

Az első tájszóközlésektől az egyesített digitális tájszótárig

1. 2008-ban lesz 300 éve annak, hogy nyomtatásban először láttak napvilágot magyar szavak régióhoz kötötten – mai értelmezés szerint tehát tájszavakként minősítve. Erre emlékezve a népryelvi szavak gyűjtéséről és publikálásáról a legfontosabb tudománytörténeti tényekként a következőket mondhatjuk.

Magyarországon a 18. század vége előtt az említett módon csak alkalmilag jegyeztek föl és tettek közzé szavakat (tájszavakat). Ez történt 300 éve is, amikor Pápai Páriz kétnyelvű szótárában húsz-egynéhány székely szót közölt 1708-ban – az „antique”(‘rég, régies’) szóval jelölve őket (magyarazatul l. Szűcs 1936: 11; Melich: 1907: 116–21). Az „első egybefoglalt tájszógyűjtemény” 1747-ből, Torkos József tollából származik (Szűcs i. h. 17.). Ahogy haladt előre a nyelvi standardizálódás folyamata és került előtérbe a magyar irodalmi nyelv szókincsbővítésének az igénye, a 18. század végétől megkezdődött a népryelvi szavak tudatos és rendszeres gyűjtése, s megszorodtak a kisebb-nagyobb tájszórajstromok s tájszavak szótárbeli közlései. Márton József volt az első rendszeres népryelvgyűjtő, aki egy évet (1799) ezzel a tevékenységgel töltött, vállalkozván a „Magyarországi külömbkülömbféle Beszédmódok és Tájjezsók” összegyűjtésére (Szűcs i. m. 17.). A Magyar Tudós Társaság 1831. évi első nagygyűlésén egyik feladatként a tájszavak gyűjtését jelölte meg (ma így mondjuk: állapot-, illetőleg korpuszfejlesztési szándékkal), s 1838-ban már meg is jelent az első magyar tájszótár (2008-ban lesz tehát az első magyar tájszótár közzétételének 170. évfordulója). A „Tájszótár”-ral indult a különböző rendű és rangú magyar népryelvi vagy tájszótárak sora. Döbrentei Gábor, az Akadémia titoknokja a bevezetőben (1835-ös dátummal) teljes joggal írta: „e’ gyűjteményből ’s folytatásából tetszendik majd ki, mellyik valósággal csak tájszó, mert még most némelly annak vétetik ez vagy amaz’ vidéken, holott másutt is és több helyen megvan, vagy épen közös. Így jönnek köztudomásra a’ szőejtés’ külömféleségei is” (VIII). A Nyelvőr megindulása (1872) új szakaszt hozott: a népryelvi szógyűjtések szorgalmazását és tájszavak rendszeres közlését. Az első modern magyar tájszótár, a Szinnyei-féle Magyar tájszótár kora színvonalán álló alpmű lett. A 20. század volt a nyelvjárási szótárak mozgalmas korszaka mind minőségi, mind mennyiségi tekintetben. Befejezés előtt áll az általános tájszótár-kategóriának 3. tagja, az Új magyar tájszótár. Új szótártípusok is megjelentek (ezeket l. Hegedűs 2001: 399–400), s jelentkeztek a további professzionalizálódás irányába mutató lexikográfiai megoldások és igények.

A táji szókészlet (különböző célú) gyűjtése és vizsgálata a nyelvjáráskutatásnak kezdettől fogva fontos feladatköre volt, s a lexikológia és a lexikográfia fejlődésének

désével párhuzamosan vált a dialektológia egyik fontos területévé. A nemzetközi dialektológiában folyamatosan napirenden van ma is különböző típusú tájszótárak készítése és kiadása (nincs olyan dialektológiai kongresszus, amelyen ne szerepelne témaként a regionális szókészlet és a nyelvjárási lexikográfia).

2. Elérkezett az idő, hogy a modern adatrögzítést, -tárolást, -továbbítást és a nyelvészeti feldolgozást nagyban könnyítő és segítő számítógépes technika kínálta lehetőségekkel élve hozzáfogjunk az egyesített digitális magyar tájszótárnak a megtervezéséhez, előkészítéséhez (vö. Kiss J. 2002: 412). Egyrészt ugyanis azt a hatalmas regionális lexikális anyagot, amelyek különböző forrásokban állnak rendelkezésre, a hagyományos papírmódszerekkel már nem lehet célravezetően tárolni, a feldolgozásról nem is beszélve. Másrészt pedig a váltást a magyar dialektológia lexikográfiai tapasztalatai és a nyelvtechnológiai fejlesztések lehetővé teszik már (vö. Balogh Lajos: „A probléma megoldásának kulcsát minden bizonnyal a modern számítástechnika jelenti” [Kiss J. szerk. 2001: 160]). Tudvalévő, hogy az utóbbi években sokat fejlődött a számítógépes lexikográfia, a hazai is („a magyar számítógépes lexikográfiában minden elem adott egy igen magas színvonalú integrált rendszer létrehozásához” [Pajzs 2007: 14]).

Szótárakról szólva korpusznyelvészeti adatbázisokról szólunk, s fölvethető a kérdés: vajon szükség van rájuk, szükség van a digitális tájszótárra is? Nyelvhasználat alapú szemléletmóddal a válasz egyértelmű igen. A tényleges nyelvhasználatot tükröző, igazoló adatok ugyanis a nyelvhasználat kutatásának legmegbízhatóbb forrásai. Gombocz mondta 1925-ben: „Ki tagadja, hogy kell anyaggyűjtés is? De ez csak eszköz” (idézve: MTud. 1996: 330). A tájszó természetesen a tájszótárak készítői számára sem cél, hanem eszköz. A cél az, hogy a tájszavakat nyelvtudományi (nyelvföldrajzi, hangtani, alaktani, nyelvtörténeti, etimológiai, lexikológiai, jelentéstani, stilisztikai) kérdések tisztázásához használhassuk, illetőleg a belőlük kihámozható művelődés- és helytörténeti, valamint néprajzi tanulságok kiaknázását lehetővé tegyük és elősegítsük. Ugyanis a szókészlet nemcsak bizonyos nyelvrendszertani szabályoknak van alávetve, hanem részben kultúrafüggő is (emlékeztessünk a nyelvhasználat kettős beágyazottságának tényére). Az említett célokat szolgálja a tájszavak meghatározott szempontok szerinti minősítése és a kiegészítő, nem csak nyelvészeti vonatkozások leírása. A mondottak azt is jelentik, hogy a szótár nem „csupán” adathalmaz. Valamilyen koncepciót feltételez ugyanis, hogy mit, miért, hogyan foglalunk szótárba. A leletmentés is a célok között van természetesen: a tájszavak nagy része ugyanis előbb-utóbb kikopik a használatból, s feledésbe merül, márpedig a kutatást a nyomtalanul eltűnt tájszó nem gazdagíthatja. A számítógépes környezet pedig olyan adatrögzítést, megjelenítést, hozzáférést, illetőleg elemzési, feldolgozási lehetőségeket biztosít, amelyekről korábban legföljebb álmodozni lehetett.

A számítógépes technikát a dialektológiában először s mindmáig legeredményesebben a nyelvföldrajzi kutatásokban, a nyelvátlasz-készítésben alkalmazták. Nyitott kapukat nem kell döngetni: szakmai beszélgetésekben már korábban fölmerült nálunk is tájszótárak számítógépes rögzítésének nem csupán a szándéka, hiszen például az ÚMTsz. III. és IV. kötetét így is rögzítették, az V. eleve így készül, hozzáférhető a Büki tájszótár is elektronikusan, a Debreceni cívis szótár

pedig olvasható a világhálón is (<http://mnytud.arts.unideb.hu/sorozat/dcsz>). Balogh Lajos és Kiss Gábor (akiknek a nevéhez fűződik A magyar nyelvjárások atlasza számítógépes feldolgozásának a megkezdése) évekkel ezelőtti szakmai beszélgetéseiken szót ejtettek elektronikus tájszói adatbázis kiépítésének gondolatáról. Támogatja a tervet Vékás Domokos dialektológiai szoftverfejlesztőként is. Szabó László szegedi orvost mint nem nyelvészt azért említem név szerint is, mert önszántából, elismerésre méltó ügybuzgalommal törekszik arra, hogy a magyar tájszótárak elektronikus megjelenítését elősegítse (távlati célként maga is a teljes magyar digitális tájszótár létrehozását tartva megvalósítandó elképzelésnek).

Az egyesített digitális tájszótárt létrehozni csak bizonyos feltételek (személyi, infrastrukturális, anyagi) megteremtésével lehet. Az első lépés azonban a szakmai célok megfogalmazása, tisztázása kell, hogy legyen: mit és miért akarunk, s hogyan kívánjuk elképzeléseinket megvalósítani. Körültekintő tervezés, átgondolt koncepció nélkül nem lehet jó szótárt csinálni. Alábbi észrevételeimmel a nyelvészeti munka előkészítéséhez kívánok hozzájárulni. (Balogh Lajosnak itt is köszönöm a tanulmányhoz fűzött észrevételeit, kritikai megjegyzéseit.)

3. A tervezett szótár tájszótár. Mint ilyen szótártipológiai szempontból egyrészt nyelvváltozat-specifikus (csak területi kötöttségű nyelvváltozatok szókészleti elemeiről, illetve ezeknek a használatáról tájékoztat). Másrészt deskriptív-értelmező (a rendelkezésre álló nyelvhasználati tények alapján, azoknak megfelelően egy bizonyos időszakaszra érvényesként írja le a lexémákat). Harmadrészt általános tájszótár, mert elvileg a nyelvterület egészéről, illetve bármelyik pontjáról, a gyakorlatban nagy részéről tartalmaz majd szavakat. A tág értelmezés szerint nincsenek időbeli korlátai sem a tájszavak felvételének: az elérhető legrégebbi tájszavaktól a ma frissen gyűjtöttekig minden tájszó lehetséges szótári címszó. (Korlátokat az idő és az a tény szab, hogy csak az írott köznyelv alakulásával és terjedésével kezd egyértelműen elkülöníthetővé válni a területi kötöttségű szókészlet. Minthogy azonban a tervezett szótár fő feladata az elmúlt körülbelül százötven-kétszáz év táji lexémáinak az egybefoglalása lenne, ez a kérdés egyelőre zárójelbe tehető.) A szótár célja értelemszerűen a magyar regionális szókészlet lehetőség szerint minél teljesebb s minősített dokumentálása a maga területi és részben társadalmi változóságában, adatokkal szemléltetve, a szükséges, illetőleg lehetséges esetekben kiegészítő információkkal ellátva. A magyar nyelvtudományban három általános tájszótár készült eddig (Tájszótár 1838, Magyar tájszótár 1893–1901, Új magyar tájszótár (1979-től, utolsó, V. kötete megjelenésre vár). Mindegyik korábban s többektől gyűjtött kéziratos és főként nyomtatott forrásokból vette adatait. A tervezett szótár lesz a sorozat negyedik, elődeitől azonban több tekintetben eltérő tagja. A különbség elsősorban az adatok rögzítésében, tárolásában, megjelenítésében, folyamatos bővítési lehetőségében, hozzáférhetővé tételének a módjában lesz, nem a lexicográfiai leírásban.

4. A bemeneti oldal néhány kérdéséről.

1. Mi kerüljön be? Olyan szavak, amelyek területi kötöttségük miatt nem közmagyar szavak. Melyek tartoznak ide? A tájszavak (tájnyelvi szavak, népnyelvi szavak: ide elsősorban a hagyományos paraszti gazdálkodással és életmóddal összefüggő szókészleti elemek tartoznak) és a regionális köznyelvi szavak mint

világos esetek. Idetartoznak továbbá két speciális kategória bizonyos szavai is. Egyrészt a területi kötöttségű mesterségyszavak. Elsősorban a népi vagy kismesterségeknek azok a szakszavai, amelyeket nem szakiskolákban, normatív tankönyvekből sajátítottak el a házi és a kisipar művelői, hanem regionális nyelvhasználatú mesterektől. Ezeknek a szavaknak egy része népnyelvvé vált azzal, hogy a helyi nem iparos felnőttek számottevő hányada is ismerte/ismeri és használta/használja őket. (Vö. ehhez első városi népnyelvi szótárunkat: Kálnási Árpád, Debreceni cívis szótár. Debrecen, 2005.) Számuk a háziipar és kisipar visszaszorulásával, illetőleg az intézményes szakoktatás általánossá válásával radikálisan visszaszorulóban van. (De vö. Fóris 2004: 43: „A mesterségek szakszókincse az olasz nyelvben még ma sem nevezhető egységesnek, regionális jellemzőket mutat, vagyis az ország különböző részein különböző megnevezései vannak azonos dolgoknak”) Sokkal több szót érint a másik szókatégória: a kisebbségi közösségek anyanyelvi nyelvhasználatában meghonosodott államnyelvi (vagy más, érintkező nyelvi) kölcsönszók és az államnyelvi mintákra született tükörszók. Míg az előbbi esetben a szakmai, tehát csoportnyelvi kötöttség, az utóbbiban az eltérő államiság ténye a sajátosságot meghatározó tényező – a területiség mellett, arra mintegy ráépülve. Lássuk az utóbbit! A kisebbségi nyelvhasználatban meghonosodott államnyelvi szavak bizonyos hányada területileg kötötten él: ezek egyértelműen tájszavak mind a magyar nyelv egésze, mind a kisebbségi magyar nyelvhasználat felől nézve – amiként azok a csak egy kisebbség nyelvhasználatában bárhogy keletkezett, s ott területi kötöttségű neologizmusok (szó- és jelentésátvételek, tükörfordítások, belső keletkezésűek). A honossá vált államnyelvi szavak egy másik hányada általánosan ismert és használt az adott kisebbségi társadalom felnőtt tagjainak a körében. Ezek a magyar szókészlet egésze felől nézve regionális köznyelvi szavak, mert köznyelvi szerepkört töltenek be regionális hatókörben (előfordulnak a nyilvános beszélt és írott nyelvhasználat különböző szinterein is). A definíció értelmében (hogy tudniillik területi kötöttségük miatt nem közmagyar elemek), a magyar nyelv egészét tekintve ezek a szavak a regionális szókészlet tagjai, sajátos alcsoportokat alkotva. (Említsük meg: a dialektológia korábban is feladatának tekintette a népi kismesterségek szókincsének vizsgálatát, l. már a Csúry-iskola gyakorlatát. A kisebbségi nyelvhasználat nem közmagyar jövevényszavait feldolgozó szótárakat pedig a „speciális regionális tájszótár” kategóriájának tagjaiként tartja számon, l. Kiss J. 2001: 399–400; az eddigiek közül kettő romániai magyarok román, egy az amerikai [Calumet-vidéki] magyarok amerikai angol jövevényszavait tartalmazza. Regionális köznyelvi szótárunk még nincs.)

2. A köznyelvi szókészlethez való viszonyítással az esetek nagy részében megnyugtatóan megállapítható, mi tájszó. A szókészlet állandó mozgása, a köznyelvi szókincs számbavételének és kodifikáltságának érthető hiányosságai miatt azonban gyakorta mégsem. A legbiztosabb tájékozódási lehetőséget ma az Értelmező kéziszótár új kiadása jelenti (Pusztai Ferenc főszerk. 2003), illetőleg – a regionális szókészlet felől – kiegészítő forrásként elsősorban az Új magyar tájszótár. Azt azonban, hogy mely szavak regionális köznyelvi, friss kutatások és felmérések nélkül egzakt módon többnyire nehéz megállapítani. Külön gond – ahogy erre Balogh Lajos figyelmeztet –, hogy az alaki tájszavak esetében az ÉKsz. semmiféle

útbaigazítást e tekintetben nem ad (nem feladata ugyanis). Azt az Imre Samutól bevezetett elvet pedig, hogy csak az adott nyelvjárás általános hangtani sajátosságaitól eltérő alaki tájszavakat közöljük tájszótárakban (Felsőöri tájszótár 11–3), csupán képzett nyelvjárásgyűjtő tudja alkalmazni. A megoldás ebben az esetben csak az lehet, hogy a szerkesztők e szempontból megrostálják majd az anyagot.

3. Honnan vehetjük a tájszavakat? Vagy valamely formában korábban rögzített adatokat használunk föl, vagy élőnyelvi új gyűjtéssel. Ami a már rögzített forrásokat illeti, a rögzítettség típusa szerint a következők vannak: a) Írott, tehát kéz- és gépírási források (a nyelvészeti kutatások során dialektológusoktól, tanítványaiktól vagy laikusoktól gyűjtött, rögzített tájszavak, idevágó témájú szakdolgozatok, doktori értekezések, pályamunkák, egyetemi és főiskolai szemináriumi dolgozatok, középiskolai szakkörös vagy más indíttatású szógyűjtések, kéziratos tájszótárak (például Sima Ferenc felvidéki gyűjtése), szójegyzékek, levelek, bármely műfajú kéziratos följegyzések és szövegek. Külön említendő a Néprajzi Múzeum Etnológiai Adattára, amely igen gazdag nyelvjárás anyagot is tartalmaz a hagyományos paraszti gazdálkodás témakörében, a korábbi, valóban eredményes néprajzi és nyelvjárás gyűjtőpályázatok eredményeként, illetőleg a szombathelyi főiskola Magyar Nyelvészeti Tanszékén a Szabó Géza és munkatársai irányításával létrehozott hatalmas nyelvjárás szókészleti adatbázis). b) Nyomtatott források: tájszójegyzékek, tájszótárak, regionális szótárak, általános tájszótárak, értelmező szótárak (mindegyik magyar értelmező szótárban vannak tájszavak is, a legújabbban, az „Értelmező szótár+”-ban mintegy 10 000, A magyar nyelv nagyszótára is, amely az 1772 és 2000 között dokumentálható magyar szókészletből válogat, „Mérték-tartóan és körültekintéssel reprezentálja [...] a nyelvjárások [...] sajátos szókincsét” (Csengery in: A magyar nyelv nagyszótára I, 18), nyelvatlaszok, néprajzi atlaszok, illetve tudományos, szépirodalmi, publicisztikai írások, szövegek, helytörténeti munkák (például Rancz Teréz, „Gyöngyeiteket ne szórjátok szét!” Kézdialmás, 2003: alaki, jelentésbeli és valódit tájszavakat külön listázva közöl, Ábrahám Imre, Écs község múltja. Vizsgálódás a település történeti és nyelvi forrásaiban. Győr, 2007: kitűnően szerkesztett tájszótári fejezete is van), néprajzi kiadványok (például Eperjessy Ernő Puszták népe a Zselicben (1900–1950), Budapest, 2006: számos, szövegbe ágyazott tájszót tartalmaz), szöveggyűjtemények magyarázó szójegyzékei, fajtánévjegyzékek stb. c) Elektronikus szótárak, számítógépes korpuszok, tudományos, szépirodalmi és publicisztikai szövegek: digitalizált szöveggörpuszok). d) Hangzó források (fonográf- és magnetofon-felvételek, nyelvjárás szövegfelvételek, képes-hangos hangfelvételek, interjúk, digitalizált beszédfelvételek).

Az élőnyelv állandó potenciális forrás: a (regionális nyelvi) beszélők nyelvhasználat és regionális nyelvi tudása mint a tájszavak éltető közege rendelkezésünkre áll. A dialektológiában a tájszavak gyűjtésére és közzétételére vonatkozóan számos tapasztalat felhalmozódott. Először laikusok gyűjtötték őket, s kezdettől fogva mindig is részt vettek a gyűjtésben laikusok is (Végh József áttekintése [1964: 17] szerint Európában a népnyelvi szókészlet gyűjtésében a franciák és a lengyelek kivételével mindenütt támaszkodtak laikusokra is). Igaz, ennek mindig megvan a kockázata: „a polihisztorok és a nem szakképzettek hasznos közreműködése mindig magában hordozza a dilettantizmus és a provincializmus veszélyét, kísértését”

(Péntek 2004: 148). Két megszorítás idekívánkozik. Az egyik: a laikusoktól csak meghatározott szakmai kívánalmak bizonyos szintű teljesítése várható el (így van ez más tudományterületeken is). Másrészt: nem általában a laikusokra, hanem többnyire szakemberektől felkészített és irányított laikus gyűjtők közreműködésére támaszkodik a szakma (vö. Végh [1964: 18] „Szakemberek segítsége nélkül ilyen munkát nem lehet kezdeményezni”). A tapasztalat másfelől azt mutatja, hogy „a nyelvjárást anyanyelvi szinten beszélő egyének közléseinek hitelessége, megbízhatósága igen magas fokú” (Balogh Lajos 1996/1997: 67, tegyük hozzá: a laikus gyűjtő helyi nyelvjárási és szociokulturális ismeretei, azaz a „benne élés és a beleélés” adottsága miatt, l. OrmSz. XL.). A helyi nyelvi és kulturális ismeretekkel rendelkező laikus a szavak fonetikai lejegyzésében, a jelentések értelmezésében és a nyelvészeti természetű sajátosságok felismerésében marad el többnyire. A laikusok irányított gyűjtéséről mégsem kell emiatt lemondanunk, sőt nem is mondhatunk le, mert ha csak nyelvészek (dialektológusok) gyűjthetnének, akkor sok-sok tájszót adnánk át az enyészetnek. Mi a teendő? Megfelelő szakmai irányítással minimálisra csökkenteni a lehetséges kockázatokat. A Bayerisches Wörterbuchnak 400–500 közötti állandó önkéntes munkatársa van jelenleg is, s őket a szótárbizottság látja el ma is szakmai tanácsokkal. Annak idején így tervezték Végh Józsefék is: „arra is gondoltunk, hogy kiépítjük vidéki levelezőink hálózatát [...] megszervezzük vidéki munkatársaink nyelvjárási adatszolgáltató munkáját” (Szépe–Végh 1959: 220). Végh József az ilyen adatközlőktől származó gyűjtésekről írta: „Egy országos nyelvjárási szókincsarchívum vagy szókincsmúzeum, ha nincs is szótárrá szerkesztve, de annyira rendezve van, hogy tudományos kutatók számára hozzáférhető, gazdag anyagával rendkívül becses forrás nemcsak a magyar nyelvtudomány, hanem a néprajz, a művelődéstörténet számára is” (1964: 26). E téren legtöbbit az egyetemek és főiskolák nyelvésztanárai, dialektológusai tehetnek. Csak egy követhető példát említek (a korábban már említett szombathelyi mellett): Vörös Ottó a révkomáromi Selye János Egyetem magyar szakos hallgatóival – választható feladatként – tájszavakat gyűjtetett lakóhelyükön vagy ahhoz közeli településen, s szótárszerűen megszerkesztve kérte be őket (a követelmény 200 címszó volt). 2007 tavaszára ily módon 30 tájszógyűjtemény készült el.

A tájszógyűjtések korábban általában feltételezték a gyűjtő személyes jelenlétét a kutatóhelyen, adatközlőinél. Már jó ideje nem így van (bár kétségtelen, a legjobb a közvetlen kapcsolat, a szemtől szembeni kommunikáció, az, ha a gyűjtő helyben s azonnal rákérdézhet, pontosíthat, ellenőrizhet, s ha az adatközlő megszo-kott, otthoni közegében beszélhet, válaszolhat). Gondoljunk a postai kérdőíves vizsgálatokra (Wenker 1876: l. Kiss J. szerk. 2001: 94), a Bayerisches Wörterbuch munkaközössége által 1958 után kiküldött kérdőívekre, a telefonos adatgyűjtésekre (Labov például erre alapozta a *The Phonological Atlas of North America* című kutatását), újabban pedig a világhálós anyaggyűjtésekre (például Chambers „Golden Horseshoe” vizsgálataira, mely utóbbiak online adatgyűjtéseken alapulnak: Kontra: 2004: 93), vagy akár egyetemi és főiskolai hallgatók szak- és szemináriumi dolgozatai egy részének nyelvi anyagára. A Székely szótár készítője, Sántha Attila egy erdélyi magyar internetes portálon fogott hozzá szótára anyagának összeállításához. Bárki küldhetett be székelynek tartott szavakat. Híre ment a vállalkozás-

nak, s az érdeklődés nagy volt. Mintegy négyezer „székelygyanús” szót küldtek be, s a szerkesztő a válogatáshoz forrásul az Új magyar tájszótárt, a Székely nyelvfeldrajzi szótárt és az Erdélyi magyar szótörténeti tárt használta. Online adatgyűjtésekre számos további példát is idézhetnék a nemzetközi gyakorlatból.

A most folyó és jövőbeli szógyűjtésekre vonatkozóan azoknak a lexikográfiai leírasi szempontoknak a követése professzionális vállalkozások esetében követelmény, amelyeket a szűkebb szakma annak tart (ezeket l. a következő fejezetben).

5. A kimeneti oldal néhány kérdéséről. A tájszótárak – nem beszélve a tájszójegyzékekről, -listákról, nem nyelvészeti tárgyú munkák, közlések tájszói mellékleteiről – a lexikográfiai szereltség, azaz a különböző szempontú minősítettség tekintetében nagy eltéréseket mutatnak. Amikor azonban a tervezett elektronikus tájszótár informatizálásának alapjául szolgáló minősítő-, illetve kódolási rendszert határozzuk meg (ez teljes mértékben nyelvtudományi feladat), akkor az optimális szintből, a mai, illetőleg a várható **felhasználói igényekből** kell kiindulnunk. A cél ugyanis a minél korszerűbb **keresőprogram** létrehozása. Mert: „... az információkérdés pontossága, gyorsasága és minősége kulcsszerepet játszik a korpuszok használatával elérhetővé váló információk elemzésében és értelmezésében” (Szirmai 2005: 100).

Lássuk a felhasználói (elsődlegesen s értelemszerűen nyelvtudományi) igényeket! A lexikográfiai elvárások, minősítési követelmények professzionális gyűjtés esetén a következők (minthogy ezeket a magyar dialektológusok nagy részének az egyetértése hordozza, tehát jelenleg közmegegyezésemércének tekinthetők, s a Nyelvörben olvashatók (Kiss J. 2002: 391–415), csupán felsorolom őket): 1. a címszó, 2. a tájszótípus megnevezése (valódi, jelentésbeli, alaki, alaki-jentésbeli, 3. a regionális szóalak (realizáció) megadása, 4. a származási hely megnevezése (település, kistáj, megye, nyelvjárascsoport vagy nyelvjárási régió, ország), 5. a gyűjtés ideje (évszámmal), 6. a társadalmi érvény jelzése (általános használatú, archaizmus, neologizmus), 7. a szófaj jelölése, 8. a jelentés(ek) közlése, 9. a stílusminősítés. Tetszőlegesenek: 1. példamondatok, szövegrészetek, 2. nyelvi-nyelvhasználati megjegyzések, 3. néprajzi, tárgytörténeti, művelődéstörténeti észrevételek, 4. utalások.

Tudvalévő, hogy a tájszótárakban nem általános a fenti szempontok szerinti lexikográfiai minősítettség (nem is lehet az, mert bizonyos szempontok csak újabban merültek föl), ezért a lexikográfiai szereltség tekintetében igencsak változatos adatok kerülnek majd be az elektronikus tájszótárba. Vannak például jelentésma-gyarázat nélkül közölt tájszavak, vannak csak nagyobb ország- vagy tájrész forrásul megnevező szóközlések (Erdély, Nyugat-Magyarország, Székelyföld, Balatonfelvidék, Rábaköz, Sármellék), a legtöbb közlés nem minősíti társadalmi érvény szerint a szavakat, gyakorta nincs stílusminősítés, nagyon változatos a kép a néprajzi tudnivalók és a példamondatok közlésében – és így tovább. Vannak kivételesen olyan források is, amelyek az adatközlő nevét, életkorát, foglalkozását is közlik. A jelen állás szerint ennek keresőprogrambeli megjelenítése ma nem látszik indokoltnak. Fontos volna viszont az azonos fogalomkörbe tartozó szavak keresésének biztosítása a szemasiológiai kutatások miatt.

A különböző szótárak, források adatainak szócikkbeli egyesítése megoldandó kérdéseket is fölvet majd (Balogh Lajos). Kézenfekvő azonban, hogy az azonos szavak egy címszó alá kerüljenek alaki és jelentésbeli eltéréseik esetén is (ahogy ez az ÚMTsz.-ban is történt, vö. **bobács** *bidibács, bobács, bobács, boblibács, bödebácsi, budabácsi, budabács, bugabácsi*). Mi legyen a frazeológiai kapcsolatokkal, táji szószerkezetekkel? Balogh Lajos fölvetése: s milyen címszót kapjanak? Követendőnek azt a gyakorlatot tartom, amelyet O. Nagy Gábor követett: kiválasztani az állandó szókapcsolat legfontosabb tartalmi jegyét hordozó szót (ez többnyire főnév), azaz a vezérszót, s azt címszóvá tenni (1976: 22).

Nagy változást jelent a korpuszalapú lexikográfia kínálta lehetőség: ahol egész szövegek vannak digitalizálva, ott keresőprogramok segítségével ki lehet emelni szavakat, szövegrészeket. A digitalizálásnak nagy szerepe van: a lexikográfusok lemásoltatják, illetve beszkenneletik, amit korábban maguk cédulázta(tta)k ki és másolta(tta)k le. A nagy, digitálisan hozzáférhető és jól kutatható szövegtörzsek a lexikográfust bőven ellátják példaadatokkal. A válogatás szempontjait meg kell határozni („csak Erdélyből”, „csak az 1950-es évekből”, „csak nyelvjárási szövegekből”). Tájszótárak esetében a lexémák kiválasztása, a példamondatok, a néprajzi-művelődéstörténeti leírások területén lehet elsősorban hasznunkra ez a lehetőség. Enciklopédikus kommentárokat, ha szükségesek, olyan forrásokból vehetünk át, amelyek túlmutatnak a tulajdonképpeni szótárbázison.

A nyelvjárási lexikográfiában is alkalmazni szokás rövidítéseket, szimbólumokat. Minthogy azonban a virtuális térben nem kell a hellyel spórolni, nem kell feltétlenül rövidítéseket és szimbólumokat használni. Vagy csak az általánosan ismerteket. Online gyűjtések és beküldések esetén nem kell figyelni a betűrendre sem, mert a szavak automatikusan a helyükre kerülnek az új technikai környezetben.

Balogh Lajos javaslatával csak egyetérteni lehet: első lépés legyen a már elektronikusan is rögzített tájszótárak anyagának egyesítése.

6. A digitális tájszótárkészítés és -használat előnyei: 1. nincs terjedelmi korlát, az összes tájszótár anyaga összefogható, egyesíthető; 2. az adatbázis idő- és térbeli korlátozás nélkül hozzáférhető, azaz bárki, bárhol, bármikor olvashatja, használhatja (ha van világhálós hozzáférési lehetősége); 3. adva van a folyamatos bővítés lehetősége (szerkesztőségi felügyelettel, irányítással); 4. megfelelő keresőprogrammal könnyen, gyorsan és sokoldalúan tájékozódhatunk a hatalmas korpuszban; 5. az anyagi költség minimális. – De tegyük hozzá: az adatbevitel pénz- és időigényes!

Végezetül: A tervezett tájszótár a magyar korpusznyelvészetnek fontos, minősített adatbázisa lesz. Megvalósításához kell majd egy szakmai központ felelős vezetővel, állandó és alkalmi (dialektológus, lexikográfus és programfejlesztő) munkatársakkal. A központ feladata lesz véglegesíteni a koncepciót, vezetni a munkálatokat, összefogni és irányítani a közreműködőket, tartani a kapcsolatot a szóba jövő intézményekkel és munkatársakkal, megteremteni a tájékoztatás és tájékozódás alkalmait (elektronikus levelezés, segédanyagok előállítás és terjesztése, előadások, tanácskozások megszervezése, pályázatok intézése, a szakma és a nagyközönség tájékoztatása), biztosítani a szervező-irányító kapacitást, az infrastrukturális feltételeket, az online beküldési lehetőségeket, megteremtve az adatok

ellenőrzésének a lehetőségét (például felkészített levelező munkatársakkal, akik ott laktak, vagy tájékozottak azon a környéken, településen, ahonnan az adatokat beküldték). Mert a digitális szótár is csak akkor jó szótár, ha adatai megbízhatóak, hitelesek, sokoldalúan vannak minősítve, s ha megfelelő keresőprogramokkal könnyen és gyorsan megtalálhatjuk benne, amit keresünk.

SZAKIRODALOM

- Atkins, B. T. S. 2002. Then and Now: Competence and performance in 35 Years of Lexicography. EURALEX 2002 Proceedings. In: Braasch, A.–Povlsen, C. (eds): *Center for Sprogteknologi*. Copenhagen, 1–28.
- Balassa Iván 1996/1997. Az Új Magyar Tájszótár és a néprajzkutatás. In: Bánki Judit (szerk.): *Emlékkönyv B. Lőrinczy Éva hetvenedik születésnapjára*. A Magyar Tudományos Akadémia Nyelvtudományi Intézete, Budapest, 166–70.
- Balogh Lajos 1996/1997. Mura-vidéki (kapcai) tájszavak egy népi adatközlő értelmezésében. In: Bánki Judit (szerk.): *Emlékkönyv B. Lőrinczy Éva hetvenedik születésnapjára*. A Magyar Tudományos Akadémia Nyelvtudományi Intézete, Budapest, 67–73.
- Computer Methods in Dialectology 1989. *Journal of English Linguistics* 22/1. A Special Issue. *Das Bayerische Wörterbuch*. <http://www.bwb.badw.de>
- Elspaß, Stephan 2005. Zum Wandel im Gebrauch regionalsprachiger Lexik. *Zeitschrift für Dialektologie und Linguistik* 72: 1–51.
- Elspaß, Stephan, König, Werner (Hrsg.) 2007. Sprachgeographie digital – die neue Generation der Sprachatlanten (mit 80 Karten). Hildesheim, Zürich, New York. Olms.
- Fóris Ágota 2002. A szótárfogalom megváltozásáról. *Modern Filológiai Közlemények* IV/2: 5–19.
- Fóris Ágota 2004. A mesterségek szótárai. In: Tóth–Földes–Fóris, 42–7.
- Fóris Ágota–Pálffy Miklós (szerk.): 2005. *A lexikográfia Magyarországon*. Tinta Könyvkiadó, Budapest.
- Gerstner Károly 2007. *A magyar nyelv nagyszótárának tájnyelvi metszetéről*. Kézirat.
- Imre Samu 1973. *Felsőíri tájszótár*. Akadémiai Kiadó, Budapest.
- Ittész Nóra főszerk. 2006. *A magyar nyelv nagyszótára*. MTA Nyelvtudományi Intézet. Budapest.
- Kálnási Árpád 2005. *Debreceni cívis szótár*. A Debreceni Egyetem Magyar Nyelvtudományi Intézetének Kiadványai 83. szám. Debrecen.
- Kiss Jenő 1998. A tájszavak és a tájszótárak – régi kérdéskör, új problémák. *Magyar Nyelvőr* 122: 427–37.
- Kiss Jenő (szerk.) 2001. *Magyar dialektológia*. Budapest, Osiris Kiadó.
- Kiss Jenő 2002. Tájszótárirás és tájszótárak. *Magyar Nyelvőr* 126: 391–415.
- Kleiber Judit 2007. Számítógépes nyelvészet Pécsen. In: Alberti Gábor–Fóris Ágota (szerk.): *A mai magyar formális nyelvtudomány műhelyei*. Nemzeti Tankönyvkiadó, Budapest, 170–88.
- Kontra Miklós 2004. Lesley Milroy–Matthew Gordon, *Sociolinguistics: Method and Interpretation*. Magyar Nyelv 100: 91–9.
- B. Lőrinczy Éva (főszerk.) 1979. *Új magyar tájszótár*. Első kötet, A–D. Budapest, Akadémiai Kiadó.
- Magay Tamás szerk. 2006. *Szótárak és használóik*. Akadémiai Kiadó, Budapest.
- Magyar tájszótár*. 1838. Kiadta a Magyar Tudós Társaság Budán.
- Möller, Robert–Elspaß, Stephan 2007. Erhebung dialektgeographischer Daten per Internet: ein Atlasprojekt zur deutschen Alltagssprache. In: Elspass, Stephan (Hrsg.), *Neue Sprachgeschichte(n)*. Themenheft der Zeitschrift *Der Deutschunterricht*. 59.3. 115–132.

- O. Nagy Gábor 1976. Magyar szólások és közmondások. Budapest.
- OrmSz. = Kiss Géza–Keresztes Kálmán 1952. Ormánysági szótár. Akadémiai Kiadó, Budapest.
- Niebaum, Hermann 1979. Deutsche Dialektwörterbücher. *Deutsche Sprache* 7: 345–73.
- Pajzs Júlia–Mártonfi Attila 2006. A magyar nyelv nagyszótárának számítógépes vonatkozásai. In: *A magyar nyelv nagyszótára*. I. MTA Nyelvtudományi Intézet. Budapest, 20–2.
- Pajzs Júlia 2007. *A jövő szótára és a magyar számítógépes lexikográfia*. Kézirat.
- Papp György (szerk.) 2004. *Mi ilyen nyelvben élünk. Nyelvészociológiai és korpuszvizsgálati tanulmányok*. Szabadka, MTT Könyvtár 9. (A határon kívüli idegen szavak, lexikális elemek adattára: 293–379.)
- Péntek János 2004. *Anyanyelv és oktatás*. Pallas-Akadémia Könyvkiadó, Csíkszereda.
- Prószycki Gábor 1997. Szótári szempontok a számítógépes nyelvi programok korában és korábban. In: Kiss Gábor–Zaicz Gábor (szerk.): *Szavak – nevek – szótárak. Írások Kiss Lajos 75. születésnapjára*. MTA Nyelvtudományi Intézete, Budapest, 326–35.
- Szabó Géza 1980. Regionális szókészleti kutatás Szombathelyen. *Vasi Szemle* XXXIV/1: 104–39.
- Szépe György–Végh József 1959. Tájékoztató a Magyar Nyelvőr nyelvjárási anyaggyűjtő versenyéről. *Magyar Nyelvőr* 83: 210–20.
- Szirmai Monika 2005. *Bevezetés a korpusznyelvészetbe*. Tinta Könyvkiadó, Budapest.
- Szűcs József 1936. *A néprajzi kutatás története*. Spitzer Aladár könyvnyomdája, Budapest.
- Tószegi Zsuzsanna 2006. A könyvdigitalizálás egyes kérdései. *Magyar Tudomány* 1376–84.
- Tóth Gergely–Földes Csaba–Fóris Ágota (szerk.) 2004. *Lexikológiai és lexikográfiai látkép: problémák, paradigmák, perspektívák*. Generalia, Szeged.
- ÚMTsz. = B. Lőrinczy Éva (főszerk.) 1979. *Új magyar tájszótár*. Első kötet A–D. Akadémiai Kiadó, Budapest.
- Végh József 1964. A nevezetesebb külföldi nyelvjárási szókincsgyűjtések és a magyar szókincsarshivum terve. Magyar Nyelvtudományi Társaság Kiadványai 109. sz. Magyar Nyelvtudományi Társaság, Budapest.
- Wandl-Vogt, Eveline 2007. *Von der Karte zum Wörterbuch*. Überlegungen zu einer räumlichen Zugriffsstruktur für Dialektwörterbücher. <http://www.wboe.at/resources/200609AewvAbstract.pdf>

Kiss Jenő

SUMMARY

Kiss, Jenő

From the earliest lists of dialect words to a unified digital dialect dictionary

The paper begins with a brief historical overview of the publication of lists of Hungarian dialect words. A detailed treatment of an ongoing project of compiling a unified digitalized dialect dictionary of Hungarian follows suit. The main issues discussed are what criteria can be used to decide what should be included in that dictionary and what would constitute an optimal system of lexicographic annotations that could serve as a basis for the search engine to be worked out. Advantages of compiling and using digitalized dialect dictionaries are listed, and organizational tasks related to the project are spelt out. The unified digital dialect dictionary, once completed, will be an important annotated data base of corpus linguistics in Hungary.