

# Nyelvjárási és helynévtörténeti anyagok geolingvisztikai kutatása

## Részletes zárójelentés

### 1. A magyar nyelvjárások atlasza hangfelvételeinek feldolgozása

#### 1.1. A feldolgozás során alkalmazott informatikai eljárások

A hanganyagok átírása a Vékás Domokos által fejlesztett Bihalbocs dialektológiai szoftver segítségével történt (Vékás 2007, honlap: <http://www.bihalbocs.hu>). A szaktudományos feladatok megoldásához speciális szoftverek használata szükséges, ugyanis a kereskedelmi forgalomban beszerezhető szoftverek (szövegszerkesztők, hangszerkesztők, adatbázis-kezelők, grafikus programok, térinformatikai alkalmazások) csupán az igények töredékének kiszolgálására lennének alkalmasak. A Bihalbocs olyan lejegyzésszerkesztő (speciális szövegszerkesztő), hangkezelő és térképgeneráló szoftver, amely egyrészt lehetővé teszi nyelvjárási adatok megfelelően kódolt rögzítését és adattárakba rendezését, másrészt az adatok sokoldalú felhasználását a kutatói igények szerint.

A nyelvjárási jelenségek kódolására olyan analitikus kódrendszert alkalmazunk, amely a teljesség igényével optimálisan biztosítja az adatrögzítés pontosságát és az informatizált anyagok hatékony felhasználását a későbbiekben (Vékás 1999).

A hanganyagok feldolgozása a Vékás Domokos által erre a célra kifejlesztett speciális BXT fájlformátumban történik, amely lehetővé teszi lejegyzett nyelvjárási adatokat tartalmazó szövegfájlok és WAV formátumú hangfájlok szinkronizálását. Vagyis a lejegyzett szövegek nem csupán magukat a nyelvjárási adatokat tartalmazzák, hanem időzítési markereket is, amelyek a vonatkozó hangfájlok megfelelő időpillanataira utalnak a másodperc töredékének pontosságával. A fájlok beolvasásakor a szövegben megjelenített markerek segítségével bármikor meghallgathatjuk a lejegyzett nyelvjárási adatokat, szövegrészleteket, illetve akusztikai elemzéseket végezhetünk, anélkül, hogy megkeresnénk és megnyitnánk a szövegfelvételt tartalmazó hangfájlt. A Bihalbocs dialektológiai szoftver segítségével lehetőségünk van az így feldolgozott adatok adatbázisba rendezésére (az adatok itt is megőrzik kapcsolatukat az eredeti hangfájlokkal), a nyelvjárási átiratokban történő keresésre, az adatok alapján más programok (szövegszerkesztők, statisztikai alkalmazások, internetböngészők) számára is kezelhető térképes kimutatások, táblázatok készítésére.

Mivel a létrehozott anyagok használatát hosszú évtizedekre biztosítani szeretnénk, a legegyszerűbb kódolási és tárolási formátumokat alkalmazzuk, hogy a rendszeresen megújuló informatikai platformok közötti hordozhatóságot minél jobban elősegítsük. („Legegyszerűbb” megoldáson természetesen a feladat összetettségéhez viszonyított egyszerűséget értjük.) A BXT-fájlok szerkezete jól strukturált és viszonylag egyszerű, így az ebben a formátumban tárolt anyagokat a jövőben esetleg más, erre a célra kifejlesztett (dialektológiai) szoftverek is kezelni tudják majd.

#### 1.2. Szövegfelvételek feldolgozása és közreadása

Kutatásunkban vállaltuk, hogy folytatjuk a 2005-ben elindított Magyar nyelvjárási hangoskönyv sorozatot, és további 6 multimédiás CD-ROM-on adunk közre hanggal szinkronizált nyelvjárási lejegyzéseket összesen 29 településről. A hangoskönyvek A magyar nyelvjárások atlasza ellenőrző gyűjtésének digitalizált anyagából olyan felvételeket adnak közre, amelyeknek a fonetikai átírása

már megjelent a Hajdú Mihály és Kázmér Miklós által készített Magyar nyelvjárási olvasókönyvben (1974).

A sorozat kiadványaihoz két változatot hoztunk létre: a Bihalbocs dialektológiai szoftverrel kialakított kutatói változatot, valamint az abból a Bihalboccsal részben automatikusan generált ismeretterjesztő változatot. A feldolgozás során az eredeti hangfelvételeket és fonetikus lejegyzésüket a szövegben elhelyezett időzítési markerek segítségével összekapcsoljuk, szinkronizáljuk. Így a lejegyzés megőrzi kapcsolatát az eredeti hanganyaggal, bármely hangrészlet elérhető, meghallgatható a megfelelő billentyű lenyomására.

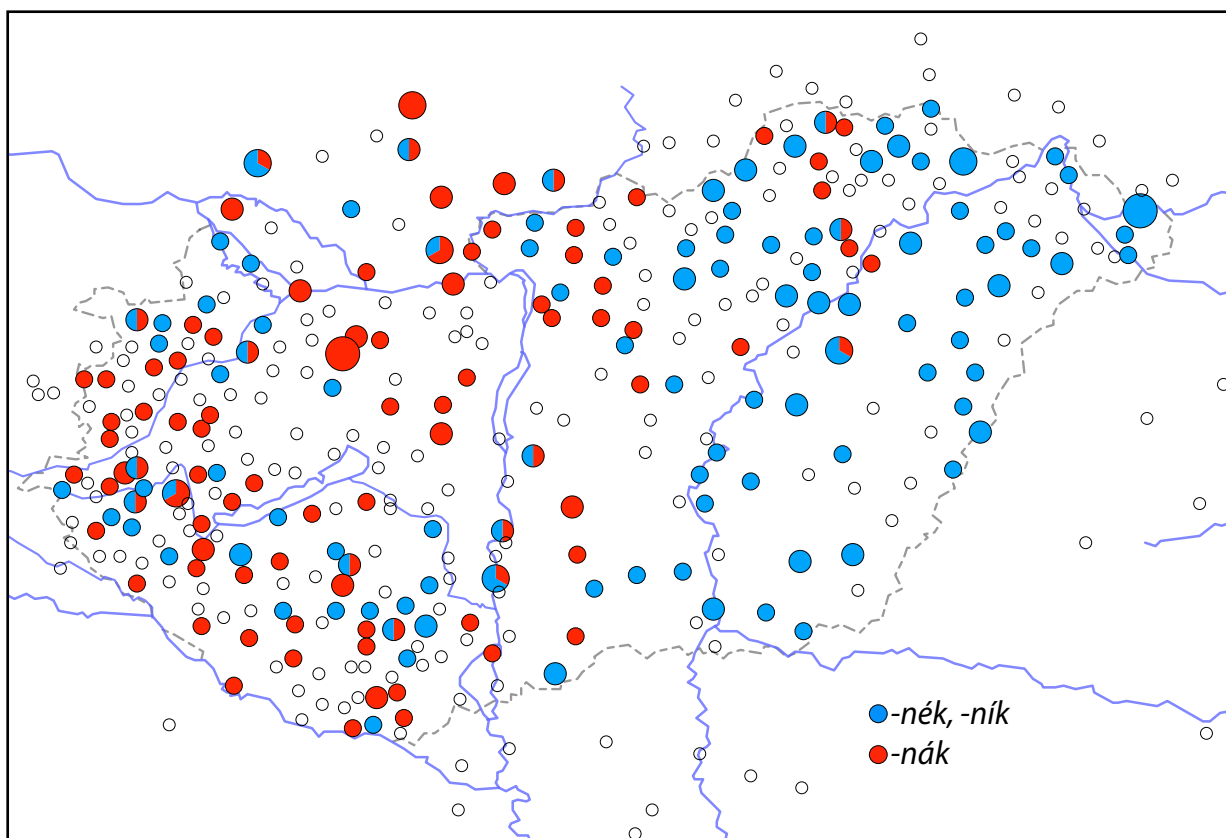
20 kutatópont anyagának kutatói változatát már egy korábbi NKFP-projektben elkészítettük, ebből jelenlegi projektünkben négy kiadványt készítettünk el, melyeket még 2008-ban megjelentettünk. 2009-ben további 9 kutatópont anyagát dolgoztuk föl, melyekből elkészítettük és két CD-ROM-on megjelentettük az ismeretterjesztő változatot is. Így vállalásainknak megfelelően összesen 29 kutatópontról adtuk közre az olvasókönyv szövegeit eredeti hangzó változatukkal szinkronizált formában, összesen 6 multimédiás CD-ROM formájában. Az ismeretterjesztő változatokat közzétettük az Interneten is, a Geolingvisztikai Műhely honlapján (<http://geolingua.elte.hu>). Vállalásainkon túl, a sorozat kiegészítéseként az ELTE digitális nyelvjárási hangarchívumában meglévő felvételek alapján elkészítettük a Magyar nyelvjárási olvasókönyv további 12 kutatópontján a lejegyzések eredeti hanganyagukkal szinkronizált változatát. Ez a 12 kutatópontnyi anyag egyelőre kutatói változatban érhető el.

A Magyar nyelvjárási hangoskönyvek Bihalbocsban kialakított speciális fájlformátuma lehetővé teszi, hogy az egységes szempontok szerint, egységes struktúrában rögzített szövegeket a későbbiekben kereshető adatbázisként működtessük. Minden átiratban rögzítjük a gyűjtés helyét, időpontját és az adatközlőre vonatkozó információkat is. Ezen információk alapján a megfelelő technológia kialakításával, továbbfejlesztésével lehetővé válhat a szövegek többszempon-tú keresése, a találatok térképezése is. A Bihalbocs konverziós képességei azt is biztosítják, hogy a hanganyagukkal szinkronizált szövegeket más formátumokba (HTML, PRAAT TextGrid) konvertáljuk az újabb és újabb kutatási céloknak, igényeknek megfelelően.

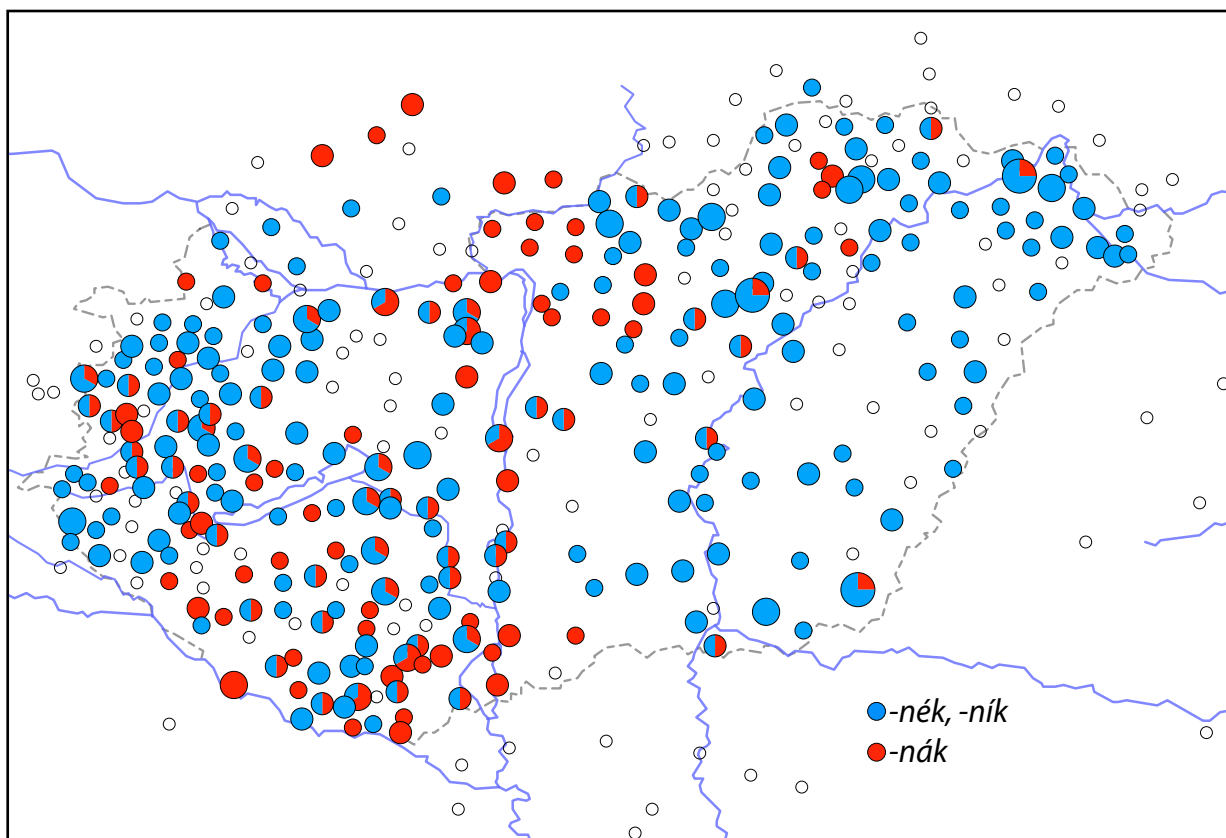
A hangoskönyveket bemutattuk a kolozsvári egyetemen, ahol Péntek János professzor meghívásának eleget téve bemutatkozó előadást tartottunk 2009 áprilisában. Kiadványaink jól hasznosíthatók oktatási segédanyagokként az egyetemeken, illetve akár a középiskolákban is. Széleskörű felhasználhatóságukat jelzi, hogy helyet kaptak a Petőfi Irodalmi Múzeum 2009-ben rendezett *Nyelvet öltünk!* című, a magyar nyelvről szóló vándorkiállításán is.

### 1.3. Beszélő atlasz

A MNyA. ellenőrző gyűjtésének hanganyaga a szövegfelvételek mellett tartalmazza annak a 41 atlaszkérdésnek a válaszait is, amelyek a papíratlasz elkészítése során nem kerültek térképre, mert lejegyzésüket az atlasz készítői problematikusnak ítélték. Tervezték ugyan a felvételen rögzített adatok lejegyzését (sőt, eszközfonetikai elemzéseket is), korszerű, a hangzó anyagot alapul vevő feldolgozásukra azonban annak idején nem volt lehetőség. Az atlaszkérdésekre adott válaszok rögzítéséhez szükséges struktúrák kidolgozását már egy korábbi projektben (NKFP 5/056/2004, vezető kutató Kiss Jenő) elvégeztük, a 352-ből összesen 50 kutatóponton lejegyeztük az adatokat is. Jelen pályázatunkban további 302 kutatópont anyagának (vagyis a teljes gyűjtés) feldolgozását vállaltuk. Terveinknek megfelelően valamennyi kutatóponton elvégeztük az adatrögzítést.



1. térkép: A nákolás területi megoszlása a 31. *szívnék* címszó adataiban, a 61 évesnél fiatalabb adatközlők válaszaiban.



2. térkép: A nákolás területi megoszlása a 31. *szívnék* címszó adataiban, a 60 évesnél idősebb adatközlők válaszaiban.

A feldolgozás alapegysége az interjú, így az adatok nem csak kutatóponthoz, hanem azon belül adatközlőhöz is kötődnek. Minden adat megőrzi kapcsolatát eredeti hangzó formájával az adatbázisba rendezés után is, így az adatok azonnal meghallgathatók, ami biztosítja a fonetikus lejegyzés egyszerű, hatékony ellenőrzését és javítását is. Mivel minden adat őrzi kapcsolatát az adatközlővel, az adatbázisban való keresés során nem csak az adatok attribútumai (adatközlői, kutatói minősítések), hanem az adatközlők neve, életkora szerint is lehet keresni. A találatok csoportosíthatók, különböző formában térképezhetők. Mintaként közlünk két térképet, amelyeket a 31., *szívnék* címszóra adott válaszokból hoztunk létre. Mindkét térképlap a nákoló (*színák*), illetve a nem nákoló (*színek*, *színik*) adatok előfordulását mutatja. Az 1. térkép a 61 évesnél fiatalabb, a 2. térkép a 61 éves vagy annál idősebb adatközlők válaszai alapján készült. A korosztályok szerinti csoportosítás markáns különbséget mutat, leginkább a nyugat-dunántúli kutatópontokon.

A feldolgozás eredményeképpen egy 43 489 adatot tartalmazó adatbázist hoztunk létre, amely a Bihalbocs dialektológiai szoftverrel kereshető, térképezhető. Az adatok csoportosítása révén létrehozott színes térképek PDF-fájlként menthetők, illetve HTML-es beszélő térképekké is konvertálhatók. A térképek internetes változatait rövidesen elérhetővé tesszük a Geolingvisztikai Műhely honlapján (link: <http://geolingua.elte.hu/maps/mnya/ma2>). Terveink szerint az adatbázist Vargha Fruzsina Sára mutatja majd be 2012 augusztusában, Szegeden, a 17. Élőnyelvi Konferencián rendezendő számítógépes dialektológiai műhelyben.

## **2. Szabó T. Attila erdélyi történeti helynévgyűjtésének informatizálása**

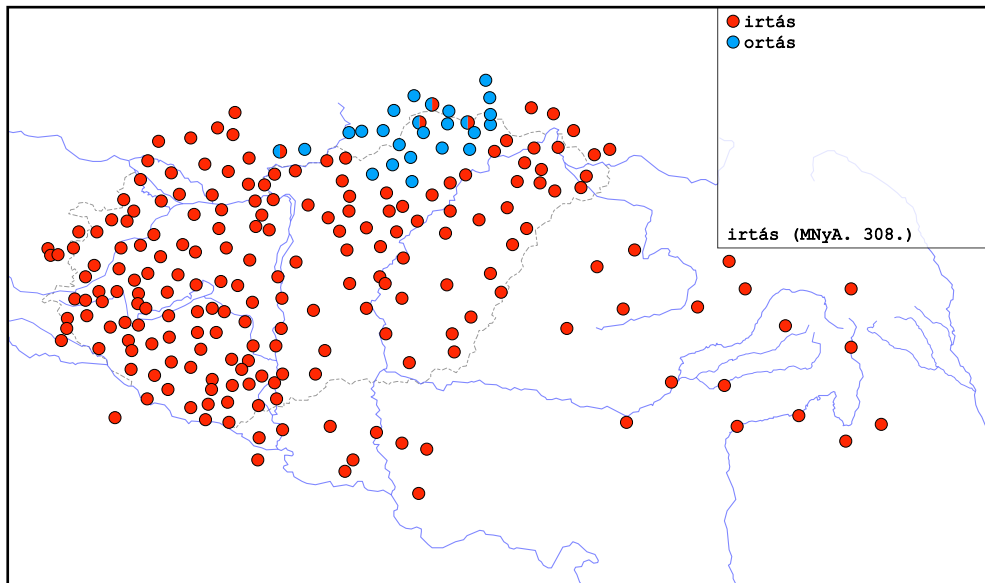
A kutatási tervnek megfelelően megtörtént Szabó T. Attila erdélyi történeti helynévgyűjtéséből 8 kötetnek (Udvarhelyszék, Csík-Gyergyó- és Kászonszék, Kis- és Nagyküüllő, Alsófehér, Szilágy, Maros-Torda A-B, Torda-Aranyos) informatizálása és helynévfajták szerinti annotálása (kb. 2400 nyomtatott oldal).

Az informatizálás módszerei lehetővé tették, hogy megőrződjenek a nyelvtörténeti-írástörténeti kutatások szempontjából kiemelkedően fontos különleges, mellékjeles karakterek, viszont mégsem akadályozzák az anyagban való keresést.

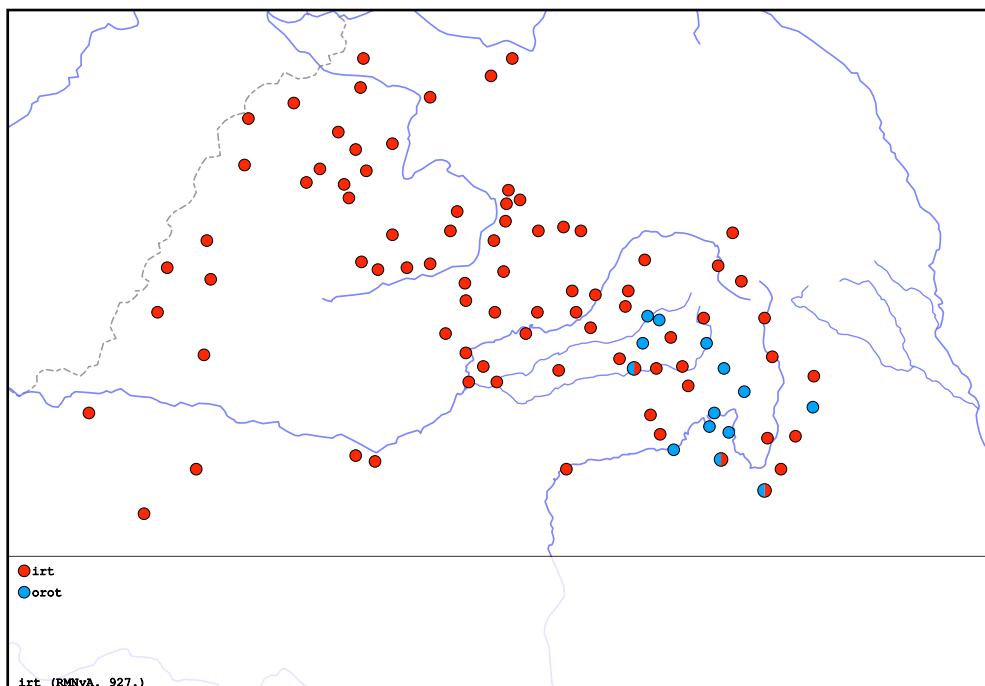
Folyamatosan fejlődött a névadatbázis kezelésére íródott szoftver: kereshetőek benne csonkolt szóalakok, betűkapcsolatok; a keresések kombinálhatók; a találatokat csoportosíthatjuk, a csoportosításokat elmenthetjük; a találati listából kizárhatók nem kívánt találatok; szűkíthetjük a találatokat névfajta, év, település, csoport szerint; listázhatjuk az eredményeket a települések, illetve az előfordulás idejének sorrendjében; elmenthetjük az adatokat külön adatbázis-fájlként; a kapott adatokat interaktív térképekre vetíthetjük, a térképezett adatokat egyszeri előfordulásként, illetve arányosítva is megjeleníthetjük; a létrehozott különböző csoportokat különféle színekkel ábrázolhatjuk a térképeken; az interaktív térképeket átméretezhetjük, az egyes kutatópontokhoz rendelt adatok számát külön szövegdobozból leolvashatjuk, és a térképeket képfájlként elmenthetjük.

Az informatizált adattár kialakítása közben a pályázatban résztvevő munkatársak több publikációban, előadásban mutatták be, hogy milyen célokra használható sikeresen a létrehozott elektronikus adatbázis és a kezelőszoftver. Mindezt legteljesebben Bárány M. János foglalta össze doktori disszertációjában, amelyben legnagyobb terjedelemben a helyneveket alkotó közzavak földrajzi elterjedésével, illetve különböző helynévtípusok földrajzi jellemzőivel foglalkozott. Ezekon kívül a dolgozat több példán keresztül mutatta be, hogy a számítógépes módszerekkel feldolgozott dialektológiai atlaszok adatainak ütköztetése a hasonló módszerekkel informatizált és

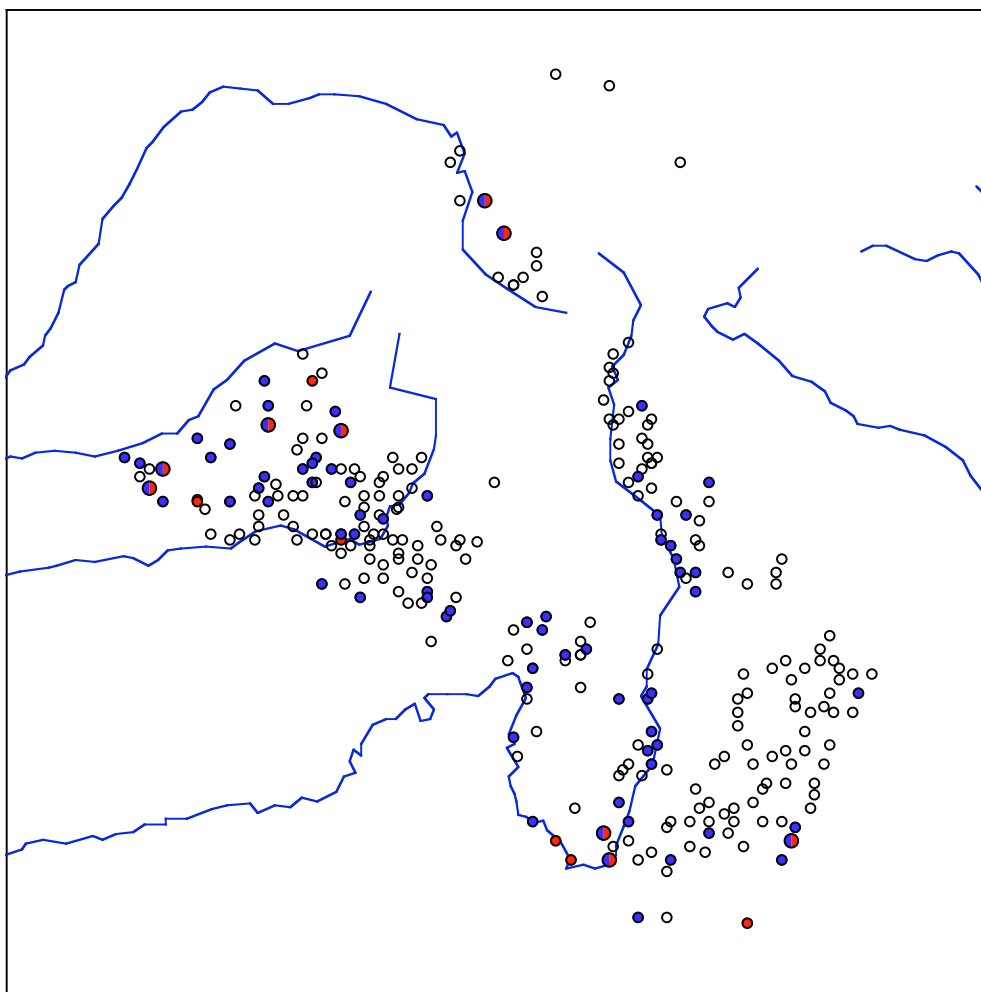
ábrázolt történeti helynévadatokkal kiválóan alkalmas nyelvjárástörténeti következtetések levonására is. Mindezt jól illusztrálja az alábbi három térkép: a 3. térkép a Magyar nyelvjárások atlasza, a 4. a Romániai magyar nyelvjárások atlasza adatai alapján mutatja, hogy irtás (pirossal) vagy ortás ~ orotás (kékkel) típusú alakok fordulnak elő a magyarországi, illetve az erdélyi nyelvjárásokban, az 5. térkép pedig a Szabó T. Attila által gyűjtött székelyföldi történeti helynévadatokban mutatja az ilyen szavakat tartalmazó névadatok területi előfordulásait (azonos színekkel). A térképek a mély hangrendű változat korábbi nagyobb elterjedését, az orotás-típusú formák fokozatos helynévi archaizmussá válását illusztrálják.



3. térkép: Az *irtás* és az *ortás* változat területi megoszlása a MNyA. adataiban



4. térkép: Az *irt* ige alakváltozatainak területi megoszlása A romániai magyar nyelvjárások atlasza adataiban



5. térkép: Az *irtás* és az *orotás* adatok területi megoszlása a helynévtörténeti adatokban a Székelyföldön

A létrejött adattár hiánypótló jellegét támasztja alá, hogy az erdélyi területről illetve a 16–19. századból kevés névgyűjtés hozzáférhető. A pályázat futamideje alatt, a Hangok-helyek műhelytalálkozót követően több olyan kezdeményezés is indult az ország különböző tudományos műhelyeiben, amely ezen adattár anyagának (és részben az informatizálás módszerének) felhasználásával egy összmagyar digitális helynévtár létrehozását tűzte ki célul.

Szabó T. Attila erdélyi történeti helynévgyűjtésének informatizálási munkálatairól több hazai és külföldi (Barcelona – ICOS, Kolozsvár – Hungarológiai Kongresszus) konferencián számoltak be a pályázat munkatársai, eredményeiket pedig több publikációban tették közzé.

### 3. Hangok – helyek műhelytalálkozó szervezése

A különböző műhelyek közti együttműködés megteremtése, erősítése érdekében 2009 októberében a magyar dialektológia és helynévkutatás több magyarországi és határon túli kutatójának részvételével Hangok – helyek címmel találkozót szerveztünk, amelynek témája a nyelvjárási adattárak és helynévtárak számítógépes feldolgozása, informatizálása volt. A rendezvényt követően több magyarországi és határon túli kutatóhelyen is új lendületet kapott a korábbi nyelvjárási anyagok feldolgozása, elindultak új számítógépes dialektológiai projektek, illetve megkezdődött a megyei helynévtárak informatizálása is.

A találkozón elhangzott előadásokból tanulmánykötet is készült, melyet 2011-ben jelentettünk meg.

#### **4. Eltérések az eredeti vállalásoktól és indoklásuk**

A kutatást eredetileg 2011. augusztus 31-ig terveztük, de kénytelenek voltunk halasztást kérni, amire az engedélyt meg is kaptuk. A halasztásra elsősorban a kutatásban résztvevő munkatársak változása miatt volt szükség (Bárth János tanársegédként kapott állást az ELTE Magyar Nyelvtörténeti, Szociolingvisztikai, Dialektológiai Tanszékén 2010 szeptemberétől, Vargha Fruzsina Sára 2011 májusától elment GyED-re, helyettük Trunchi Péter folytatta a lejegyzői munkát tudományos segédmunkatársként), másrészt A magyar nyelvjárások atlasza ellenőrző gyűjtésének atlaszadataiból kialakított adatbázis ellenőrzési munkálatai a tervezettnél időigényesebbnek bizonyultak (itt is megemlítjük, hogy a nagyatlasz készítői annak idején éppen ezeket az atlaszkérdéseket ítélték legproblémásabbnak).

A kutatás zárásaként terveztük a hanggal szinkronizált atlaszadatok multimédiás, “beszélő atlasz”-ként történő megjelentetését DVD-n, illetve egy interaktív helynévtörténeti térképeket tartalmazó CD-kiadvány közzétételét. Mivel azonban az említett adathordozók időközben elavultak, és a modern számítógépekbe már nem is építenek be CD-, illetve DVD-olvasókat, az Internetes közzététel mellett döntöttünk ([http://geolingua.elte.hu/proj\\_frameOTKA\\_hu.html](http://geolingua.elte.hu/proj_frameOTKA_hu.html)). Technikai okok miatt (a hanganyagok nagy tárhelyigénye) a beszélő térképeket röviddel a projekt zárását követően fogjuk tudni fölteni a Geolingvisztikai Műhely honlapjára, egyelőre az első három címszó anyaga érhető el interaktív, multimédiás térkép formájában. Az Internetes közzététel előnye, hogy az adatok lejegyzésének későbbi esetleges változtatása, finomítása esetén újragenerálhatjuk az érintett térképek internetes változatát is, illetve a későbbiekben is lesz lehetőségünk az adatok csoportosításába új szempontokat bevezetni, és további színes térképekkel kiegészíteni a honlapot.

#### **Hivatkozások**

Hajdú Mihály – Kázmér Miklós 1974. *Magyar nyelvjárasi olvasókönyv*. Budapest.

Vékás Domokos 1999. Informatikai lehetőségek a dialektológiában, különös tekintettel a fonetikai szempontokra. Kézirat. Budapest: ELTE Fonetikai Tanszék.

Vékás Domokos 2007. Számítógépes dialektológia In: Guttmann Miklós – Molnár Zoltán (szerk.): *V. Dialektológiai Szimpozion*. Szombathely. 289–293.