

A hármas határ menti korpusz informatizálásának előkészítési munkáiról¹

1. Azok a kutatásaink, amelyekről több olyan konferencián számoltunk be, amelyeken kedves ünnepeltünkkel, Sándor Annával közösen vettünk részt, amelyek szakmai, módszertani problémáiról, eredményeiről – közös szívügyünknek tekintve – sokat beszélgettünk, új szakaszához érkeztek. Az OTKA K 76239 sz. pályázat keretében lehetőségünk nyílik a hármas határ két-két oldalán gyűjtött korpusz informatizálására és a régóta tervezett többdimenziós térképlapok létrehozására. (Vö.: P. Lakatos szerk. 2002, T. Károlyi – P. Lakatos 2006: 83–95; P. Lakatos – T. Károlyi 2007: 300–308.)

Köszöntő munkánkban – az együttgondolkodás reményében is – Sándor Annával szeretnénk megosztani az informatizálás előkészítésének problémáit, egyben beszámolni arról, honnan hová jutottunk a jelenlegi szakaszig.

Az informatizálás alapját az összesen 18 kárpátaljai, romániai és magyarországi településen aktív indirekt kérdőíves módszerrel, több száz különböző szociokulturális helyzetű, különböző független változókkal jellemezhető adatközlőtől gyűjtött morfológiai és lexikológiai korpusz jelenti.

Kutatási (és térképlap-alkotási) szempontból legfontosabb célunk a trianoni határ nyelvhasználatot befolyásoló szerepének érzékeltetése, a szinkron dinamizmus azonosságainak és különbségeinek feltárása, az eddigi értékelésekkel már igazolt fáziskésés bemutatása, ezzel együtt elsősorban a standard–dialektális változatok mozgásának tükröztetése. Ugyanakkor figyelembe kellett és kell vennünk azt is, hogy a variabilitás – különösen a szókincs területén – nem csak a standard–dialektális változatok használatából adódik, éppen ezért már kutatásaink megindulásakor az első kérdőívek összeállításával is a számos (nyelvi és nem nyelvi) tényezőtől befolyásolt nyelvhasználati variabilitás megragadására törekedtünk.

2. Előkészítő munkánk során alapelvnek tekintettük, hogy minden objektív és szubjektív nyelvi adatot informatizálunk. (Ekkor nem is számoltunk még azzal, hogy mennyi nehézséggel kell majd megküzdenünk az ún. beviteli felület kialakításakor.)

Adataink a vártnál is nagyobb változatosságról tanúskodtak, már korábban is nyilvánvalóvá vált, hogy a feldolgozásnál, így majd az informatizálásnál is új elemzési–értékelési szempontok bevonására is szükség lesz. Ugyanakkor le kellett szögeznünk, hogy az informatizált adatbázist kezelő szoftver ún. *elemzési szempontok* törzstábláját, az adatok kódolását, illetve a beviteli és lekérdezési felületeket is (ezekről lásd még lentebb) elsődleges célkitűzéseink, a tervezett térképlapok elkészíthetősége szempontjából kell megterveznünk, tehát a standard–dialektális változatok megoszlása és ezek független változókhöz rendelése alapján. Munkánkban alapegységnek a különböző morfológiai és lexikológiai *nyelvi változókat* tekintjük, azonosítható alternatív *változataiknak/variánsaiknak* a vizsgált területen megragadható (térképre vetíthető) eloszlása az, amely a nyelvi változatosság és a társadalmi rétegződés

¹ Készült az OTKA K 76239 sz. pályázat keretében.

összefüggéseinek feltárását segíti. A nyelvi változó tehát – a szociolingvisztikában használatos alapfogalomnak megfelelően – variabilis egység, alternatív lehetőségek legalább kéttagú együttese egyazon dolog, jelenség, funkció kifejezésére (vö. pl. Trudgill 1997: 88; Kontra 2003: 33; Lanstyák–Szabó Mihályi 1997: 15–17). A nyelvi változók nagy részének használatát kizárólag nyelvi tényezők befolyásolják, vannak azonban olyan nyelvi változók is, amelyek változatainak a használata választható, illetve társadalmi tényezőtől függ. Ezeket nevezzük szociolingvisztikai változóknak. Mindezeket is figyelembe véve az előre feltételezettnél mégis nagyobb problémát jelentett, hogy mit tekinthetünk változatnak/variánsnak (V), vizsgálati szempontból relevánsnak, s mit további adatnak (T), jelenlegi vizsgálatunk szempontjából nem relevánsnak. Egyértelmű pl. hogy az E/2. személyű –sz változó variánsainak (*adsz–adol; tudsz–tudol* stb.) variabilitását a standard–dialektális kontinuumok mentén való elhelyezkedés adja. Ugyanez jellemezheti bizonyos tőtípusba sorolható változók variánsait (pl. *tehén–tehen*) vagy a szókincs területéről is több lexémát (pl. *kukorica–málé*). Ugyanakkor más tőtípus tagjai csak alakváltozatokban élnek (pl. *alkuszik–alkudik–alkudozik; lélegzeni–lélegezni*), a variabilitást nem a köznyelvi–nyelvjárási változatok adják, hiszen utóbbi variánsokhoz nem társíthatók ezek a jegyek. A lexikológiai kérdőív alapján ilyen pl. a *lekvár–dzsem–íz* stb.

3. Általában elmondhatjuk, hogy a morfológiai korpusz ilyen szempontú előkészítése nem okozott nagy gondot, ezzel szemben számos problémát vetett föl a lexikológiai kérdőív anyagának átgondolása. (Természetesen ez a probléma összefügg a morfológiai és lexikológiai jelenségek nyelvi rendszerbeli sajátosságaival, eltérő variabilitási lehetőségeivel.) Nyilvánvalóvá vált, hogy a hívómondatokra megadott lexémák sokfélesége nem mindig vagy nem kizárólag az alaktani korpusz előkészítésekor alkalmazható, „szűkebb” értelemben vett variabilitásból – adott nyelvi változó változataiként értelmezett nyelvi egységek realizálódásából –, hanem több, csak tágabb keretek között magyarázható megnevezési szemléletből eredhet. (Erről lásd lentebb.) Mindenképpen szükségessé vált tehát az ún. *további adatok* kategória felvétele. Értelmezésünk szerint a további adatok azok, amelyek a vizsgálat szempontjából nem relevánsak. A további adatok kategóriájának felvételét a fenti okon kívül egyéb tényezők is indokolják: az informatizálás ún. beviteli szakaszában nem mindig dönthető el egyértelműen, hogy a kérdőíves adatfelvétel során kapott „bejövő változatot” variánsként vagy további adatként kezeljük-e. Előfordulhat, hogy majd az adatok összegzése után derül csak ki, hogy valamely eredetileg további adatként kezelt változat az adott beszélőközösség nyelvhasználatában szociálisan általános érvényűnek mutatkozik, tehát variánssá „léphet elő”. (Ilyen lehet anyagunkban az *üst* denotátum *bogrács* lexémával való jelölése, illetve pl. a tésztafélék, esetleg bizonyos bútordarabok nem várt megnevezése pl. a *laska* használata a *cérnametél*tre és a *vastagmetél*tre; a *galuska* használata a *szélesmetél*tre; hasonlóan mutatkozik a *pad–lóca* felcserélése is.) E problémák felvetődésekor értékelhettük igazán azt, hogy alapelvünket, vagyis hogy minden adatot informatizálunk, következetesen betartottuk. Az adatok előkészítésének kezdeti szakaszában ugyanis még felvettük az ún. *téves adatok* kategóriát, vagyis a vizsgálati szempontunkból „használatatlan” adatok kategóriáját. A fentiek miatt jelenlegi bázisunkban nincs „téves adat”, hiszen már eddigi munkánk is igazolta, hogy a tévesnek vélt adatok ugyanakkor megbízható nyelvi anyagot nyújthatnak más szempontú elemzésekhez, megközelítésekhez pl. a pszicholingvisztika vagy a kognitív nyelvészet területeihez.

Egyértelmű számunkra már az is, hogy munkánk összegzésekor, a szöveges értékelésben majd ki kell térnünk ezekre a szempontokra.

A további adat kategóriája a fentiek értelmében hipotetikus, ezért a variánsok és a további adatok kategóriája közötti mozgással számolnunk kell. A programot tehát alkalmassá kellett tenni arra, hogy szükség esetén ezt a hipotetikus kódolást módosítani tudjuk.

Az elemzés szakaszában a további adatok listázása azért is lényeges, mert a tévesként értelmezett válaszok mellett olyan adatokat is felölel, amelyek például *orvadtok* is lehetnek, vagyis a hívómondatra nem a vizsgált jelenséghez, jelenségtípushoz tartozó adatot kapunk. Az adat azonban vagy a kérdőív valamely másik vizsgált jelenségéhez szolgáltat adatot, vagy esetleg a nyelvjárási terület jelenségeihez ad újabb adalékot. Pl. a *málé* szemben a konkrét hívómondatához kapcsolódó válaszokkal orvadtokként lényegesen gyakoribb használatot mutat; a hangtani megfelelésekhez köznyelvi *i*: nyelvjárási *ü* (*üdő*); köznyelvi *ö*: nyelvjárási *e* (*szűkesen*) vagy az egyes és többes szám 3. személyű birtokos személyjeles alakok nyíltabb realizációi (*közepi, kötőjök, gyerekök*).

A lexikológiai kérdőív korpuszának informatizálásra való előkészítése egyéb problémákat is felvetett (vö. még P. Lakatos – T. KÁROLYI – IGLAI 2010). A kérdőív címszavait először a hagyományos nyelvjárási szakirodalomra is támaszkodva kizárólag fogalmi körök szerint próbáltuk csoportosítani (vö. Deme –Imre szerk. 1975 és Imre 1987 alapján), a tájszavak kódolásának esetében pedig a valódi, jelentésbeli és alaki tájszó kategóriáit használni (vö pl. Menyhárt–Presinszky–Sándor 2009: 103–105); ill. a variabilitás okaihoz már a beviteli szakaszban jelentéstani (főleg metaforikus, metonimikus) pragmatikai, esetleg (a vizsgált területen egyébként szintén felvetődő) kontaktológiai kategóriákat rendelni. Már az előkészítő munkák során kiderült azonban, hogy az ún. beviteli táblákon ezek kódolási szempontként nem alkalmazhatók, esetleg csak hipotetikusan értelmezhetők és a jövőben inkább a szöveges értékelésben kamatoztathatók.

4. Alapvető elvnek tekintettük azt is, hogy a százalékos arányokat semmilyen adat nem torzíthatja, ezért az előkészítés során ki kellett szűrniük és külön csoportban szerepeltetni az ún. *nem értékel* (E) adatokat. Ide soroltuk azokat az adatokat, amelyek csak a kérdőív kikérdezését folyamatossá, kevésbé merevvé tevő ún. kapcsolókérdésekre adott válaszok (pl. a *tud, olvas, eszik* –re kérdező hívómondatok, amelyeket csak azért teszünk fel, hogy paradigmatajajukat megkapjuk) vagy a kérdőív megtapasztalt hiányosságaiból fakadó, rosszul fogalmazott hívómondatokra adott feleletek. [Pl. *Józi bácsi mindig felnagyítja a dolgokat, ő szívesen... (túloz/túlozik)*].

Külön jelölést kapott az *adathiány*, amelyen belül kétféle hiánnyal számolhatunk: a gyűjtés körülményeiből (pl. az adatközlő életkora, mentális állapota) vagy a szóban forgó denotátum kiszorulásából, nem ismeretéből adódó adathiány. A jelölésben (H; [-]). Ezek felvétele és jelölése azért is szükséges, mert technikailag lehetővé válik megjegyzés írása pl. arra vonatkozóan, hogy mi az adathiány oka: az adatközlő nem ismeri, nem hallotta stb. a kifejezést. A beviteli felületre tehát egy ún. *megjegyzés* kategóriát is fölvtünk, amely a fentiekén kívül számos egyéb szubjektív adat , sőt pl. nonverbális megnyilvánulás rögzítésére is lehetőséget ad.

A kérdőíves kikérdezés megadja azt a lehetőséget, hogy az adatközlő nemcsak saját, hanem az adott beszélőközösség vagy más beszélő/beszélőközösség nyelvhasználatára vonatkozó közlést tegyen. Éppen ezért fontosnak tartottuk a beviteli felületen az ún. *közölt*

adat jelölésének lehetőségét is. (Pl. a *gulya* címszóról: „A környéken nem használják.” A *kasornyáról*: „Az öregek használták.” A *disznója/disznaja* változóhoz: „*régen disztouja*”.)

A kérdőíves kikérdezés számos módszertani problémát vet föl, minden gyűjtő tudja, hogy az adott nem természetes gyűjtési szituációban (vö. megfigyelői paradoxon) nem mindig kapjuk meg a kívánt adatot. Ilyenkor a *rákérdezéses*, direkt módszert alkalmazzuk. Ezt azonban szintén feltüntetjük az adatok mellett.

5. A legalapvetőbb problémák felvetése után tekintsük át az eddig elkészült szoftver felületét és használatát!

A szoftver fejlesztését három nagyobb ütemben végeztük. Első feladatunk a beviteli felület tervezése és kialakítása volt, ideértve a nyelvi adatok erre való előkészítését, azaz kódolását is. A szoftver kezelőfelületén több ún. törzstáblát alakítottunk ki a korpusz felvételezésére (vö. települések, adatközlők, elemzési szempontok, minősítések, értékelések, orvadtatok). A *települések* törzstábla értelemszerűen a vizsgált településekről tárol információkat: tartalmazza azok listáját és országkódját. Egy másik törzstáblában pontosan rögzítjük az *adatközlők* rendelkezésre álló adatait, hiszen ez a sikeres lekérdezés és egyben az azonosíthatóság egyik legfontosabb feltétele. Monogramhoz rendeljük a lakóhelyet, országkódot, születés időt, nemet, tehát olyan paramétereket amik az egyén életútja során nem változnak. [Az olyan – főként követéses vizsgálat kapcsán szerephez jutó – időben változ(hat)ó kategóriákat, mint életkor vagy iskolai végzettség, mindig az éppen rögzített kérdőív felvételezésénél kell manuálisan beállítanunk, illetőleg annak generálása automatikusan megtörténik pl. az életkor, az adatközlők törzstáblába bejegyzett születési idő és a terepmunka dátumának különbségéből.] Az *elemzési szempontok törzstábla* már a kérdőívvel kapcsolatos nyelvészeti szempontú szakmai kérdéseket érinti. Itt két nézőpontot is figyelembe veszünk. Egyrészt, hogy adott változó (illetve a hozzá tartozó változatok) a standard–dialektális kategorizálást tekintve relevánsnak vagy irrelevánsnak tekintendők, másrészt hogy a hívómondatok hogyan csoportosíthatók abból a szempontból, hogy a várt adatok milyen jelenség vizsgálatát szolgálják vagy milyen nagyobb jelentéstani kategóriákba sorolhatók. A terminológiai elkülönítés azért szükséges, mert míg az előbbi rendezési szempont (jelenségtípusok szerinti csoportosítás) az morfológiai, utóbbi (fogalmi körök szerinti csoportosítás) a lexikológiai kérdőív címszavaira vonatkozik. A fogalmi körök kialakításakor a szómező-elmélet szemléletén alapuló, ismert szakirodalmakra támaszkodtunk (vö. Deme–Imre szerk. 1975; Imre 1987.) Ilyen fogalmi körök pl.: kapásnövények, vadon termő növények, a ház részei stb.

6. Adott nyelvi változó – bizonyos szempontból mint „invariáns elem”, címszó vagy szűkebb értelemben jelenséghordozó – többféleképpen realizálódhat aminek rendszerezése (és ezzel együtt kódolása) több szinten, a *minősítés-és az értékelés-törzstábla* segítségével történik a kérdőívkeretben. (A kérdőívkeret fogalmáról lásd alább.)

A minősítés (első szint) tágabb kategória, mely elsősorban az egyes variánsok megterheltségének (gyakoriságának) kimutatásához, a szóföldrajzi vizsgálatok számadataihoz szolgál alapul. Ezen a szinten történik a variáns (V) és további adat (T) szerinti kategorizálás. Az egymástól eltérő variánsokat és további adatokat indexszámokkal jelöljük és vezetjük be a kérdőívkeretbe (pl. V1, V2 stb.; T1, T2 stb.).

Az értékelés (második szint) szűkebb kategória, ami bizonyos szabályokat követ attól függően, hogy a nyelvi változó realizációi milyen jelenségtípust vizsgálnak (vö. Elemzési

szempontok törzstábla: jelentéstípusok/fogalmi körök) vagy milyen vizsgálati lehetőségeket kínálnak (a variabilitás a standard–dialektális kontinuum mentén való mozgás és/vagy egyéb okok következménye). Az erre vonatkozó elveket a kérdőívkeretben következetesen alkalmaztuk, de technikai kikötést, szabályt a programban nem tettünk. A minősítés során indexszámmal megkülönböztetett variánsokat (V1, V2 stb.) az értékelés-törzstábla szerint S (standard), S_i (ikes ragozás paradigmasorához tartozó standard), D (dialektális), SPS (szuperstandard), SBS (szubstandard) vagy I (irreleváns) adatként értékelhetünk a vizsgált jelenségtípustól függően. A szuperstandard és a szubstandard terminus (vö. Lanstyák–Szabó Mihályi 1997: 16) az ikes ragozás paradigmasorába tartozó variánsokra áll, a standard, dialektális megnevezés pedig a területi kötöttség jelölésére vonatkozik. Értékelés törzstábla szerint irreleváns csak az a nyelvi adat lehet, amely pusztán alakváltozatokban él (szociolingvisztikai változó), területi különbség nem vagy nem a standard–dialektális kontinuum mentén jelentkezik. [Várhatóan pl. a *faluk–falvak*; az *erdője–erdeje* vagy a *dobog-ver* (a szív) szociokulturális kötöttséget mutat]. Az indexszámmal megkülönböztetett további adatokat (T1, T2, stb.) minden más értékelési szinten ugyanígy (T) értékeljük.

E két szintű minősítő–értékelő rendszer adja az adatkódolás alapját:

I.	II.	III.
Minősítés	Értékelés	Összesített ért.
V1, V2...	S	S (standard)
V1, V2...	D	D (dialektális)
V1, V2...	SPS	SPS (szuperstandard)
V1, V2...	SBS	SBS (szubstandard)
V1, V2...	I	I (irreleváns)
T1, T2...	T	T (további adat)
[-]	[-]	[-] (adathiány)
nincs adat	nincs adat	H (kérdőív hiány)
nincs adat	nincs adat	E (nem értékelt)

1. ábra

7. A beviteli felület további ismertetése előtt szükséges rövid kitérőt tennünk a fejlesztés során kialakított ún. *kérdőívkeret-rendszer* működési elvéről, funkciójáról. A kérdőívkeret egy egyszerre statikus és dinamikus rendszer, ami tartalmazza a hívómondatok listáját (kérdőív típusokra lebontva: AT= alaktani kérdőív; SZT= lexikológiai kérdőív); a hívómondatok és ezzel együtt a címszavak elemzési szempontok szerinti válogatását; az egyes hívómondatoknál valaha előforduló (minősítéssel és értékeléssel ellátott, vagyis kódolt) adatokat. Adatfeldolgozás szempontból a statikusság azt jelenti, hogy a bevétel során tulajdonképpen mindig egy legördülő listából választunk azok közül a lehetőségek közül,

amelyeket előzetesen a keretben rögzítettünk. A dinamikusság lényege pedig, hogy bármilyen kérdőívkeretben végzett módosítás hatással van a már informatizált adatok természetére. A kérdőívkeret kialakításakor a megbízható működés érdekében az alább megfogalmazott technikai kikötéseket alkalmazzuk.

8. A fentiekén kívül az informatizált adatok besorolása kiegészül egy harmadik szintű – kérdőívkereten kívüli – osztályozással is, ezzel együtt az értékelést összesen három szinten (minősítés, értékelés, összesített értékelés) végezzük (vö. 1. ábra), ez elsősorban a többdimenziós lekérdezés és kimutatás megalapozását szolgálja. Az első és második szintről a törzstáblák kapcsán már fentebb volt szó, s mint utaltunk rá, az adatok kérdőívkeretben történő rögzítése is ezekkel együtt történik. Az összesített értékelés egy adott hívómondatra felvett adat(ok) együttes értékelése (harmadik szint). A kérdőív sokszínűségét szem előtt tartva ezen a szinten is tovább vittük a korpusz kutatási célok szerinti tagolását, egyes elemek kizárhatóságát egyik-másik adathalmazból. Alapesetben (azaz egy adat megadása esetén) az összesített értékelést az értékelés törzstábla irányítja (vö. 1. ábra), a programszintű szabályok megfogalmazása is ezt szem előtt tartva történik. Mivel kutatási céljaink között az egyik legfontosabb a dialektális–standard szembenállás, illetve ennek mozgása, változása, ezért ebből a szempontból nem releváns, s ezért elkülönítendő adatokat, következésképpen ezen a szinten is irreleváns (I) adatoknak nevezzük. Természetesen ez nem jelenti azt, hogy a változásvizsgálat egyéb típusaihoz vagy szóföldrajzi vizsgálatokhoz nem használhatók eredményesen. Az eltérő terminológia miatt ugyancsak külön kezelendők és jelölendők az ikes igék ragozásának vizsgálatához gyűjtött adatok, ezek jelölése (Si, SPS, SBS) az összesített értékelésben is ugyanaz marad. Több adat megadása esetén ezen a szinten további jeleket és szabályokat alkalmazunk. Már korábbi munkáinkban is kiemelt figyelmet fordítottunk az ún. kettős, a hívómondatokra együttesen megadott köznyelvi és nyelvjárási adatokra, hiszen „pillanatfelvételeinkben” azok éppen a szinkrón dinamizmust érzékeltetik. Eszerint a DS (standard és dialektális) és K (standardhoz tartozó ikes és szubstandard/szuperstandard) együttállásokat különböztetünk meg. Ha a variánsok mellett megjelenik egy vagy több további változat, akkor az összesített értékelés szempontjából az előbbieket a mérvadóknak. Az összesített értékelésben a minősítés törzstábla szerinti további adatokat (T) a variánsok (V) mellett figyelmen kívül hagyjuk, ha pedig önmagában fordul elő további változat, azt T-ként értékeljük az összesített értékelésben. [Pl. ha a *fő/fől* változó esetében *fől* (V2/D) és *készít* (T1/T) válaszokat kapunk egyszerre, akkor az összesített értékelés szerint D, azaz dialektális besorolást kap].

9. Mint fentebb erről már volt szó, rá kellett jönnünk arra, hogy a szókinccs vizsgálatához alkalmazott klasszikus szempontok nem minden esetben kielégítőek, a variabilitás standard–dialektális kontinuumra való visszavezetése csak egy része a szókinccs variabilitását adó tényezőknek. Az új szemlélet iránti érdeklődésen túl tehát egy feloldhatatlannak látszó gyakorlati problémakör vezetett el bennünket a kognitív nyelvészeti szempontok alkalmazásának lehetőségéig. Távlati kutatási kérdéseink tehát, hogy az alapvetően nyelvi variabilitás vizsgálatához gyűjtött – szociolingvisztikai szempontok szerint is tagolt – korpuszt kognitív keretben szemlélve milyen tanulságokra, továbblépési lehetőségekre lehet számítani. A nyelvjárási területek lexikai heterogenitása mögött milyen, a kognitív nyelvészet segítségével magyarázható jelenségek húzódnak, mi adja a variabilitás alapját, mi hívja elő a variánsokat. Az újabb szemlélet főként a változatosság mögötti okokra fókuszálva, a

szemantikát előtérbe helyezve adhat újabb lendületet a dialektológiának, s ezzel együtt kutatásainknak. A dialektológia számára legfontosabb hozadék, hogy az eddigi megközelítések szerint értékelhetetlen, téves adatként értelmezett változatok egy része kognitív keretben új értelmet nyerhetnek, így még jobban kihasználhatjuk a korpusz adta lehetőségeket. Mindezek a problémák ráirányították a figyelmünket arra, hogy milyen fontos a hívómondatok precíz, a vizsgálandó jelenségre, lexémára fókuszáló megfogalmazása (láthattuk, hogy a kérdőívünkhöz alapot adó nagyatlasz hívómondatai is mennyi problémát hordoznak). Szükség van a hívómondatok területi, pragmatikai, jelentéstani stb. továbbdifferenciálására.

A további, még hosszú ideig tartó együttműködés reményében sok szeretettel köszöntünk, Anna!

A felhasznált irodalom

- Deme László–Imre Samu szerk. 1975. A magyar nyelvjárások atlaszának elméleti-módszertani kérdései. Akadémiai Kiadó, Bp.;
- Imre Samu, Nyelvjárási szókészletünk néhány szerkezeti kérdése a magyar nyelvatlasz anyaga alapján. Akadémiai Kiadó, Bp., 1987.
- Kontra Miklós 2003. Nyelv és társadalom a rendszerváltás kori Magyarországon. Osiris Kiadó, Budapest.
- Lanstyák István–Szabómihályi Gizella 1997. Magyar nyelvhasználat–iskola–kétnyelvűség. Nyelvi változók a szlovákiai és magyarországi középiskolások néhány csoportjának magyar nyelvhasználatában. Kalligram Könyvkiadó, Pozsony.
- Menyhárt József–Presinszky Károly–Sándor Anna 2009. Szlovákiai magyar nyelvjárások. Konstantin Filozófu Egyetem, Nyitra.
- Peter Trudgill 1997. Bevezetés a nyelv és társadalom tanulmányozásába. JGYTF Kiadó Szeged.
- P. Lakatos Ilona szerk. 2002. Mutatványok a hármas határ menti nyelvhasználat kutatásából. Bessenyei György Könyvkiadó, Nyíregyháza.
- P. Lakatos Ilona–T. Károlyi Margit 2007. Társadalmiság a területiségben. In: Zelliger Erzsébet szerk. Nyelv, területiség, társadalom MNyTK. 228. szám. Budapest, 2007: 300–308.
- P. Lakatos Ilona–T. Károlyi Margit–Iglai Edit: A hármas határ mentén gyűjtött korpuszok feldolgozási lehetőségeiről. In: Hári Gyula (szerk.), „Végetlen a tér, mely munkára hív” Köszöntő kötet Révay Valéria 60. születésnapjára. Pannon Egyetem Magyar Nyelvtudományi Tanszék, Veszprém. 139–146.
- T. Károlyi Margit–P. Lakatos Ilona 2006. A többdimenziós nyelvföldrajzi ábrázolás lehetőségei a hármas határ menti korpuszban. Magyar Nyelvjárások 83–95.