni, a munkafolyamatról Hochbauer Mária számolt be rövidesen megjelenő tanulmányában (Hochbauer, megjelenés előtt).

Eddig összesen hat romániai nyelvjárási adattár (legalább részleges) informatizálása valósult meg, ezek közül a *Szamosháti atlasz* és *A moldvai csángó nyelvjárás atlasza* harmadik kötetének jóval az első két kötet publikálása után előkerült anyaga jelenleg csak számítógépen kutatható (az utóbbira lásd Bodó–Vargha 2008). Befejeződött *A székely nyelvjárás atlasza* háromszéki térképlapjainak rögzítése is, így erről a területről

1 Hochbauer Mária személyes közlése.

már nem csak a *Székely nyelvöldrajzi szótárban* publikált válogatás áll rendelkezésre (Cs. Nagy 2011). Az informatizálásba eddig bevont adattárak kutatópont-hálózatait összesítve az 1. térkép szemlélteti.



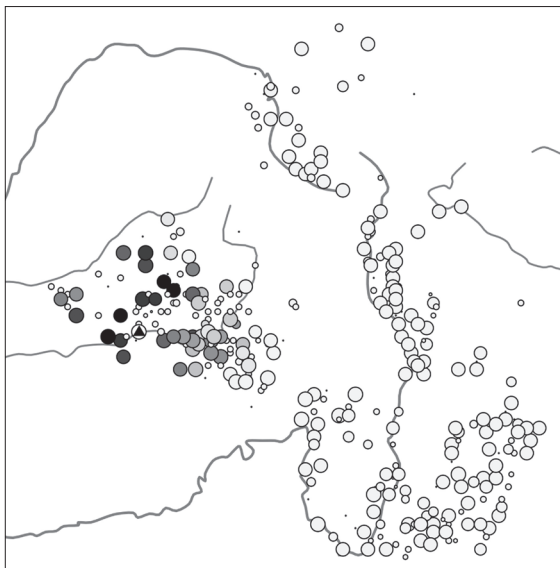
1. térkép. A kolozsvári iskolában készült, már (legalább részben) informatizált nyelvjárás adatárak. SzamA. = Szamosháti atlasz, SzilA. = Szilágysági nyelvátlasz, SzNySz. = Székely nyelvöldrajzi szótár, HétfA. = Hétfalu nyelvjárás atlasza, MCsNyA. = A moldvai csángó nyelvjárás atlasza I-III., fekete szín jelöli A romániai magyar nyelvjárások atlasza kutatópontjait, fehér szín jelöli, ha egy kutatópont egyúttal A magyar nyelvjárások atlaszában is szerepel.

Az 1. térképen is jól látszik, hogy az abszolút sűrűségű kutatópont-hálózzal készült *Székely nyelvjárás atlasz*nak van a legtöbb kutatópontja, az adatok számát tekintve pedig ez lenne a legnagyobb *A romániai magyar nyelvjárások atlasza* után az itt bemutatott nyelvjárás adatárak közül, egyelőre azonban csak az a néhány, főleg lexikai változatosságot mutató címszó kutatható belőle, amelyek helyet kaptak a *Székely nyelvöldrajzi szótárban*. A háromszéki kutatópontok adatainak feldolgozása már megtörtént a kéziratos térképlapok segítségével, az udvarhelyszéki, csíki és gyergyói kutatópontokon azonban még mindig csak a szótár keretein belül kutathatunk. A továbbiakban néhány dialektometriai térkép segítségével azt mutatom be röviden, milyen, eddig kiaknázatlan lehetőségek rejlenek az adattárban a székely nyelvjárások kutatására, illetve rámutatok arra is, mennyire hiányos a szótár anyaga, milyen nagy szükség volna egy teljesebb és pontosabb elemzés elvégzéséhez a teljes adattár földolgozására.

A kvantitatív nyelvöldrajzi kutatások körébe tartozó dialektometria lényege nem egy-egy nyelvi változó elemzése, hanem egy nyelvátlasz (vagy annak legalább száz-kétszáz térképlapja) összes adatának kutatópontonkénti összevetése, ahogyan ezt 2009-ben bemutattuk a *Somogy-zalai nyelvátlasz* és *A magyar nyelvjárások atlasza* (a továbbiakban MNyA.) informatizált részének felhasználásával (Vargha-Vékás 2009). Az adatelemzés többféleképpen is elvégezhető, az itt látható térképek Levenshtein algoritmusának használatával készültek. Ezzel a módszerrel – térképlaponként haladva – páronként összevetjük egymással egy nyelvátlasz kutatópontjainak adatait, mindig

megállapítva a két adat közötti különbözőség (illetve hasonlóság) mértékét, akár több száz kutatópont és térképlap esetében. Az összevetések számszerűsített végeredménye egy hasonlósági mátrix, amely megmutatja, átlagosan milyen arányban mutatnak hasonlóságot egymással az egyes kutatópontok adatai. Így bármelyik kutatópontról megállapíthatjuk, hogy adatai átlagosan mely kutatópontok adataival mutatnak nagyobb, és melyekkel kisebb hasonlóságot. A különféle elemzési módok ismertetésére itt most nem térünk ki, erről l. bővebben Goebel (2006) és Nerbonne–Heeringa (2010).

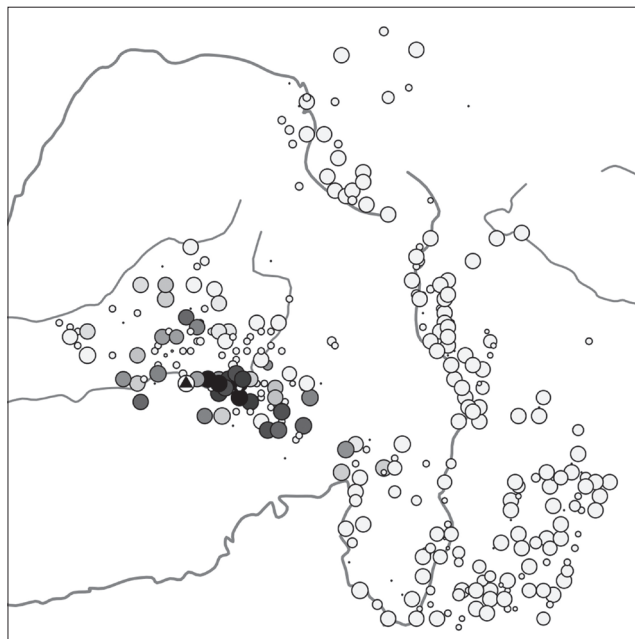
A *Székely nyelvöldrajzi szótár* adataiból készített hasonlósági mátrix alapján interaktív térképet hoztunk létre. A térkép színezése aszerint változik, melyik kutatópontot jelöljük ki az egérrel kattintva. A színek megmutatják, mely kutatópontok mennyire állnak közel nyelvileg a kijelölt kutatóponthoz az adatok összevetésével kialakított hasonlósági mátrix szerint. Az itt bemutatott térképek mindig egy adott kutatópont nyelvi kapcsolatrendszerét mutatják: a kijelölt kutatóponttal nagyobb nyelvi hasonlóságot mutató települések sötétebb, a kisebb nyelvi hasonlóságot mutatók világosabb árnyalatokkal látszanak. A dialektometria mint módszer megbízhatóan alkalmazható bizonyult nyelvjáráshatárok feltérképezésére: ha két szomszédos település nyelvi hasonlósági súlypontja máshová esik, vagyis jellemzően más-más területekkel mutatnak nagyobb nyelvi hasonlóságot, akkor feltehető, hogy köztük nyelvjáráshatár húzódik (Vargha 2010). A dialektometriai elemzéseket sikerrel alkalmazhatjuk a nyelvjárásszigetek nyelvi kapcsolatainak elemzésében is (erről l. Vargha 2011).



2. térkép. Székelykeresztúr (a fekete háromszöggel jelölt kutatópont) dialektometriai térképe a *Székely nyelvöldrajzi szótár* adatai alapján

A 2. térképen Székelykeresztúr nyelvi hasonlósági viszonyait láthatjuk. A sötét színezetű kutatópontok elhelyezkedése alapján leginkább a Székelykeresztúri járáshoz tartozó, tőle nyugatra, északnyugatra fekvő településekkel mutat erősebb nyelvi hasonlóságot. A 3. térképen, ahol már a mintegy 6 kilométerre fekvő Kisgalambfalva a

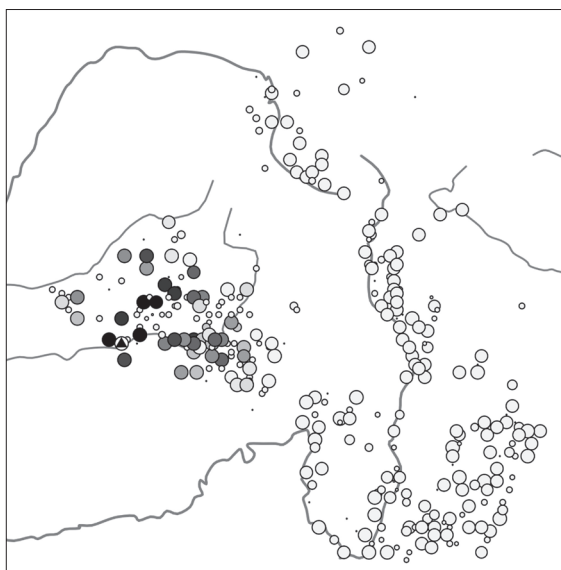
kijelölt kutatópont, egészen más a helyzet, a hasonlósági kapcsolati súlypont határozottan keleti irányba esik. A térképek alapján feltételezhetjük, hogy Székelykeresztúr és Kisgalambfalva között egyfajta nyelvjárási határ húzódik, annak ellenére, hogy mindkét település az egykori Székelykeresztúri járáshoz tartozik. A még pontosabb határ egyébként valószínűleg Kisgalambfalva és az attól Székelykeresztúr felé mintegy másfél kilométerre található Betfalva között húzódhat a rendelkezésünkre álló kevés betfalvi adat alapján, de éppen az adatok csekély száma miatt Betfalva nyelvi hasonlósági viszonyai az itt közölt térképekből nem olvashatók ki.



3. térkép. Kisgalambfalva (a fekete háromszöggel jelölt kutatópont) dialektometriai térképe a Székely nyelvöldrajzi szótár adatai alapján

A térképek színezése természetesen csak akkor tér el jelentősebben egymástól, ha a kijelölt kutatópontok nyelvi hasonlósági viszonyai eltérőek. Újszékely dialektometriai térképe (4. térkép) jól láthatóan eltér az előző (3.) térképtől, amely Kisgalambfalva nyelvi hasonlósági viszonyait ábrázolja, de nem tér el számottevően az Újszékely nyelvjárására inkább hasonlító Székelykeresztúr dialektometriai térképétől (2. térkép).

Az itt bemutatott térképek azonban nem csak az egyes kutatópontok adatai közti összevetés szempontjából tanulságosak, hanem fontos információkkal szolgálnak magáról az adattárról, vagyis a *Székely nyelvöldrajzi szótár*ról is. A kutatópontok mérete az itt bemutatott térképen nem azonos. Egyes kutatópontok apró, mindig világos színű körként, némelyek pedig egészen kicsi pontként jelennek meg a térképen. Ennek magyarázata, hogy az összevetéseknél csak azokat a kutatópontokat vesszük figyelembe, amelyek legalább 150 címszónál szerepelnek. Jellemzően 200–250 adatpáron végezhető el az összevetés. Minél nagyobb egy kutatópont, annál több az összevethető adat.



4. térkép. Újszékely (a fekete háromszöggel jelölt kutatópont) dialektometriai térképe a Székely nyelvfeldrajzi szótár adatai alapján

Készítettem két táblázatot a MNyA. és a Székely nyelvfeldrajzi szótár három-három kutatópontja közti nyelvi hasonlóság mértékéről ezrelékben kifejezve. A táblázatokból kiolvasható, milyen arányban mutatnak egyezést a Székely nyelvfeldrajzi szótár adatai azon kutatópontok esetében, amelyek dialektometriai térképét fentebb bemutattam, illetve milyen mértékű a nyelv(járás)i hasonlóság a MNyA. három székelyföldi kutatópontja között az atlasz 476 fonetikai és morfológiai térképlapja alapján.

	Székelykeresztúr	Újszékely	Kisgalambfalva
Székelykeresztúr	1000		
Újszékely	708	1000	
Kisgalambfalva	688	642	1000

1. táblázat. A Székely nyelvfeldrajzi szótárban szereplő székelykeresztúri, kisgalambfalvi és újszékelyi adatok közti hasonlóság ezrelékben kifejezve

	Kobátfalva	Kászontalíz	Csíkrákos
Kobátfalva	1000		
Kászontalíz	786	1000	
Csíkrákos	720	780	1000

2. táblázat. A magyar nyelvjárások atlasza három kutatópontjának adatai közti hasonlóság ezrelékben kifejezve az atlasz 476 fonetikai és morfológiai térképlapja alapján

A táblázatokban szereplő számok összevetéséből az látszik, hogy az egymástól távolabb fekvő, így minden bizonnyal nyelvileg is kisebb mértékben hasonló MNyA.-kutatópontok nagyobb arányban mutatnak egyezést, mint a *Székely nyelvfeldrajzi szótár* térben és valószínűleg nyelvjárásában is közelebb eső három kutatópontja. A számok alakulása elsősorban az elemzett térképlapok, illetve szócikkek eltérő jellegével magyarázható. Míg a *Székely nyelvfeldrajzi szótár* lexikai szempontból érdekes címszók gyűjteménye *A székely nyelvjárás atlaszából*,<sup>2</sup> addig a MNyA. sűrű kutatópont-hálózatú térképei fonetikai, illetve morfológiai jelenségeket ábrázolnak. A lexikai különbségek pedig nagyobb eltérésekkel járhatnak együtt az adatok összevetésekor, mint a fonetikai és a morfológiaiak: ezrelékben kifejezve is nagyobbak lesznek a különbségek az egyes kutatópontok adatai között.

Valószínű, hogy *A székely nyelvjárás atlasza* adatainak feldolgozásával számos, a szótárból kimaradt adat pótolható volna. Erre engednek következtetni Cs. Nagy Lajos összevetései is, amelyeket a szótár háromszéki adatai és a Háromszéki atlasz (*A székely nyelvjárás atlaszának kéziratos háromszéki térképlapjai*) alapján végzett el (Cs. Nagy 2011). Így a térképeken most apró karikák nagyobbak lehetnének, az adatelemzés pontosabb képet mutatna.

A nyelvjárási tipológiák, nyelvjárások közötti összevető vizsgálatok alapja elsősorban a hangtan. A dialektometriai elemzéseknél is szerencsésebb lenne, ha informatizált formában rendelkezésünkre állnának a székely atlasz hangtani szempontból érdekes térképlapjainak adatai is.

Bízom benne, hogy a még földolgozatlan kérdőfüzetek, cédulák nem pihennek sokáig a tanszék szekrényeiben, lesz, aki fölkarolja őket, és az informatizálás révén használhatóvá teszi a nyelvjárások iránt érdeklődők számára.

## Hivatkozások

- Hochbauer Mária (megjelenés előtt). A Szamosháti nyelvatlasz informatizálásának munkálatai. In: Szoták Szilvia – Vargha Fruzsina Sára (szerk.): *Változó nyelv, nyelv-változatok, területiség. A VII. Hungarológiai Kongresszus nyelvészeti tanulmányai*. Bolyai Társaság, Kolozsvár.
- Bodó Csanád – Vargha Fruzsina Sára 2008. Régi nyelvatlaszok – új módszerek. *Magyar Nyelv* 104: 335–351.
- Gálffy Mózes – Márton Gyula 1987. *Székely nyelvfeldrajzi szótár*. Akadémiai Kiadó, Budapest.
- Goebel, Hanns 2006. Recent Advances in Salzburg Dialectometry. *Literary and Linguistic Computing*, 21: 411–435.
- Cs. Nagy Lajos 2011. Úton a Háromszéki nyelvatlasz informatizálása felé. In: Bárh M. János – Vargha Fruzsina Sára (szerk.): *Hangok – helyek*. ELTE Magyar Nyelvtudományi és Finnugor Intézet, Budapest. 61–73.

2 „Itt hívjuk föl a figyelmet arra, hogy ebben a kötetben a felgyűjtött anyagnak csupán a lexikális részét tesszük közzé. A hang- és alaktani anyag kiadásra vár még!” (Gálffy–Márton 1987: 9)

- Nerbonne, John – Heeringa, Wilbert 2010. Measuring Dialect Differences In: Schmidt, Jürgen Erich – Auer, Peter (eds.): *Language and Space. An International Handbook of Linguistic Variation: Theories and Methods*. Mouton de Gruyter, Berlin. 550–567.
- Vargha Fruzsina Sára 2010. A dialektometria alkalmazása és történeti helynevek nyelvföldrajzi vizsgálata a Székelyföldön. *Helynévtörténeti Tanulmányok* 5: 223–233.
- Vargha Fruzsina Sára 2011. A romániai magyar nyelvjárások atlasza informatizált térképlapjainak kvantitatív nyelvföldrajzi vizsgálata. Előadás az *Egy elkészült és egy készülő magyar nyelvátlasz – kutatási tapasztalatok és perspektívák* című tudományos ülésen. MTA, 2011. november 23. [http://www.bihalbocs.hu/eloadas/varghafru\\_rmnya\\_kvantitativ\\_20111123.pdf](http://www.bihalbocs.hu/eloadas/varghafru_rmnya_kvantitativ_20111123.pdf)
- Vargha Fruzsina Sára – Vékás Domokos 2009. Magyar nyelvjárási adattárak vizsgálata interaktív dialektometriai térképekkel. Előadás a Magyar Nyelvtudományi Társaság felolvasóülésén. [http://bihalbocs.hu/eloadas/dialektometria\\_20090324.pdf](http://bihalbocs.hu/eloadas/dialektometria_20090324.pdf)

### Summary

#### *Linguistic Atlases in the Cupboard*

The scholars of the linguistic school of Kolozsvár collected dialect data in the second half of the 20th century for several regional and local linguistic atlases. The digitization of six of these databases has been started or already accomplished, opening the way to quantitative analyses. The analysis of aggregate data, called dialectometry, makes dialect classification more objective as compared to traditional approaches where the number of variables involved is limited. In the present study, based on the analysis of data coming from the Székely Linguistic Geography Dictionary, it is illustrated that dialect borders can be traced using dialectometry as a tool for measuring dialect distances between language varieties. The outcome of the aggregate analysis, represented on interactive maps, reveals also some limitations of the dictionary that is a selection from the lexical data of the still unpublished Székely Linguistic Atlas. Comparing degrees of similarity between three nearly neighboring localities of the dictionary and those between three more distant localities on the dialectometric map of The Atlas of Hungarian Dialects, the absolute values of similarity prove to be higher in the case of the more distant investigation points of the atlas. The different nature of the data may account for this unexpected difference: while only lexical data could be used from the dictionary, in the case of The Atlas of Hungarian Dialects the analysis is based on data coming from phonetic and morphological maps. As dialect classification is usually based on phonetic rather than lexical correspondences, the digitization of the whole Székely Linguistic Atlas, including the still unpublished phonetic and morphological maps, is needed as a precondition for a more reliable dialectometric analysis and for assigning dialect boundaries.

