

NKFP

Azonosító szám: 5/056/2004

ELTE

Magyar Nyelvtörténeti, Szociolingvisztikai, Dialektológiai Tanszék

Szerződésszám: OM-00217/2004

MTA Nyelvtudományi Intézete

Nyelvművelő Osztály

Szerződésszám: OM-00219/2004

A MAGYAR NYELVVÁLTOZATOK GEOLINGVISZTIKAI KUTATÁSA

**A regionális nyelvi örökség
gyűjtése, informatizálása és korpusznyelvészeti integrálása**

KUTATÁSI BESZÁMOLÓ A 4. MUNKASZAKASZRÓL

(2007. JANUÁR–OKTÓBER)

PROJEKTVEZETŐ: KISS JENŐ, AKADÉMIKUS

TARTALOMJEGYZÉK

I. A vállalt részfeladatok listája és státusa	3
II. A munkaszakasz feladatainak és eredményeinek bemutatása	5
1. Hagyományos dialektológiai adattárak informatizált változatának létrehozása	5
1.1. Az adatfájlok formátuma és kódolása	5
1.2. A magyar nyelvjárások atlasza VI. kötetének informatizálása	6
1.3. A romániai magyar nyelvjárások atlasza VIII. kötetének informatizálása	7
1.4. A Jelenségtérképek A moldvai csángó nyelvjárás atlaszából című kiadvány	7
2. A moldvai diakrón multimédiás nyelvatlasz	8
2.1. A diakrón nyelvatlasz terepmunkája	8
2.2. A felvett anyag feldolgozása	8
3. A magyar nyelvjárások atlasza munkálatai során készített hangfelvételek feldolgozása	10
3.1. A MNyA. szövegfelvételeinek digitalizálása, archiválása és informatizálása	10
3.2. A kérdőíves adatok beillesztése az atlasz informatizált változatába és egy szövegfelvétel-válogatás megjelentetése	11
4. A Geolingvisztikai Műhely honlapja	13
III. A munkaszakasz alatt megjelent publikációk listája ..	14
IV. Összefoglaló táblázat a munkaszakasz költségeiről	16

I. A VÁLLALT RÉSZFELADATOK LISTÁJA ÉS STÁTUSA

A kutatás negyedik szakaszában (2007. január–október) konzorciumunk összesen három fő feladat részleges vagy teljes megvalósítására vállalkozott:

1. hagyományos dialektológiai adattárak informatizált változatának létrehozása;
2. a moldvai diakrón multimédiás nyelvátlasz létrehozása;
3. A Magyar nyelvjárások atlasza hangfelvételeinek feldolgozása.

A jelzett részfeladat státusáról az alábbi táblázatokban megadott adatok azt mutatják, hogy kutatásunk utolsó szakasza terveinknek megfelelően fejeződött be.

4. beszámoló	Konzorciumi tag neve: ELTE Magyar Nyelvtörténeti Tanszék	2007. január– október
Részfeladatok megnevezése	A részfeladatok szakmai tartalma az adott beszámolási időszakban	A részfeladatok státusa
1. feladat 1. részfeladata	Elkészül A magyar nyelvjárások atlasza VI. kötetének informatizált változata.	Elkészült.
1. feladat 2. részfeladata	Elkészül A romániai magyar nyelvjárások atlasza VIII. kötetének informatizált változata.	Elkészült.
2. feladat 1. részfeladata	A diakrón nyelvátlasz terepmunkájának utolsó szakasza	Teljesült.
2. feladat 2. részfeladata	A diakrón nyelvátlasz hangfelvételeinek feldolgozása	Teljesült.
3. feladat 2. részfeladata	A magyar nyelvjárások atlasza ellenőrző gyűjtése során felvett szövegfelvételek digitalizált változatának átírása, egy szövegfelvétel-válogatás CD-n való megjelentetése	Elkészült.

4. beszámoló	Konzorciumi tag neve: MTA Nyelvtudományi Intézet, Nyelvművelő Osztály	2007. január– október
Részfeladatok megnevezése	A részfeladatok szakmai tartalma az adott beszámolási időszakban	A részfeladatok státusa
3. feladat 1. részfeladata	A magyar nyelvjárások atlasza ellenőrző gyűjtése során felvett szövegfelvételek digitalizálása, archiválása és dokumentálása	Elkészült.
3. feladat 2. részfeladata	Az ellenőrző gyűjtés során felvett kérdőíves adatok részleges átírása és dokumentálása	Elkészült.

A továbbiakban az egyes feladatok fent megadott sorrendjében tekintjük át a kutatás negyedik szakaszának eredményeit. Ezt követően az **5.** fejezetben röviden bemutatjuk azokat a változásokat, amelyeket a kutatócsoportunk tevékenységét és eredményeit szemléltető honlapon hajtottunk végre a kutatásunk előző szakasza óta eltelt időszakban.

II. A MUNKASZAKASZ FELADATAINAK ÉS EREDMÉNYEINEK BEMUTÁSA

1. Hagyományos dialektológiai adattárak informatizált változatának létrehozása

Az 1. feladatban meghatározott célkitűzéseink két részfeladatot foglalnak magukba. Ezek egy-egy hagyományos adathordozón (papíron) kiadott nyelvatlaszkötet informatizált változatának létrehozását jelentették. A két atlaszkötet a következő volt:

1. A magyar nyelvjárások atlasza VI. kötete;
2. A romániai magyar nyelvjárások atlasza VIII. kötete.

A két kiadványok informatizálása számos eltérő problémát vetett fel, ennél fogva érdemes áttekinteni, hogy milyen megoldásra váró nehézségek adódtak a részfeladatok során. Ezt megelőzően azonban röviden ismertetjük azt a nyelvészeti eljárást, amelyet a nyelvatlaszadatok informatizálásban felhasználtunk az adatfájlok formátumának és kódolásának meghatározásához. Már korábban informatizált atlaszunk, A moldvai csángó nyelvjárás atlasza (a továbbiakban: MCsNyA.) anyagából a kutatás jelen szakaszában készített kiadványunkat e fejezet utolsó részében mutatjuk be.

1.1. Az adatfájlok formátuma és kódolása

A projekt keretében végzett informatizálási munka a Bihalbocs dialektológiai szoftverrel történt. A szaktudományos feladatok megoldásához speciális programok szükségesek, ugyanis a kereskedelmi forgalomban beszerezhető szoftverek (szövegszerkesztők, hangszerkesztők, adatbáziskezelők, 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 megfelelően informatizált adattárak óriási előnye, hogy integrálhatók – ez a világos stuktúrából és a kompatibilis kódrendszerből következik. A nyelvjárási jelenségek kódolására olyan analitikus kódrendszert alkalmaztunk, amely a teljesség igényével optimálisan biztosítja az adatrögzítés pontosságát és az informatizált anyagok hatékony felhasználását. A szöveges anyagokat szövegfájlokban, illetve BXT-fájlokban tároljuk. 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 konkrét feladat összetettségéhez viszonyított egyszerűséget értjük.)

A már kiadott, illetve korábbi gyűjtésekből származó anyagok kódolása és tárolása szövegfórmátumban történik. Az adatok szigorúan rögzített struktúrába rendeződnek, és fonetikai kódok sorozatából állnak. Az alapjeleknek (például *e*, *é*) vagy mellékjeleknek (például a félhosszúság jele, a palatalizáltság jele) egyaránt külön kódok felelnek meg.

Kutatásunk jelen szakaszában vállalt feladatainknak megfelelően ebben a formátumban fejeztük be A magyar nyelvjárások atlasza VI. kötetének informatizálását,

valamint elvégeztük A romániai magyar nyelvjárások atlaszának VIII. kötetének informatizálását. A következőkben az egyes kiadványokhoz kapcsolódóan tekintjük át az informatizálás folyamatát.

1.2. A magyar nyelvjárások atlasza VI. kötetének informatizálása

A Magyar Tudományos Akadémia Nyelvtudományi Intézetében már korábban elkészült A magyar nyelvjárások atlasza (a továbbiakban: MNyA. vagy Nagyatlasz) első három kötete jelentősebb részének (pontosabban: az első 450 térképlapnak) a számítógépes változata. Ezt a munkát Balogh Lajos és Kiss Gábor irányította. Megemlítendő, hogy munkálataik kizárólag az atlaszadatok informatizálását tűzték ki célul, és ehhez nem kapcsolódott olyan – az általunk alkalmazotthoz hasonló – szoftver, amely a már informatizált adatok elemzését (statisztikai lekérdezések, keresés, analitikus térképek generálása stb.) is lehetővé tette volna. Balogh Lajos emellett részt vett abban a kutatásunkat megelőző folyamatban is, amely az atlasz soron következő, IV. kötetéhez kapcsolódott, és immár az előző pontban említett Bihalbocs lejegyzőszoftverrel történő informatizálásnak az előkészítését segítette elő.

Az előkészítést követően a IV. kötet közel 200 térképlapjának informatizálása már a Bihalbocs szoftverrel valósult meg projektünk első szakaszában. A második szakaszban az V. kötet informatizálását kezdtük meg, ebben periódusban fejeztük be az V. kötet adatainak rögzítését. Az adatbevitel után a projekt harmadik szakaszában végeztük el a kötet tételes ellenőrzését és a hibák javítását. A kötet adatainak teljes körű informatizálását (rögzítését, ellenőrzését és javítását) 2006 októberére fejeztük be. A VI. kötet informatizálását a már bevált „menetrendet“ követve korábbi munkatársainkkal végeztük el 2007-ben, a kutatás utolsó szakaszában.

Az adatrögzítéshez kölcsönkértük az atlasznak a Nyelvtudományi Intézetben archivált kéziratát, és az adatokat ez alapján vittük számítógépre. Ezzel a kiadott térképlapok és az informatizált fájlok egyaránt a gépírási kéziratlapok alapján jöttek létre. Az informatizált fájlok az atlasz egy-egy térképlapjának felelnek meg, elnevezésük is utal a kiadott kötet lapjainak számozására (például a ma*0571.txt fájl a Nagyatlasz 571. térképlapjának adatait tartalmazza).

A kézirat és a kiadott térképlapok struktúrája egyetlen pontban tér el az informatizált változatétól: a fő adatsorokat követő megjegyzésekben. A kiadott térképlapok terjedelmi megfontolásból egyes adatokat csak megjegyzésként adtak közre (ezeket „a térképről leesorult adatok“ címet követően a térképlapok legendájában tüntették fel). Az informatizált változatban ilyen terjedelmi megszorítások értelemszerűen nincsenek, ennél fogva ezeket az adatokat beillesztettük a fő adatsorba. Ugyancsak így jártunk el azokkal az adatokkal, amelyek azért kerültek csak a megjegyzések közé, mert a próbagyűjtés, a gyűjtés fő szakasza és az ellenőrző gyűjtés közül csak egyikben fordultak elő. Ezeket az adatokat azonban minden esetben lábjegyzettel láttuk el, a gyűjtésről szóló információt tehát más formában ugyan, de megőriztük.

A pályázatban vállaltaknak megfelelően legnagyobb nyelvátlaszunk három kötetének (IV–VI.) összesen közel 600 térképlapját (az 571.-től az 1162-ig) hoztuk korszerűen kutatható, elemezhető formába. Mivel a számítógépes rögzítés során a kéziratos eredeti példányokat használtuk, adatbázisunk a térképen megjelent kiadvánnyal összemérhető pontosságú adatokat tartalmaz.

1.3. A romániai magyar nyelvjárások atlasza VIII. kötetének informatizálása

A romániai atlasz informatizálási folyamatát projektünk első szakaszában az V. kötettel kezdtük meg, majd a másodikban a VI., a harmadikban a VII. kötettel folytattuk. Az előző kötetek informatizálásához kialakított módszerrel rögzítettük és javítottuk A romániai magyar nyelvjárások atlasza VIII. kötetét, megőrizve az adatokra vonatkozó minősítéseket, megjegyzéseket és az adattár nyomtatott változatában alkalmazott csoportosítást. Az így létrehozott adattár térképre vetíthető, kereshető és integrálható a munkánk során rögzített más adattárakkal. A VIII. kötet 300 térképlapját a Bihalbocs dialektológiai szoftverrel rögzítettük. (Az egyes fájlok elnevezése itt az rma1200.txt szerkezethez hasonlóan hordozza a kiadott atlaszkötet térképlapjainak számozását.)

A pályázatban vállaltaknak megfelelően négy atlaszkötet (V–VIII.) összesen 1200 térképlapjának anyagát hoztuk korszerűen kutatható, elemezhető formába. Az ellenőrzés során az adatokat összevetettük Murádin László kéziratos térképeivel is, így lehetőségünk volt az adatbevitel alapjául szolgáló nyomtatásban megjelent változat esetleges hibáinak a javítására. A RMNyA. adataiból létrehozott adattár nem zárt rendszer: szükség szerint bármikor módosítható, tovább javítható, bővíthető. Az informatizált adatok kereshetők, tetszőleges szempontok szerint csoportosíthatók, integrálhatók más adattárakból származó, megfelelően rögzített nyelvjárási adatokkal.

1.4. A Jelenségtérképek A moldvai csángó nyelvjárás atlaszából című kiadvány

A MCsNyA. három kötetének informatizált változata alapján Bodó Csanád és Vargha Fruzsina Sára készítette el a fenti című CD-kiadványt (bibliográfiai adatait lásd a III. fejezetben). E kiadvány ötven olyan jelenségtérképet ad közre, amelyek valamilyen nyelvi szempont alapján értelmezik a papíron megjelent térképek adatait, többnyire két-három térképlap adatainak integrálásával. A CD-ROM részeként a MCsNyA. három kötetének teljes anyagáról is közlünk információkat: közreadjuk a címszavak listáját a megjelenés sorrendjében és betűrendben, valamint az atlasz teljes anyagát összesítő Excel-táblázatokat. Mivel az atlasz mostanáig lappangó és általunk informatizált III. kötetének nem készült hagyományos adathordozón (papíron) hozzáférhető változata, ezzel a CD-kiadvánnyal tesszük közzé ezt az anyagot.

2. A moldvai diakrón multimédiás nyelvátlasz

A moldvai diakrón nyelvátlasz létrehozásának két fő feladata – a kutatás megtervezését és a próbagyűjtéseket követő szakaszban – a következő volt:

1. a diakrón nyelvátlasz terepmunkája;
2. a felvett anyag folyamatos feldolgozása.

2.1. A diakrón nyelvátlasz terepmunkája

Kutatásunk első szakaszában hat településen készítettük el 18–18 adatközlővel a nyelv földrajzi interjút. Ezek a települések Klézse, Pusztina, Frumósza, Újfalu (a Tatros mentén), Trunk és Bahána. Közülük a legutolsó település kivételével mindegyik falu a régebbi atlasz kutatópontja is volt. Csak 10 adatközlőnk volt Szolonckán, ahol a középkorú korosztályban sem teljes körű már a magyar nyelv használata, valamint 8 interjú készült Pokolpatakon. Egy további kutatóponton, Szőlőhegyen a felmérés megkezdődött, de csak négy beszélővel végeztünk adatfelvételt.

A projekt második szakaszában a terepmunka Somoskán és Szitáson folytatódott, mindkét településen 18–18 adatközlővel, valamint befejeződött Szőlőhegyen (14 további adatközlővel).

A projekt harmadik szakaszára esett a moldvai terepmunka második nagyobb szakasza. Ebben a kutatási periódusban a következő településeken folyt terepmunka: Szabófalva (eddig 5 interjú készült el), Lészped, Lujzikalagor, Gajdár, Bogdánfalva, Gyoszény, Ketris, Szalánc (Ciresoia), Onyest, Gorzafalva és Vizánta.

A kutatás utolsó szakaszában terepmunkát folytattunk Nagypatakon, Esztufujban, Lábnikon és Dózsában.

A projekt során összesen 26 települést kerestünk fel, és mindenhol a korábbi beszámolókból kifejtett elveknek megfelelő számú adatközlőt kérdeztünk ki (kivéve Szabófalvát, ahol csak öt interjú készíthetett el, valamint Pokolpatakot, ahol 8 felvétel készült). Kutatópontjainkon összesen 408 adatközlővel készítettünk interjút.

2.2. A felvett anyag feldolgozása

A terepmunka alatt felvett anyag feldolgozása folyamatosan zajlott. Itt két részfeladatot kellett megoldani: egyrészt standardizáltuk az adatok és kísérő információik kódolását, másrészt pedig kialakítottuk azt az informatikai keretet, amelyben a nyelvjárási adatok hangzó és átírt formája egymással összekapcsolódik.

Az előbbi feladathoz a MCsNyA. eljárásával azonos módon a magyar egyezményes átírást választottuk, de az egyes adatok minősítését már a korábbi atlaszhoz képest részletesebben oldottuk meg. Ez azt jelenti, hogy egyrészt minden egyes adathoz hozzátartoznak az adatközlő szociológiai és nyelvészeti jellemzői (pl. életkora, neme, esetleges magyarországi tartózkodásának időtartama; valamint nyelvválasztási szokásai, nyelvtudása, a magyar nyelv helyi és más változataihoz fűződő attitűdjei), másrészt pedig azt, hogy az adatközlő által adott minősítések ugyancsak megjelennek a lejegyzett szövegben az adatot követően.

A projekt keretében folytatott terepmunkából származó adatokat olyan formátumban tároljuk, amely lehetővé tette, hogy a Bihálbocs dialektológiai szoftver további fejlesztésével kattintható multimédiás nyelvjárási atlaszt hozzunk létre – ebben

a kutatóponthoz nem csak a lejegyzett adatot, hanem annak hangzó formáját is hozzárendeljük.

Az adatok lejegyzéséhez egy általunk kialakított fájlformátumot használtunk, amelyet BXT-nek nevezünk. A BXT-t abból a célból fejlesztettünk ki, hogy lehetővé tegyük 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. Az ebben a formátumban tárolt anyagok tehát lehetővé teszik a markerekkel megjelölt adat gombnyomásra történő meghallgatását az adott hangfájl megkeresése és megnyitása nélkül.

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.

A kutatás harmadik szakaszában készített felvételek teljes körű feldolgozása 2007 áprilisára fejeződött be. Ezt követően a felvételek ellenőrzése mellett a kutatás utolsó szakaszában felvett nagypataki, esztufuji, lábni és dózsai interjúk átírása következett.

A kutatásunk során gyűjtött anyag feldolgozásának lezárultával lehetővé vált a MCsNyA. adatainak és az újabb gyűjtésű anyagnak az egyetlen közös adatbázisban való egyesítése. Ezzel létrejött a multimédiás diakrón moldvai nyelvatlasz. Az atlasz bemutatására Bodó Csanád vállalkozott az V. Dialektológiai Szimpozionon, Szombathelyen, 2007 augusztusában. Az előadás demonstrációs anyaga megtekinthető a http://geolingua.elte.hu/about/eload/bodocs_VDialSzimp.ppt.htm címen.

3. A magyar nyelvjárások atlasza munkálatai során készített hangfelvételek feldolgozása

A MNyA. ellenőrző gyűjtései során, az 1960-as évek első felében készített hangfelvételek feldolgozását két részfeladatra osztva végeztük el. Ezek a következők voltak:

1. A szövegfelvételek digitalizálása, archiválása és dokumentálása;
2. A kérdőíves adatok beillesztése az atlasz informatizált változatába; valamint egy szövegfelvétel-válogatás CD-n való megjelentetése.

A két részfeladat egyaránt folyamatosan, a projekt teljes időtartama alatt zajlott; az utóbbi részfeladatban egy CD-sorozat első három kiadványának vállaltuk évenként egy-egy kiadvánnyal. A sorozat harmadik darabja 2007 nyarára készült el.

3.1. A MNyA. szövegfelvételeinek digitalizálása, archiválása és dokumentálása

Az első részfeladat elvégzése során a beszámolási időszakban folytatódott és befejeződött *A magyar nyelvjárások atlaszának* a múlt század 60-as éveiben végzett ellenőrző gyűjtése során magnetofonszalagra felvett archív szövegfelvételek digitalizálása és CD-re írása. A munka sikeres elvégzését az MTA Nyelvtudományi Intézete és az MTA Zenetudományi Intézete közötti megállapodás tette lehetővé azért, hogy az elvégzendő feladathoz igénybe vehettük a Zenetudományi Intézet elektronikus berendezéseit és az ezek működtetéséhez szükséges infrastruktúrát.

A digitalizált hanganyagok fájlformátuma (WAV, 44100-as mintavételezés) a tudományterület legigényesebb elvárásainak is megfelel, és biztosítja, hogy a megismételhetetlen hangfelvételek minősége kicsit se csorbuljon. A viszonylag nagyobb tárhelyigény napjainkban – a nagy kapacitású adattároló eszközök árának egyre gyorsuló ütemű csökkenésével – már a legtöbb intézmény számára nem jelent megoldhatatlan problémát. További, legalább ennyire releváns szempont, hogy az akusztikai fonetikai elemzésekhez használatos szoftverek (például a Praat) nem olvasnak be tömörített (például mp3) formátumú fájlokat, a wav-formátumot ellenben minden elemző program támogatja. A veszteségesen tömörítő fájlformátumok (pl. mp3) alkalmazása nem ajánlatos a terepmunkával gyűjtött adatokat is felhasználó tudományterületeken (pl. fonetika, dialektológia, szociolingvisztika). Mi is a wav-változatot használjuk fel az interjúk lejegyzésének és hanganyagának összekapcsolásakor.

Az oktatási és ismeretterjesztő szintű internetes anyagok esetében viszont egészen más szempontok érvényesek. Itt fel sem merül az akusztikai vizsgálat igénye, annál fontosabb viszont, hogy ezek az anyagok könnyen hozzáférhetők, megfelelő sávszélesség esetén pillanatok alatt letölthetők, meghallgathatók legyenek. Ennélfogva ezekben az anyagokban mi is mp3-formátumú hangfájlokat használunk.

A digitalizálási munkálatok lezárultával kifejezetten oktatási és népszerűsítő céllal olyan internetes felületet hoztunk létre, amelyen keresztül az érdeklődők részleteket hallgathatnak meg a MNyA. ellenőrző gyűjtése során készített interjúkból. A közzétett részleteket az egyes kutatópontokhoz kapcsolva, kattintható térkép segítségével lehet meghallgatni. Így az ellenőrző gyűjtés teljes kutatóponthálózatát

alapul véve mind a 352 kutatópontról közlünk egy-egy hangfelvétel-részletet mp3-formátumban. Reményeink szerint az így létrehozott internetes nyelvjárási hangmintatár hatékonyan segíti majd az egyetemi és főiskolai hallgatókat (sőt, esetleg a közoktatásban résztvevőket is) a magyar nyelv területi sokszínűségének és az egyes nyelvjárások jellemzőinek megismerésében.

A hangminták megtekintéséhez a Geolingvisztikai Műhely honlapján a Térképek részt kell felkeresni, illetve közvetlen kattintással a következő címen találhatjuk meg az adatokat: http://geolingua.elte.hu/maps/MNyA_hangosterk/index.html.

3.2. A kérdőíves adatok beillesztése az atlasz informatizált változatába és egy szövegfelvétel-válogatás megjelentetése

A részfeladat elvégzése során a restaurált és digitalizált hangfelvételeken rögzített kérdőíves kérdések bekerültek az atlasz informatizált változatába. Emellett elkészült egy szövegfelvételeket tartalmazó – oktatási és kutatási célokra egyaránt alkalmas – CD-ROM is, a Magyar nyelvjárási hangskönyv című sorozatunk harmadik darabja (bibliográfiai adatait lásd a III. pontban).

A kiadvány az eddig digitalizált anyagból olyan felvételeket ad közre, amelyeknek a fonetikai átírása már megjelent a Hajdú Mihály és Kázmér Miklós által készített Magyar Nyelvjárási Olvasókönyvben (Bp., 1974.). Az Olvasókönyv átírt szövegeit és a hangfelvételeket együtt közreadó multimédiás CD-ROM Banáról, Bakonyszentlászlóról, Somlóvásárhelyről, Patakról és Nagyölvendről származó adatokat tartalmaz. Mint korábban a sorozat első kiadványainak esetében, most is két változatot hoztunk létre: a kutatói változatot, valamint az abból a Bihalbocs szoftverrel részben automatikusan generált ismeretterjesztő változatot.

A sorozat első három darabja így összesen 14 kutatópontról ad közre nyelvjárási szövegfelvételeket. Mindhárom kiadvány internetes változata megtalálható a Geolingvisztikai Műhely honlapján (<http://geolingua.elte.hu>), a Hangskönyv rovatban. Rögzítettük további 20 település anyagát is, szintén az eredeti hanganyaggal szinkronizálva, ezekhez a szövegekhez egyelőre kutatói változatban lehet hozzáférni.

A Hangskönyv lényeges jellemzője, hogy az eredeti hangfelvételeket és fonetikus lejegyzésüket a szövegben elhelyezett időzítési markerekkel szinkronizáljuk. Hanganyag és lejegyzés kapcsolata az Interneten is közzétett oktatási és népszerűsítő változatban is megőrződött. Lényeges különbség van azonban a két változat között a struktúrát és a hanganyag minőségét, valamint már említett formátumát illetően. A kutatói változat interjúként egyetlen, speciálisan erre a célra kialakított, a lejegyzés mellett minden más szükséges információt is tartalmazó fájlból és a hozzá kapcsolódó (egy vagy több hosszú) hangfájlból áll, ahol a kapcsolódásokat a tárolt információk alapján a geolingvisztikai szoftver (Bihalbocs) biztosítja. A platformfüggetlen, internetes változat, amelyet a kutatói változathoz hozunk létre, ennél jóval bonyolultabb szerkezetű (az internetes böngészők lehetőségeinek függvényében), mivel itt a (rövid) hangrészletek meghallgatását a hangfájlok és az eredeti lejegyzés feldarabolásával tudjuk csak biztosítani. Így interjúként egyrészt közreadjuk a teljes lejegyzést és a hozzá tartozó, meghallgatható interjút (mp3-formátumban), másrészt annyi szövegrészletet tartalmazó HTML-fájlt és hozzájuk kapcsolódó, csak az adott részletet tartalmazó (mp3-formátumú) hangfájlt, ahány részletet korábban (a kutatói változat elkészítésekor) szinkronizációs pontok megadásával kijelöltünk. Így az internetes változat sokszáz vagy akár sokezer hangfájl és sokszáz más szövegfájl rendszeréből áll.

Kialakításához a geolingvisztikai szoftver – megfelelő beállítások után – automatikusan végzi el a feladatok tekintélyes részét a kutatói változat alapján (például a hangfájlok feldarabolását, a lejegyzések feldarabolását és html-formátumba való konvertálását). Nem speciális szoftverekkel (hangszerkesztővel, szövegszerkesztővel, html-szerkesztővel) szinte vállalhatatlanul nagy feladat lenne ilyen termékek előállítását.

A kutatói változat alapján elkészített népszerűsítő (oktatási és ismeretterjesztő) változatot honlapunkon is hozzáférhetővé tettük, megtalálható a következő címen: http://geolingua.elte.hu/hkonyv/Hangoskonyv3_index.html. A mellékjeles lejegyzések megtekintéséhez a Firefox, Opera vagy Safari internetes böngésző használata szükséges.

A MNyA. ellenőrző gyűjtésének hanganyaga a szövegfelvételek mellett tartalmazza annak a 41 atlaszkérdésnek a válaszait is, amelyek a papíratlasz elkészítése során nem kerültek térképre, korszerű, a hangzó anyagot alapul vevő feldolgozásukra annak idején nem volt lehetőség. Az atlaszkérdésekre adott válaszok rögzítésére jelen pályázat keretein belül mi is csak részlegesen vállalkozhattunk. Az adatok feldolgozását a Magyar nyelvjárási hangoskönyv sorozat kialakításának szempontjaihoz is igazodva olyan településekről származó anyagokkal kezdtük meg, amelyek egyaránt kutatópontjai voltak a Magyar nyelvjárási olvasókönyvnek és a Nagyatlasznak is.

Az atlaszadatok rögzítése során a moldvai multimédiás atlasz interjúinak feldolgozásához kialakított technológiát alkalmaztuk. A feldolgozás alapegysége az interjú, így az egyes adatok nem közvetlenül kutatóponthoz, hanem adatközlőhöz kötődnek. Minden adat megtartja kapcsolatát eredeti hangzó formájával. Ezt úgy biztosítjuk, hogy szinkronizációs pontokat (időzítési markereket) helyezünk el a szövegben, amelyek egy adott hangfájl adott időpillanatára mutatnak. Az eredeti hangzó forma elérhető marad az adatok adatbázisba rendezése után is, így valamennyi adat azonnal meghallgatható, a fonetikus lejegyzés bármikor ellenőrizhető és javítható.

Az eredeti szövegfájlokból automatikusan generált adatbázishoz kialakított keresőfelület segítségével általunk megadott szempontok szerint csoportosíthatjuk, térképezhetjük az adatokat. Lehetőségünk van az adatok és a keresendő kifejezés lejegyzési finomságát pontosan meghatározni. Ez azt jelenti, hogy bár a lejegyzés során a MNyA. gyakorlata szerint járunk el, az adatok feldolgozásakor általunk megadott szempontok szerint tekinthetünk azonosnak vagy eltérőnek például egy diftongust és egy hosszú monoftongust, egy adott hang nyíltabb vagy zártabb változatát. Mivel az adatok megőrzik kapcsolatukat az adatközlővel az adatbázisba rendezés után is, lehetőségünk van az adatközlők paramétereire szerinti keresésre. Egyetlen beállítás megváltoztatásával készíthetünk térképet az általunk vizsgált nyelvi jelenség nemek vagy kor szerinti megoszlásáról.

A MNyA. ellenőrző gyűjtésének atlaszadataiból kialakított adattárról és elemzési lehetőségeiről Vargha Fruzsina Sára tartott előadást az V. Dialektológiai Szimpozionon, Szombathelyen, 2007. augusztus 22-én. Az előadás HTML-változata megtekinthető a http://varghafr.web.elte.hu/VarghaFr_Szombathely2007.htm címen.

4. A Geolingvisztikai Műhely honlapja

Szükségesnek tartjuk, hogy az általunk végzett kutatások, létrehozott nyelvjárási adattárak, nyelvészeti technológiák szélesebb körben is ismertté váljanak. Ennek érdekében kutatásunk előző szakaszában internetes honlap kialakítását kezdtük meg, és a honlap tartalmi fejlesztését a jelen beszámolási szakaszban is folytattuk. A honlapot felkeresve az érdeklődőknek lehetőségük van bepillantani kutatócsoportunk tevékenységébe, felvenni a kapcsolatot kutatócsoportunk tagjaival, az egyes részfeladatokat irányító kutatókkal.

Az általunk létrehozott nyelvjárási anyagokat (hangoskönyveket, nyelvjárási térképeket) a Bihalbocs szoftverrel olyan formába konvertáljuk, amely alkalmas a világhálón való megjelenítésre.

Honlapunkon teljes terjedelmében elérhetővé tettük korábbi kutatási beszámolóinkat. Előző kutatási szakaszunk végére a honlapon az általunk készített munkák közül a következő anyagokat helyeztük el: (a) a Magyar Nyelvjárási Hangoskönyv című sorozatunk I–II. kiadványának HTML-változatát, (b) néhány mintatérképet az általunk informatizált nyelvföldrajzi adattárak anyagából, (c) A magyar nyelvjárások atlasza ellenőrző gyűjtése során készített hangfelvételek listáját, (d) a Háromszék megyei történeti névtani anyag informatizált változatából készített mintatérképeket, valamint (e) mutatványt a MNyA. ellenőrző gyűjtése során felvett szövegfelvételekből (teljes kutatóponthálózattal).

E kutatási szakaszban a honlap a következő anyagokkal bővült: (a) a Magyar nyelvjárási hangoskönyv című sorozat III. kiadványának HTML-változata, (b) további mintatérképek az általunk informatizált nyelvföldrajzi adattárak anyagából, (c) a kutatási szakaszban elhangzott előadások demonstrációs anyagai, valamint (d) teljesebbé vált a honlap angol nyelvű változatának tartalma.

A honlap a következő címen érhető el: <http://geolingua.elte.hu>.

III. A munkaszakasz alatt megjelent publikációk listája

1. Önálló kiadványok. A munkaszakasz alatt két önálló szakmai kiadvány jelent meg kutatócsoportunk tagjaitól geolingvisztikai projektünk részeként:

Magyar Nyelvjárási Hangoskönyv III. Szövegek Banáról, Bakonyszentlászlóról, Somlóvásárhelyről, Patakról és Nagyövedről. Multimédiás CD-ROM. Szerk. Hajdú Mihály és mtsai. Budapest: ELTE Magyar Nyelvtörténeti, Szociolingvisztikai, Dialektológiai Tanszék. 2007.

Bodó Csanád – Vargha Fruzsina Sára: *Jelenségtérképek A moldvai csángó nyelvjárás atlaszából.* Budapest: ELTE Magyar Nyelvtörténeti, Szociolingvisztikai, Dialektológiai Tanszék. 2007.

2. Tanulmányok, cikkek. Előző kutatási beszámolóinkban még publikálatlan dolgozatként említettük a következő, időközben megjelent tanulmányokat:

Bárth M. János: Szabó T. Attila Erdélyi Helynévtörténeti Adattára. In: Bárth M. János (szerk.), *Emlékkönyv Szabó T. Attila születésének 100. évfordulójára.* Bp. 2006. 63–66.

Bárth M. János: Háromszéki helynevek nyelvészeti elemzése informatikai módszerekkel. In: Hoffmann István – Tóth Valéria (szerk.), *Helynévtörténeti tanulmányok 2.* Debrecen. 2006.

Heltai János Imre: Új lehetőségek a moldvai magyar beszélőközösségek nyelvi viselkedésének vizsgálatában. Heltai Pál (szerk.), *Nyelvi modernizáció. Szaknyelv, fordítás, terminológia.* I. kötet. MANYE – Szent István Egyetem, Pécs – Gödöllő. 2007. 72–77.

Vargha Fruzsina Sára: Állatok kicsinyeinek megnevezése a keleti magyar nyelvjárásokban. In: Hoffmann István – Juhász Dezső szerk. *Nyelvi identitás és a nyelv dimenziói.* Nemzetközi Magyarágtudományi Társaság, Debrecen–Budapest, 2007. 237–248.

3. Megjelenés előtt álló dolgozatok. Elsőként azokat a tanulmányokat mutatjuk be, amelyek a 4. részfeladat során készültek el, közülük elsőként az V. Dialektológiai Szimpozion előadásából készült tanulmányokat. Az újabban elkészült dolgozatokat követően a már korábban publikálásra benyújtott és elfogadott dolgozatokat soroljuk fel.

Az V. Dialektológiai Szimpozionon elhangzott előadások alapján készített tanulmányok közül Bodó Csanád *Követéses geolingvisztikai vizsgálat Moldvában* című dolgozata a multimédiás diakrón nyelvatlasz általános jellemzését nyújtja és egy konkrét nyelvi jelenség változásának elemzésén keresztül mutatja be az atlasz hasznosításának lehetőségeit. Az előadás demonstrációs anyaga megtekinthető a következő címen: http://geolingua.elte.hu/about/eload/bodocs_VDialSzimp.ppt.htm.

Heltai János Imre *Vélekedések a magyar nyelv területi változatairól Moldvában* című dolgozata az előző tanulmányhoz hasonlóan a moldvai multimédiás atlasz adatait használta fel.

Vargha Fruzsina Sára ugyancsak Szombathelyen *Nyelvi változók A Magyar nyelvjárások atlasza hangfelvételeiben* címmel tartott előadást, ennek ppt-változata itt érhető el: http://varghafr.web.elte.hu/VarghaFr_Szombathely2007.htm.

Vékás Domokos *Számítógépes dialektológia* című szombathelyi előadásában átfogó képet adott a tudományterület nemzetközi eredményeiről, a leginkább újszerű megoldások bemutatásához elsősorban a pályázatunkban létrehozott informatizált adattárakra támaszkodott.

Vargha Fruzsina Sára a helynévtörténeti anyagok dialektológiai felhasználásának lehetőségeiről tartott előadást *Nyelvjárási és helynévtörténeti anyagok számítógépes feldolgozása* címmel a Kontextus–Filológia–Kultúra II. nemzetközi konferencián, Besztercebányán, 2007. május 30-án. Az előadás írott változata a konferenciakötetben fog megjelenni.

Bárth M. János *Székelyföldi történeti helynevek névföldrajzi vizsgálata* című előadása Balatonszárszón, a VI. Névtudományi Kongresszuson hangzott el 2007. június 23-án. (A konferenciakötet várható megjelenése 2007 decembere.)

Bárth M. János *Vápa ~ lápa* című előadása a 2. Helynévtörténeti Szemináriumon hangzott el (Síkfőkút, 2007. október).

Bodó Csanád *A moldvai magyar nyelvjárások román kölcsönszórétégeinek területisége* című dolgozata a Péntek János-Emlékkönyv számára készült, a kötet megjelenése még ez évben várható.

Továbbra is megjelenés előtt állnak olyan tanulmányok, amelyek már kutatásunk korábbi szakaszaiban elkészültek. Ezek a következők:

A Magyar Nyelv c. folyóiratban jelenik meg Bodó Csanád és Vargha Fruzsina Sára *Régi nyelvatlaszok – új módszerek* c. tanulmánya, amely a Magyar Nyelvtudományi Társaság 2005. október 25-ei felolvasóülésén elhangzott előadás átdolgozott változata. Az előadás PowerPointtal készült prezentációja megismerhető honlapunkon a következő címen: <http://geolingua.elte.hu/about/eload/regiterk.htm>.

A XIV. Élőnyelvi Konferencián kutatócsoportunk tagjaitól két előadás is elhangzott: Bodó Csanád tanulmánya *A magyar nyelv vitalitásának területisége Moldvában* címmel, valamint Vargha Fruzsina Sára *A Magyar nyelvjárások hangoskönyv és elemzési lehetőségei* című tanulmánya. Az előbbi előadás PowerPoint-prezentációja honlapunkon a http://geolingua.elte.hu/about/eload/bodocs_buk.ppt.htm cím alatt található. A konferenciakötet megjelenése jövő év első felében esedékes.

IV. Összefoglaló táblázat a munkaszakasz költségeiről

A munkaszakasz összesített – tervezett és tényleges – költségeit a következő táblázat mutatja be. A két konzorciumi tag költségeit együttesen adjuk meg. Nem tüntetjük fel a táblázatban a saját és egyéb forrásokat, mivel ezek mind a szerződésben foglalt költségvetésben, mind a felhasznált támogatások megoszlásában 0 Ft-ot mutattak.

Kölségtípusok	Tervezett támogatás (Ft)	Felhasznált támogatás (Ft)
Működési költségek (1+2+3+4)	11262000	11262000
Személyi juttatások (1)	6700000	7541762
Munkaadókat terhelő járulékok (2)	1562000	1515562
Külső megbízások (3)	0	126000
Egyéb dologi kiadás (4)	3000000	2078676
Működési költségekből (1+2+3+4) rezsiköltség	2252000	1126200
Felhalmozási költségek (5+6)	0	0
Immateriális javak beszerzése (5):	0	0
Gépek, berendezések, felszerelések beszerzése (6)	0	0
Költségek összesen: (1+2+3+4+5+6)	11262000	11262000

A szerződésben foglalt költségvetéstől a személyi juttatások és a dologi kiadások rovatban jelentkezett nagyobb eltérés. A kutatás költségvetésének megtervezésekor a rezsiköltséget jóval nagyobb összeggel terveztük meg, mint amekkorát ténylegesen erre fordítottunk a munkaszakaszban. Az így felszabaduló összeget főként a gyermeke születése miatt a kutatócsoportból kiváló kollegánk helyettesítésének költségeire fordítottuk. A helyettesítés relatíve magasabb költségeit azzal ellensúlyoztuk, hogy nappali tagozatos egyetemi hallgatókat és doktorandusokat vontunk be a munkába (a részükre kifizetett személyi juttatások nem jelentettek járulékterhet a kutatás költségvetése számára, mivel havi jövedelmük a járulékmentes összeg felső határa alatt maradt).