

*Bárth M. János*

## **Háromszéki helynevek nyelvészeti elemzése informatikai módszerekkel**

Tanulmányomban SZABÓ T. ATTILA által gyűjtött erdélyi történeti helynevek elemzési, felhasználási lehetőségeit igyekszem bemutatni, és az e célra létrehozott számítógépes szoftver elméleti alapjairól, működéséről és széleskörű alkalmazhatóságáról fogok szólni. Témám tehát leginkább módszertani jellegű.

A történeti és jelenkori helynevek folyamatos gyűjtése és közzététele évtizedek óta változatlan feladata a névtudománynak. De egyre időszerűbb az összegző, elemző, rendszerező munka, amihez szükség van olyan korszerű eszközökre, melyek a napjainkra összegyűlt hatalmas adatmennyiség kezelésére, átfogó következtetések levonására alkalmasak. Az informatika lehetőségeivel él például a Debreceni Egyetem helynévkutató műhelye, ahol évek óta folyik GYÖRFFY GYÖRGY történeti földrajzának névtani adatbázissá alakítása, amelynek felhasználásával már jó néhány tanulmány született.

Az ELTE Névkutató Munkaközösségében folyik HAJDÚ MIHÁLY irányításával SZABÓ T. ATTILA kéziratban őrzött erdélyi történeti helynévgyűjtésének feldolgozása és kiadása immár hat éve. (SZABÓ T. ATTILA Erdélyi helynévtörténeti adattáráról és ennek kiadásáról lásd: BÁRTH M. 2006). Egy éve indult el az adattár informatizált változatának elkészítése egy NKFP-projekt keretében, amely elsősorban digitális nyelvföldrajzi adatbázisok létrehozását tűzte ki célul (a projektről bővebb információk a Geolingvisztikai Műhely honlapján olvashatók: <http://geolingua.elte.hu>).

Ez a munka arra irányul, hogy SZABÓ T. ATTILA könyv formában kiadott történeti helynévgyűjtésének informatizált szótárszerű feldolgozását elkészítsük, ehhez egy olyan számítógépes felületet hozunk létre, amely maximálisan alkalmas a tulajdonnevek és alkotóelemeik több szempontú keresésére, csoportosítására, és a kapott csoportok elemzésére, rendezésére időbeli-térbeli elhelyezkedésük alapján.

Az ehhez szükséges szoftver hosszas tervezés, javíthatóság után 2006 elejére született meg, a program névtani, elméleti alapjaihoz HAJDÚ MIHÁLY nyújtott fontos tanácsokat e sorok írójának, a számítógépes nyelvészeti technológia megalkotása VARGHA FRUZSINA és VÉKÁS DOMOKOS érdeme. SZABÓ T. ATTILA gyűjtéséből Háromszék anyagán kezdtük el tesztelni, alkalmazni a szoftvert, amelyet ennek példáján keresztül mutatok be a következőkben.

A helynévgyűjtemény kötetait lapozgatva könnyen föltűnhet az úgynevezett különleges karakterek nagy száma: a tördelés közben 20–30 betűkészletre volt szükség a több száz régi karakter megjelenítésére. Az informatizálás első fontos részfeladatákként ezek helyett elkészült egy olyan kódrendszer és egy hozzá kapcsolódó betűkészlet, amelyben minden szükséges alapjel és mellékjel megtalálható. Ebben a kódolásban a történeti szövegekben előforduló, a mai helyesírásból azonban hiányzó ékezeteknek külön kódok felelnek meg. Így a kódrendszer viszonylag kevés elemből áll, egyetlen betűkészletben is elfér.<sup>1</sup>

A munkafolyamat második lépéseként az anyagot olyan kicsi egységekre kellett bontani, amelyekből egy-egy adat locusa már könnyedén megkereshető, a megyére, településre, forrásra, évre utaló információkkal együtt. Ezek az egységek tulajdonképpen megfeleltethetők egy-egy SZABÓ T. ATTILA-féle cédulának.

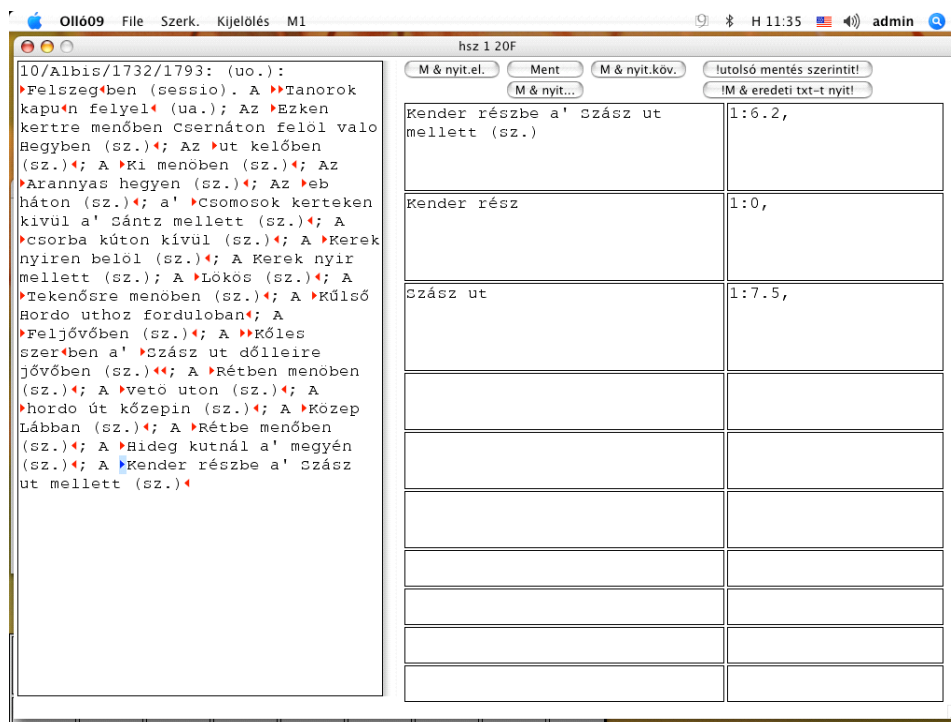
bugyogo 1,hsz 69 9: 1☐	Catharina 1,hsz 51 3: 1☐
Bugyogo 4,hsz 69 8: 1,hsz 81 12: 1,hsz 89 1: 2☐	cenosum 1,hsz 16 6: 1☐
Bugyogonál 1,hsz 89 1: 1☐	úgieo[n] 1,hsz 108 21: 1☐
Bugyogonál(k.) 1,hsz 69 8: 1☐	úgy 5,hsz 74 1: 1,hsz 75 1: 1,hsz 82 6: 1,hsz 111 6: 1,hsz 112 4: 1☐
Bukor 1,hsz 89 1: 1☐	úgynevezett 1,hsz 38 16: 1☐
Buksa 1,hsz 58 7: 1☐	chegel 2,hsz 63 1: 1,hsz 68 1: 1☐
Buksorom 2,hsz 7 8: 2☐	Chiher 1,hsz 5 2: 1☐
Buna 1,hsz 79 23: 1☐	Chÿernaton 1,hsz 4 4: 1☐
Bundok 1,hsz 9 7: 1☐	Circa 1,hsz 53 2: 1☐
Burde 1,hsz 16 24: 1☐	circiter 2,hsz 89 1: 1,hsz 92 58: 1☐
burdeján 1,hsz 5 2: 1☐	Circum 1,hsz 7 15: 1☐
Burdé 1,hsz 16 24: 1☐	Citigiosam 1,hsz 16 6: 1☐
Burdéba 2,hsz 71 16: 1,hsz 71 18: 1☐	új 2,hsz 74 1: 1,hsz 88 10: 1☐
Burdéj 1,hsz 79 23: 1☐	újfalú 1,hsz 74 6: 1☐
burdejoknál 1,hsz 89 1: 1☐	Clastrom 1,hsz 36 5: 1☐
Burdusél 1,hsz 79 23: 1☐	Coffalvy 1,hsz 19 7: 1☐
Burjanosb[an] 1,hsz 106 26: 1☐	Colonialis 1,hsz 14 64: 1☐
Burjánosb[an] 2,hsz 106 21: 1,hsz 106 25: 1☐	Coloniális 1,hsz 82 7: 1☐
Burjanosnak 1,hsz 106 25: 1☐	colonialis 1,hsz 79 23: 1☐
burukban 1,hsz 7 8: 1☐	com 1,hsz 56 4: 1☐
Busara 1,hsz 100 18: 1☐	Comitatu 1,hsz 16 7: 1☐
Busi 1,hsz 108 130: 1☐	comitatu 1,hsz 56 1: 1☐
Busnya 3,hsz 7 20: 2,hsz 7 21: 1☐	Commoditások 1,hsz 7 15: 1☐
buta 1,hsz 74 2: 1☐	Commoditásokban 1,hsz 7 15: 1☐
Buta 2,hsz 74 5: 1,hsz 74 7: 1☐	commorantes 1,hsz 67 10: 1☐
Butak 1,hsz 92 11: 1☐	communi 1,hsz 6 5: 1☐
butarét 1,hsz 74 2: 1☐	Communitás 1,hsz 84 7: 1☐
Butarétje 1,hsz 74 2: 1☐	Communitáse 1,hsz 67 9: 1☐
	Compositio 2,hsz 66 19: 2☐

### 1. ábra. A lexémák előfordulásait vizsgáló művelet

A fenti 1. ábra egy próbaművelet eredményét mutatja. Egy olyan parancsét, amely az anyag valamennyi lexémájának előfordulási gyakoriságát és helyét hívja elő. De mivel fő feladatunk a tulajdonnevek mutatójának elkészítése, a névtestek egyéb szövegrészekből való elkülönítését manuálisan kell megoldanunk, egyedi döntések segítségével választjuk ki és minősítjük a szövegben ta-

<sup>1</sup> A történeti írásjelek közti eligazodásban nyújtott segítségért köszönet illeti Korompay Klárát.

lálható adatokat. Az „Olló”-nak elkészített szoftver kijelölő alkalmazása ezt a célt szolgálja. Kezelőfelületét az alábbi 2. ábrán láthatjuk. Alkalmazási lehetőségeit a SZABÓ T. ATTILA-anyag háromszéki adatain mutatom be, de remélhetőleg más jellegű szövegek feldolgozására, mutatózására is alkalmas lesz.



2. ábra. Az „Olló” kijelölő-felülete

Ha megfigyeljük (az eredetivel betűhíven megegyező) bal oldali adatsort, látható, hogy nemcsak egyértelműen tulajdonnévnek tekinthető nyelvi formákat, „tőhelyneveket” tartalmaz, hanem (nagyobb részt) „helymegjelölő körülírásokat”, olyan ragos vagy névutós szerkezeteket, amelyeket „laza szerkezetű helynevek”, „szokatlan alaki szerkezetű helynevek” terminusokkal illetett a szakirodalom. A kiemelt részletekből az is látható, hogy az efféle alakulatok mindig magukba foglalnak, felhasználnak „valódi” helyneveket. Ezért kérdésünk — Mi kerüljön a majdan összeálló mutatóba? Mi számít egy helynévadatnak, és mi nem? — elvezet a névtani irodalomban gyakorta előkerült névvé válás, illetve a valódi vagy nem valódi helynevek problematikájához.

Elhatárolhatjuk ettől a kérdéstől azokat a szöveghelyeket, amelyek kétségkívül nem helynévi jellegű szöveghelyeket is tartalmaznak, és logikailag kiszűrhetőek belőlük a helynevek: „Az *Albisi* erdős teritoriumba vagyon egy darab Nyírfás erdőm, a mely *Pintze helynek* neveztetik”. Szintén nem jelent gondot

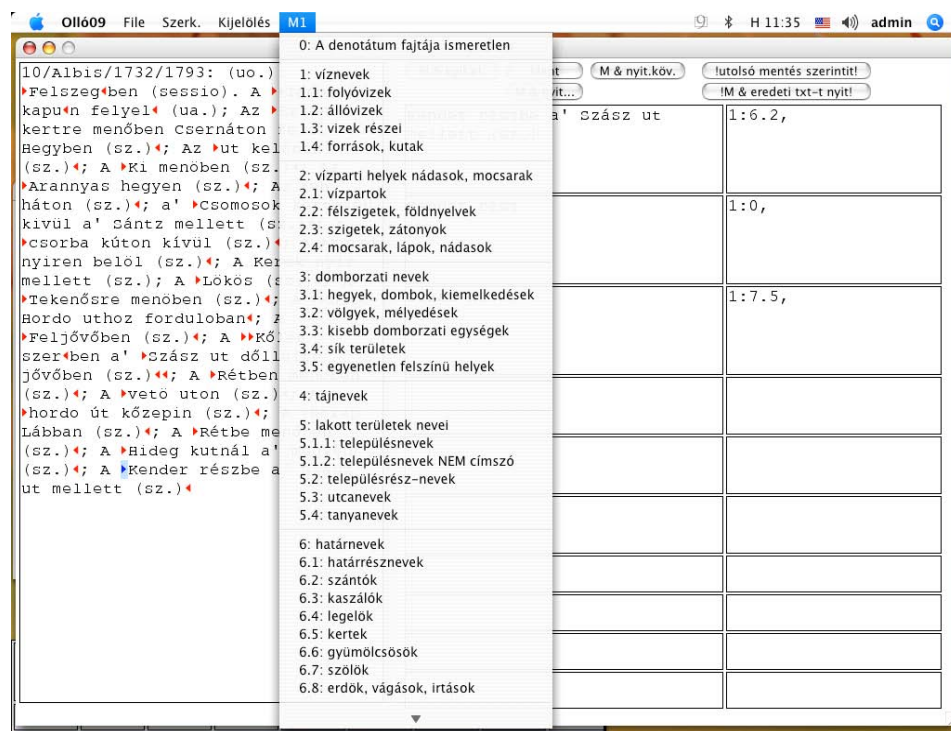
az ugyancsak sűrűn előforduló helységnévadatok kijelölése, mert alakjuk mindig állandó és „helynév-szerű”. Nagyon gyakoriak a szövegben azonban a fent számoltatottal illetett, leggyakrabban körülírásoknak, laza szerkezetű helyneveknek nevezett alakulatok, amelyeknek sokan szenteltek figyelmet írásaikban, helynév-voltuk eldöntése a jelenkori földrajzinév-gyűjtések módszertani kérdései között és a névvé válás folyamatának kutatásában is előkerült. Elsőként SZABÓ T. ATTILA helynévgyűjtési és közlétevési módszerét és még életében megjelent, példaértékű helynévtárának előszavában leírt megjegyzéseit idézhetjük föl. Ő azt hangoztatta, hogy a történeti adatok gyűjtőinek a teljességre kell törekedniük, elsősorban azért, mert a körülírási formákban, a nevek szöveggörnyezetében olyan adatok rejtőznek, amelyek a társtudományok számára, kiváltképpen a történeti tudományok számára fontosak (SZABÓ T. 1937). A jelenkori (20. századi) földrajzinév-gyűjtésekben ritkák a laza szerkezetű formák, de ennek szemléletbeli, módszertani oka is lehet.

Arra a kérdésre, hogy miképpen keletkeznek a körülírási helymeghatározások, a történeti források eredeti célját vizsgálva vagy éppen a mindennapi beszédhelyzeteket elemezve könnyedén felelhetünk: még névtelen helyek pontos lokalizálása céljából. SZABÓ T. ATTILA gyűjteményében az adatokat tartalmazó okmányok nagyrészt birtokösszeírások, amelyekben nemcsak a birtokok részeit, hanem az azt határoló felszíni formákat, tereptárgyakat, pl. utakat kellett minél pontosabban körülírni a hajdani írástudóknak. Így tehát talán egyszerinek, nem maradandónak tekintendők. Névszerű tulajdonságuk, hogy egy jól körülhatárolható objektumra utalnak. Viszont nem toldalékolhatók, és valószínűleg nem hangzanának el kétszer azonos formában. SEBESTYÉN ÁRPÁD rámutat, hogy általában csak magát a helyet, nem az objektum egészének fogalmát idézik föl, és mindig egy-egy kiegészítést kívánnak maguk után: az *Arannyas hegyen...* lévő szántó (SEBESTYÉN 1998).

A szakirodalomban található vélekedések zöme épp a változó nyelvi forma miatt zárja ki a helynevek köréből a helymegjelölő körülírásokat. HAJDÚ MIHÁLY a névtárakban kapott helyük jogosságát elsősorban azzal indokolja, hogy a diakrón szemléletű kutatások számára az ilyen formák a névvé válás útjának állomásait jelenthetik. A Csepel-sziget helynevei közül olyan példákat mutat be, ahol a különböző korú adatok között tetten érhető a kapcsolat, ám a változás iránya nem mindig az egyszerűsödés (HAJDÚ 1985).

Funkcionális szempontból a nevek felé, formai szempontból a közszoói helymegjelölés felé „húzó” nyelvi alakulatokkal van tehát dolgunk. De mivel a „100% helynév” — „100% körülírás” végpontok között nehéz egy-egy helymegjelölést pusztán nyelvérzékünk alapján elhelyezni, meg kell vizsgálnunk a kérdéses adatok szerkezeti tulajdonságait. A mutatókészítők feladata ezzel a kérdéssel kapcsolatban csak az lehet, hogy a nevekhez hasonlóan kiválasszák, a névtárba fölvetett névtetek közül megjelöljék azokat, amelyek körülírásnak tekinthetők. SZABÓ

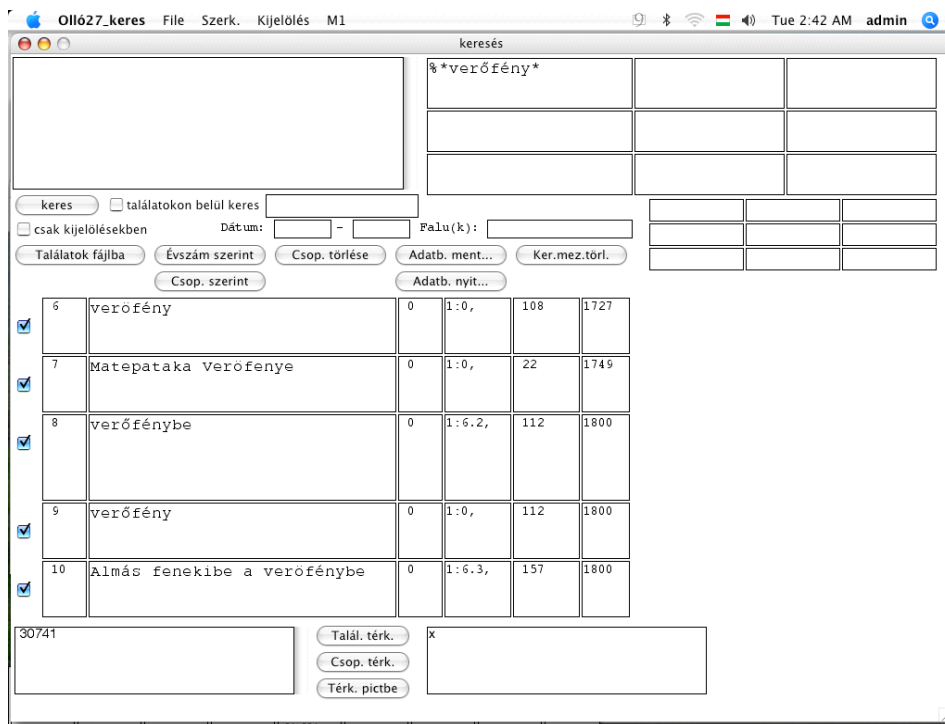
T. ATTILA adattárában olyan tömegben fordulnak elő ragos vagy névutós formák, körülírások, hogy éppen az elkészülő mutató lehet alkalmas részletes vizsgálatukra. Egy ilyen témájú tanulmány pedig nemcsak a helynevek kialakulása, a helymegjelölések jellemzése szempontjából lenne tanulságos, hanem a régi magyar hivatalos nyelvhasználat, írott jogi nyelv történetéről alkotott képet is árnyalná.



3. ábra. Az „Olló” használata

Visszatérve az „Olló” használatához és gyakorlati problémáihoz: a szövegekben előforduló névtesteket manuálisan (egy kattintással) jelöljük ki, majd ha a kijelölésben is találunk helyneveket, azokat ugyancsak „kiemeljük”. Ezután a névfajta megjelölését rendeljük hozzá a kijelölt névhez. A névfajta rendszerét HOFFMANN ISTVÁN műve alapján állítottuk össze (1993). Az eredeti anyagban lévő zárójelbe tett megnevezések (sz. = szántó, k. = kaszáló) alapján több ezer ilyen névfajtakód automatikusan bekerült a név mellé. A névfajtakódot a legördülő menüből (M1) is kiválaszthatjuk (lásd a 3. ábrát), vagy billentyűkód segítségével is megadhatjuk. Fontos megjegyeznünk, hogy a program használója újabb szempontrendszereket, osztályozási csoportokat is beilleszthet az M1 menühöz hasonlóan (M2, M3) és annak kódjaival elláthatja a kijelölt részeket. A névtipológiai vizsgálatok esetén ez a leggyorsabb módszer lehet a névanyag feldolgozására.

A 4. ábrán az „Olló” keresőfelületét látjuk. A háromszéki anyagban több mint 30 000 adatot sikerült kijelölni. Keresésükre számtalan lehetőség nyílik. Előhívhatjuk az adatokat a hozzájuk rendelt denotátumfajta alapján, település, dátum szerint és a bennük található szavakra, betűkapcsolatokra keresve. A felsorolt szempontokat kombinálva pedig leszűkíthetjük a keresést.



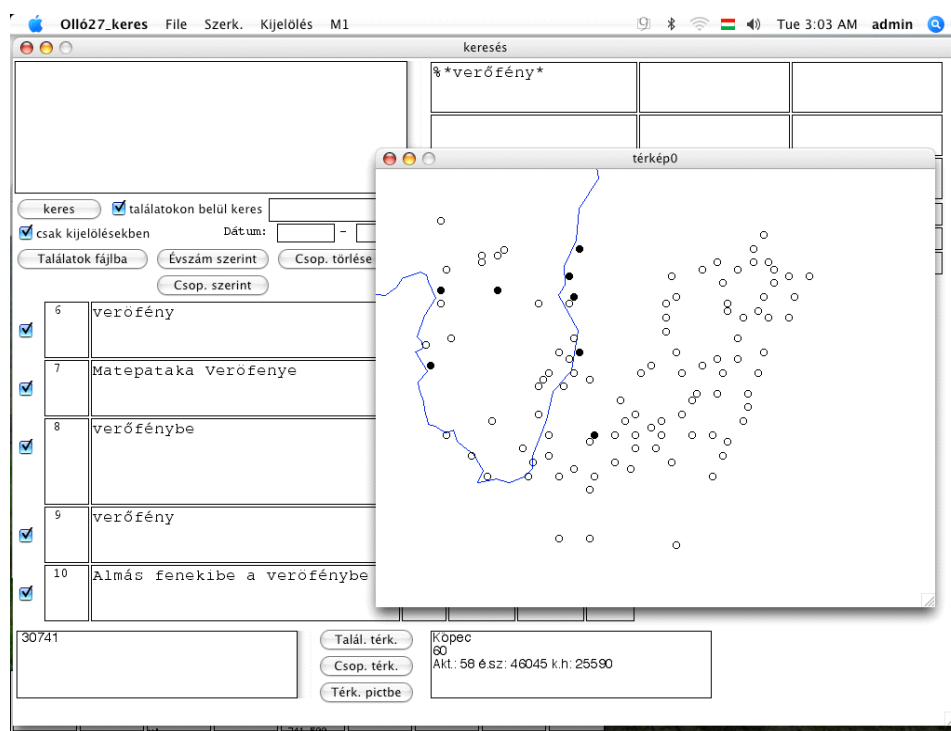
4. ábra. Az „Olló” keresőfelülete

A jobb felső sarokban lévő üres mezőkbe több keresőkérdést is beírhatunk és/vagy viszonyban egyaránt. Ha a keresett szó vagy betűkapcsolat elé beírjuk a %-jelet, akkor megkapjuk a más mellékjelekkel írott változatokat (pl. *út* — *ut*, *út* stb.) is. A keresés eredményét vagy az egész adatbázist a képernyő alsó részén látható találati lista középső oszlopában csoportosíthatjuk (számjegyek beírásával). A kapott csoportok elmenthetők az adatbázissal együtt, és később is előhívhatók. Az adatbázisok és találati listák Excel programban is megnyithatóak, ott készíthetjük el az adatok sorrendezését, például ott jön létre az egész gyűjtemény tulajdonképpeni mutatójának nyomdai változata is.

A program talán legizgalmasabb részlete abban rejlik, hogy segítségével térképeket generálhatunk, névföldrajzi elemzéseket valósíthatunk meg. A feldolgozott anyag jellegéből adódik, hogy miután hely, idő alapján és egyéb szempontból csoportosítottuk az adatokat, kíváncsiak lehetünk az eredmény földrajzi el-

oszlására. Mivel minden név egy településhez, annak névrendszeréhez kapcsolódik, így egy digitális térkép kutatópontjaihoz köthető. Igyekeztünk történeti források segítségével lokalizálni minden települést, azokat is, amelyek más helységekkel olvadtak össze vagy elpusztultak (ezúton köszönjük MIKESY GÁBOR-nak, a FÖMI munkatársának a segítségét). Az elkészült térképen szereplő pontok ezredfok pontossággal jelzik a települések helyét. A program ezen része a nyelvföldrajzi informatika eredményeire támaszkodik (a legfontosabb előzmény VÉKÁS DOMOKOS „Bihalbocs” néven ismert nyelvészeti technológiája).

Milyen esetekben lehet hasznos adataink térképes vizsgálata? Például a névalkotó földrajzi közszavak dialektológiai elemzésében, mely a magyar névföldrajzi kutatások egyik jellegzetes irányzata. Erre számtalan lehetőségünk nyílik. Példaképpen álljon itt a 4. ábrában szereplő „napos, déli helyoldal” jelentésű *verőfény* szó térképe, amelyet egyetlen kattintással előhívhatunk (a Találatok térképre! gomb segítségével). Íme:

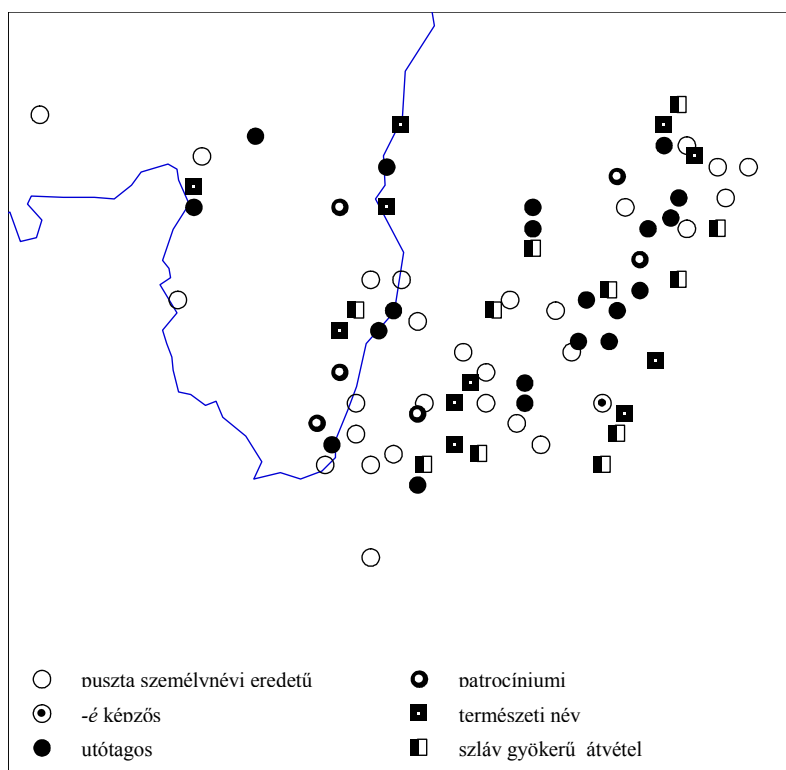


5. ábra. A verőfény előfordulása térképen

Az 5. ábrán látható térképen a kurzor mozgásával tájékozódhatunk, a helységek neve és adatai az alul jobbra lévő szövegdozban jelennek meg. Ha a keresés témája alkalmas további kutatásokra, a térképet más formátumokban is elmenthetjük. Az itt megjelenített térképi ábrázolásból kiindulva további vizsgálá-

lattal talán biztonsággal megállapíthatjuk — ami a térképről, és az egyes adatok szövegkörnyezetét vizsgálva valószínűnek tűnik —, hogy a *verőfény* korábban dokumentált jelentéséhez ezen a területen a 'vízparton lévő' jelentésmozzanat is hozzátartozik.

A névföldrajzi kutatások egy másik hagyományos formája elsősorban történettudományi igény miatt született. A helynevek, kivált a településnevek kronologikus típusai a történeti tudományok élénk érdeklődésére tartanak számot. Az ilyen munkákban a térképen jelölt helységnevtípusok a megtelepülés egymáshoz viszonyított idejét jelzik. Példaként álljon itt egy olyan művelet, amelyben az anyagban szereplő helységneveket kérdeztem le az adatbázisból, és a kapott adatokat csoportosítottam a KNEZSA–BÁRCZI-féle történeti tipológia szerint. A 6. ábrán látható egyes csoportok különböző színekkel vagy szimbólumokkal jelennek meg egy kattintásra a térképen (a Csoportokat térképre! gomb segítségével). Az eredmény elemzése természetesen hosszadalmas, itt nem vállalkozom rá, de azt mindenképpen jelzi, hogy a hasonló időigényes munkálatokhoz képest hatalmas lehetőségeket nyit ez az eljárás.

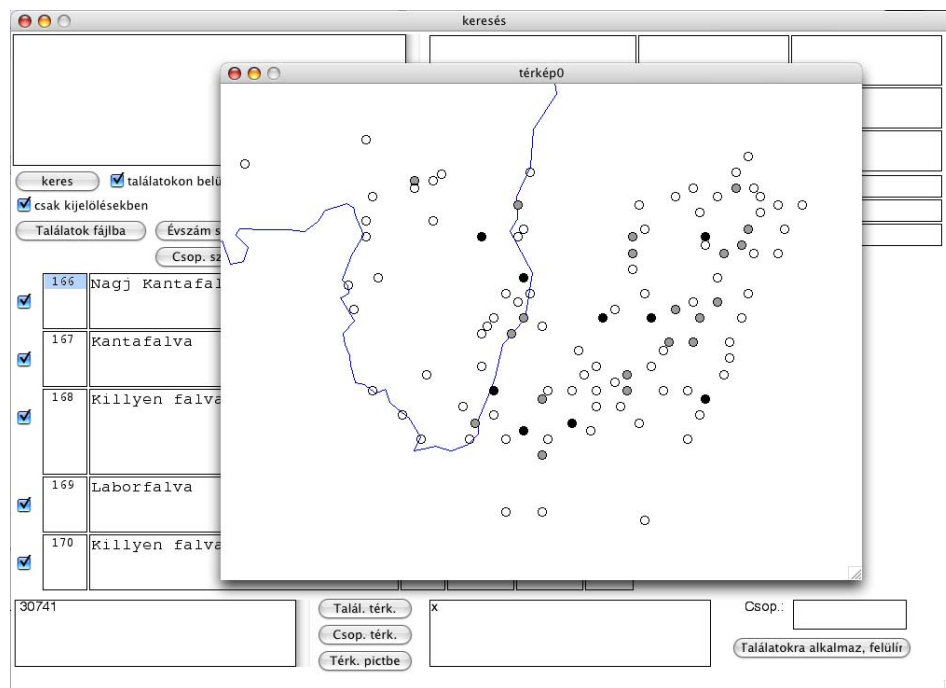


6. ábra. Településnév-típusok a térképen



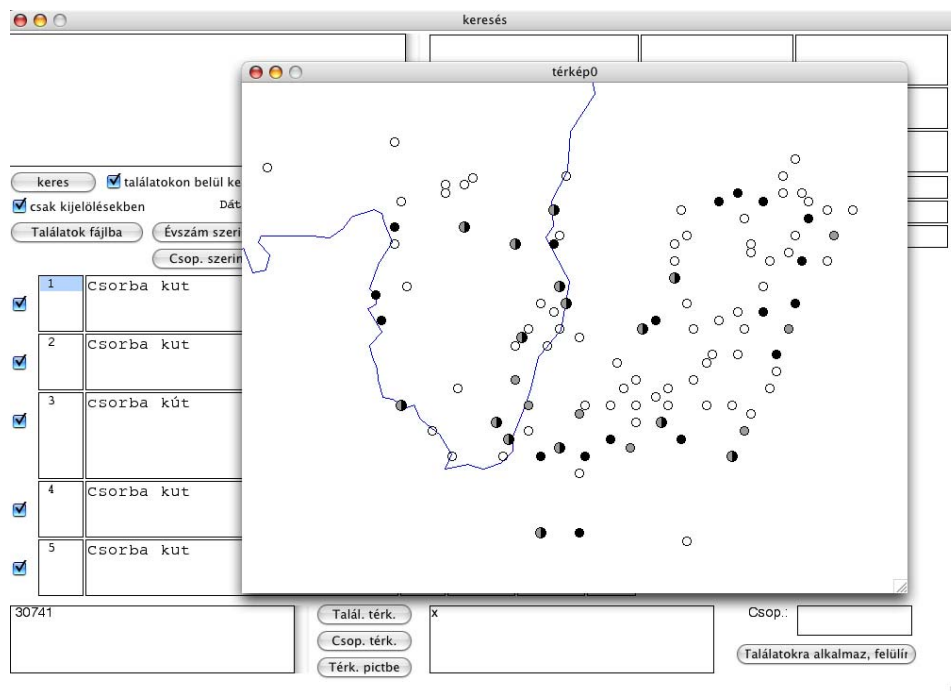
Ezután áttérek a névföldrajz egy lehetséges harmadik ágára, az „önelvű” névtani vizsgálatok térképes leképezésére. Ez a terület volt ezidáig a legkevésbé termékeny, és ebben a témában válik a bemutatott program igazán hasznossá. Sok névszerkezeti változás, névrendszertani típus földrajzi vizsgálata villámgyorsan elvégezhető, így közelebb kerülünk egy-egy terület „névjárásának” megismeréséhez, illetve egy-egy jelenség földrajzi elterjedésének megállapításához. A program haszna épp abban rejlik, hogy akár ezer és ezer olyan térképet generálhatunk pillanatok alatt, amelyek nem mutatnak releváns eredményt, viszont közben rálelhetünk nagy jelentőségű ábrázolásokra is.

Példaként a névtan és a névföldrajz egyik klasszikusának, KÁZMÉR MIKLÓSNak fontos művéhez (1970) kapcsolódva azt a jelenséget vizsgáltam, amely a 16–17. században fejtette ki hatását a Székelyföldön. A *-falva* utótagos nevek mintájára más típusú nevek (pl. puszta személynévi eredetűek) is *-falva* utótagot kaptak, de a 18. század közepétől újra eredeti alakjukban váltak általános használatúvá. Ez utóbbiakat látjuk a 7. ábrán lévő térképen feketével, az előbbieket pedig szürkével jelölve.



7. ábra. A *-falva* utótagos nevek és a *-falva* utótaggal kiegészülő nevek

Egy másik adatbázisban a kutak neveit kerestettük ki az adatbázisból. Aszerint csoportosítottuk a neveket, hogy sajátosságot kifejező névrészüket tulajdonságot, illetve valamilyen külső dologhoz, körülményhez (leggyakrabban birtokoshoz) való viszonyt fejez ki. Ezt mutatja a 8. ábra.



8. ábra. A kútnevek típusai

A sort szinte a végtelenségig folytathatnánk. Külön tanulmány témája lehetne a *patak* és *pataka* utótagú víznevek kronológiai és területi vizsgálata, a nevek szemantikai osztályozásával. A keresőkérdés térképre vetítését el lehet végezni (és ugyancsak érdemes) 25–50 évenkénti szűkítéssel (az anyag jellegéből adódóan kb. 1500-tól 1850-ig), és a kapott térképek együttes vizsgálata jelezhet egy-egy fontos történeti változást: ebben az esetben a külső dologra (birtokosra) utaló nevek számának gyarapodását.

Bármily kecsegtetőek is az eddig felvillantott vizsgálati lehetőségek, le kell szögeznünk, hogy a háttérben meghúzódó adatbázist még jócskán bővíteni kell, és ehhez (a névadatok elhatárolásához és besorolásához) emberi erőfeszítés szükséges. A jövőben elsődleges cél lehet SZABÓ T. ATTILA teljes helynévgyűjtésének feldolgozása, különösen a nagy adatmennyiségű közép-erdélyi és székelyföldi területekről.

Emellett azonban kísérletet teszünk a névföldrajzi adatok integrálására a nyelvföldrajzi adattárakkal, és eredményeinket egyrészt nyomtatott mutató, másrészt digitális kiadvány formájában tesszük közzé. Célkitűzéseink között szerepel, hogy ha ezek a munkálatok az eddigiekhez hasonló sebességgel folyhatnak tovább, akkor kiterjesztjük a program felhasználását több adattárra is. Reményeink szerint ez a 21. század tempójához méltó lendületet adhat a magyar helynévkutatásnak.

**Irodalom**

- BÁRTH M. JÁNOS (2006), Szabó T. Attila Erdélyi Helynévtörténeti Adattára. In: *Emlékkönyv Szabó T. Attila születésének 100. évfordulójára*. Szerk. BÁRTH M. JÁNOS. Bp. 63–5.
- HAJDÚ MIHÁLY (1985), A névvé válás folyamatáról. *MNyTK*. 170: 29–34.
- HOFFMANN ISTVÁN (1993), *Helynevek nyelvi elemzése*. Debrecen.
- KÁZMÉR MIKLÓS (1970), *A »falu« a magyar helynevekben. XIII–XIX. század*. Bp.
- SEBESTYÉN ÁRPÁD (1998), A névutós helynevek kérdéséhez. *ETFTK*. 24: 66–88.
- SZABÓ T. ATTILA (1937/1988), Dés helynevei. In: *Nyelv és település*. Bp. 103–83.

