

Rendkívül rövid bevezetés a dokumentációs nyelvészetbe

Szeverényi Sándor

SZTE Finnugor Nyelvtudományi Tanszék

1. Bevezetés

Ennek az írásnak nem az a célja, hogy tudománytörténeti jellegű anekdotázás keretében mutassa be a dokumentációs nyelvészet és azon belül a leíró nyelvészeti terepmunka szépségeit és nehézségeit, hanem hogy napjaink nyelvészeti dokumentációját mint önmagát gyakran a(z alkalmazott) nyelvészet egyik önálló területeként meghatározó diszciplínát pozícionálja.

„Kicsi”, „egzotikus” – az idézőjelek e fogalmak relatív mivoltát jelölik – nyelveket, kultúrákat évszázadok óta dokumentálnak, a tudatos nyelvleírásnak (benne a nyelvtan-írásnak), a nyelv dokumentálásának pedig évezredes hagyományai vannak. Ha ez így van, akkor vajon mi az oka annak, hogy az elmúlt évtizedekben, különösen az elmúlt 15–20 évben új területként tűnt fel a nyelvi dokumentáció, amely dokumentációs nyelvészetként határozza meg magát? Az, hogy a dokumentációs nyelvészet új szemléletet, új területeket, új eszközöket ötvöz, s ami a világ nyelvi diverzitását illeti: a leírásban teljességre, egységességre és összehasonlíthatóságra való törekvést jelenti.

A nyelvészeti dokumentáció természetesen sokféle lehet. Amiről én itt és most szólni fogok, az mindennek az alapja: egy ismeretlen vagy alig ismert nyelv leíró szempontú dokumentálása. Hogy ez egyben valóban új tudományágat is jelent-e, ebben megoszlanak a vélemények. Ami biztos, az a kiindulópont: az alap a terepmunka, amely ebben az esetben a „descriptive linguistic fieldwork”, magyarul kb. leíró nyelvészeti terepmunka, definíciója szerint „...egy nyelv szerkezetének a vizsgálata elsődleges nyelvi adatokon keresztül a nyelvet anyanyelvként beszélő konzulensek segítségével” (de Reuse–Chelliah 2011: 7), hiszen a gyűjtés célja, módja és körülményei meghatározzák a további kutatási lehetőségeket.

Az elmúlt években nagyon sok összefoglaló munka született (l. a Bibliográfia), nagymértékben támaszkodom ezek megállapításaira, illetve mivel ez egy valóban nagyon rövid összefoglalása egy nagyon nagy témának, így ebből következik, hogy e szöveg pusztán az általam legfontosabbnak vélt elemeket emeli ki.

2. Veszélyeztetett nyelvek, kis nyelvek

A mai dokumentációs nyelvészetet az a felismerés hívta életre, mely szerint a 20. század második felében – részben a gyorsuló globalizációs folyamatok következtében – a nyelvek jelentős része veszélybe került, beszélőinek száma radikálisan csökkent és továbbra is csökken, s számos nyelv mára már el is tűnt. S mivel a világ mintegy 7000 nyelvének a 60%-át százezernél is kevesebben beszélik (természetesen nem összesen...), így sok nyelv nagyon gyorsan, akár pár év alatt az eltűnés határára kerülhet. (Ugyanakkor egy nyelv nem feltétlenül veszélyeztetett azért, mert kevesen beszélik. A veszélyeztetettség akkor lép fel, amikor a nyelv átadása generációról generációra megszűnik, vagy annak lehetősége korlátozódik.)

A tudománytörténet szeret korszakolni, korszakhatárokat kijelölni, a mi esetünkben ez 1992. 1992 előtt is beszéltek már veszélyeztetett nyelvekről, azok dokumentációjának fontosságáról, megmentéséről, arról a bizonyos 24. óráról, miszerint ha most nem áll neki mindenki nyelveket leírni, akkor nagyon hamar azt fogjuk észrevenni, hogy sorra tűnnek el majd nyelvek, s a nyelvekkel együtt kultúrák is. 1992-ben az történt, hogy az egyik vezető nyelvészeti folyóiratban, a *Language*-ben megjelent egy tanulmánycsokor *Endangered languages* (Veszélyeztetett nyelvek) címmel (Hale et al. 1992). A tanulmányt Ken Hale, Michael Krauss és több neves nyelvész jegyezte, és a nyelvi veszélyeztetettség felismerése okán a dokumentáció fontosságára hívták fel a figyelmet programszerűen, mert ahogy Krauss fejezetének címében szerepel: „The world’s languages are in crisis” (A világ nyelvei válságban vannak.). S noha vannak, akik szerint Kraussék túlságosan pesszimista képet vázoltak fel (pl. Simons–Lewis 2013), azt kell mondanunk, hogy a nyelvvesztés, nyelvhalál a legtöbb uráli nyelv esetében is valós veszély, és egyre gyorsuló képet mutat. Nem mellékesen pedig ennek a tanulmánygyűjteménynek valóban mozgósító ereje volt.

Eljutni távoli népekhez, hosszú időt ott tölteni (esetleg többször visszamenni), majd éveken keresztül feldolgozni a gyűjtött anyagot, a feldolgozott anyagot közzétenni elektronikus vagy hagyományos formában: ez mind – mint minden – időbe és pénzbe, nem kevés pénzbe kerül. Több, különböző területeken működő szakemberek együttműködése szükséges hozzá. Mindez azt vonja maga után, hogy dokumentációs kutatások leggyakrabban kutatási projektek keretében valósulnak meg, s fontos, hogy ezek a projektek tudjanak egymásról, annak érdekében, hogy a források a lehető legokosabban legyenek felhasználva. A legismertebb veszélyeztetett nyelvek dokumentációját támogató szervezetek a DOBES (a Volkswagen alapítványon keresztül) és a The Hans Rausing Endangered Languages Project (SOAS, University of London) is az 1990-es években kezdték meg működésüket, e felismerések hatására.

3. Nyelvészeti terepmunka és dokumentáció

A nyelvi dokumentáció egy nyelv adatainak rögzítését, annotálását, megőrzését és disszeminációját jelenti, azaz gyűjtést, feldolgozást, annotációt, közzétételt, archiválást és tárolást (Woodbury 2011: 159), s mindezt a bevezetőben említett egységesség, összehasonlíthatóság és teljességre való törekvés jegyében.

Az egységesség és összehasonlíthatóság azt jelenti, hogy a nyelveírás során a nyelvtipológiai keretet nem lehet figyelmen kívül hagyni (pl. Evans 2006). A tipológia egyik feladata, hogy általános fogalmi keretet biztosítson nyelvek leírásához, nyelvi szerkezetek felismeréséhez. A veszély természetesen mindig fennáll: ha eltérő, különböző fogalmakat használunk egy-egy jelenségre, akkor azt kockáztatjuk, hogy más nyelvek hasonló jelenségeit nem ismerjük fel. Viszont ha hasonló fogalmakat használunk, akkor a különbségek felismerése maradhat el (vö. Dryer 2006).

A tipológia segítségével megismert strukturális sokszínűség segít abban, hogy félretegyük anyanyelvünk és idegen nyelvek tanulásakor megszerzett és megtanult mintákat, előítéleteinket és elvárásainkat. Egy példa: terepmunka szemináriumon ismeretlen nyelv lejegyzésekor a hallgatók automatikusan szuffixumnak minősítenek olyan jelentésű elemeket, amelyek megfelelői a magyarban szuffixumok (pl. tárgyeset, a birtokos jelölése stb.), ugyanis nem merül fel bennük, hogy az adott elem státusza más is lehet (önálló lexéma, névutó, klitikum stb.), és fordítva: amikor pl. a tárgyjelölés feltérképezése a feladat, szinte kizárólag a tárgyrag „megtalálására” koncentrálnak.

A teljességre törekvés azt jelenti, hogy lehetőség szerint minél több nyelvről minél több információval rendelkezünk. A terepmunkát ez annyiban befolyásolja, hogy a terepmunkásnak abban a szellemben kell gyűjtenie, hogy az anyaga a későbbiekben minél szélesebb közönség (kutatók, a helyi közösség, érdeklődő laikusok) számára hasznos legyen, már csak azért is, mert nem lehet tudni, hogy az adott nyelvet fogja-e még bárki a későbbiekben kutatni, gyűjteni, dokumentálni, ezért a rendelkezésre álló időt és eszközöket a lehető legjobban kell kihasználni. Ezért, noha az elicítált, kérdőívvel, célirányosan gyűjtött adatok is fontosak, a legsokoldalúbban felhasználható anyagokat a szövegek jelentik, amelyek aztán a későbbi, egymástól különböző kutatási célok teljesítéséhez is megfelelőek lehetnek.

A prototipikus dokumentációs nyelvészet kulcsa maga a terepmunka. Ideális esetben a terepmunka a helyszínen (a beszélő lakóhelyén), a közösségben történik, és a kutató a lehető legtöbb időt tölti velük. Az alábbi táblázat a terepmunka alapvető elemeit tartalmazza a legideálisabb állapottól a legkevésbé ideálisig (de Reuse–Chelliah 2011: 8; Hyman 2001: 21):

	prototipikus terepmunka	a prototipikus terepmunka ellentéte	legkevésbé terepmunkaszerű
az elicitált (az adatközlő)	más (nem a kutató)	maga	introspekció
az elicitor / megfigyelő / kutató	maga	más	másodlagos adat
távolság	messze	közel	saját lakhely
település, közösség	kicsi	nagy	város, egyetem
időtartam	hosszú	rövid	rövid megszakításos
nyelv	egzotikus	ismert	saját
az anyag	természetes/kult. kontextus	mint formális rendszer	absztrakt szintaxis
az adat	naturalisztikus	ellenőrzött	szintetikus beszéd
motiváció	nyelv-vezérelt	elmélet-vezérelt	

E szerint a „tipikus” terepmunka „egzotikus” (nem vagy alig ismert) nyelvre irányul, amelyet nagyon messze, nagyon kevesen beszélnek, s a kutató sok időt (akár egy évet) tölt közöttük.

Dokumentációs nyelvész az is, aki otthon (vagy a munkahelyén) a számítógépe előtt ülve mások által gyűjtött nyelvi anyagokat rendszerez, elemez, annotál, nyelvtant ír, publikál... Van egy egészen extrémnek tűnő ága ennek a műfajnak: a misszionárius nyelvészeti is tekinthető dokumentációs tevékenységnek: régi, misszionáriusok által évszázadokkal ezelőtt gyűjtött nyelvi anyagok tudományos vizsgálata.

A dokumentáció, a terepmunka, amely nem csak „tudományként” tekint a nyelvre, a huszadik század első felében – kb. az újgrammatikus időszak és Chomsky fellépése között – kapott hangsúlyt elsősorban az amerikai strukturalizmus/deskriptivizmus keretében, ezt nevezi az utókor Boas–Sapir–Bloomfield hagyománynak. Ennek a hagyománynak az egyik célja a nyelvi dokumentáció teorizációja, és eredménye az ún. boasi trilógia. A boasi trilógia, mint a neve is mutatja, három pilléren épül, és Frank Boas (1858–1942) amerikai kutató jellemző módszere volt, amelyet többek között a *Handbook of American Indians* című kötetében valósított meg. A három pillér a következő:

1. szöveg
2. szójegyzék
3. nyelvtani vázlat.

A gyűjtött szövegek jellemzően folklórszövegek – nem véletlenül, Boas antropológus, etnológus volt, a nyelvészetet az etnológián belül kezelte –, és ezekhez készítettek szójegyzéket és egy nyelvtani vázlatot, amelyek segítségével többé-kevésbé bárki el tudja olvasni és értelmezni a szövegeket. „Vannak szótáraink, de eltekintve a régi misszionárius munkáktól, nagyon kevés a rendszeres munka. Még ha vannak is nyelvtanaink, nincsenek őslakosoktól lejegyzett szövegek... miközben nyilvánvalóvá vált, hogy a nyelv szerkezetének megismerése csak szövegeken keresztül lehetséges” – írja Boas, kiemelve a szövegek fontosságát (1917: 1). Magyarországon ilyen szellemben készült az egykori Tankönyvkiadó Chrestomathia sorozata, amelyben kilenc¹ uráli nyelv leírása látott napvilágot az 1970–80-as években.

Ha megnézzük ezeket a hagyományos szövegközléseket, és összehasonlítjuk valamely elektronikus archívumban (pl. ELAR, <http://www.elar-archive.org/index.php>) hozzáférhető anyaggal, akkor több alapvető különbséget is észrevehetünk: a hagyományos chrestomathiákban nem találunk hétköznapi témákról beszélgetést, nem találunk csúnya szavakat, a beszélők nem használnak töltelékszavakat, nincs megakadás, hezitáció, és kódváltással sem találkozunk. Tényleg ennyivel szebben beszéltek volna régen az adatközlők? Aligha. A különbség a gyűjtés és lejegyzés szemléletmódjában keresendő. Ma a nyelvi dokumentáció nemcsak a nyelv tartós, többcélú rögzítése, hanem a nyelvhasználaté is, amelyen valóban a tényleges nyelvhasználatot értjük. További alapvető különbség, hogy ma elvárás, hogy a lejegyzett szöveg mellett hozzáférhető legyen a hanganyag is, így bárki ellenőrizheti, hogy valóban azt mondta-e a beszélő, mint ami a szövegben szerepel. Ennél fogva a terepmunka nem más, mint

- (1) Adatgyűjtés nyelvi dokumentáció és leírás céljából,
- (2) Adatgyűjtés beszélők közreműködésével.
- (3) Adatgyűjtés olyan szituációkban, amelyekben elvárható, hogy a beszélők természetesen használják a nyelvet.

Ez azt jelenti, hogy nem minősül terepmunkának az introspektív adatgyűjtés, az adatgyűjtés csak írott dokumentumokból és korpuszokból.

A kutatás középpontjában az anyanyelvi beszélő, a „native speaker” áll. Egy dokumentációs projekt megtervezésekor alapvető kérdés, hogy hány és milyen tudású beszélővel való együttműködés tűnik megvalósíthatónak. Ebből az is következik, hogy kulcsfontosságú az adatközlővel kialakított viszony. A kutatás céljaitól, lehetőségeitől függően kell meghatároznunk, hogy hol keressük, mi a szerepük, milyen tulajdonságok szükségesek, továbbá hány és milyen nyelvtudású beszélőre van szükségünk, kell-e

¹ Manyisi, hanti, udmurt, komi, mordvin, mari, számi, valamint nyenyec és szölkup.

engedély, és nekünk mint kutatóknak milyen etikai kötelezettségeink vannak a beszélőkkel szemben.

4. A dokumentáció céljai

Vaux és Cooper (1999:17) szerint a terepmunka célja „a lehető legnagyobb mennyiségű adat elicitálása a lehető legrövidebb idő alatt”. A szakirodalomban előfordul, hogy a dokumentációt megkülönböztetik a leíró terepmunkától, mások ezt elvetik azzal, hogy a kettő között a gyakorlatban nagyon nehéz meghúzni a határt. Ebben az értelemben a dokumentáció pusztán a nyelvi adatok gyűjtését jelenti (de Reuse–Chelliah 2011: 15 alapján):

	dokumentációs cél	leíró nyelvészeti cél
szó adatok	szavak rögzítése	szótárak
mondat adatok	mondatok rögzítése	elemzett mondatok
diskurzus	szövegek rögzítése	elemzett szövegek
integrálás	–	grammatikák

A dokumentációs nyelvészet elsődleges céljai így e táblázat harmadik oszlopában olvashatók, ezek jelenthetik az alapot további célok megvalósítására.

Másodlagos célok közé tartoznak a nyelv megőrzésére, ill. stabilizációjára való törekvések, mint pl. a nyelvi revitalizáció, a megújítás, és a nyelv újratanulása írott források alapján. Revitalizációval akkor érdemes próbálkozni, amikor a nyelvnek van még néhány (tucat) anyanyelvi beszélője (akik jellemzően az idősebb generációk tagjai). Ebben a szituációban a nyelvész feladata a még összegyűjthető nyelvi tudásanyag dokumentálása. Megújításról akkor beszélünk, amikor a nyelvnek már nincs anyanyelvi beszélője, de az embereknek maradt látens nyelvtudása. Ebben a szituációban a nyelvész terepmunkás a nyelvre vonatkozó emlékek összegyűjtésében segíthet, például rokon nyelvi tudás és példaanyag alapján. Visszanyerésről akkor beszélünk, amikor a nyelv már nem része az élő tudásnak, a visszatanítás alapját történeti dokumentumok képezik. Ebben a terepmunkás nem tud célzott segítséget nyújtani. Ami nagyon fontos: a szándéknak valamilyen szinten meg kell lennie az érintett közösségekben is. A revitalizációval kapcsolatos célok együtt jelentkeznek egyéb gyakorlati célokkal – hiszen maga a revitalizáció is alkalmazott terület, és a hozzá kapcsolódó nyelvtechnológiai eszközök kidolgozása (pl. morfológiai elemzők, helyesírás-ellenőrzők, szótárak, párhuzamos korpuszok, fordítóprogramok, appok stb.), amelyek további alkalmazott célok megvalósítását teszik lehetővé (pl. tananyagok létrehozása). További célként szerepelhetnek nem összehasonlító teoretikus célok, összehasonlító elméleti célok (areális, történeti stb.), valamint dialektológiai, szociolingvisztikai, nyelvtörténeti, kulturális célok.

5. A nyelvészeti terepmunka történetének jelentősége

A nyelvészeti terepmunka különösen „tapasztalati tudomány”, de a dokumentáció további fázisai esetében is nagy fontossággal bír az elődök tapasztalata. Különösen fontos, hogy az elődök milyen módszerekkel dolgoztak, milyen eredményeket értek el, és milyen hibákat követtek el, és nem utolsósorban az etikai kérdések változása: mivel a terepmunka – csúnya szóval – tárgya maga az ember, az emberhez, a „native speaker”-hez való viszonyulásunk is óriási változáson ment keresztül.

A keresztény misszionáriusok terepmunkája (jellemzően a 15–18. században) a portugál és a spanyol kolonizáció, majd a brit, francia, holland, német gyarmatosítás velejárója volt (más vallások nem voltak „érdekeltek” az ismeretek ilyen típusú terjesztésében). A protestáns vagy éppen katolikus misszionáriusok hosszú ideig éltek a közösségben vagy érintkeztek a közösséggel, így gyakran maguk is megtanulták a nyelvet. Ezek a korai lejegyzések nagyon értékesek és fontosak, még akkor is, ha mai szemmel nézve módszertanilag és nyelvészetileg értelemszerűen ritkán pontosak.

A misszionáriusok tevékenységével némileg átfedésben van az ún. „gentleman scholars” időszak. Ők azok a világi utazók, akik ugyan nem nyelvészek voltak, hanem kereskedők, katonák, utazók, felfedezők, és természetesen ebben az időszakban mind férfi... Az ő elsődleges céljuk nem a nyelvi dokumentáció volt, hanem az, amiért éppen úton voltak, oda, ahová a munkájuk vagy éppen a kényszer hajtotta őket. Ennélfogva a lejegyzésük jellemzően fonemikus lejegyzés (jó vagy éppen rossz „füllel”, ahogy hallották), leírtak annyit és úgy, amennyit és ahogy csak bírtak vagy tudtak.

Régi közhely, hogy terepmunkát bárhol lehet végezni, nem kell feltétlenül messzire elutazni. A mai módszertani tankönyvekben is megemlítik, hogy soknyelvű területeken érdemes „körülnézni” a szociális élet helyszínein, úgymint iskola (különösen ahol internátusrendszer működik), kórház, ahová akár messzi falvakból is beszállítják azokat a betegeket, akiket a lakóhelyükön nem tudnak ellátni, de ilyen helyszínek a múzeumok, sőt a börtönök, régen a rabszolgák közösségei, vagy háborús időszakban a hadifogolytáborok. Klasszikus szemináriumi kérdés, hogy Magyarországon az 1910-es években hol gyűjtöttek finnugor nyelveket. Hát az orosz hadifogolytáborokban. Például Munkácsi Bernát az Esztergom melletti táborban gyűjtött mordvin, udmurt, komi anyagot. A másik példa az 1880-as évek, amikor szamojéd családokat utaztattak Európában, etnográfiai látványosságként, s Budapesten az állatkertben (!) szállásolták el őket. A magyar tudósok, köztük pl. Budenz József, szinte sorban álltak, hogy tanulmányozhassák őket. Aztán a fordított eset, amikor a kutatót küldik akarata ellenére messzi vidékre: Waldemar/Vladimir Bogorast (1865–1936) és Waldemar Jochelsont (1852–1937), akiket forradalmi tevékenységük miatt Szibériába száműztek az orosz cári időkben, és akik végeztek szibériai gyűjtéseket többek között a csukcsok, a jukagirok

és az eszkimók között, majd a későbbiekben már amerikai támogatással folytatták munkájukat.

A mai nyelvészeti terepmunka közvetlen előzményeként az amerikai strukturalistákat kell említeni, az ő módszereiket az 1960-as évektől kezdődően a Moszkvai Állami Egyetem kezdte átvenni, és indította el azt a máig tartó gyakorlatot, hogy a nyelvészeti képzésnek alapvető eleme a terepmunka, és ez az uráli nyelvek dokumentációja szempontjából is nagy jelentőségű, hiszen éveken keresztül gyűjtöttek így uráli nyelvi anyagokat.

6. A veszélyeztetett uráli nyelvek dokumentációja

Nem részletezve, de nem is megfélekezve arról, hogy az uráli nyelvek dokumentációjának, különösen az oroszországi őshonos uráli nyelvek esetében több évszázados hagyománya van, s amelyet nem lehet figyelmen kívül hagyni, ha a mai dokumentációs tevékenységet elemezzük. Most csak azokat említjük meg a teljesség igénye nélkül, amelyek beleillenek napjaink dokumentációs mainstreamjébe.

A globális, azaz nem egy-egy régióra vagy nyelvcsoporra fókuszáló támogató szervezetek közül az elmúlt években a legjelentősebb szerepet a DOBES és a HRELP játszotta, melyek támogatásával számos kis uráli nyelv dokumentációja is megvalósíthatott, illetve jelenleg is folynak ilyen projektek. Ezek a nyelvek: kola számi (DOBES, Michael Rießler), enyec és erdei nyenyec (HRELP, Florian Siegl), enyec (HRELP, Andrej Sluinszkij és Oleszja Hanyina), tundra nyenyec (HRELP, Tapani Salminen és Irina Nikolaeva), manysi (HRELP, Székely Gábor), keleti hanti és déli szölkup (HRELP, Balthasar Bickel és Andrej Filcsenko).

Természetesen egy-egy nemzeti kutatási alap is támogat számos dokumentációs projektet vagy olyan projektet, melynek része a terepmunka vagy a dokumentáció egy másik stádiuma (pl. annotált nyelvi korpusz építése), így a finn KONE, az orosz Akadémia, vagy éppen a magyar OTKA. Utóbbira jó példa a Nganaszan nyelv számítógépes morfológiai elemzője (Wagner-Nagy Beáta, Várnai Zsuzsa), vagy Ruttkay-Miklián Eszter hanti nyelvet és kultúrát érintő projektjei.

Ezen projektek számos eredményt produkáltak/produkálnak, függően a kitűzött céloktól és lehetőségektől. Ha pusztán a deskriptív nyelvészeti célokat nézzük, akkor ki lehet emelni a leíró nyelvtanokat (Siegl 2013; Nikolaeva 2014; Rózsanskij 2011), Eredményként szaktanulmányok, modern nyelvészeti elméleti keretben tisztázott grammatikai kérdések, tipológiai szempontú elemzések, tanulmánykötetek jelennek meg. Az uráli nyelvek kutatása egyre inkább kilép a hagyományos keretből, és az általános nyelvészet porondján is megmutatják magukat, s ebben ezeknek a projekteknek fontos szerepe van.

A leíró (deskriptív) nyelvtanok az elsődleges forrást jelentik egy adott nyelvhez. A világ több ezer nyelvének jelentős része nem rendelkezik pontos és részletes leírással. Egy leíró nyelvtan a leglátványosabb gyümölcse egy dokumentációs projektnek, ugyanakkor a legnehezebben beérő gyümölcs is egyben: hatalmas kihívás, és gyakran több évtizedes dokumentációs tevékenység áll mögötte. Ennek oka, hogy egy nyelvten a nyelvészek széles közösségének (elméleti nyelvészek, történeti nyelvészek stb.), nyelvtanároknak és a beszélők közösségének készül. Ez azt jelenti, hogy egyszerre kell modernnek lennie és praktikusnak, ezek között kell megtalálni a középutat: figyelembe venni az adott nyelvcsalád leírási hagyományait, tekintettel lenni a tipológiai szempontokra, kiemelni a sajátos jelenségeket, ugyanakkor nyitottnak kell lenni az újszerű elemzésekre. És mindezt közérthető formában és stílusban...

Az egyik legnagyobb nyelvészettel (is) foglalkozó könyvkiadó, a Mouton de Gruyter sorozatában (Mouton Grammar Library) mintegy 70 ilyen nyelvten jelent meg, jelentős része az elmúlt négy évben.

7. Gyűjtés, feldolgozás

A mai deskriptív nyelvészeti dokumentációnak leglényegesebb vonása, hogy leírásban, technikában, módszerben egységességre törekszik.

7.1. Adatok és korpuszok

A nyelvi dokumentáció azt jelenti, hogy anyanyelvi kompetenciánk sosem lesz az adott nyelv vizsgálatához (természetesen előfordul, hogy egy kis nyelvi közösség valamely tagja nyelvesszé válik, de általában ez a ritkább eset). A boasi időkhöz képest – melynek szemlélete sokáig tartotta magát – nagyjából úgy lehet tömören megfogalmazni, hogy MINDEN (KIMONDOTT) ADAT FONTOS. Mivel a dokumentációs nyelvészet legfontosabb célja, hogy a (fel)használók széles csoportjainak készítsen elsődleges adatokat, így nincsenek kulturális vagy történeti szempontból „fontos” dokumentumok vagy adatok!

A korpuszadatoknak Himmelmann (2006) szerint háromféle típusa van: (1) Kommunikációs események. Lehetőleg tartalmazza a hétköznapi nyelvhasználat minél több területét. Lehetnek dialógusok, elbeszélések, élettörténetek stb. (2) Elicitált adatok. Standard szólisták, tematikus szólisták, kérdőívvel lekérdezett mondatokból kinyert adatok, paradigmák. (3) Elemző beszélgetések, amelyeket a kutató irányít.

Milyen a jó a korpusz? Woodbury hat szempontot sorol fel (2003: 46–47):

- (1) változatos: mindenféle tartalmú és műfajú nyelvi mutatóanyagot tartalmazzon;
- (2) nagy: a digitális korpusz terjedelmes, és tartalmazhat képet, hangot, filmet, szöveget;

- (3) folyamatban lévő, megosztott és opportunistá: bármikor és bármivel bővíthető legyen;
- (4) transzparens: a korpuszt úgy kell strukturálni, hogy az más kutatók számára is használható legyen;
- (5) megőrizhető és szállítható: oly módon kell előállítani, hogy biztosítva legyen a hosszú távú megőrzés;
- (6) etikus: a gyűjtés és az elemzés során tekintettel kell lenni az etikai elvekre, ez vonatkozik a használatra és a hozzáférésre is.

7.2. Metaadatok

A metaadatolás az egyik legfontosabb, de – főleg kezdő terepmunkások esetében – a leginkább elhanyagolt tevékenység, mivel egy-egy körülményről nem biztos, hogy a gyűjtés pillanatában azt gondoljuk, hogy „egyszer majd még fontos lehet”, vagy „majd felírom, ha majd hazamentem”. A metaadat egyfajta információcsomag, tulajdonképpen adat az adatról. A metaadatok legfontosabb csoportjai:

- (1) katalogizálást segítő: az adat azonosításához hasznos információk, pl. nyelvi kód, fájlazonosító, a felvétel körülményeinek adatai (hely, dátum stb.), a beszélő adatai stb.;
- (2) leíró: az adat típusához szolgáló információk, pl. összefoglaló a file tartalmáról;
- (3) szerkezeti: a file szerkesztésének módja, ami lehet speciális, pl. egy adott szöveg egy kétnyelvű szótár;
- (4) technikai: milyen típusú szoftver szükséges a dokumentum megjelenítéséhez, adatok a formátumról, az adatok megőrzése/tárolása;
- (5) adminisztratív: „munkabejegyzések” (pl. utolsó mentés, szerző, morális jogok, terjesztési, hozzáférési megszorítások).

7.3. Eszközök

A mai dokumentációs nyelvészetnek rendkívül fontos eleme a nyelvtechnológiai, korpusznyelvészeti háttér, ebben az értelemben az egységesség a kompatibilitásra is vonatkozik: lehetőleg ugyanolyan szoftverrel, ugyanolyan formátumban dolgozzuk fel anyagainkat (ezt bizonyos támogató szervezetek elő is írják). Aki húsz évvel ezelőtt is járt már terepre, és próbált adatbázist, korpuszt építeni régen és mai szoftverekkel, az tudja igazán értékelni, hogy milyen az, amikor akár már a helyszínen egyszerűen és gyorsan tudja lejegyezni, elemezni és rendszerezni anyagát.

A dokumentációs szoftverek célja a munka leegyszerűsítése, felgyorsítása és az elérhetőség, közzététel megkönnyítése. Ez nem azt jelenti, hogy magát az elemzési/annotálási műveletet automatizáljuk.

A rendelkezésre álló ingyenes dokumentációs nyelvészeti szoftver kínálat széles, és a dokumentáció céljait tekintve sokszínű.

Ezért a nagy „kiszolgáltatók” mindezekre előírják, ugyanakkor ingyenessé teszik, maguk fejlesztik a szoftvereket. Ebben élenjár a SIL (SIL International, www.sil.org) és a TLA, a régi-új DOBES archívum (The Language Archive, <https://tla.mpi.nl/>). Szöveg-lejegyzésre fordítással, lexikális és morfológiai annotációra a Toolbox és a Flex a legelterjedtebb, fonetikai vizsgálatokhoz a Praat, komplexebb hang- vagy videoanyag elemzéséhez az ELAN, s a sort még hosszan lehetne folytatni. A lényeg, hogy ebben is igaz a mondás: a gyakorlat teszi a mestert. Egy-egy programmal való ismerkedés hosszú időt vehet igénybe, viszont ha az ember kellő gyakorlatot szerez, akkor az nagyon nagy mértékben megkönnyíti a munkáját.

8. Végezetül...

...amit ki kell emelni: úgy tűnik, hogy a dokumentációs nyelvészet újabb kihívások elé néz. Manapság a hangsúly a terepmunkáról, a gyűjtésről kezd eltolódni a feldolgozás, tárolás és archiválás felé. Az elmúlt években óriási mennyiségű adatot gyűjtöttek a kutatók és terepmunkások, amely gyűjtéseknek a mai napig csak töredéke vált hozzáférhetővé. Egyfelől ezt a folyamatot kell felgyorsítani, másfelől a tárolásnak és disszeminációnak megint csak személyi, anyagi és technikai konzekvenciái vannak, amivel nagyon nehéz a lépést tartani.

Hivatkozások

- Boas, Franz 1917. Introduction. *International Journal of American Linguistics* 1:1–8.
Reprinted in Franz Boas: Race, Language and Culture, Franz Boas (ed.), 199–210. Chicago: The University of Chicago Press.
- de Reuse, Willem J. – Shobhana L. Chelliah 2011. *Handbook of Descriptive Linguistic Fieldwork*. Dordrecht – Heidelberg – London – New York: Springer.
- Dryer, Matthew S. 2006. Descriptive theories, explanatory theories, and Basic Linguistic Theory. In Felix K. Ameka – Alan Dench – Nicholas Evans (szerk.) *Catching Language. The Standing Challenge of Grammar Writing*. Berlin: Mouton de Gruyter. 207–234.
- Evans, Nicholas – Alan Dench 2006. Introduction: Catching language. In Felix K. Ameka – Alan Dench – Nicholas Evans (szerk.) *Catching Language. The Standing Challenge of Grammar Writing*. Berlin: Mouton de Gruyter. 1–40.
- Himmelman, Nikolaus P. 2006. Language documentation: What is it and what is it good for? In Jost Gippert – Nikolaus P. Himmelman – Ulrike Mosel (szerk.) *Essentials of Language Documentation*. Berlin – New York: Mouton de Gruyter. 1–30.

- Hale, Ken – Michael Krauss – Lucille J. Watahomigie – Akira Y. Yamamoto – Colette Craig – LaVerne Masayesva Jeanne – Nora C. England 1992. Endangered languages. *Language* 68:1–43.
- Hyman, Larry M. 2001: Fieldwork as a state of mind. In Paul Newman – Martha Ratliff (szerk.) *Linguistic Fieldwork*. Cambridge: Cambridge University Press. 15–33.
- Nikolaeva, Irina 2014. *A Grammar of Tundra Nenets*. (Mouton Grammar Library 65.) Berlin – Boston: De Gruyter Mouton.
- Rozsanszkij, Fedor [Рожанский, Фёдор] 2011. *Современный водский язык. Тексты и грамматический очерк*. I–II. Санкт-Петербург: Нестор-История.
- Siegl, Florian 2013. *Material on Forest Enets, an Indigenous Language of Northern Siberia*. (Mémoires de la Société Finno-Ougrienne 267.)
- Simons, Gary F. – M. Paul Lewis 2013. The world's languages in crisis: A 20-year update. In Mihás, Elena – Bernard Perley – Gabriel Rei-Doval – Kathleen Wheatley (szerk.) *Responses to language endangerment. In honor of Mickey Noonan*. (Studies in Language Companion Series 142.) Amsterdam: John Benjamins. 3–19.
- Vaux, Bert – Justin Cooper 1999: *Introduction to Linguistic Field Methods. Coursebooks in Linguistics*. Muenchen: LINCOM.
- Woodbury, Anthony C. 2003. Defining documentary linguistics. In Peter K. Austin (szerk.) *Language Documentation and Description, Vol. 1*. London: School of Oriental and African Studies. 35–51.
- Woodbury, Anthony C. 2011. Language documentation. In Peter K. Austin – Julia Sallabank (szerk.) *The Cambridge Handbook of Endangered languages*. Cambridge: Cambridge University Press. 159–186.