

A MAGYAR JELNYELVI KORPUSZ: A (SZOCIOLINGVISZTIKAI) KUTATÁSOK ÚJ LEHETŐSÉGEI ÉS KIHÍVÁSAI

1. A korpusznyelvészet lehetőségeiről

A korpusznyelvészet nagy múltra tekint vissza, Kädling 1897-ben már egy 11 millió szóból álló korpuszon vizsgálta a német nyelvben előforduló betűk és betűsorok gyakoriságát, és már az 1870-es években is használtak olyan, szülők által készített naplót, melyek segítségével a gyermekek nyelvelsajátítását vizsgálták (McEnery–Wilson 2001). Chomsky színre lépésével ugyan a (korpuszok segítségével vizsgált és vizsgálható) performanciáról a nyelvi kompetenciára helyeződött a hangsúly, a korpuszalapú vizsgálatokat pedig néhány további (többé-kevésbé helytálló) kritikával is illették (pl. egy természetes nyelv sohasem véges, a korpuszok ezért mindig csak részlegesen lehetnek), melyek a korpuszalapú vizsgálatokat egy időre háttérbe szorították, a különböző technikai fejlődések és új lehetőségek eredményeként a korpuszok használata egyre elterjedtebbé vált. Leech (1992) úgy véli, hogy a korpusznyelvészet nem hasonlítható össze a többi úgy nevezett *kötőjeles nyelvészettel*, mivel „nem egy kutatási területre utal, hanem inkább egy módszertani alapra a nyelvészeti kutatások lefolytatásához” (Leech 1992: 105), amely a fonetikától a szintaxison át a szociolingvisztikáig a nyelvészet minden területén alkalmazható.

Korpuszvizsgálatok segítségével olyan esetekben is eredményekhez juthatunk, amikor az önmegfigyelés nem működik (pl. a nyelvelsajátítás vizsgálatakor, hiszen kisgyermeknek még nincs olyan kiépült metanyelvi tudatossága, amelynek segítségével nyilatkozni tudnának saját nyelvhasználatukról), emellett míg önmegfigyeléskor az adatközlőt befolyásolhatjuk, így manipulálhatók az adatok is, addig a korpuszadatok objektív, megbízhatóbb eredményt szolgáltathatnak (McEnery–Wilson 2001). Gyakorisággal kapcsolatosan sincsenek megbízható ismereteink, a korpuszok segítségével azonban kvalitatív adatokat nyerhetünk az egyes szavak, kifejezések használatáról.

Baker (2010: 8–9) 5 fő érvel támasztja alá, hogy miért lehet hasznos a korpuszokra támaszkodó módszerek felhasználása a szociolingvisztikában. Úgy véli, hogy **a**) egyrészt mindkettő nagy hangsúlyt fektet a természetes nyelvi adatokra, a valós nyelvhasználatra és a társadalmi kontextusra, **b**) másrészt mindkettőben fontos a kvalitatív megközelítés is, **c**) harmadrészt pedig a mintavételi technikák fontosságát emeli ki közös jellemzőként. Hangsúlyozza továbbá, hogy **d**) mindkettő a változatosságot és a változást vizsgálja, széles körű nyelvi jellemzők figyelembevételével, végül pedig azt, hogy **e**) mindkettő megpróbál magyarázatot találni a kutatási eredményekre.

Baker rávilágít számos olyan konkrét kutatási kérdésre is, melyekre a korpusznyelvészeti vizsgálatok választ adhatnak, így például arra, hogy olyan szociolingvisztikai változathoz kapcsolódó mintázatokat is felfedezhetünk, melyeket korábban nem is feltételeztünk, hiszen minden nyelvi elem esetében megnézhetjük a hozzákapcsolódó változókat (nem, kor stb.) és azt is, hogy milyen különbségekhez mely változók kötődnek. A nyelv

változása is nyomon követhető különböző időpontokból származó gyűjtések összevetésével, a gyakoriságot mutató kvalitatív adatainkat pedig kiegészíthetjük többek között a kontextus vizsgálatával, például konkordancialisták segítségével.

A nyelvi változás és változatosság feltárásán túl a korpuszok szótárak és grammatikák alapjait is képezhetik (l. többek között Jones–Waller 2015), ugyanakkor, ahogy Jones és Waller is rámutat, habár számos módon használhatunk egy korpuszt, és „sokféle adattal szolgálhat, de azt nem tudja megmondani nekünk, hogy miért használunk egy adott min-tázatot” (Jones–Waller 2015, 1.7. fejezet 1. bekezdés, oldalszám nélkül).

Az alkalmazott szociolingvisztika számára is fontosak a megfelelő módszerekkel gyűjtött korpuszanyagok, ugyanis erőforrásul szolgálhatnak a valós problémák kezelésében, platformot szolgáltathatnak a közösségi bevonódás megteremtéséhez, és a tudományos kutatás számára is forrásul szolgálhatnak (Clark et al. 2016). Clark és munkatársai a 2010–11-es új-zélandi földrengést követően gyűjtöttek „földrengéstörténeteket”, melyek később a katasztrófavédelmi kutatások számára is fontos információkat biztosítottak, emellett az oktatásban is felhasználták őket tananyagként, a diákok bevonódásának növelésére.

Megkülönböztethetünk korpuszalapú (*corpus-based*) és korpuszvezérelt (*corpus-driven*) kutatásokat, habár ezek között nem mindig könnyű különbséget tenni (Tognini-Bonelli 2001). Általában elmondható, hogy a korpuszalapú megközelítés már korábban megfogalmazott elméletek vagy leírások tesztelésére, alátámasztására használja a korpuszokat. A korpuszvezérelt megközelítés túlmutat ezen, a leírások kizárólag a korpuszból származó bizonyítékokon alapulnak.

Viszonylag új kutatási terület a multimodális korpusznyelvészet, amely arra a kiindulási pontra épít, hogy a kommunikáció önmagában is multimodális, a természetes nyelv pedig egy beágyazott jelenség, és amennyiben a kommunikációt átfogó módon szeretnénk értelmezni, akkor szükség van a testi cselekvések és a nyelvi produkció együttes értelmezésére (l. Adolphis 2013). A jelnyelvek multimodális vizsgálata, melyekben a nonmanuális összetevők (testtartás, tekintet stb.) egyébként is különösen fontos szerepet játszanak, hozzájárulhat az emberi kommunikáció jobb megértéséhez is.

2. Jelnyelvi korpusznyelvészet

2.1. A jelnyelvi korpusznyelvészet kialakulása és kihívásai

Korai jelnyelvi gyűjtések már az 20. század elejéről is maradtak fenn, így például egy történelmi jelnyelvi archívum, melyet az amerikai Siketek Nemzeti Szövetsége készített 1910–20 között (l. Konrad 2012), Johnston és Schembri (2013) azonban úgy vélik, hogy mivel a 2008-ig létrejött jelnyelvi „korpuszok” csupán videófelvételekből és esetleg ezekhez kapcsolódó átírásokból álltak, más esetekben felvételek nélküli átírásokból, melyek módszertanukat tekintve se természetes, se reprezentatív adatoknak nem tekinthetők, ezeket modern értelemben nem nevezhetjük *korpusznak*. „Jelnyelvi korpuszok létrehozása a modern értelemben ugyanis többet jelent, mint videó alapú szövegek felvételeinek készítése, digitalizálása, szerkesztése, katalogizálása és archiválása” (Johnston–Schembri 2013: 1216), ezeknek bármely más korpuszhoz hasonlóan reprezentatívoknak, géppel olvashatóknak és jól dokumentáltaknak kell lenniük.

Johnston és Schembri 3 fő okot emel ki, amiért a modern jelnyelvi korpusznyelvészeti vizsgálatok csak a 2000-es években indulhattak el. 1) A generatív nyelvészet hatására a

nyelveírások kis számú adatközlő grammatikai ítéletei alapján születtek meg. 2) A jelnyelveknek nincs széles körben elfogadott és használt írásrendszere, és habár már léteznek speciális notációs rendszerek, a háromdimenziós jelnyelvek ilyen módon történő „átírása” továbbra is nehézséget okoz. 3) Sokáig nem állt rendelkezésre a megfelelő technikai háttér ahhoz, hogy időben össze lehessen kapcsolni az átírt megnyilatkozásokat a jelnyelvi felvételekkel, így valójában nem léteztek forrásszövegek. Először tehát a 21. században, a számítógépek fejlődésével és multimédiás annotációs szoftverek létrejöttével vált lehetővé valódi jelnyelvi korpuszok létrehozása.

A korpuszalapú kutatások a jelnyelvek szempontjából ugyanakkor talán még nélkülözhetlenebbek, mint hangzó nyelvek esetében. Jelelő közösségekben ugyanis nagyon nagy a lexikai és a grammatikai változatosság, beszélt nyelvekhez képest kisebb a konszenzus a fonetikai és grammatikai jellemzők, csakúgy, mint az elfogadhatóság tekintetében, így kis számú adatközlő esetében nagyon félrevezető eredményeket kapunk, amennyiben csupán a jelelők intuíciói és önmegfigyelése alapján vonunk le következtetéseket (Johnston–Schembri 2013). Amennyiben a jelnyelvi megnyilatkozásokat valamilyen hangzó nyelvekre támaszkodó lejegyzési formában rögzítjük, keveset vagy szinte semmit nem tudunk meg magáról a jelnyelvi formáról, a vizsgálatok során az elsődleges nyelvi adat pedig hozzáférhetetlenné, visszafejthetetlenné válik. Fenlon és munkatársai (2015) szintén a korpuszalapú nyelvi vizsgálatok fontosságát hangsúlyozzák, miközben ők is kiemelik, hogy a kutatások során felül kell vizsgálni a korábbi, kevés adatközlőt bevonó vizsgálatok érvényességét, éppen a nagy változatosság miatt. A korpuszokat továbbá dokumentálásra, megőrzésre is használják, különös tekintettel arra, hogy egyre több jelnyelvet nyilvánítanak veszélyeztetetté.

A korpuszvizsgálatok az oktatás szempontjából is fontos információkat tárhatnak fel, a különböző nyelvfejlődési mérföldkövek meghatározása többek között a különböző tantervek és mérések kidolgozásában segíthet (Johnston–Schembri 2013), de a nyelvek feltárása, jobb megismerése az oktatási anyagok minőségét is emeli (Fenlon et al. 2015).

2.2. Nemzetközi projektek

Az elmúlt körülbelül egy évtizedben számos helyen indultak meg jelnyelvi korpuszépítési munkálatok, melyek közül fontos kiemelnünk a Trevor Johnston és Adam Schembri által vezetett, az ausztrál jelnyelv (Auslan) dokumentálását és vizsgálatát célzó kutatásokat, a holland jelnyelvvél (NGT) kapcsolatban a Radboud Egyetemen Onno Crasborn vezetésével illetve a Jelnyelvi Centrumban Trude Schermer vezetésével folyó munkálatokat, a Hamburgban zajló 15 éves korpuszépítési és szótárfejlesztési munkálatokat, valamint a brit jelnyelvhez (BSL) kapcsolódó változatos témákat felölelő korpuszkutatásokat.¹ A felsorolt központokhoz kötődő korpuszok jellemzőit (összehasonlítva a JelÉsély projekt keretében létrejött első Magyar jelnyelvi korpuszal) az 1. táblázat szemlélteti. A felsorolt nyelveken kívül számos más jelnyelv esetében is megkezdődtek a korpuszalapú vizsgálatok (pl. svéd jelnyelv, dán jelnyelv, lengyel jelnyelv stb. [lásd még Konrad 2012]).

1 A projektek honlapjai: Auslan Corpus: <http://www.auslan.org.au/about/corpus/>; Corpus NGT: <http://www.ru.nl/corpusngtuk/>; Nederlands Gebarententrum: <https://www.gebarententrum.nl/>; Corpus-DGS: <http://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/welcome.html>; BSL Corpus Project: <http://www.bsllcorpusproject.org/>

Módszertanukat tekintve nagyon változatos képpel találkozunk. Jellemzően fontos szempont a területi változatosság lefedése, illetve a női-férfi adatközlők aránya, illetve több korcsoport bevonása. Gyakran dolgoznak célzott elicitációs feladatokkal (pl. fogalmakat kell eljelelni), illetve más olyan stimulusokkal, amelyek eredményeként monológ jellegű jelnyelvhasználatot rögzítenek, melyeket általában valamilyen másik siket személlyel folytatott konverzáció egészít ki.

Nyelv	Idő-tartam	Célok	Adat-közlők	Adatmennyiség jelenleg	Adatgyűjtés
német jelnyelv (DGS)	2009–2023	Referenciakorpusz, korpuszalapú szótár	330, 13 régióból	1159 óra	15–20-féle elicitációs feladat (dialogus, monológ, elicitált lexikai elemek stb.)
brit jelnyelv (BSL)	2008–2011	BSL-korpusz Grammatikai és szó-kincset érintő kutatás, változás és változatosság kutatása	249, 8 helyszínen	~ 180 óra Személyenként 5 perc narratíva személyes élményről, 30 perc konverzáció, 15 perc interjú + jelelicitáció	Konverzáció másik siket személlyel, interjúkérdések, történetmesélés, 102 fogalom eljelelése
holland jelnyelv (NGT)	2006–2008	Kutatási forrás létrehozása	92, ország-szerte	71,5 óra Páronként ~ 1,5 óra	Auslan feladatai NGT-hez igazítva, monológok különféle stimulusok alapján (rajzfilm, mese, képregény, TV-klipek, továbbá félspontán konverzáció)
ausztrál jelnyelv (Auslan)	2003–2004 2004–2007	Szociolingvisztikai változatosság vizsgálata Dokumentáció, archívum, általános és kutatási célok	210, 5 helyszínen 100, 5 helyszínen	140 óra 300 óra	Interjú a névjelről és a jelelő háttéréről, szabad konverzáció, 2 narratíva. Interjúk, elicitált narratívák, csoportos konverzáció, elicitált lexikai elemek
magyar jelnyelv (HSL)	2013–2015 JelÉsély	A siket közösség nyelvhasználati mintázatának szociolingvisztikai leírása, a magyar jelnyelvi változatok vizsgálata. Bizonyítékalapú, digitális jelnyelvi szótár és valós nyelvhasználaton alapuló grammatikai leírások.	142, 7 régió, 9 helyszínen (szociolingvisztikai interjú) + további csoportos és interjúfelvételek	~ 2000 óra	Szociolingvisztikai interjúk, grammatikai tesztek, konferenciák és workshopok felvételei stb.

1. táblázat. A vezető nemzetközi korpusznyelvészeti központokhoz kapcsolódó jelnyelvi korpuszok és az első Magyar jelnyelvi korpusz – Összehasonlítás (Konrad 2012; Blanck et al. 2016; Blanck et al. 2010; BSLCP é. n.; Crasborn–Zwitserslood 2008; Schembri et al. 2012; Schembri–Johnston 2004)

3. A Magyar jelnyelvi korpusz

3.1. A korpusz elsődleges céljai

A korpusz létrehozásának legfőbb célja – azon „öncélú” indíttatásán túl, hogy létrejöhessen az első magyar jelnyelvi korpusz, amely dokumentálja a magyar jelnyelvet és autentikus szövegekhez biztosít hozzáférést – a magyarországi siket közösség nyelvhasználati mintázatainak szociolingvisztikai leírása, illetve a magyar jelnyelvi változatok vizsgálata volt. További célkitűzés volt egy korpuszalapú vizsgálatokra és eredményekre építő, bizonyítékalapú, digitális jelnyelvi szótár létrehozása, valamint a magyar jelnyelv émikus kategóriáinak grammatikai leírása.

3.2. A korpuszt alkotó anyagok

A korpusz legnagyobb részét a szociolingvisztikai interjúk felvételei adják, melyeket 7 mintavételi pont 9 helyszínén vettünk föl (Budapest, Szeged/Hódmezővásárhely, Békéscsaba, Debrecen, Kaposvár, Sopron/Győr, Vác). Összesen 147 interjú (67 budapesti és 80 vidéki) készült 16 siket terepmunkás bevonásával. Minden terepmunka alkalmával két terepmunkás volt jelen, miközben az egyik terepmunkás irányította az interjút, és folyamatos interakcióban volt az adatközlővel, a másik terepmunkás rögzítette a válaszokat. Ez azért volt fontos, hogy a jegyzetelés, illetve a kérdőív tartása ne akadályozza és/vagy befolyásolja a jelelést. A 345 kérdést tartalmazó interjú (melyek között voltak nyitottak és zártak is) átlagos hossza 3-4 óra volt, a felvételeket három kamerával készítettük el (egy az első számú terepmunkást, egy az adatközlőt, egy pedig a teljes interjúhelyzetet rögzítette). Kiemelten fontos volt, hogy a felvételek során hallók nem lehettek jelen.

A szociolingvisztikai interjúkat megelőzően készült egy online nyelvészociológiai kérdőív is, melynek fő célja a szociolingvisztikai interjúk előkészítése és véglegesítése volt. Ezt azért tartottuk szükségesnek, mivel korábban nem készült a szociolingvisztikai interjúhoz hasonló, átfogó felmérés a jelelő siket közösségben. Az online feltett kérdések a siketek és nagyothallók legfontosabb nyelvhasználati szokásaira, nyelvvel és nyelvhasználattal kapcsolatos jellemzőire, demográfiai adataira, magyar nyelvvel és magyar jelnyelvvvel kapcsolatos attitűdjeire vonatkoztak. Összesen 66 kérdésből állt a kérdőív, de attól függően, hogy valakinek, például, van-e gyermeke vagy hány gyermeke van, a kérdések száma változhatott. Az online felület teljes mértékben kétnyelvű volt, az instrukciók, a kérdések és az esetleges válaszlehetőségek is megjelentek magyar jelnyelven, csak úgy, mint magyar nyelven is. Összesen 238 teljes kitöltés érkezett.

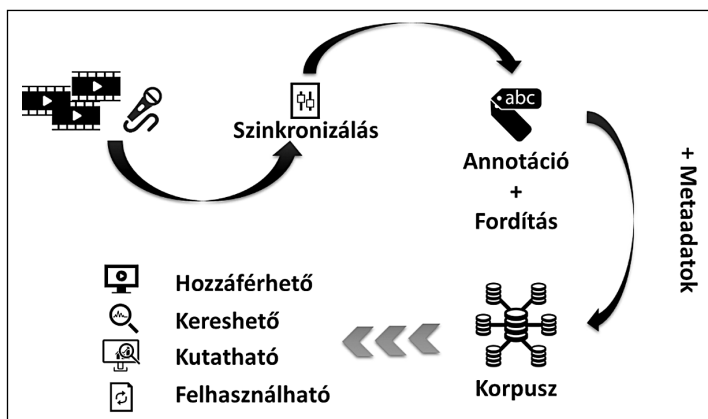
A grammatikai tesztek felvételeihez a magyar jelnyelv alapgrammatikájának megírásához szükséges elicitációs tesztsorok szolgáltatták az alapot. Ezek közül a terepmunkák közül összesen 27 készült (11 vidéki, 16 budapesti), interjúként 2-2 adatközlővel (tehát összesen 54-gyel). Olyan személyek vettek részt a grammatikai tesztekben, akik korábban már a szociolingvisztikai interjúkban is, így esetükben lehetőség nyílik a kétféle módszertannal nyert nyelvi adatok összevetésére. Ezek a felvételek már 5 kamerával készültek (a 2 adatközlőt egyenként 2 kamera vette, egy szemből-oldalról, egy pedig felülről, az ötödik kamera pedig a teljes teszthelyzetről készített átfogó felvételeket), átlagos hosszuk két óra volt.

A fent ismertetett anyagok kiegészülnek továbbá diákokkal, tanárokkal és szülőkkel készült interjúkkal, illetve workshopok, konferenciák, trainingek stb. felvételeivel is.



1. ábra. Részlet egy szociolingvisztikai interjúból

3.3. A korpusz feldolgozása

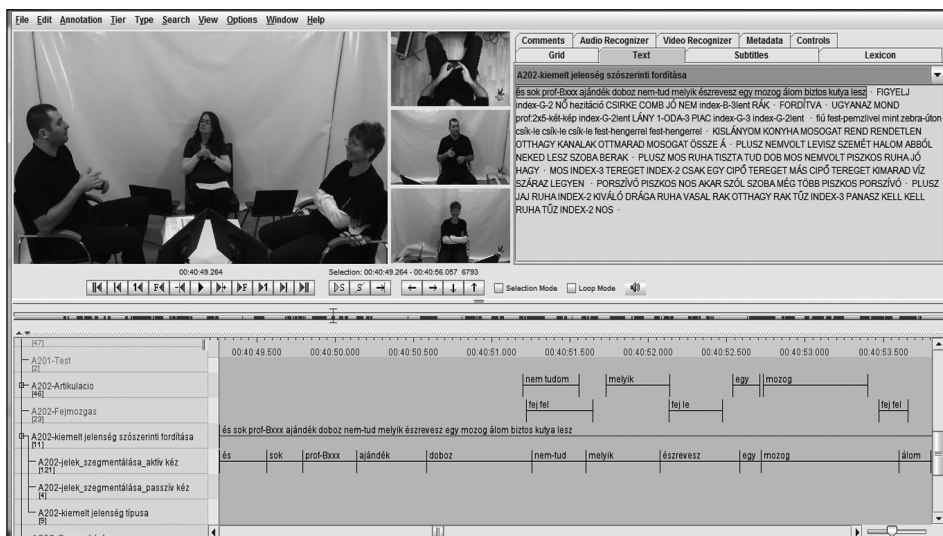


2. ábra. A jelnyelvi korpuszépítés vázlatos áttekintése²

Ahhoz, hogy a felvételekből egy olyan korpusz jöhessen létre, amely nemcsak videóarchívumnak tekinthető, hanem egy hozzáférhető, kereshető, kutatható és további célokra is felhasználható kutatási bázisnak is, számos lépésre van szükség. Az előkészítő munkálato-
kat (konvertálás, vízjelzés, rendszerezés stb.) követően a felvételek szinkronizálása szükséges, továbbá a fordítások és metaadatok hozzárendelésén túl további annotációs folyamatok. Annotációnak nevezünk „minden olyan információt és jelet, amelyet az eredeti szöveg [...] nem tartalmazott, de a korpusz készítésekor vagy feldolgozása során a szövegbe bekerült” (Szirmai 2005), a metaadatok pedig „adatok az adatról”, rendszerint a videófelvétel helyéhez és idejéhez, valamint az adatközlőhöz kapcsolódó információk (Crasborn–Hanke 2010). A munkafolyamatokat az ELAN (EUDICO Linguistic Annotator)

2 Az ábra a Flaticon oldalról származó képek felhasználásával készült. Designed by Freepic from Flaticon.

nevű programmal végeztük, mivel ez lehetővé teszi négy videó (és két hangsáv) egy időben való megjelenítését, az összekapcsolható/összeszinkronizálható fájlok száma azonban ennél magasabb is lehet. A program által biztosított legfontosabb funkció, hogy a fordítás, illetve további annotációk időben összekapcsolódnak a videóval, amely ezt követően kereshetővé válik. Akár több fájlban is kereshetünk egyszerre, a konkordancialisták elemeire kattintva pedig azonnal lejátszhatjuk a kapcsolódó videószegmenseket. A program jó átjárhatóságot biztosít, hiszen nemcsak szövegfájlok (pl. Microsoft Word), de akár Praat TextGrid is exportálható a számtalan lehetőség között. A konkordancialistákon túl az ELAN alapvető statisztikákhöz is hozzáférést biztosít, emellett egy kiegészítő programmal példamondatok és segédanyagok is exportálhatók.



3. ábra. Annotált grammatikai tesztelés részlete

A munkálatok során a legfőbb alapelv a jelnyelvközpontúság volt, hogy alulról felfelé építkezve, a jelnyelv belső, énikus kategóriáit megfigyelve dolgozzunk. Ennek érdekében, hogy a hangzó nyelvi hatásokat minél inkább ki tudjuk zárni, kizárólag siket terepmunkásokkal és annotátorokkal dolgoztunk, és ahol lehetett, vizuális annotációt használtunk.

Kihívást jelentett az írásbeliség hiányának áthidalása (bővebben l. a szintén ebben a kötetben szereplő *Jelnyelvi-élnyelvi szótárépítés a 21. században* című tanulmányt), továbbá különösen a magyar jelnyelv lehetséges manuális és nonmanuális összetevőinek az azonosítása. Ugyan korábban már készültek leírások a magyar jelnyelv fonológiai komponenseiről, ezek nem széles körű korpuszalapú vizsgálatok eredményeként születtek meg, így nem vehettük át őket felülvizsgálat nélkül, a külföldi notációs rendszerek elemei (különösen a kézformák esetében) pedig túl részletesnek és a gyakorlatban alkalmazhatatlannak bizonyultak a munkálatok során. Megkezdődött a nonmanuális komponensek körének azonosítása korábbi jelnyelvtudományi, illetve nem nyelvészeti, de kapcsolódó kutatások (pl. emócióelemzés és gesztuskutatás) alapján, illetve a kézformaállomány meghatározása percpációs tesztek és korpuszadatok segítségével.

A fordítások esetében jelnyelvalapú fordítást kértünk, amely követi a jelek egymásutánosságát, amely habár a formáról nem ad információt, de a kivitelezésről (sorrendiség) igen, emellett tartalmilag visszaadja a jelelést, így hozzáférést biztosítva a halló kutatók számára is.

3.4. A szótár és grammatika létrejöttének korpuszalapú támogatása

A szótári annotáció megalapozását elsőként öt régió 6-6 interjúján kezdtük el, melyben fiatal és idősebb generációs jelelők egyenlő mértékben voltak reprezentálva. A magyar megfelelően túl a jel típusa, a domináns kéz kézformája, a nem domináns kéz kézformája, a kivitelezési hely és a mozgás került annotálásra. Fontos volt, hogy elkerüljük a hangzó nyelvi hatásokat, ezért a magyar megfelelő megadásán kívül a jelnyelvi szinteken kizárólag vizuális alapon dolgoztunk (bővebben l. Holecz–Bartha 2017). A szótári annotáció részletét a 4. ábrán láthatjuk.

A grammatikai leírások megalapozásához 15 grammatikai terepmunkán felvételenként átlagosan 120 ún. „fontos interjúszakaszt” jelöltünk ki. A vizsgálatok első fázisában a kérdések és kérdőszavak, másodlagos predikátum, felszólítás, birtokos szerkezet, tagadás, előidejűség, utasítás, főnévi kifejezés, szórend, kontrasztív topik, kvantifikátorok stb. grammatikai kategóriáira koncentráltunk, melyhez a szó szerinti fordításon és a kiemelt jelenség típusán túl az artikuláció, fejmozgás, aktív kéz, passzív kéz, orca, szem, szemöldök, száj, tekintet és testszintek kerültek annotálásra.

Szám	Típus1: magyar megfelelő	Típus2: jel típusa	Típus3: Kézforma	Típus4: Kézforma	Típus5: Kivitelezés	Típus6: mozgás ir.
1	kérdés	○	✋		👤	—
2	igen	○	✋		👤	—
3	születési	○/p	✋	✋	👤	↷
4	év	○	✋		👤	—
5	férfi	○	✋		👤	—
6	legmagasabb	△○	✋	✋	👤	—
7	iskolai	○	✋		👤	↷
8	végzettségű	△○	✋	✋	👤	—
9	szakközépiskola	△○	✋	✋	👤	↷
10	gimnázium	○	✋		👤	↷
11	érettségi	○/p	✋	✋	👤	—

4. ábra. A szótári annotáció részlete

4. Összefoglalás

Láthatjuk, hogy habár korábban nem állt rendelkezésre megfelelő technikai háttér és módszertan modern jelnyelvi korpuszok létrehozására, az elmúlt körülbelül egy évtized-

ben több helyen indultak meg jelnyelvi korpuszépítési munkálatok, melyek számos lehetőséget biztosítanak a nyelvek megőrzésén túl a nyelvészeti kutatások számára is.

A JelEsély projekt során egy nemzetközi viszonylatban is jelentős méretű, módszertanát tekintve előremutató jelnyelvi korpusz jött létre. Nemcsak az anyaggyűjtés módszertana építkezik szociolingvisztikai alapokra, de a projekt további kimeneteinek munkálatai (online szótár és grammatika) is erre a szemléletre épülnek. A már jelenleg is több mint 2000 órányi felvétel feldolgozása és a kutatások körének szélesítése folyamatos, a már említett manuális és nonmanuális komponensek vizsgálatán túl további nyelvhasználati mintázatok feltárása zajlik (pl. az artikulációt és a jelnyelvi–hangzó nyelvi kétnyelvűséget illetően). A következő szakaszban egy ID-Glossokon (egy jelforma egyedi azonosítója [l. Johnston 2008]) alapuló lemmatizációs fázis veszi kezdetét.

Köszönetnyilvánítás

A tanulmányban leírtak nem valósulhattak volna meg a JelEsély projekt (TÁMOP-5.4.6/B-13/1-2013-0001, kutatásvezető: Bartha Csilla), valamint az MTA NYTI Többsnyelvűségi Kutatóközpont NyelvEsély Szakmódszertani Kutatócsoportjában zajló *Nyelvileg tudatos iskola, kétnyelvű sikertoktatás és a nyelv által kiaknázható tudás innovatív módszereinek, eszközeinek fejlesztése* című projekt támogatása nélkül (SZ-007/2016, kutatásvezető: Bartha Csilla). Köszönetet mondunk a JelEsély projekt megvalósítóinak, valamennyi siket és halló munkatársnak.

Hivatkozások

- Adolphis, Svenja 2013. Corpora: Multimodal. In: Chapelle, Carol A. (ed.): *The Encyclopedia of Applied Linguistics*. Wiley-Blackwell, Chichester, West Sussex. 1188–1190.
- Baker, Paul 2010. *Sociolinguistics and Corpus Linguistics*. Edinburgh University Press, Edinburgh.
- Blanck, Dolly–Bleicken, Julian–Hanke, Thomas–Hanß, Andreas–Hofmann, Ilona–Jeziorski, Olga–König, Lutz–König, Susanne–Konrad, Reiner–Langer, Gabriele–Rathmann, Christian–Salden, Uta–Wagner, Sven 2016. *DGS-Korpus – Sprachressourcen für Deutsche Gebärdensprache. Poster präsentiert auf dem CLARIN-Forum CA3 2016 in Hamburg vom 07–08.06.2016*. http://www.sign-lang.uni-hamburg.de/dgs-korpus/files/inhalt_pdf/2016_CLARIN_FORUM_CA3_Sprachressourcen_Poster.pdf (2017. 02. 25.)
- Blanck, Dolly–Hanke, Thomas–Hofmann, Ilona–Hong, Sung-Eun–Jeziorski, Olga–Kleyboldt, Thimo–König, Lutz–König, Susanne–Konrad, Reiner–Langer, Gabriele–Nishio, Rie–Rathmann, Christian–Vorwerk, Stephanie–Wagner, Sven 2010. *The DGS Corpus Project. Development of a Corpus Based Electronic Dictionary German Sign Language – German*. Poster presented at the Theoretical Issues in Sign Language Research (TISLR) 10 Conference, Sept 30 – Oct 2, 2010 at Purdue University, Indiana, USA. http://www.sign-lang.uni-hamburg.de/dgs-korpus/files/inhalt_pdf/DGS_Corpus_Projekt.pdf (2017. 02. 25.)

- BSLCP é. n. BSLCP information. <http://www.bslcorpusproject.org/project-information/> (2017. 02. 25.)
- Clark, Lynn–MacGougan, Helen–Hay, Jennifer–Walsh, Liam 2016. „Kia Ora. This is My Earthquake Story”. Multiple Applications of a Sociolinguistic Corpus. *Amper sand* 3: 13–20.
- Crasborn, Onno–Hanke, Thomas 2010. *Metadata for Sign Language Corpora*. Background document for an ECHO workshop May 8 + 9, 2003, Radboud University Nijmegen. http://sign-lang.ruhosting.nl/echo/docs/ECHO_Metadata_SL.pdf (2017. 02. 25.)
- Crasborn, Onno–Zwitserslood, Inge 2008. The Corpus NGT: An Online Corpus for Professionals and Laymen. In: Crasborn, Onno–Hanke, Thomas–Efthimiou, Eleni–Zwitserslood, Inge–Thoutenhooft, Ernst D. (eds.): *Construction and Exploitation of Sign Language Corpora*. ELDA, Paris. 44–49.
- Fenlon, Jordan–Schembri, Adam–Johnston, Trevor–Cormier, Kearsy 2015. Documentary and Corpus Approaches to Sign Language Research. In: Orfanidou, Eleni–Woll, Benjie–Morgan, Gary (eds.): *Research Methods in Sign Language Studies. A Practical Guide*. Wiley Blackwell, Chichester, West Sussex. 156–172.
- Holecz Margit–Bartha Csilla 2017. Vizuális fonológia: Jelnyelvek fonológiája és a vizuális annotáció támogatása. *Beszédkutatás* 25: 205–221.
- Johnston, Trevor 2008. Corpus Linguistics and Signed Languages: No Lemmata, no Corpus. In: Crasborn, Onno–Hanke, Thomas–Thoutenhoofd, Ernst D.–Zwitserslood, Inge–Efthimiou, Eleni (eds.): *Construction and Exploitation of Sign Language Corpora. Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages*. ELRA, Paris. 82–87.
- Johnston, Trevor–Schembri, Adam 2013. Corpus Analysis of Sign Languages. In: Chapel, Carol A. (ed.): *The Encyclopedia of Applied Linguistics*. Wiley–Blackwell, Chichester, West Sussex. 1312–1319.
- Jones, Christian–Waller, Daniel 2015. *Corpus Linguistics for Grammar. A Guide for Research*. E-book változat. Routledge, New York.
- Konrad, Reiner 2012. *Sign Language Corpora Survey*. http://www.sign-lang.uni-hamburg.de/dgs-korpus/files/inhalt_pdf/SL-Corpora-Survey_update_2012.pdf (2017. 02. 25.)
- Leech, Geoffrey 1992. Corpora and Theories of Linguistic Performance. In: Svartvik, Jan (ed.): *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82, Stockholm, 4–8 August 1991*. Mouton de Gruyter, Berlin, New York. 105–122.
- McEnery, Tony–Wilson, Andrew 2001. *Corpus Linguistics. An Introduction*. Edinburgh University Press, Edinburgh.
- Schembri, Adam–Fenlon, Jordan–Rentelis, Ramas–Reynolds, Sally–Cormier, Kearsy 2012. Towards a British Sign Language Corpus: A Short Report. *Japanese Journal of Sign Linguistics* 21: 5–14.
- Schembri, Adam–Johnston, Trevor. 2004: Sociolinguistic Variation in Auslan (Australian Sign Language). A Research Project in Progress. *Deaf Worlds* 20/1: 78–90.
- Szirmai Mónika 2005. *Bevezetés a korpusznyelvészetbe. A korpusznyelvészet alkalmazása az anyanyelv és az idegen nyelv tanulásában és tanításában*. Tinta Könyvkiadó, Budapest.
- Tognini-Bonelli, Elena 2001. *Corpus Linguistics at Work*. John Benjamins Publishing Company, Amsterdam, Philadelphia.

Summary

Corpus linguistics can be used in all areas of linguistics from phonetics through syntax to sociolinguistics, it provides objective and reliable research results even in such cases where other methods (e.g., self-observation) do not work. Though early sign language data were already collected at the beginning of the twentieth century, the modern sign language corpus linguistic studies have started only in the 2000s, the main reasons being the lack of suitable technical background and the difficulties due to the lack of literacy. However, sign language corpus building projects have started in the past decade at several places, and the first modern sign language corpora have been created.

The main goals of creating the first Hungarian Sign Language Corpus were the documentation of Hungarian sign language (HSL), providing access to authentic texts, sociolinguistic description of the language use patterns of the deaf community in Hungary, examination of the HSL variations, the creation of an evidence-based digital dictionary building on corpus-based examinations and results, and the grammatical description of the emic categories of HSL. During the SIGNificant Chance project, a sign language corpus was created with approximately 2000 hours of video recordings, the main principle being the focus on sign language during processing.