

FERENCZI GÁBOR

**NYELVI ADAT ÉS ADATINFORMATIZÁLÁS: ADALÉKOK EGY NYELVFÖLDRAJZI  
KUTATÁS ELMÉLETÉHEZ****Kivonat**

A tanulmány a nyelvi adat modern értelmezése, továbbá az adatinformatizálás és a kvantitatív nyelvföldrajzi vizsgálatok típusainak vázlatos áttekintése után a Kemenesalja és vidéke nyelvföldrajzi kutatásának (röviden: KNyK.-projekt) helyét és szerepét ismerteti a kvantitatív nyelvföldrajzi vizsgálatok sorában. A KNyK.-projekt fókuszában a területi kötöttséget mutató, élő nyelvi adatok szociodialektológiai elemzése, valamint azok korszerű informatizálási módszerekkel való térképes ábrázolása áll.

**Kulcsszavak:** nyelvi/nyelvjárási adat, adatinformatizálás, kvantitatív nyelvföldrajz, dialektometria, KNyK.-projekt

**1. Bevezetés**

Tanulmányom a nyelvi adat modern értelmezése, továbbá az adatinformatizálás és a kvantitatív nyelvföldrajzi vizsgálatok típusainak vázlatos áttekintése után a Kemenesalja és vidéke nyelvföldrajzi kutatásának (röviden: KNyK.-projekt) helyét és szerepét ismerteti a kvantitatív nyelvföldrajzi vizsgálatok sorában. A KNyK.-projekt fókuszában a területi kötöttséget mutató, élő nyelvi adatok szociodialektológiai elemzése, valamint azok korszerű informatizálási módszerekkel való térképes ábrázolása áll. A dialektológia, szociolingvisztika, geolingvisztika, szociálpszichológia, számítógépes nyelvészet együttállásával a tér, idő és társadalom hármasságának koordináta-rendszerében megsokszorozódik a nyelvléirási szempontok, illetőleg a nyelvi változások mögöttes okainak feltárását lehetővé tevő magyarázatok száma is. A projekt interdiszciplináris jellegű: mindez megmutatkozik a vizsgálatba bevont nyelvi változók megválasztásában, a módszertani alapelvek kijelölésében, az adatgyűjtés és -felhasználás technológiájában, továbbá a nyelvi változatosság okainak feltérképezésében (a KNyK.-projekt elvi-módszertani alapvetéseit l. Ferenczi 2021: 11–46).

A nyelvi adat újszerű értelmezése kapcsán létrehozott adatbázisok részben az újonnan gyűjtött korpuszok sokrétű kutatását veszik célba, részben gazdagíthatják a hagyományos, nyomtatott formában kiadott atlaszok felhasználási lehetőségeit, részben új utakat nyithatnak meg a kéziratban fennmaradt

nyelvatlaszok előtt: gyorsíthatják, hatékonyabbá tehetik a nyelvtudományban való hasznosításukat. Nemzetközi és hazai szinten egyaránt számos kutató hangsúlyozza a számítógépes dialektológia lehetőségeit: a központi jelentőségű eljárást, az adatok informatizálását pedig egyhangúlag az adatelemzés elengedhetetlen feltételének tartják (Chambers–Trudgill 2004: 20; Bodó–Vargha 2008: 336). A modern nyelvtechnológiai eljárasmódok általános elterjedésével a *számítógépes* jelző szükségtelenné válhat a közeljövőben a dialektológia mint tudományszak megnevezésében (Vékás 2007).

## 2. A nyelvi adat mibenlétéről

A múlt század kilencvenes éveitől kezdve nemzetközi szinten megfigyelhető, hogy kezdetét vette a korábban gyűjtött nyelvjárási adatok informatizálása, nyelvjárási adattárakba rendezése (többek között az amerikai nyelvatlasz, az *American Linguistic Atlas Project*, ALAP átírási projektje kapcsán l. LAMSAS, illetve Kretzschmar–McDavid–Lerud–Johnson 1993, Anglia nyelvatlaszának számítógépes feldolgozására l. Viereck 1997, a Wenker-nyelvatlasz digitalizálásához l. DiWA 2001–2009). A számítógépes nyelvföldrajz azonban – szűkebb értelmezési keretben – a Los Angeles-i Kaliforniai Egyetemen (UCLA) tevékenységéhez köthető. A nyolcvanas évek végén – a multimédiás számítógépek megjelenésével – ugyanitt született meg a nyelvi adat modern értelemben vett fogalma. Vagyis: terepen gyűjtött, digitálisan tárolt, konkrét földrajzi területhez, kutatóponthoz kötött, fonetikus lejegyzett, mindamelllett hangzó formában is elérhetővé tett, akár akusztikailag is vizsgálható, méghozzá önmagában a számítógépen futó programmal s nem külső eszközzel (Vékás 2007: 289; továbbá a nyelvészeti adatok és az evidencia problémaköréről, továbbá a nyelvészeti adatok korpusznyelvészeti és generativista megközelítéséről l. Kertész–Rákos 2008). Napjainkban egyre inkább a számítógépen rögzített adatot kell dialektológiai adatnak tekintenünk, mivel az egyre modernebb nyelvtechnológiai módszerek az informatizált adatok felhasználási lehetőségének újabb fejezeteit nyitják meg, s ezek minőségi ugrást hoznak a korábbi, papír alapon megjelentetett adatok nyújtotta kutatási lehetőségekhez képest. Következésképpen a számítógépes nyelvföldrajz a nyelvi adatnak korábbiól eltérő felfogását hozta el. A nemzetközi szakirodalmat tekintve (Nerbonne–Heeringa 2001; Gooskens–Heeringa 2004) szinte alapkövetelményként mutatkozik az újonnan készülő nyelvjárási adatbázisok létrehozása kapcsán az adatok számítógépen való rögzítése. Ami pedig az adatok természetét illeti: (1) elengedhetlenül fontossá vált a fonetikai nézőpont, vagyis a hangzó adat elsődlegessége; (2) a hangfelvétel és lejegyzés összekapcsolását lehetővé tevő számítógé-

pes adatrögzítés, illetve (3) az adatok ellenőrzését, elemzését, feldolgozását biztosító lejegyzés, amely egyben lehetővé teszi a különböző adatbázisok közötti adatmozgást, az adatok integrálását (a hangzó adat elsődlegességére l. CorPho egyezmény, Ladefoged–Maddieson 1996, illetve a hazai kutatásokban l. Vékás 2007; Vargha 2013a; Bodó–Vargha 2008; Vargha 2017; a nemzetközi szakirodalom ide vágó, legfrissebb eredményeire többek között l. SweDia 2000, ANAE Labov–Ash–Boberg 2006; Kendall 2008; ScanDiaSyn; Lanchart, ALD, ehhez kapcsolódóan l. Bauer 2012: 95–120; a hangzó/digitális adatok tárolásáról, archiválásáról l. Kontra 2013: 253–69).

Nem sokkal a Los Angeles-i kezdeményezés után – a technikai fejlődés következtében – Magyarországon is erőre kaptak a modernizációs törekvések: a dialektológiai eredmények nyelvföldrajzi ábrázolása terén Balogh Lajos és Kiss G. Gábor 1992-ben elsőként számolnak be a nyelvjárási adatbázisok számítógépes feldolgozhatóságáról, majd megtervezték s megkezdték az MNyA. informatizálását (Balogh–Kiss G. 1992, a számítógépre vitt anyag elemzési lehetőségeiről l. Balogh–Kiss G. 2002). Megfelelő technológiai háttértámogatás hiányában az MNyA. adatainak számítógépes rögzítése elakadt ugyan, ám a nyelvjárási adattárak informatizálása 1996-ban nagy erővel beindult Vékás Domokos kezdeményezésére, aki meghatározta a számítógépes dialektológia mint átfogó nyelvészeti koncepció alapvetéseit, s ezeknek alapján hozzálátott a nyelvjárási adatok informatizálásához szükséges nyelvészeti technológiák tervezéséhez és fejlesztéséhez. A különböző egyetemeken, kutatóhelyeken e technológiákat használó kutatók informális hálózata Geolingvisztikai Műhely néven is ismert volt. A folyamatos fejlesztés alatt álló nyelvjárási lejegyzésszerkesztő, hangkezelő és térképgeneráló szoftver, a Bihalbocs révén valósulhat meg a különféle adattárakból származó nyelvjárási adatok informatizálása, integrálása, térképekre vetítése és analízisa (a program honlapját l. <http://www.bihalbocs.hu>, illetve Vékás 1999; Vékás 2000; Kiss 2001a, Juhász–Radványi–Vékás 2000; Vékás 2007; továbbá Vargha 2017). A lefektetett alapelvek illeszkednek a nemzetközi irányelvekhez: mind az informatizálási tevékenység, mind a hangzó adat dominanciájának elve kiemelt jelentőséggel bír a (szocio) dialektológiai kutatásokban. Minthogy korábban nem volt hagyománya a magyar nyelvjáráskutatásban a dialektológiai hangfelvételek eszközfontikai vizsgálatának, tudománytörténeti jelentőséggel bír az MNyA. ellenőrző gyűjtései (MNyAEGy. 1960–1964) során készített magnetofonfelvételek digitalizálása. A hangfelvételek alapján egyrészt a már korábbi lejegyzéseket kapcsolták össze az eredeti hanganyagokkal, másrészt elkezdték feldolgozni a hangfelvételen rögzített, de az atlaszban nem publikált kérdések válaszait. Ennek alapján született meg a *Magyar nyelvjárási hangoskönyv*

(MNYHK., Hajdú–Kázmér–Bodó–Vargha szerk. 2005–2009), amely mind a tudományos kutatás, mind az ismeretterjesztés szintjén új távlatokat nyit a hazai nyelvjáráskutatásban (Vargha 2007a, a hangfelvételekhez kapcsolódóan a suksükölés vizsgálatáról l. Vargha 2007b: 279–88, az MNYHK. korpusza alapján végzett eszközfonetikai vizsgálatokra l. Vargha 2013a; továbbá Fazakas (Gál) Noéminek Vöő István nyelvjárási gyűjtése alapján készített hangszinkronizációját: [http://www.bihalbocs.hu/hkonyv\\_kolozsvar/rev/index.html](http://www.bihalbocs.hu/hkonyv_kolozsvar/rev/index.html), valamint l. még a *Szlovákiai magyar nyelvjárási hangoskönyvet*: Presinszky 2017, 2020, 2021). A MNYAEGy. alapján kilátásba került az első „beszélő nyelvatlasz” megalkotása, hiszen a Bihalbocs program révén a hangzó anyag és a lejegyzés hatékonyan összekapcsolható (Vargha 2013c: 459–74). Továbbá az MNYA. ellenőrző gyűjtéseiből származó hangfelvételek alapján kísérelte meg Vargha (2007b) a nagyatlasz térképlapjai alapján kevésbé vizsgálható nyelvi jelenségek területi szempontú elemzését.

### 3. A nyelvi adatok informatizálásáról

Először a kétezres évek táján, majd a későbbiekben többször is rávilágít Kiss Jenő arra, hogy „igencsak időszerű problémakör a gyűjtés–közzététel–feldolgozás dialektológiaiabeli viszonya” (Kiss 2002: 13). Meglátása szerint a legutóbbi idők nyelvjáráskutatásra vonatkozó kritikái javarészt abból fakadnak, hogy „a nagy és a kicsi nyelvjárási korpuszok (atlaszok, szótárak, szövegfelvételek stb.) kínálta lehetőségek kiaknázása, a bennük rejlő nyelvtudományi problémák megoldása általában messze elmarad az optimálistól” (Kiss 2002: 13). E megállapítás érvényes a regionális atlaszok többségére is Bodó és Vargha (2008: 335–6) meglátása szerint, sőt vallják, hogy a feldolgozás lehetőségét alapjaiban határozza meg a közreadás mi-kéntje. Éppen ezért hamar kiütközött, hogy – a nemzetközi tendenciákhoz hasonlóan – a magyar számítógépes dialektológia legsürgetőbb feladatai közé tartozik (1) a már meglévő, nyomtatott formában kiadott atlaszok informatizálása, (2) a különféle forrásokból eredő, integrált adatokból létrehozott s egységesített számítógépes magyar nyelvatlasz megalkotása és (3) a kiadatlan nyelvjárási korpuszok modern, költséghatékony megjelenítése (többek között l. Vargha 2015, a dialektológia számítógépes háttértámogatására l. még Juhász 2014).

Az 1990-es évek elején Balogh Lajos és Kiss G. Gábor módszerével (1992) rögzítették az MNYA. első és második kötetét teljesen, a harmadikat részben a Nyelvtudományi Intézetben, majd munkásságuk nyomán a Geolingvisztikai Műhely kutatói 2014-re teljeskörűen informatizálták az MNYA. adatbázisát, a *Székely nyelvöldrajzi szótár* (SzNySz.) anyagát (az informatizálásra

1. Krémer 2002), *A moldvai csángó nyelvjárás atlasza* (MCsNyA.) nyomtatásban megjelent első és második, valamint a kéziratban fennmaradt harmadik kötetének adatait, továbbá *A romániai magyar nyelvjárások atlasza* (RMNyA.) 5–7., hagyományos térképek formájában kiadott, de eredetileg digitálisan előállított kötetének anyagát, továbbá Király Lajos *Somogy-zalai nyelvatlaszát* (S–ZA.). E munkálatok révén az eredetileg nyomtatásban megjelent adatbázisok sokkal jobban hasznosíthatók informatizált adatok formájában, továbbá könnyedén, tetszőleges irányelvek alapján kereshetők, rendszerezhetők és a kutatói kíváncsnak megfelelően újrendezhetők. Más szempontból is számottevő a nyelvi adatok informatizálása. Meghatározott szempontok szerint összesíthetők, vagyis integrálhatók a különféle adattárakból, nyelvföldrajzi szótárakból, térképlapokról származó adatok. Az eljárásnak köszönhetően számos módon hasznosíthatók az informatizált adatok: megjeleníthetők egyetlen térképen; létre lehet hozni belőlük csoportosított térképeket vagy több adatbázisból merítő, de egy-egy kutatópontra vonatkozó adatsorokat; illetve készíthetők olyan térképes kimutatások is, amelyek egy-egy fonéma előfordulási arányait mutatják (Bodó–Vargha 2008: 344; továbbá az adatintegrálás nyelvatlaszok közötti megvalósíthatóságára és az *Egységesített számítógépes magyar nyelvatlasz*, az ESZA. módszertanára l. Hegedűs 2008, illetve az ESZA. módszertani elveit követve az atlaszegyítés lexikográfiai-filológiai kérdéseire l. Szabó 2015, 2018). Az informatizálási tevékenység további előnye, hogy az eljárás révén az adatok közzététele költségghatékony formában valósulhat meg, s ez különösen jelentős előrelépés, ha a kéziratok térképlapok tartalmazta nyelvi adatok elérhetővé válásáról van szó. Hochbauer Mária és Gaál Noémi munkássága révén indult meg az eddig publikálatlan *Szamosháti atlasz* részleges, 22 kutatópontra kiterjedő anyagának feldolgozása (Hochbauer 2013: 164–71), továbbá végéhez ért *A székely nyelvjárás atlasza* háromszéki, kéziratban fennmaradt térképlapjainak feldolgozása (Cs. Nagy 2011). Így 2013-ra sor kerülhetett hat romániai nyelvjárás adatbázis (legalább részleges) informatizálására (Vargha 2013b), noha lehetővé válhatna történeti nyelvjárás térképek megalkotása is az Erdélyi magyar szótörténeti tár adatainak informatizálásával, illetve földrajzi elhelyezésével (Vargha–Vékás 2012).

Az intézményi keretek között zajló, illetőleg egyéni kutatói kezdeményezésen alapuló geolingvisztikai vállalkozásokon túl jelentős eredmények köthetők a Bihalbocs fejlesztőinek munkásságához. A Varga Fruzsina Sára által művelt dialektometriai vizsgálatok (összefoglalóan többek között Vargha 2017) révén nemzetközi mércével mérve is újszerű számítógépes nyelvjárás kutatások honosodtak meg hazánkban: ezek elsősorban a nyelvjáráshatárok, illetve a nyelvi kapcsolatok feltérképezését célozzák. De az újítások körét bő-

víti a Bodó Csanád neve fémjelezte, követéses vizsgálaton alapuló moldvai multimédiás „beszélő” atlasz projektje is (Bodó 2007a; Vargha 2021: 325–50), továbbá a készülőben lévő csallóközi nyelvatlasz projektje is, amelynek révén elsőként nyílik lehetőség a hasonló fonotaktikai helyzetben lévő magánhangzók minősítésének összehasonlítására (Menyhárt–Presinszky 2013; Presinszky 2016: 209–19). Módszertani szempontból ezeknek sorába kíván illeszkedni a KNyK.-projekt is (Ferenczi 2021: 11–46).

#### 4. A kvantitatív nyelvföldrajzi kutatások fajtáiról

A kvantitatív nyelvföldrajzi vizsgálatok lényege, hogy több száz térképlap sok százezernyi vagy milliónyi nyelvi adata vizsgálható előre meghatározott szempontok szerint objektíven, számszerűsítve (Vargha 2016: 151). Az efféle vizsgálatoknak több típusa van. Az egyik, amikor előre kiválasztott nyelvi változóknak, illetve a hozzájuk tartozó kódolt változatoknak számszerű kimutatására kerül sor informatizált nyelvatlaszok vagy területi szempontok alapján felgyűjtött, szöveges adatbázisok felhasználásával (néhány korábbi nyelvjárásméréssel kapcsolatos cikk bemutatását l. Kiss 2001b: 138–40). Többek között jó példája ennek a hazai nyelvföldrajzi vizsgálatok sorában az RMNyA. és a *Szilágysági nyelvatlasz* hangtani szempontú vizsgálata (Vékás 2000; Hegedűs 2007 vagy Bodó Csanádnak hangstatisztikai térképekkel ellátott, különféle fonémák előfordulási gyakoriságát vizsgáló kutatása az MCsNyA. adattára alapján, illetve a román kölcsönszók területiségét felvázoló kutatása [Bodó 2007b]) a Bihalbocs segítségével történt, ÚMNyA.-hoz kapcsolódó részfeldolgozások (többek között l. Iglai 2013; 2017; Kocsis 2013). Ebbe a sorba illeszkedik továbbá Vargha (2007b) kutatása is, amely az MNyA. szövegfelvételei alapján a suksükölés hatvanas évekbeli területi kiterjedtségének feltárását célozza; illetve – a számítógépes dialektológia, a kvantitatív nyelvföldrajz és a modern szociodialektológiai szempontok ötvözése révén – ide tartozik a Nyíregyházi Kutatóműhely által létrehozott, informatizált adatbázis (P. Lakatos szerk. 2012).

A kvantitatív nyelvföldrajz másik típusa, a dialektometria nem egy-egy nyelvi változót helyez a fókuszba, hanem egy nyelvatlasz valamennyi adatának kutatópontonkénti összemérését célozza meg. A *dialektometria* kifejezést az irányzat alapítója, Jean Séguy használta először az 1970-es években. Szomszédos kutatópontok megfelelő nyelvi adatai közötti különbségeket számszerűsített atlaszadatok alapján, és eszerint próbált meg – a korábbiaknál pontosabban – nyelvjáráshatárokat megállapítani (többek között Chambers–Trudgill 1998: 137–40, továbbá az eljárásról l. még Milroy–Gordon 2003:

223–6). Hans Goebel (2006) szintén az egyes nyelvjárások közötti azonosságok és különbségek mérésére fókuszál, azonban már nem csak a szomszédos kutatópontok adatait méri össze: egy atlasz minden kutatópontját minden más kutatópontjával egyenként egybeveti, és egy ún. hasonlósági mátrixot hoz létre. A Levenshtein-algoritmuson alapuló módszere révén minden egyes kutatópont valamennyi másik kutatóponthoz mért nyelvi hasonlóságának mértéke kifejezhető egy szám formájában, amelyekhez különböző színek társíthatók a térképezés során (erről és egyéb dialektometriai elképzelésekről l. Gooskens–Heeringa 2004, a hazai szakirodalomban többek között Vékás 2007; Bodó–Vargha 2008; Vargha–Vékás 2009; Vargha 2010, 2017).

A módszertani megújuláshoz kapcsolódóan érdemes felidézni Kiss Jónának (2007: 13) a magyar dialektológia kihívásaival és távlataival kapcsolatos meglátását: „A legtöbb olyan eredményt, amellyel a dialektológia leginkább tud nyelvészeti és tudományközi érdekelttségű és fontosságú eredményeket produkálni – új ismereteket adva a nyelvtudomány érvényes tételeihez –, a nyelvföldrajzi és szociodialektológiai kutatásoktól (a kettő részleges fedésével s dialektometriai vizsgálatok alkalmazásától) várhatjuk.” – A dialektometriai kutatások hazai eredményeire l. még Vargha–Vékás 2009; Vargha 2010, 2013a, 2014; Ditrói–Vargha 2016: 113–30; Kocsis–Vargha 2016: 193–207, továbbá átfogóan l. Vargha 2017. A nemzetközi kutatási eredmények kapcsán: a LAMSAS informatizált, adatbázisba rendezett anyagának dialektometriai vizsgálatára l. <http://www.lap.uga.edu/Site/LAMSAS.html>, továbbá Kretzschmar 2001; Nerbonne–Kretzschmar 2003.

## **5. A Kemenesalja és vidéke nyelvföldrajzi kutatási projekt helye a kvantitatív nyelvföldrajzi vizsgálatok keretrendszerében**

A KNYK.-projekt a napjainkban folyó, hazai dialektológiai, nyelvföldrajzi és szociolingvisztikai vizsgálatok sorát bővíti, és pedig úgy, hogy ötvözi részben a klasszikusnak mondható, részben az újabb kutatási módszereket. A kutatás dimenzióit a tér, idő és társadalom hármas koordináta-rendszere határozza meg (Juhász 2002: 149–53): ennek révén a nyelvreírás szempontok mellett a nyelvi változások mögöttes okaira adott magyarázatok száma is gyarapodik, továbbá jóval nagyobb lehetőség nyílik a nyelvi változatosság minél pontosabb leírására. A területi, az ebben megragadható társadalmi különbségek, továbbá mindezek időbeli változásának feltárása mellett kiemелendő, hogy a gyűjtött nyelvi adatok informatizálása teret ad nyelvészeti technológiák alkalmazásával végzett elemzésükre és térképes ábrázolásukra is. A kutatás egészét – elméleti-módszertani megalapozottságát,

az ábrázolástechnika és az adatelemzés módját – áthatja a dialektológia, nyelvöldrajz és szociolingvisztika összekapcsolásának szándéka, mélyen gyökerező hagyományának egyesítése: szociodialektológiai megközelítésű, nyelvöldrajzi módszerű (modulárisan) változásvizsgálatként helyezi el magát a szakirányú kutatások körében.

Alapvető különbség a nyelvöldrajzi és a szociolingvisztikai természetű vizsgálatok között, hogy az első a térbeliséget hagyományosan a különféle települések nyelvváltozatának (egy-egy viszonylag homogén közösség „nyelvjárás-átlagának”) összevetéseként értelmezik (vagyis a nyelvi adatokat elsősorban nem beszélőkhöz, hanem egy-egy település közösségéhez kötik); a második azonban a közösség nyelvhasználatának sokféleségét úgy szemléli, hogy a nyelv leírása során kisebb teret enged a területiséghez köthető vonatkozásoknak. Összegezve: természetükből fakadóan a geolingvisztikai vizsgálatok jóval nagyobb hangsúlyt helyeznek a területiség szempontjának vizsgálatára, mint a klasszikus szociolingvisztikai kutatások; ez utóbbiak pedig kisebb területre fókuszálnak, s – a különféle társadalmi csoportok nyelvhasználati különbségeire fókuszálva – vizsgálatuk középpontjában inkább egy-egy település, településrész vagy intézmény nyelvhasználati sokrétűségének feltárása áll. A kutatási projekt mindkét vizsgálati módszer sajátosságából merít: a területiség szempontjából nézve a kutatópont-hálózat a Kemenesalja és vidéke számos településére kiterjed (majdnem teljes sűrűségű kutatópont-hálózatot adva a térben), ugyanakkor valamennyi kutatóponton – jelen adatgyűjtési keretben a legidősebb generáció nyelvjárására koncentrálna – a szociolingvisztikai szempontokat figyelembe vevő módszerrel történt a mintavétel. Ez részint alkalmassá teszi a vizsgálatot, hogy modulárisan összemérhetőek legyenek az eredmények az MNyA. adataival, részint alapot teremt a kutatás – trend-, esetleg panelvizsgálat keretében való – folytatására. Mindezek ábrázolása nyelvöldrajzi módszerrel valósul meg: a klasszikus dialektológiai hagyományokból merítve újszerűsége az adatábrázolás technikájában és annak módszertani vetületében (informatizált adatbázis, illetve informatizált térképlapok) áll. A területi kötöttséget mutató adatok egy-egy adatközlőhöz rendelhetők, s így a nyelvjárás adatok térképen ábrázolhatók, s a különféle független változó szerint – szinkron szemlélettel egyéni és közösségi, diakron megközelítésben közösségi szinten – elemezhetők.

A kutatás dimenziói közül elsőként a területit szemlélve megállapítható, hogy a KNyK.-projekt – szinte teljes sűrűségű kutatópont-hálózata révén – lehetőséget terem egy, nyelvjárás szempontból szinte alig kutatott terület településeinek nyelvi, nyelvtörténeti „mélyfűrésére”, beszélőközösségeinek vizsgálatára. Így csökken a magyar nyelvterület azon részeinek, illetve (néprajzi szempontból) azon kistájak száma, amely – a nagyatlasz koncepciójá-



nak megalkotói szerint – regionális atlással még nincs lefedve (a táji atlasz szükségességéről, nagyatlaszhoz való viszonyáról többek között l. ÓHA.; Végh 1959: 21–2). A másodikat, a „humán” dimenziót a gazdasági, társadalmi, politikai tényezők nyelvi hozadékának feltérképezésén túl az adatközlők nyelvi mentalitásának feltárása és összevetése biztosítja. A vizsgálat harmadik síkja, az időbeliség alapján a kutatási projekt a nyelvi változásvizsgálatok közül többet is lehetővé tesz. „Valóságos” abban az értelemben, hogy az MNyA. koncepciójához illeszkedve, kérdőívének speciális szempontú redukciója által a kérdőíves kikérdezés a korábbi mintavételnek részleges megismétlése, s az adatközlők életkori rétegzettsége által lehetőség nyílik az egyazon időben a különböző életkorú beszélők nyelvhasználatának összevetésére. „Közösségi” abban a tekintetben, hogy viszonyítási alapja, az MNyA. a nyelvi adatokat – a gyűjtőfüzeteket nem tekintve – nem konkrét beszélőkhöz, hanem kutatópontokhoz köti a nyomtatásban megjelent, papíralapú és az informatizált térképlapokon. A korpusznak konkrét adatközlőkhöz köthetősége révén e projekt kiindulópontként szolgálhat akár egy jövőbeni trend-, esetleg panelvizsgálathoz. Egy ilyen előremutató kutatás terve természetszerűleg magában hordozza „a diakroniában való elmélyülés biztosítását”, illetve „a területiségben is áttetsző társadalmi és humán szempontok” kiemelt érvényre juttatását (Bokor 2007: 48).

A KNYK.-projekt terepmunkájának teljes nyelvi adatbázisát hangfelvételek sora rögzíti, így is illeszkedve a nyelvi adatnak napjainkban releváns fogalmához (l. 2. fejezet). A modern kori dialektológia adatrögzítéssel és -feldolgozással kapcsolatos kíváncsiságainak megfelelően a hangfelvételek tartalmazta adatbázis feldolgozása a nyelvi adatok informatizálásával valósult meg, az adatok térképre vitelének ábrázolástechnikai feltételeit pedig a Bihalbocs magyar nyelvjárási lejegyző, adatbázis-kezelő, térképező és elemző program biztosította. Az informatizált adatokon alapuló kutatás – a napjaink nyelvföldrajzi kutatásaival szemben támasztott elvárásoknak megfelelően – figyelmet fordít az „integráció és közzététel” alapelvére (Vargha 2017).

## Szakirodalom

- ALAP = Kretzschmar, William: *Linguistic databases of the American Linguistic Atlas Project*. <http://ldc.upenn.edu/annotation/database/papers/kretzschmar/11.3.kretzschmar.pdf> (Letöltve: 2023. 03. 23.)
- ALD = *Linguistic atlas of dolomitic ladinian and neighbouring dialects*. Honlap: <https://www.micura.it/en/activities/ald-linguistic-atlas> (Letöltve: 2023. 03. 23.)
- ANAE = Labov, William – Ash, Sharon – Boberg, Charles 2006. *The Atlas of North American English*. Mouton De Gruyter.

- Balogh Lajos – Kiss G. Gábor 1992. A magyar nyelvjárások atlaszának számítógépes feldolgozása. In: Kontra Miklós (szerk.): *Társadalmi és területi változatok a magyar nyelvben*. Linguistica. Series A. Studio et dissertationes, 9. A MTA Nyelvtudományi Intézete. Budapest. 5–17.
- Bauer, Roland 2012. Zur Dialektometrisierung des ALD (I und II): Ein Arbeits- und Erfahrungsbericht (2000–2012). In: Tosques, Fabio hrsg. *20 Jahre digitale Sprachgeographie Tagungsband*. Berlin 02. bis 03. November 2012. Humboldt-Universität. Berlin. 95–120.
- Bihalbocs = Magyar nyelvjárási lejegyző, adatbázis-kezelő, térképező és elemző program. Fejlesztő: Vékás Domokos (1996–) és Vargha Fruzsina Sára (2005–). <http://www.bihalbocs.hu/> (Letöltve: 2023. 03. 23.)
- Bodó Csanád – Vargha Fruzsina Sára 2008. Régi nyelvátlaszok, új módszerek. *Magyar Nyelv* 335–51.
- Bodó Csanád 2007a. Követéses geolingvisztikai vizsgálatok Moldvában. In: Guttmann Miklós – Molnár Zoltán (szerk.): *V. Dialektológiai szimpozion*. Szombathely, 2007. augusztus 22–24. A Berzsenyi Dániel Főiskola Magyar Nyelvészeti Tanszékének Kiadványai VIII. Szombathely. 37–47.
- Bodó Csanád 2007b. A moldvai magyar nyelvjárások román kölcsönszórétegének területisége. In: Benő Attila – Fazakas Emese – Szilágyi N. Sándor (szerk.): *Nyelvek és nyelvváltozatok*. Köszöntőkötet Péntek János tiszteletére. I–II. A Szabó T. Attila Nyelvi Intézet Kiadványai 4., Anyanyelvápolók Erdélyi Szövetsége. Kolozsvár. 160–74.
- Bokor József 2007. Töprengés egy új nagyatlasz lexikai anyagán és gyűjtésmódján. In: Guttmann Miklós – Molnár Zoltán (szerk.): *V. Dialektológiai Szimpozion*. Szombathely, 2007. augusztus 22–24. A Berzsenyi Dániel Főiskola Magyar Nyelvészeti Tanszékének Kiadványai VIII. Szombathely. 48–52.
- Chambers, J. K. – Trudgill, Peter 1998, 2004. *Dialectology*. Cambridge University Press. Cambridge.
- Cs. Nagy Lajos 2011. Úton a Háromszéki nyelvátlasz informatizálása felé. In: Bárány M. János – Vargha Fruzsina Sára (szerk.): *Hangok – helyek*. ELTE Magyar Nyelvtudományi és Finnugor Intézet. Budapest. 61–73.
- Ditrői Eszter – Vargha Fruzsina Sára 2016. Települések közti nyelvjárási és névföldrajzi hasonlósági viszonyok összevető elemzése Vas és Zala megyében. *Névtani Értesítő* 38: 113–130.
- DiWA = *Digitaler Wenker Atlas*, 2001–2009. <http://www.diwa.info/>
- Fazakas (Gál) Noémi: *Nyelvjárási szövegek Vöő István gyűjtéséből*. [http://www.bihalbocs.hu/hkonyv\\_kolozsvar/kemence/index.html](http://www.bihalbocs.hu/hkonyv_kolozsvar/kemence/index.html) (Letöltve: 2023. 03. 23.)
- Ferenczi Gábor 2021. A Kemenesalja és vidéke nyelvföldrajzi vizsgálata: Alapvetések. In: Pomozi Péter – Rási Szilvia – Tóth Zsolt (szerk.): *Magyar nyelvjárások napja 2020*. Örökség és jövőkép. Magyarországtudató Intézet. Budapest. 11–46.
- Goebel, Hanns 2006. Recent Advances in Salzburg Dialectometry. *Literary and Linguistic Computing* 21: 411–35. <https://doi.org/10.1093/lilc/fql042>

- Gooskens, Charlotte – Heeringa, Wilbert 2004. Perceptive evaluation of Levenshtein dialect distance measurements using Norwegian dialect data. *Language Variation and Change* 16: 189–207. <https://doi.org/10.1017/S0954394504163023>
- Hegedüs Andrea 2007. A kétféle é fonéma megléte és kiterjedtségének mértéke a Szilágyságban. In: Guttmann Miklós – Molnár Zoltán (szerk.): *V. Dialektológiai szimpozion*. Szombathely, 2007. augusztus 22–24. A Berzsenyi Dániel Főiskola Magyar Nyelvészeti Tanszékének Kiadványai VIII. Szombathely. 120–5.
- Hegedüs Andrea 2008. *Nyelvföldrajzi vizsgálatok a romániai magyar nyelvterületen az egyesített atlaszok felhasználásával*. ELTE BTK, Budapest. PhD-értékezés.
- Hochbauer Mária 2013. A Szamosháti nyelvatlasz informatizálásának munkálatai. In: Szoták Szilvia – Vargha Fruzsina Sára (szerk.): *Változó nyelv, nyelvváltozatok, területiség*. A VII. Hungarológiai Kongresszus nyelvészeti tanulmányai. Bolyai Társaság. Kolozsvár. 164–71.
- Iglai Edit 2013. A nyelvi mentalitás és a határmentiség szerepe a területi nyelvváltozatok változásában. In: Kontra Miklós – Németh Miklós – Sinkovics Balázs (szerk.): *Elmélet és empiria a szociolingvisztikában: válogatás a 17. Élőnyelvi Konferencia előadásaiból*. Gondolat Kiadó. Budapest. 417–32.
- Iglai Edit 2017. *Nyelvjárásközi egybevető vizsgálat A magyar nyelvjárások atlasza és az Új magyar nyelvjárások atlasza alapján*. Doktori disszertáció. ELTE BTK, Nyelvtudományi Doktori Iskola.
- Juhász Dezső 2002. A dialektológia hármass feladata és a nyelvföldrajz. In: Szabó Géza – Molnár Zoltán – Guttmann Miklós (szerk.): *IV. Dialektológiai szimpozion*. Szombathely. 149–53.
- Juhász Dezső 2014. A dialektológia számítógépes támogatása. Rövid hazai körkép. *Magyar Nyelvjárások* 52: 57–64.
- Juhász Dezső – Radványi Péter – Vékás Domokos 2000. Egy dialektológiai adatbázis és elektronikus prezentációs rendszer alapvetése. *Magyar Nyelv* 47–58.
- Kendall, Tyler 2008. Ont he History and Future of Sociolinguistic Data. *Language and Linguistic Compass* 2: 332–51. <https://doi.org/10.1111/j.1749-818X.2008.00051.x>
- Kertész András – Rákosi Csilla 2008. Megjegyzések a nyelvészeti adatok és evidencia problémájáról folyó vita jelenlegi állásához. *Magyar Nyelv* 274–86.
- Kiss Jenő 2001a. A számítógépes dialektológia. In: Kiss Jenő (szerk.): *Magyar Dialektológia*. Osiris Kiadó. Budapest. 142–4.
- Kiss Jenő 2001b. A kvantitatív dialektológia. In: Kiss Jenő (szerk.): *Magyar Dialektológia*. Osiris Kiadó. Budapest. 135–40.
- Kiss Jenő 2002. Dialektológia és nyelvtudomány: hagyomány és korszerűség. *Magyar Nyelvjárások* 40: 3–20.
- Kiss Jenő 2007. Kihívások és lehetőségek előtt a magyar dialektológia. In: Guttmann Miklós – Molnár Zoltán (szerk.): *V. Dialektológiai szimpozion*. Szombathely, 2007. augusztus 22–24. A Berzsenyi Dániel Főiskola Magyar Nyelvészeti Tanszékének Kiadványai VIII. Szombathely. 8–16. <https://doi.org/10.1111/j.1749-818X.2008.00051.x>

- Kocsis Zsuzsanna – Vargha Fruzsina Sára 2016. A magyar nyelvjárások atlasza és A romániai magyar nyelvjárások atlasza integrált dialektometriai elemzése. In: Czetter Ibolya – Hajba Renáta – Tóth Péter (szerk.): *VI. Dialektológiai szimpozion*. Szombathely, 2015. szeptember 2–4. NYME Magyar Nyelvtudományi Intézeti Tanszék – UKF Közép-európai Tanulmányok Kara – Szlovákiai Magyar Akadémiai Tanács. Szombathely – Nyitra. 193–207.
- Kocsis Zsuzsanna 2013. Nyelvjárási *ë* hangok vizsgálata, különös tekintettel a labializáció mértékére. In: Kontra Miklós – Németh Miklós – Sinkovics Balázs (szerk.): *Elmélet és empiria a szociolingvisztikában*. Válogatás a 17. Élőnyelvi Konferencia előadásaiból. Gondolat Kiadó. Budapest. 433–43.
- Kontra Miklós 2013. „A nyelvészet eltűnhet” (Walt Wolfram) – és a nyelvi adatok? In: Kontra Miklós – Németh Miklós – Sinkovics Balázs (szerk.): *Elmélet és empiria a szociolingvisztikában*. Válogatás a 17. Élőnyelvi Konferencia előadásaiából. Gondolat Kiadó. Budapest. 253–69.
- Krémer Zsófia 2002. *A Székely Nyelvöldrajzi Szótár informatizálása*. Szakdolgozat. ELTE BTK. Budapest.
- Kretzschmar, William 2001. *Linguistic databases of the American Linguistic Atlas Project (ALAP)*. <http://ldc.upenn.edu/annotation/database/papers/kretzschmar/11.3.kretzschmar.pdf> (Letöltve: 2023. 03. 23.)
- Kretzschmar, William A. Jr. – McDavid, Virginia G. – Lerud, Theodore K. – Johnson, Ellen 1993. *Handbook of the Linguistic Atlas of the Middle and South Atlantic States*. The University of Chicago Press. Chicago.
- Ladefoged, Peter – Maddieson, Ian 1996. *The Sounds of the World's Languages*. Blackwell Publishers. Oxford.
- LAMSAS = *Linguistic Atlas of Middle and South Atlantic States*. <http://www.lap.uga.edu/Site/LAMSAS.html> vagy <http://us.english.uga.edu/lamsas/> Letöltve: 2023.03. 23.
- LANCHART = Lanchart project on language change in 20th century Danish. Honlap: <http://lanchart.hum.ku.dk/> (Letöltve: 2023. 03. 23.)
- MCsNyA. = Gálffy Mózes – Márton Gyula – Szabó T. Attila (szerk.): 1991. *A moldvai csángó nyelvjárás atlasza. I–II*. A Magyar Nyelvtudományi Társaság Kiadványai 193. Budapest.
- Menyhárt József – Presinszky Károly 2013. A szórványos *ö*-zés vizsgálata a Csallóközben generációk és iskolázottság szerint. In: Kontra Miklós – Németh Miklós – Sinkovics Balázs (szerk.): *Elmélet és empiria a szociolingvisztikában*. Válogatás a 17. Élőnyelvi Konferencia előadásaiból. Gondolat Kiadó. Budapest. 444–59.
- Milroy, Lesley – Gordon, Matthew 2003. *Sociolinguistics*. Method and Interpretation. Blackwell Publishing. Oxford. <https://doi.org/10.1002/9780470758359>
- MNyA. = Deme László – Imre Samu (szerk.) 1968–1977. *A magyar nyelvjárások atlasza I–VI*. Akadémiai Kiadó. Budapest.
- MNyHK. = *Magyar nyelvjárási hangoskönyv*. Hajdú Mihály – Kázmér Miklós, Bodó Csanád – Vargha Fruzsina Sára (szerk.) 2005–2009. *Magyar nyelvjárási hangoskönyv I–IX*. ELTE Magyar Nyelvtudományi és Finnugor Intézet. Budapest.

- Nerbonne, John – Heeringa, Wilbert 2001. Computational Comparison and Classification of Dialects. *Dialectologia et Geolinguistica* 9: 69–83.  
<https://doi.org/10.1515/dig.2001.2001.9.69>
- Nerbonne, John–Kretzschmar, William 2003. Introducing Computational Techniques in Dialectometry. In: Nerbonne – Kretzschmar eds. *Computational Methods in Dialectometry*. Special issue of Computers and the Humanities. 37/3: 339–57.  
<https://doi.org/10.1023/A:1025064105053>
- ŐHA. = Végh József 1959. *Őrségi és hetési nyelvatlasz*. Akadémiai Kiadó. Budapest.
- P. Lakatos Ilona (szerk.) 2012. *Változó nyelvhasználat a hármashatár mentén*. Többdimenziós nyelvföldrajzi térképlapok tanúságai. Tinta Könyvkiadó. Budapest.
- Presinszky Károly 2016. A csallóközi határ színei. Az á előtti asszociatív illabiális á hang akusztikai elemzése a Csallóközben. In: Czetter Ibolya – Hajba Renáta – Tóth Péter (szerk.): *VI. Dialektológiai szimpozion*. Szombathely, 2015. szeptember 2–4. NYME Magyar Nyelvtudományi Intézeti Tanszék – UKF Közép-európai Tanulmányok Kara – Szlovákiai Magyar Akadémiai Tanács. Szombathely – Nyitra. 209–19.
- Presinszky Károly 2017. *Szlovákiai Magyar Nyelvjárási Hangoskönyv I*. Nyitrai Konstantin Filozófus Egyetem, Nyitra. <http://www.umjl.fss.ukf.sk/szmnyhk/>
- Presinszky Károly 2020. *Szlovákiai Magyar Nyelvjárási Hangoskönyv II*. Nyitrai Konstantin Filozófus Egyetem, Nyitra. [http://www.umjl.fss.ukf.sk/szmnyhk\\_2/](http://www.umjl.fss.ukf.sk/szmnyhk_2/)
- Presinszky Károly 2021. *Szlovákiai Magyar Nyelvjárási Hangoskönyv III*. Nyitrai Konstantin Filozófus Egyetem, Nyitra. [http://www.umjl.fss.ukf.sk/szmnyhk\\_3\\_HTML/](http://www.umjl.fss.ukf.sk/szmnyhk_3_HTML/)
- RMNyA. = Murádin László gyűjt. – Juhász Dezső (szerk.) 1995–2006. *A romániai magyar nyelvjárások atlasza I–X. kötet*. Magyar Nyelvtudományi Társaság. Budapest.
- ScanDiaSyn = *ScandianavianDialectSyntax*. <http://www.tekstlab.uio.no/nota/scandiasyn/index.html> (Letöltve: 2023. 03. 23.)
- SweDia = *The phonetics and phonology of the Swedish dialects around the year 2000*. <http://swedia.ling.gu.se/> (Letöltve: 2023. 03. 23.)
- S–ZA. = Király Lajos 2005. *Somogy–zalai nyelvatlasz*. MNyTK. 223. Budapest.
- Szabó Panna 2015. Pillanatkép a nyelvatlaszok egyesítési háttérmunkálatainak állásáról. In: Bárh M. János – Bodó Csanád – Kocsis Zsuzsanna (szerk.): *A nyelv dimenziói*. Tanulmányok Juhász Dezső tiszteletére. ELTE BTK, Budapest. 221–8.
- Szabó Panna 2018. *Nyelvföldrajzi vizsgálatok a magyar nyelvterület nyugati felén a dialektológiai atlaszok alapján*. PhD-értekezés. ELTE BTK, Nyelvtudományi Doktori Iskola.
- SzNySz. = Gálffy Mózes – Márton Gyula 1987. *Székelnyelvföldrajzi szótár*. Akadémiai Kiadó. Budapest.
- ÚMNyA. = *Új magyar nyelvjárási atlasz*. <http://www.umnya.hu> (Letöltve: 2023. 03. 23.)

- Vargha Fruzsina – Vékás Domokos 2009. *Magyar nyelvjárási adattárak vizsgálata interaktív dialektometriai térképekkel*. Elhangzott a Magyar Nyelvtudományi Társaság felolvasóülésén 2009. március 24-én.
- Vargha Fruzsina – Vékás Domokos 2012. Lokalizálható nyelvtörténeti adatok informatizálása és térképezése. *Erdélyi Múzeum* 74: 160–5.
- Vargha Fruzsina Sára 2007a. A magyar nyelvjárási hangoskönyv és elemzési lehetőségei. In: Zelliger Erzsébet szerk. *Nyelv, területiség, társadalom*. A 14. élőnyelvi konferencia (Bük, 2006. október 9–11.) előadásai. MNyTK. 228. Budapest. 429–43.
- Vargha Fruzsina Sára 2007b. Nyelvi változók A magyar nyelvjárások atlasza hangfelvételeiben. In: Guttmann Miklós – Molnár Zoltán (szerk.): *V. Dialektológiai Szimpozion*. Szombathely, 2007. augusztus 22–24. A Berzsényi Dániel Főiskola Magyar Nyelvészeti Tanszékének Kiadványai VIII. Szombathely. 279–88.
- Vargha Fruzsina Sára 2010. A dialektometria alkalmazása és történeti helynevek nyelvföldrajzi vizsgálata a Székelyföldön In: Hoffmann István – Tóth Valéria (szerk.): *Helynévtörténeti Tanulmányok* 5. Debreceni Egyetemi Kiadó. Debrecen. 223–33.
- Vargha Fruzsina Sára 2013a. A hangzó adat szerepe a magyar dialektológiában. In: Szoták Szilvia – Vargha Fruzsina Sára (szerk.): *Változó nyelv, nyelvváltozatok, területiség*. A VII. Hungarológiai Kongresszus nyelvészeti tanulmányai. Egyetemi Műhely Kiadó. Kolozsvár. 194–204.
- Vargha Fruzsina Sára 2013b. Nyelvatlaszok a szekrényben. In: Benő Attila – Fazakas Emese – Kádár Edit (szerk.): „...*hogyan legyen a víznek lefolyása...*” Köszöntő kötet Szilágyi N. Sándor tiszteletére. Erdélyi Múzeum-Egyesület. Kolozsvár. 523–9. [http://frufu.web.elte.hu/publikaciok/VarghaFS\\_SzilagyNS\\_kotet.pdf](http://frufu.web.elte.hu/publikaciok/VarghaFS_SzilagyNS_kotet.pdf) (Letöltve: 2023. 03. 23.)
- Vargha Fruzsina Sára 2013c. Az adatok kor és nem szerinti megoszlása A magyar nyelvjárások atlasza ellenőrző gyűjtésében. In: Kontra Miklós – Németh Miklós – Sinkovics Balázs (szerk.): *Elmélet és empiria a szociolingvisztikában*. Válogatás a 17. Élőnyelvi Konferencia (Szeged, 2012. augusztus 30. – szeptember 1.) előadásaiból. Gondolat Kiadó. Budapest. 459–74.
- Vargha Fruzsina Sára 2014. *A magyar nyelvjárások távolsága a köznyelvtől*. Elhangzott a 18. Élőnyelvi Konferencián 2014. szeptember 19-én Nyitrán.
- Vargha Fruzsina Sára 2015. Atlaszintegrálás és kvantitatív adatelemzés. In: Báth M. János – Bodó Csanád – Kocsis Zsuzsanna (szerk.): *A nyelv dimenziói*. Tanulmányok Juhász Dezső tiszteletére. ELTE BTK, Budapest. 242–9.
- Vargha Fruzsina Sára 2016. A romániai magyar nyelvjárások atlasza informatizált térképlapjainak kvantitatív nyelvföldrajzi vizsgálata. *Magyar Nyelv* 152–63. <https://doi.org/10.18349/MagyarNyelv.2016.2.151>
- Vargha Fruzsina Sára 2017. *A nyelvi hasonlóság földrajzi mintázatai*. Magyar nyelvjárások dialektometriai elemzése. Magyar Nyelvtudományi Társaság. Budapest.
- Vargha Fruzsina Sára 2021. Különböző nyelvjárások magánhangzó-rendszereinek találkozása: a meg igekötő esete Moldvában. *Nyelvtudományi Közlemények* 117: 325–50.

- Vékás Domokos 1999. *Informatikai lehetőségek a dialektológiában, különös tekintettel a fonetikai szempontokra*. Budapest. <http://www.btk.elte.hu/fonetika/p/s/fonszem.html> (Letöltve: 2023. 03. 23.)
- Vékás Domokos 2000. Magánhangzó-rendszerek elemzése informatizált nyelvjárási korpuszon. In: Gósy Mária (szerk.): *Beszéd kutatás 2000. Beszéd és társadalom*. MTA Nyelvtudományi Intézete. Budapest. 75–86.
- Vékás Domokos 2007. Számítógépes dialektológia. In: Guttmann Miklós – Molnár Zoltán (szerk.): *V. Dialektológiai Szimpózium*. Szombathely, 2007. augusztus 22–24. A Berzsényi Dániel Főiskola Magyar Nyelvészeti Tanszékének Kiadványai VIII. Szombathely. 289–93.
- Viereck, Wolfgang 1997. The Computer Developed Linguistic Atlas of England, Volumes 1 (1991) and 2 (1997): Dialectological, computational and interpretative aspects. *ICAME Journal* 21: 79–90.

Ferenczi Gábor

PhD-hallgató

Eötvös Loránd Tudományegyetem

E-mail: [g.ferenczi88@gmail.com](mailto:g.ferenczi88@gmail.com)

<https://orcid.org/0000-0002-2522-9210>

## Abstract

FERENCZI, GÁBOR

### LINGUISTIC DATA AND DATA INFORMATISATION: CONTRIBUTIONS TO THE THEORETICAL FRAMEWORK OF A LINGUISTIC GEOGRAPHY RESEARCH PROJECT

Followed by an outline of the modern interpretation of linguistic data, data informatisation and the types of quantitative linguistic-geographical studies, the paper describes the place and role of Kemenesalja and the surrounding region's linguistic-geographical research (Project KNYK) in the series of quantitative linguistic-geographical studies. The focus of the KNYK project is on the socio-dialectological analysis of living linguistic data showing spatial boundaries and their mapping using modern computer methods.

**Keywords:** language/linguistic data, data informatisation, quantitative linguistic geography, dialectometry, Project KNYK