

A kézirat leadásának időpontja: 2018. január 18.

Az elfogadás időpontja: 2018. október 16.

CS. NAGY Lajos

nyugalmazott egyetemi docens

Veresegyház, Magyarország

lcsnagy@gmail.com

## DIGITALIZÁLÁS ÉS INFORMATIZÁLÁS A MAGYAR NYELVJÁRÁSKUTATÁSBAN

Az alábbi tudománytörténeti tanulmány bemutatja a számítógépnek a magyar dialektológiában betöltött szerepét, jelentőségét, nélkülözhetetlen voltát. Az első kísérletezésektől – *A magyar nyelvjárások atlasza* félbemaradt digitalizálási munkálataitól – kezdve a „Beszélő atlaszig” vezet az út. A közben létrehozott nyelvatlások – *A romániai magyar nyelvjárások atlasza*, a *Szilágysági nyelv atlasz*, a *Medvesalja magyar nyelvjárás atlasza*, *Többdimenziós nyelvföldrajzi térképlapok* – az előrehaladás egy-egy fontos állomását jelzik.

*Kulcsszavak:* digitalizálás, informatizálás, elektronikus prezentációs rendszer, digitális archívum, beszélnyelvi korpusz, *BihalBocs* szoftver, jelentésföldrajz, nyelvjáráskutatás.

### 1. A kezdetek

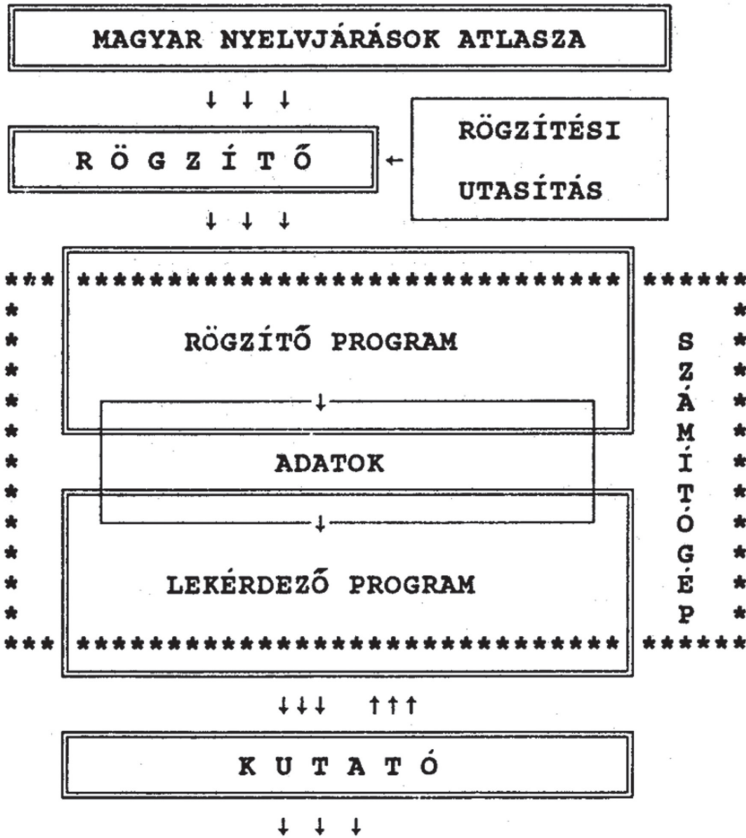
1.1. A magyar nyelvtudomány különböző területein is egyre gyakrabban találkozunk informatikai megoldásokkal, számítógépes adatrögzítésekkel és feldolgozásokkal. *A magyar nyelvjárások atlaszának* (továbbiakban MNyA.) a számítógépre vitele az 1960-as évek közepén vetődött föl először Kelemen József részéről. „Én magam [= Balogh Lajos] is többször beszélgettem vele erről a kérdésről, azokról az előnyökről, amelyek a gépi elemzésből adódnak. Az akkori kezdetleges technika – egy félszobányi gép, amely lyukszalagra dolgozott – azonban megkérdőjelezte ennek a vállalkozásnak a komolyságát, illetőleg azt, hogy a sok munkát igénylő befektetés hosszú távon is hasznosíthatóvá, gyümölcsözővé válik” – emlékezik vissza Balogh Lajos (Balogh – Kiss 1992, 6), aki tudomásom

szerint a MNyA. egyetlen ma is élő munkatársa.<sup>1</sup> A magyar nyelvjárások első számítógépes elkülönítése Simon Zoltán nevéhez fűződik, aki Vancouverben számítógépes elemzéssel hasonlított össze a nagyatlaszból kiválasztott 100 ősi magyar szót. 395 kutatópont ejtészváltozatait egybevetve 38 nyelvjárasi egységet állapított meg (vö. Hankó 1990). Ez lényegében megegyezik Imre Samunak ugyancsak a nagyatlasz anyaga alapján kézi feldolgozással kialakított főbb típusaival (Imre 1971, 329–373).

A 80-as évek végén, a számítástechnika rohamos fejlődésének következtében, belátható távolságba került a nagyatlasz teljes anyagának gépi adatbevitel. Kiss Gábor az atlasz számítógépbe vitelét több tényezővel indokolta. Az atlasz 6 kötete 48,5 kg tömegű, ezért fizikailag szinte kezelhetetlen, különösen akkor, ha egy adott kutatópont adatait szeretnénk elemezni, ugyanis minden kötetet (dobozt) kézbe kell venni. Az atlasznak ebből a tulajdonságából következik az, hogy a magyar tudomány (nyelvtudomány, néprajz, helytörténet stb.) vérkeringésébe alig került be. Számítástechnikai szempontból az adatbázis jól strukturált, viszonylag könnyen készíthető program a feldolgozásához. Az adattár újabb kutatópontok adataival is kiegészíthető. Az adatok kézi csoportosítása, összeszámlálása, összehasonlítása hosszadalmas. A számítógép alkalmazásával (új, csak géppel megvalósítható) elemzési szempontok érvényesítésével újabb tudományos eredményekhez jut a nyelvtudomány (vö. Balogh – Kiss 1992, 11–12). A nagyatlasz mellékjeles karakterkészletét kiadáskor a nyomda egyedileg alakította ki. A betűk fölött és alatt is gyakran előfordulnak olyan mellékjelek a fonetikai jelenségek jelölésére, amelyek a számítógépen akkor még nem voltak meg. Ezért az egyes mellékjelekhez számokat rendeltek (az alapjel után kitett 3-as például a félhosszúságot jelentette), vagy a számítógépen egyébként meglévő, nem fonémákat jelölő egyéb szimbólumokat használtak (a diftongusok kapcsos zárójelbe kerültek). Így készítette el Balogh Lajos és Kiss Gábor a MNyA. 1–3. kötetének informatizált változatát. Módszerük elméletileg teljes értékű, gyakorlatilag azonban az adatrögzítés során alkalmazott különböző platformok kompatibilitási nehézségeket okoztak (vö. Bodó – Vargha 2008, 336). A következő ábrán a MNyA. digitalizálási folyamatának a lépéseit követhetjük.

---

1 Balogh Lajos egyrészt az atlasz munkálatainak egyik közreműködője, másrészt pedig Végh Józseffel, Kovács Istvánnal, Szépe Györggyel és Farkas Vilmostal magnetofonfelvételeket gyűjtött a kutatópontokról. Ezeknek a hanganyagoknak a feldolgozását a közelmúltban kezdték meg a Geolingvisztikai Műhely munkatársai (lásd [http://geolingua.elte.hu/index\\_hu.html](http://geolingua.elte.hu/index_hu.html)).



1. ábra

1.2. Az elmúlt két évtizedben a számítógép magyar dialektológiai alkalmazását több tényező is motiválta. Az 1990-es években az informatika és technikai háttere olyan gyorsan fejlődött, hogy az alkalmazások lehetősége meghatározódott. A nyelvtudomány különböző ágaiban egyre elterjedtebbé váltak az informatikai megoldások. Nemzetközi dialektológiai kongresszusokon már külön szekció foglalkozott a számítógépes adatfeldolgozással, s már szoftverbemutatókat is tartottak.<sup>2</sup> Kiefer Ferenc az informatikai forradalom nyelvi hatásairól megállapítja: „Ez a változás a könyvnyomtatás felfedezésénél is komolyabb következményekkel jár. Az Európai Közösség szakembereinek véleménye szerint

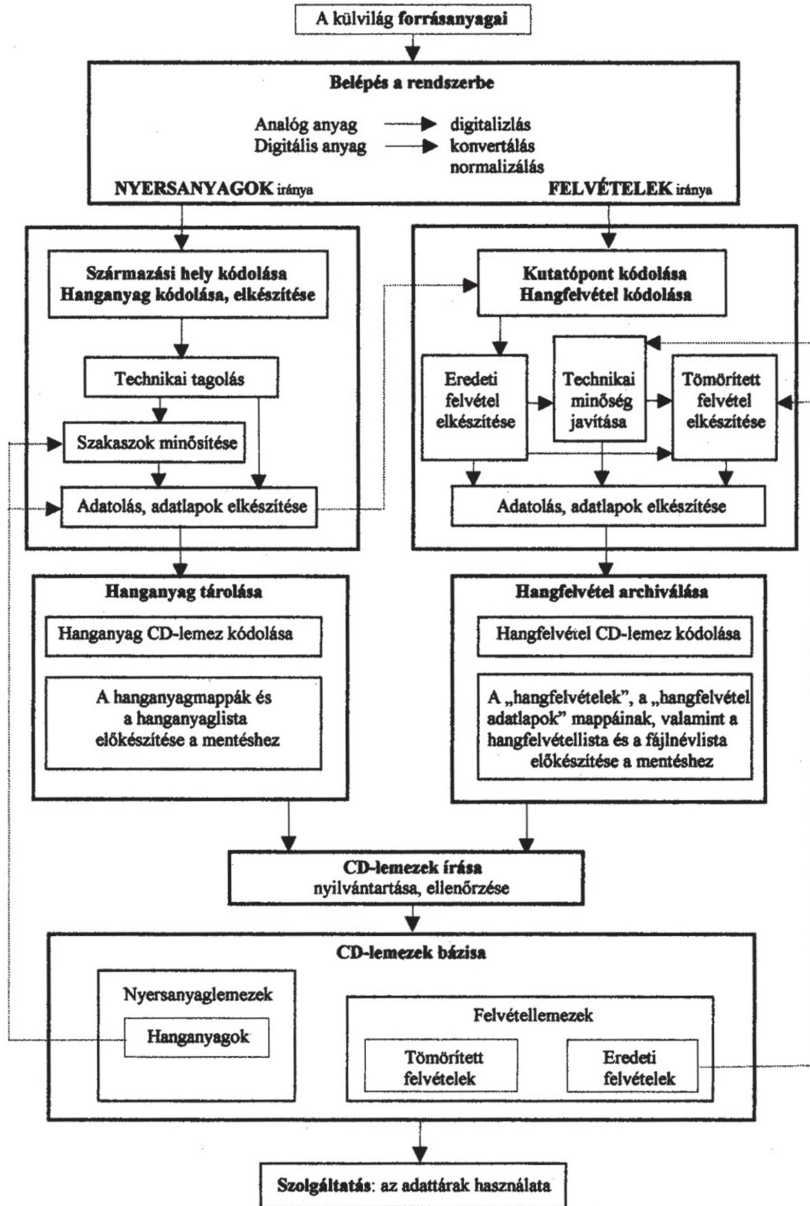
2 Az 1982-ben megjelent *Dialektologie. Ein Handbuch zur deutschen und allgemeinen Dialektforschung* (Hrsg. Werner Besch, Ulrich Knopp, Wolfgang Putschke, Herbert Ernst Wiegand. Walter de Gruyter, Berlin – New York) már külön fejezetben foglalkozott a számítógép nyelvtudománybeli alkalmazási lehetőségeivel.

az új nyelvi technológiák olyan hatással lesznek a kultúra fejlődésére, mint amilyen szerepet játszott az írásbeliség a kultúra megőrzésében és terjesztésében. Jelentőségük a könyvnyomtatásénál is nagyobb” (Kiefer 1994, 635). Motiváló tényezőként szerepelt az is, hogy a Nyelvtudományi Intézetben, a Zenetudományi Intézetben, a Néprajzi Múzeumban, valamint egyetemi és főiskolai tanszékeken, sőt nyelvjáráskutatók, néprajzosok birtokában is rengeteg archiválható nyelvjárási írott és hangos anyag található. Ezeknek az összegyűjtése, technikai megmentése, egy hatalmas adattárba rendezése a tudományos kutatás számára elengedhetetlen. Ezt a munkát a korábbi, manufaktúris módszerrel nem lehet elvégezni. Főszereplővé vált tehát a számítógép, melynek fejlettségi foka már lehetővé tette ezeknek a kezdeményezéseknek a megvalósítását.

1996-ban az ELTE Magyar Nyelvtörténeti és Nyelvjárástani Tanszéke, valamint a Fonetikai Tanszéke a fenti cél érdekében összefogott, és kidolgozott egy elektronikus prezentációs rendszert. Juhász Dezső témavezetőként főként a dialektológiai szempontok érvényesülését biztosította, Vékás Domokos a fonetikai szempontok megvalósításával, valamint nemzetközi példák alapján az informatizálási koncepció kialakításával foglalkozott. Radványi Péter a számítógépes nyelvi archívum kialakítása és a hanganyagok digitalizálása mellett technikai feladatok megoldását vállalta magára (vö. Juhász – Radványi – Vékás 2000, 48–49).

Ismerkedjünk meg legalább nagy vonalakban az elektronikus prezentációs rendszer tervével, melynek nagyon sok eleme napjainkra már valósággá vált. A munka első fázisa a digitális archívum feltöltése, vagyis a hagyományos adathordozókon lévő adattárak bevitele a számítógépbe. Hagyományos adathordozók (primer adattárak) a papír alapú, vagyis nyomtatott vagy kézíratos szöveglejegyzések, szó- és szólásgyűjtemények, tájszótárak, nyelvatlaszok, nyelvtanok, néprajzi és honismereti közlemények, illetve magnó-, esetleg videoszalagra fölvetett nyelvjárási, néprajzi gyűjtések. Az ezekből a forrásokból nyert adatokat digitalizálják, technikailag tagolják, kialakítják az ún. nyersanyagtárat. Az anyagokat fonetikailag, nyelvföldrajzilag, lexikailag, grammatikailag, statisztikailag és bibliográfiailag kódolják is. A folyamat következő lépésében tartalmi válogatás, hangfinomítás, zajszűrés, restaurálás, válogatás, rendezés, utalozás és a tárolás megoldása történik. A tárolás CD-lemezen, újabban nagyobb kapacitású tárolókon vagy nyomtatásban (nyelvatlasz, szótár stb.) történik. A digitális archívumra operatív rendszerek épülnek, melyekkel lehetővé válnak fonetikai, fonológiai, lexikai, grammatikai stb. keresések a teljes korpuszban. Különböző statisztikák készíthetők, egyedi és összesítő térképlapok szerkeszthetők nyelvjárási jelenségek bemutatására, hogy csak néhány lehetőséget említsünk az alkalmazási területekből (vö. Juhász – Radványi – Vékás 2000, 48–49).

A következő ábra azt szemlélteti, hogyan lesz a hangfelvételekből nyelvileg és tartalmilag jól tagolható informatikai anyag.



A dialektológiai archívum hangfájlfelkészésének technikai blokk-sémája

2. ábra

## 2. Digitalizált nyelvatlaszok

2.1. *Az első digitalizált nyelvatlasz.* A fent vázolt elméleti alapokra épülve indult meg a digitalizált atlaszok kiadása. Elsőként 1995-ben Juhász Dezső irányítása mellett a 11 kötetes *A romániai magyar nyelvjárások atlaszának* (RMNyA.) a közzétételét kezdte el az ELTE Magyar Nyelvtörténeti és Nyelvjárástani Tanszékének néhány fiatal oktatója hallgatók közreműködésével. Az utolsó kötet 2010-ben látott napvilágot. Ez azt jelenti, hogy a számítógépes technika dialektológiai alkalmazásának kidolgozása és kipróbálása már a 90-es években elkezdődött. A munkacsoport minden mellékjeles betűt megtervezett, kódszámmal látott el. Mellettük több tagoló és szemléltető funkciójú grafikai jelet is készített, majd az adatrögzítés programját egy szövegszerkesztős programból alakította ki. Ezen lépések következtében váltak a térképek egyformává, esztétikailag is magasabb szintűvé, az adatok jól olvashatóvá.

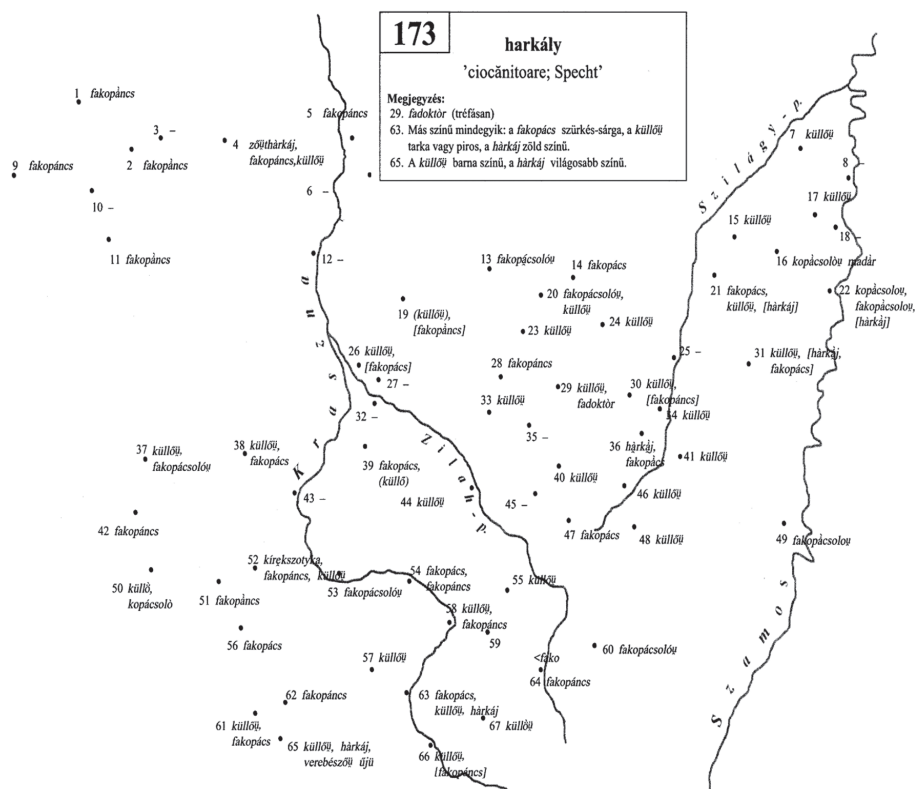
A számítógép használatának a lehetőségét az atlasz teljes anyaga gyűjtőjének, Murádin Lászlónak a gondos előszerkesztő munkája teremtette meg. A RMNyA. történetének az első fontos dátum 1955, akkor ugyanis a Román Tudományos Akadémia Kolozsvári Nyelvtudományi Intézete fölvette kutatási tervébe e nyelvatlasz elkészítését. Nagy szerepe volt ebben a nemzeti elfoglaltságtól mentes akkori igazgatónak, Emil Petrovicinak. A gyűjtés előkészítő munkálatai (a kérdőívzet összeállítása, a kutatópontok kijelölése) 1956-ban kezdődtek. A munka vezetésével Szabó T. Attilát bízták meg, s két főállású kutatót, Gazda Ferencet és Nagy Jenőt kérték föl munkatársnak. A kezdeti lendület azonban hamar megtorpant, mert 1957-ben Gazda Ferencet, 1959 elején pedig Nagy Jenőt politikai okokból eltávolították az intézetből, majd később Szabó T. Attila intézeti állását is megszüntették. 1957 végén Gazda Ferenc helyére Murádin Lászlót nevezték ki, s Nagy Jenővel kettesben folytatták a munkát, igaz, csupán egy évig. 1959-től Murádin László egyedül végezte mindazt a munkát, amelyet korábban egy egész munkacsoport számára terveztek. A terepmunka egy évtizeden keresztül, 1957-től 1967-ig tartott: 136 kutatóponton 3379 kérdést tartalmazó kérdőív kikérdezése, lejegyzése. Murádin László szerkeszteni kezdte *A romániai magyar nyelvjárások nyelvöldrajzi adattárát*. Köznyelvi címszók alá sorolta be az adatokat, külön bekezdésekbe szerkesztette a morfémaavázásokat, jelölte előfordulási helyüket. A 80-as évek végére készült el a teljes anyag, mely mintegy 2000 oldalt tesz ki. Ez az adattár lett az alapja az első digitális nyelvatlasznak (vö. Murádin 1995, 5–8).

A 11 kötet összesen 3297 szóföldrajzi térképet tartalmaz. Minden kötetben közli a szerkesztő a térképre nem került szócikkek listáját, melyekben a címszóhoz tartozó adatok kutatópontonkénti felsorolása kapott helyet.



vetkezett halála megakadályozta. A hagyaték Magyarországra kerülése után, 1993-ban kezdődtek meg a közzététel munkálatai. A kézirat összeállítását és szerkesztését Hegedűs Attila végezte. A romániai magyar nyelvjárások atlaszának a karakterkészletével és lényegében azonos digitalizálási technikával készült el ez az atlasz. A Magyar Nyelvtudományi Társaság és a Pázmány Péter Katolikus Egyetem közös kiadványaként jelent meg 2000-ben.

Az alábbi térképről megállapítható, hogy viszonylag kis területen mennyire gazdag például a *harkály* fogalmának megnevezési rendszere.



4. ábra

### 3. Informatizálás

„Az informatizált adatok előnye, hogy integrálhatók (vagyis a különböző atlaszok és szótárak egyesíthetők), és hogy viszonylag könnyen konvertálhatók a gyorsan változó informatikai platformok között, amire néhány évenként szükség is lehet. Amikor az egységes európai hangzó atlasz összeáll az egyes nyelvek



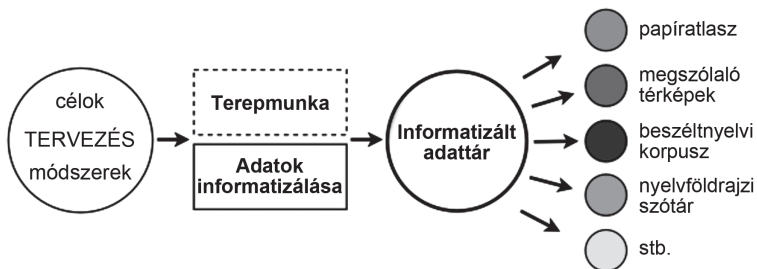
különböző területi változatait dokumentáló beszéltnyelvi korpuszokból, csakis a megfelelően informatizált, tehát integrálható nyelvjárási anyagokkal járulhatunk hozzá e közös vállalkozáshoz is” – hangsúlyozza az informatizálás tudományos jelentőségét Vékás Domokos a *BihalBocs* honlapján (<http://www.bihalbocs.hu/utmutatok.html>. 2017. okt. 24.).

3.1. *Geolingvisztikai Műhely*. Juhász Dezső *A romániai magyar nyelvjárások atlasza* utolsó, XI. kötetének előszavában 2010-ben szól arról, hogy elektronikus adatbázisukat komplex multimédiás dialektológiai rendszerbe integrálják a *BihalBocs* néven készült számítógépes program segítségével.

„A *BihalBocs*<sup>3</sup> egy nyelvészeti és problémamegoldási szemlélet, illetve az ezen a szemléleten alapuló szoftver. Vékás Domokos 1996-ban kezdte meg egy ilyen, speciális számítógépes program fejlesztését, amely hatékonyan teszi lehetővé nyelvjárási anyagok rögzítését (kódolását), tárolását és felhasználását” ([http://geolingua.elte.hu/bihalbocs/bb\\_hu.html](http://geolingua.elte.hu/bihalbocs/bb_hu.html) 2017. okt. 24.).

A *BihalBocs* legfontosabb tulajdonsága többek között az, hogy támogatja a magyar egyezményes hangjelölést, lehetővé teszi a lejegyzés és a hanganyag szinkronizálását, a kutatói igényeknek megfelelően továbbfejleszthető, igen bonyolult és összetett feladatok megoldására is alkalmas, oktatási változatok készíthetők vele. A vele készült hangoskönyvek, térképek viszont átalakíthatóak olyan platformfüggetlen változatra, amely a PC-ken is megnyithatóvá (láthatóvá és hallhatóvá) teszi a nyelvjárási anyagokat (vö. [http://geolingua.elte.hu/bihalbocs/bb\\_hu.html](http://geolingua.elte.hu/bihalbocs/bb_hu.html) 2017. okt. 24.). A *Geolingvisztikai Műhely* honlapján található informatizált nyelvjárási adattárak, térképek és hangoskönyv mind ezzel a szoftverrel készültek.

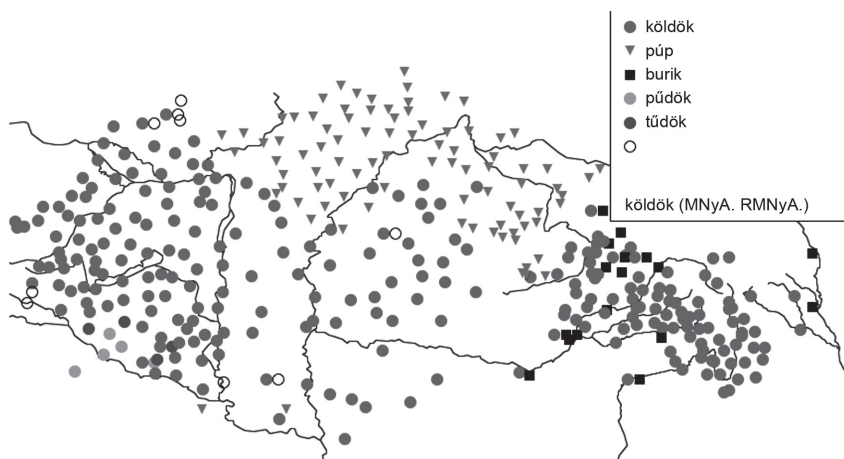
A következő ábrán láthatjuk a *BihalBocs* szoftver legfontosabb alkalmazási lehetőségeit. Az utolsó az adatoknak a legkülönbélebb nyelvészeti megszólaltatását teszi lehetővé. Például korpuszegyesítések, azok alapján fonetikai, lexikológiai statisztikák készítése, jelenséghatárok megállapítása, fogalmak megnevezési rendszerének vizsgálata, újabban pedig jelentésföldrajzi kutatások végzése.



5. ábra

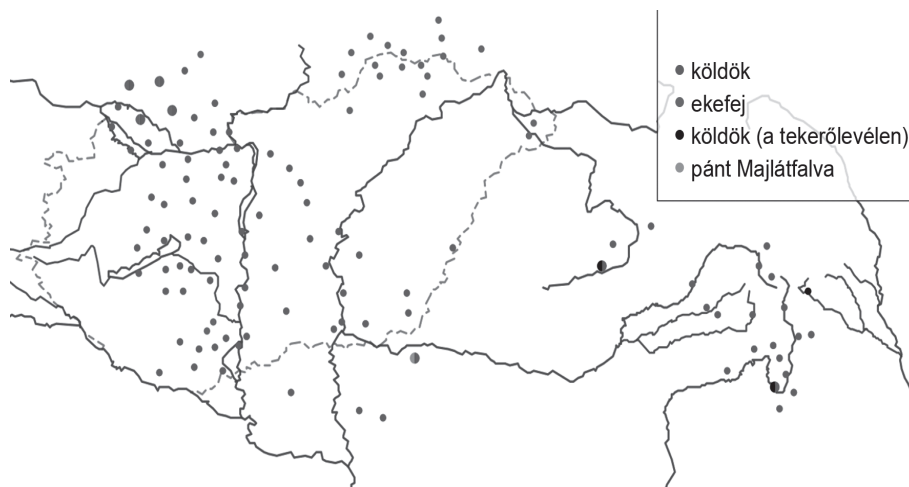
3 <http://www.bihalbocs.hu/bihalbocsrol.html> (2017. okt. 24.)

3.2. *Jelentésföldrajzi kutatás.* A szóföldrajzi vizsgálatban azt keressük, hogy egy adott fogalmat – például *köldök* –, jelenséget a különböző tájegységeken hogyan neveznek meg. Arra a kérdésre, hogy: „Mi a neve ennek a kis gödörnek az ember hasa közepén?” (MNyA. 503. köldök, RMNyA. 1213. köldök) a következő válaszokat kapták a gyűjtők: *köldök, püddök, tüddök / púp, burik*. Ezt szemlélteti a következő térkép.



6. ábra

A jelentésföldrajzi vizsgálat alapkérdése, hogy az illető lexémával a környező világ mely elemét nevezi meg egy beszélőközösség.



7. ábra

A *köldök* lexéma a nyelvterület nagy részén az 'ember hasa közepén levő kis gödör', északon, északkeleten 'ekefej', Erdélyben 'köldök' (a tekerőlevélen) [a fonálgombolyítónak az a része, amely körül a tekerőlevél forog], egy kutatóponton pedig 'pánt'. Itt jegyzem meg, ha nemcsak az atlaszokban keresünk, hanem megtörténhetne legalább az *Új magyar tájszótár* (ÚMTsz.) teljes anyagának az informatizált korpuszba illesztése, akkor az előbbi valóságselemek kívül másokat is fölvehetnénk: 'dinnye része', 'a kenyér közepén ujjal nyomott bemélyedés', 'a szalmakalap tetejének középső kis része', 'az ing elejét lezáró kis téglalap alakú rész', 'fazekaskemence tetején lévő kis nyílás'.

A fenti jelentésföldrajzi térképet a MNyA. 451–1162. térképlap, a RMNyA. V–VIII. kötet, a *Székely nyelvöldrajzi szótár* (SzNySz.), a *Moldvai csángó nyelvatlasz* (McsNyA.) I–III. kötet és a *Medvesalja magyar nyelvjárási atlasza* (MedvA.) egyesített adattára összesen 1 006 574 nyelvi adata alapján készítettem (vö. Cs. Nagy 2007, 63–68; 2013, 153–163).

Ezen a kutatási területen elengedhetetlen a nyelvatlaszoknak legalább a lexikai szómutatóját elkészíteni, bevinni az adattárba, mert másképp csak ötletszerű lehet még mindig azon lexémák kiválasztása, amelyek jelentéseinek a területi megoszlását valóban érdemes vizsgálni. Megjegyzendő, hogy eddig csak a nagy-atlaszhoz és a medvesalji atlaszhoz készült mutató.

3.3. *Az első informatizált nyelvatlasz. A Medvesalja magyar nyelvjárási atlasza* (MedvA.) anyagának a gyűjtése Cs. Nagy Lajos vezetésével, családtagok és nyitrai hallgatók közreműködésével 1993-ban kezdődött Medvesalja és közvetlen környéke 12 felvidéki településén. Az 1204 tételből álló kérdőfüzet a MNyA. kérdései közül az alaktani és a lexikai jellegű kérdéseket tartalmazza, valamint a teljes palóc kérdőívet (vö. Balogh – Pelle 1972). A további adatgyűjtést szélesebb néprajzi tájékozódás előzte meg, s az OTKA 60379. számú pályázatának támogatásával végeztük. A vidék táplálkozási, építészeti és népi szokásait leíró munkákból, valamint Sándor Anna Nyitra-vidéki magyar nyelvjárások atlaszából [= NyvA.] (Sándor 2004) újabb kérdőív készült. Az összes (1305) kérdés megoszlása a következő: MNyA.-kérdés: 978 (617 lexikai, 356 morfológiai és 5 fonetikai), Palóc kérdőív: 200 (65 lexikai, 44 morfológiai és 91 fonetikai), pótgyűjtés 120 (120 lexikai), NyvA. 7 (7 lexikai). A hangfelvételes kérdőíves gyűjtés mellett több szövegfelvételt is készítettünk. A teljes anyagból 750 címszót választottunk ki, melyeknek adataiból a *BihalBocs* szoftver segítségével adatbázist építettünk. Ezt követően az adatokból térképeket generáltunk: 675 (90%) lexikai, 69 (9,2%) morfológiai és 6 (0,8%) fonetikai térképlapot, s ez jelent meg 2011-ben papír változatban.

Az informatizált változat a *BihalBocs* szoftver segítségével lehetővé teszi a teljes korpusz többoldalú elemzését, hangtani és egyéb statisztikák készítését,

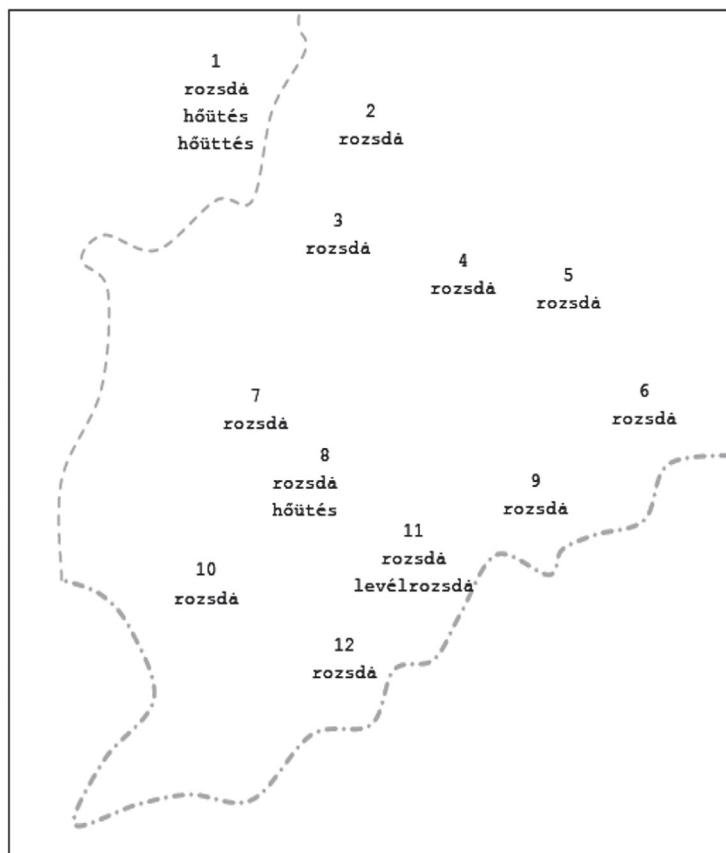
a kutatói igényeknek megfelelő új térképek generálását stb. Ezt a sajátosságot szemlélteti az atlasz 32 színes térképe.

Az első térképlap (*rozsdá a [gabonán]*) egyszerűen, kutatói csoportosítás nélkül mutatja a címszóhoz tartozó adatokat. Ebben papír alapú végtermék létre megegyezik a korábbi adatbeírásos nyelvatlazsokkal, vagyis minden kutatópontnál az onnan származó nyelvi adatok szerepelnek. Természetesen az ilyen típusú szóföldrajzi térkép is alkalmas az elemzésre, csak lényegesen hosszabb időre van szükség az azonos adatok kiválogatására, különösen akkor, ha gazdagabb kutatópont-hálózatú lapot elemzünk. (Erre más-más szempontból jó példa a *köldök* fenti szó- és jelentésföldrajzi térképe.)

**12. rozsdá (a gabonán) LEX.**

hrdza (na obili); blight

12. Van egy betegsége a búzának: a szárán meg a levelén vörösesbarna foltok lesznek. Mi ennek a neve?



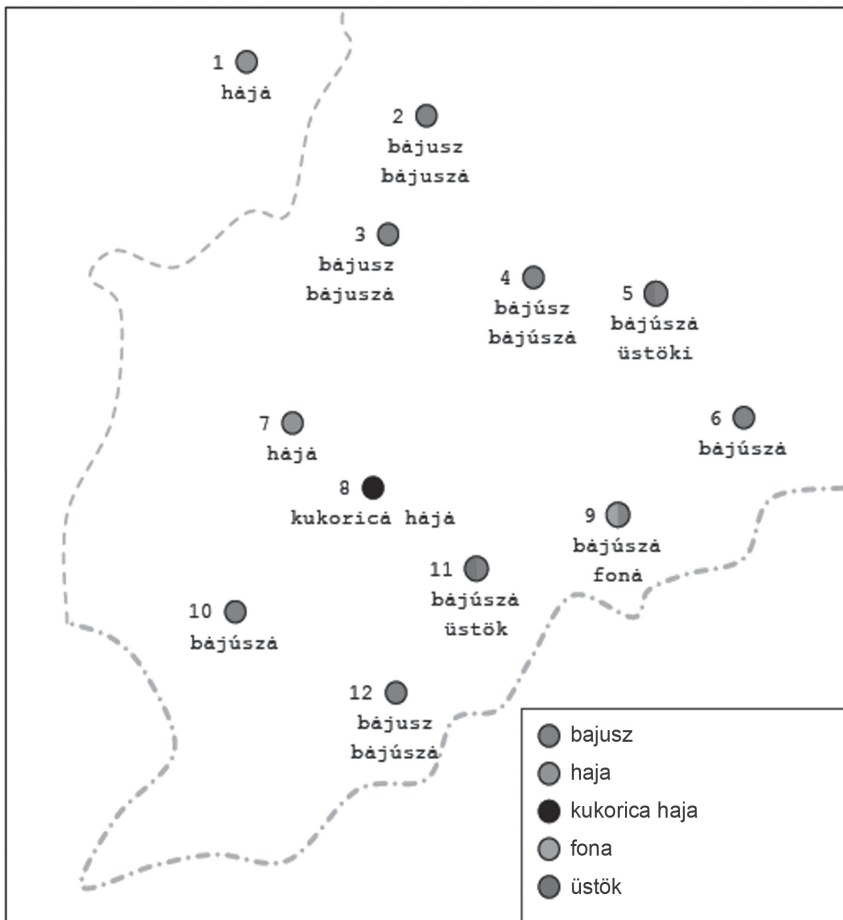
8. ábra

A második térképlap (*bajusza [a kukoricáé]*) ún. adatszoportosítási megoldású. Az azonos fogalmi megnevezések azonos színnel jelennek meg. Ebből első látásra megállapítható, hogy területünkön melyik név a főváltozat. A *bajusz(a)* ~ *bajúsz(a)* a főváltozat, ugyanis összesen 9 településen fordul elő. Az *üstök(i)*, a *haja*, a *fona* és a *kukorica haja* pedig mellékváltozatok. Az egyes megnevezések érintkezése és területisége is azonnal követhetővé válik. Ha a reprezentatív nagyatlaszok és a regionális atlaszok adatait egyesítjük, akkor természetesen nagyobb összefüggések is föltárhatók, térképen ábrázolhatók.

**46. bajusza (a kukoricáé) LEX.**

fűz (kukurice); corn-silk

42. A cső végén, mikor még száron van, vékony szálak lógnak ki. Mije ez a kukoricának?



9. ábra

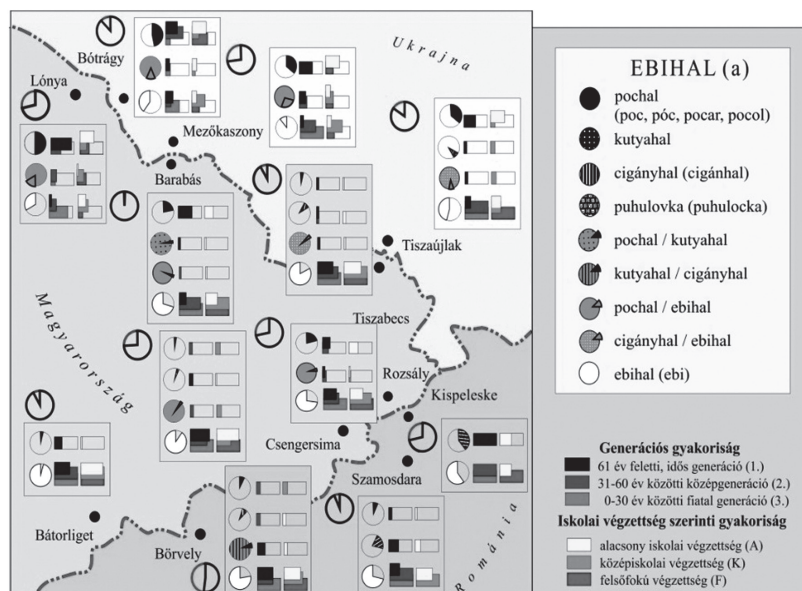
#### 4. Többdimenziós nyelvatlasz

A többdimenziós (polisztratikus) dialektológia arra a kérdésre keresi a választ, hogyan lehet a területi alapú és az abban is megmutatkozó társadalmi vetületű nyelvi variációkat mennyiségi szempontok alapján térképeken bemutatni és dokumentálni (vö. Kiss 1999, 421).

A Nyíregyházi Főiskola Magyar Nyelv- és Irodalomtudományi Intézete keretében működő Szociodialektológiai Műhely munkatársai (P. Lakatos Ilona, T. Károlyi Margit) 1997-től végeznek a magyar, ukrán, román határ mentén szociodialektológiai változásvizsgálatokat.<sup>4</sup> A kutatócsoporthoz a 21. század első évtizede második felében kapcsolódott Iglai Edit. „Legfontosabb célkitűzéseinket (az országhatár nyelvi elkülönítő szerepének mennyiségi mutatókkal való bemutatása; a különböző nyelvi szituációkban élő, más-más beszélőközösségekhez tartozó, azonos szociokulturális státuszú csoportok nyelvhasználati mutatóinak összevetése; a nyelvi változások szinkrón dinamizmusának érzékeltetése; a lehetséges változási irányok előrevetítése) többdimenziós nyelvföldrajzi térképeken ábrázoljuk” – írják a szerzők (P. Lakatos – T. Károlyi – Iglai 2012, 122–123). A *Változó nyelvhasználat a hármas határ mentén* című könyvükben az alkotók 8 magyarországi, 6 kárpátaljai és 4 romániai, összesen 18 településről 537 adatközlőtől különböző módszerekkel gyűjtöttek adatokat. A 143 000 nyelvi adatból válogattak, s a kutatási céljaiknak legjobban megfelelő 6 településpárt kiemelve szerkesztették meg a többdimenziós morfológiai és lexikai térképlapokat. Az adatok földolgozásához (informatizálásához) olyan speciális szoftvert fejlesztettek ki, amely lehetővé tette a hatalmas nyelvi anyag gépbe vitelét, több szempontú (pl. mennyiségi, gyakorisági, a megfelelő adatok egymásra vonatkozhatóságának stb.) lekérdezhetőségét (vö. P. Lakatos – T. Károlyi – Iglai 2012, 126). A szoftver segítségével 141 többdimenziós lexikai és morfológiai térképet generáltak, melyeket a kötethez mellékeltek CD formájában. A könyv a szerzők több évtizedes kutatómunkájának elméleti és módszertani eredményeivel is megismertet. Szemléltetésül az *ebihal* lexikai térképlapját mutatjuk be. Ezt a térképet látva megérthetjük, hogy ilyen sokrétű elemzésre alkalmas többdimenziós lapok kizárólag számítógéppel hozhatók létre.

---

4 A Szociodialektológiai Műhely honlapja: <http://szocial.nyf.hu/>.



10. ábra

## 5. „Beszélő atlasz”

A *Magyar nyelvjárások atlasza* ellenőrző gyűjtésekor, 1960 és 1964 között a hangtani tekintetben problematikus kérdésekről és a rájuk adott válaszokról hangfelvétel készült. Az informatikai kutatások újabb eredményei a válaszok lejegyzett formájának a hangzó változattal való összekötését tették lehetővé, az interneten pedig digitális formában elérhetővé. Megnyílt az út a „Beszélő atlasz” létrehozása előtt (<http://geolingua.elte.hu/maps/mnya/ma2/> 2017. okt. 24.).

## 6. Befejező megjegyzések

Jelen tudománytörténeti tanulmányban azt kívántam bemutatni, hogy milyen nagy szerepet tölt be a számítógép a dialektológiában. A különböző műhelyekben kifejlesztett speciális szoftverek alkalmazásával korábban elképzelhetetlen gyorsasággal és mélységben tárhatók föl a nyelvhasználat területi és társadalmi összefüggései. (Nem ünnepontási szándékkal jegyzem meg, de kívánatos lenne az egyes műhelyek tevékenységének az integrálása.) Új dialektológiai kutatási területek (pl. jelentésföldrajzi) nyílhatnak meg. Óriási adatbázisok építhetők ki, folyamatosan bővíthetők a korábbi (kéziratban és nyomtatott forrásban lévő) gyűjtések anyagaival: regionális tájszótárakkal, regionális nyelvatlaszokkal. A magyar dialektológia ismét bekapcsolódhat a nemzetközi kutatásokba.

## Irodalom

- Balogh Lajos – Kiss G. Gábor. 1992. A magyar nyelvjárások atlaszának számítógépes feldolgozása. In Kontra Miklós, szerk. *Társadalmi és területi változatok a magyar nyelvben*. 5–17. Budapest: MTA Nyelvtudományi Intézete, Linguistica Series A Studia et Dissertationes, 9.
- Balogh Lajos – Pelle Béláné. 1972. *Kérdőív a palóc nyelvjárásai sajtósági felkutatásához*. Eger: Heves Megyei Tanács.
- Bodó Csanád – Vargha Fruzsina Sára. 2008. Régi nyelvatlaszok – új módszerek. Nyelvtchnológiai eljárások a nyelvföldrajzban. *Magyar Nyelv* 104 (3): 335–351.
- Deme László – Imre Samu szerk. a munkaközösség közreműködésével. 1968–1977. *A magyar nyelvjárások atlasza I–VI*. Budapest: Akadémiai Kiadó.
- Hankó Ildikó. 1990. Kanadai magyarok vállalkozása. *Magyar Nemzet* 53 (291): dec. 10.
- Imre Samu. 1971. *A mai magyar nyelvjárások rendszere*. Budapest: Akadémiai Kiadó.
- Juhász Dezső – Radványi Péter – Vékás Domokos. 2000. Egy dialektológiai adatbázis és elektronikus prezentációs rendszer alapvetése. *Magyar Nyelv* 86 (1): 47–58.
- Kiefer Ferenc. 1994. A magyar nyelv mint a modern kommunikáció eszköze. *Magyar Tudomány* (101) 39 (1): 631–642.
- Kiss Jenő. 1999. A dialektológia kettős feladata és a nyelvföldrajz. *Magyar Nyelv* 95 (4): 418–425.
- Lakatos Ilona, P., szerk. 2012. *Változó nyelvhasználat a hármasthatár mentén. Többdimenziós nyelvföldrajzi térképlapok tanúságai*. Budapest: Tinta Kiadó.
- Lakatos Ilona, P. – Károlyi Margit, T. – Iglai Edit. 2012. Többdimenziós nyelvföldrajzi ábrázolás: dinamikus nyelvi struktúra. In É. Kiss Katalin és Hegedűs Attila szerk. *Nyelvelmélet és dialektológia 2*. Piliscsaba: PPKE BTK, Elméleti Nyelvészeti Tanszék – Magyar Nyelvészeti Tanszék.
- Márton Gyula gyűjt., Hegedűs Attila szerk. 2000. *Változó nyelvhasználat a hármasthatár mentén. Szilágysági nyelvatlasz*. Budapest – Piliscsaba: Magyar Nyelvtudományi Társaság, Pázmány Péter Katolikus Egyetem Bölcsészettudományi Kara.
- Murádin László gyűjt., Juhász Dezső szerk. 1995–2010. *A romániai magyar nyelvjárások atlasza I–XI*. Budapest: Magyar Nyelvtudományi Társaság.
- Nagy Lajos, Cs. 2007. Jelentésföldrajzi vizsgálatok a nagyatlasz, az Új magyar tájszótár és a medvesalji gyűjtés alapján. In Guttmann Miklós, Molnár Zoltán szerk. *V. Dialektológiai Szimpozion*. 63–68. Szombathely: A Berzsényi Dániel Tanárképző Főiskola Magyar Nyelvészeti Tanszéke.
- Nagy Lajos, Cs. 2011. *Medvesalja magyar nyelvjárásai atlasza*. Nagykanizsa: Luminosus n.o. Kiadó.
- Nagy Lajos, Cs. 2013. Jelentésföldrajzi vizsgálatok informatizált nyelvatlaszok korpuszaiban. In Szoták Szilvia – Vargha Fruzsina Sára szerk. *Változó nyelv, nyelvváltozatok, területiség*. A VII. Nemzetközi Hungarológiai Kongresszus szekcióelőadásai. Kolozsvár, 2011. augusztus 22–27. Kolozsvár: Egyetemi Műhely Kiadó, Bolyai Társaság.
- Sándor Anna. 2004. *A Nyitra-vidéki magyar nyelvjárások atlasza*. Pozsony: Kalligram Könyvkiadó.



Lajoš Č. NAĐ

## DIGITALIZACIJA I INFORMATIZACIJA U ISTRAŽIVANJU DIJALEKATA MAĐARSKOG JEZIKA

Ovaj istorijsko-naučni rad prikazuje ulogu i značaj korišćenja kompjutera u dijalektologiji mađarskog jezika, njegov značaj i činjenicu da se kompjuteri ne mogu izostaviti iz dijalektoloških istraživanja. Prvi pokušaj primene računarske tehnike je bio eksperimenat sa *Dijalektološka karata mađarskog jezika*, koja se, na žalost, nije završila. Radovi koji slede su značajne stanice uvođenja kompjutera u obradu: *Karta mađarskih dijalekata u Rumuniji*, *Dijalektološka karta područja Silađ*, *Dijalektološka karta područja Medvešalja*, *Višedimenzionalne jezičko-geografske karte*. Poslednje ostvarenje je *Zvučni atlas*.

*Ključne reči:* digitalizacija, informatizacija, sistem elektronske prezentacije, digitalni arhiv, korpus razgovornog jezika, softver *Bihalboč*, geografija značenja, istraživanje dijalekata.

CS. NAGY Lajos

## DIGITALIZATION AND INFORMATIZATION IN HUNGARIAN DIALECTOLOGY

This study presents the role, importance and necessity of the computer in Hungarian dialectology from a historical viewpoint. The road leads from the first experiment – the unfinished process of digitalization of the *A magyar nyelvjárások atlasza* (*The atlas of Hungarian dialects*) – to the “Speaking atlas”. The atlases created between these two – *A romániai magyar nyelvjárások atlasza* (*The atlas of Hungarian dialects in Romania*), *Szilágysági nyelvátlasz* (*The linguistic atlas of Szilágyság*), *Medvesalja magyar nyelvjárési atlasza* (*The atlas of Hungarian dialects in Medvesalja*), *Többdimenziós nyelvföldrajzi térképlapok* (*Multidimensional linguistic maps*) – represent important stages in the development of digital and computational techniques in the research of Hungarian dialects.

*Keywords:* digitalization, informatization, electronic presentation system, digital archive, spoken language corpus, *BihalBocs* softver, language geography, dialectology.