

***Baumann Tímea – Majoros Judit – Pelcz Katalin – Schmidt Ildikó –
Szita Szilvia – Vermeki Boglárka***

**Bemutatkozik a Korpusznyelvészeti és Szakmódszertani
Munkacsoport**

A 2020 februárjában alakult Korpusznyelvészeti és Szakmódszertani Munkacsoport a következő célokat tűzte maga elé: a munkájával hidat kíván építeni a korpusznyelvészeti eredmények és a MID szakmódszertana közé. E cél elérése érdekében almunkacsoportokat hoztunk létre. A csoportok az elmúlt hónapok alatt jelentős mennyiségű anyagot gyűjtöttek és készítettek. Elindult az autentikus és félautentikus pedagógiai korpuszok, illetve a nyelvtanulók produktumait rendszerező tanulói korpusz építése, valamint a nagy nyelvi korpuszok alapján az újabb tananyagok létrehozása is. Terveink szerint egyrészt olyan adatbázisokat hozunk létre, amelyeket a kutatók és a gyakorló nyelvtanárok is használhatnak munkájukban, másrészt olyan konkrét tananyagokat készítünk, amelyek a korpuszhasználat többletértékét a gyakorlatban is láttatják.

1. Bevezetés

A Korpusznyelvészeti és Szakmódszertani Munkacsoport (KorSzak) 2020 februárjában jött létre azzal a céllal, hogy hidat építsen a korpusznyelvészeti kutatások és a magyar mint idegen nyelv tanításának szakmódszertana között. A csoport létrehozását alapvetően az a felismerés motiválta, hogy az empirikus nyelvészeti kutatások eredményei és a MID-módszertan között igen kevés az átjárás, holott a nagy nyelvi adatbázisok elemzése a szakmódszertan megújítására is rendkívül kedvező hatással lehet.

A nyelv korpuszalapú feltárásának elsődleges előnye, hogy a valóban használt nyelvről ad képet. A korpuszelemző szoftverek statisztikai mutatói felvilágosítással szolgálnak arról, mely elemeket milyen gyakorisággal használjuk, és jellemzően milyen szöveggörnyezetben fordulnak elő (ld. pl. Hanks 2013, Hoey 2005, Taylor 2009). A vizsgálatok irányulhatnak a grammatika egyes aspektusainak feltérképezésére (szórend, két ragozás, ragok használata, igekötők stb.), vagy a lexika egyes területeire (kollokációk, gyakran használt nyelvi elemek és szöveggörnyezetük), de ugyanúgy pragmatikai kérdésekre vagy regiszterek jellemzőinek, egy-egy szövegtípus nyelvi eszközeinek kikutatására is.

Ezeknek a jelenségeknek a feltérképezése és tananyagokba való beépítése igen nagy változást hozhat a MID-tanítás módszertanába. Arra, hogy a korpuszvizsgálatok eredményeit így módon felhasználhatóvá tegyük, most történik először nagyobb szabású, közösségi kísérlet. Módszertani megközelítésünk egyik alapköve, hogy amennyiben a korpuszok a magas szintű nyelvi kompetenciával rendelkező beszélők (expert users) nyelvi tapasztalatát rögzítik (Hunston 2009, Hoey 2005), akkor az ott kiemelt gyakorisággal előforduló elemekkel a tanuló az órán kívül szintén kiemelt gyakorisággal fog találkozni, ezeknek az elemeknek tehát a nyelvtanításban hangsúlyos szerepet kell kapniuk (O’Keeffe et al. 2007).

Amellett, hogy a természetes nyelvhasználat egyes sajátosságait már meglévő korpuszok alapján kívánjuk vizsgálni, a korpuszépítés is céljaink között szerepel. Ebből fakadóan jött létre a gyermeknyelvi, az élőnyelvi és a tanulói korpuszt építő három munkacsoport. Itt elsősorban pedagógiai korpuszokról van szó, tehát olyan adatbázisokról, amelyek kialakításakor a tanulók igényeit, nyelvtudási szintjét tartjuk szem előtt (ezekről lásd még ugyanebben a kötetben Szita). S mivel a KorSzak munkacsoport alapvető célkitűzése a

magyar mint idegen nyelv szakmódszertanának fejlesztése, a korpuszokból nyert információkat természetesen újabb tananyagokba is be kívánjuk építeni (ld. alább korpuszalapú füzetek).

A csoport feladatának érzi a korpusznyelvészet és a szakmódszertan összekapcsolásáról és a korpusznyelvészet legújabb eredményeiről való ismeretszerzést, rendszeres beszámolót és gondolatcserét is, s ennek megvalósítására rendszeresen sort is kerít.

2. Adatok a KorSzak munkacsoportról

A munkacsoport 2020 februárjában tartotta alakuló ülését a Pécsi Tudományegyetem Nemzetközi Oktatási Központjában. A második ülésre pedig 2020. november 28-án került sor az online térben. Az alakuló ülésen 29 tag vett részt, a tagok száma azóta 43-ra emelkedett. A tagok többsége egyetemi alkalmazott – kutató, tanár –, de vannak a munkacsoportban egyetemisták, egyetemhez nem kötődő kutatók, gyakorló nyelvtanárok is. A rendes tagok zömében magyar vagy külföldi egyetemeken (ELTE, SZTE, PTE, KRE, Corvinus Egyetem, Szófiai Szent Kelemen Egyetem, Varsói Tudományegyetem, Zágrábi Tudományegyetem), illetve más, magyar nyelvet tanító intézményekben, nyelviskolákban tevékenykednek, és szép számmal akadnak olyanok is, akik a lektori hálózat kiküldöttjeként dolgoznak külföldi egyetemeken.

Az eddig létrejött almunkacsoportok az alábbi témák köré szerveződtek:

- (1) Gyermeknyelvi korpusz. Csoportvezető: Vermekei Boglárka
- (2) Tanulói korpusz. Csoportvezető: Baumann Tímea
- (3) Élőnyelvi korpusz. Csoportvezető: Pelcz Katalin
- (4) Korpuszalapú füzetek. Csoportvezető: Szita Szilvia

Az alábbiakban az egyes munkacsoportok feladatkörét, eddigi tevékenységét és terveit ismertetjük részletesebben.

2.1. Gyermeknyelvi munkacsoport

A KorSzak gyermeknyelvi korpuszt építő munkacsoportjának célja olyan pedagógiai korpuszok építése, amelyek különböző, a gyermekek számára is releváns és érdekes témákat dolgoznak fel. Emellett ezek nemcsak korpusznyelvészeti kutatások számára, hanem egy részük tananyagok formájában a magyart idegen vagy származásnyelvként tanuló diákok és tanáraik részére is elérhetővé és hasznossá válhatnak.

Az eddigi gyermeknyelvre irányuló korpusznyelvészeti kutatások elsősorban az első nyelv elsajátítási folyamatának megértése és tanulmányozása céljából készültek. Számos ilyen jellegű kutatás korpusza található a Sketch Engine-en, illetve a TalkBank rendszer CHILDES (*Child Language Data Exchange System*) elnevezésű részén, ahol különböző nyelvek írott és beszélt nyelvi adatai is elérhetőek már. A magyar nyelv vonatkozásában készültek felvételek egynyelvű (MacWhinney 1970, Réger 1990, Bodor 1994) és többnyelvű (Navracsics 1994–1997, Bíró 2013) gyermekekkel is. A legnagyobb magyar gyermeknyelvi korpuszok azonban – mint a 62 óvodás bevonásával készült MONYEK, vagyis a Magyar Óvodai Nyelvi Korpusz (vö. Mátyus–Orosz 2014), a 3–18 évesek felvételeivel folyamatosan épülő GABI (vö. Bóna 2017) vagy a 16–18 évesekről készült felvételeket tartalmazó Beszélt nyelvi Adatbázis Tini Bea része (vö. Gyarmathy–Neuberger 2015) – sajnos nem érhetőek el ezeken a felületeken. Az említett magyar nyelvű korpuszok alapján készült kutatások szintén elsősorban az első nyelv elsajátításának fázisait vizsgálják.

A 2020 februárjában létrejött gyermeknyelvi korpuszt építő munkacsoport két korpusz fejlesztését kezdte el, melyek közül a beszélt nyelvi korpusz építése már szinte azonnal a megalakulás után elindult, az írott nyelvi korpusz építése pedig a kezdeti szakaszban tart.

A beszélt nyelvi korpusz adatgyűjtésének megtervezésekor az általánosiskolás-korú gyermekeket több korcsoportra osztottuk (6–7, 8–9, 10–11, 12–13, 14–15 évesek), valamint egy felső tagozatos tanulókkal készült felmérés során összegyűjtöttük, majd kiválasztottuk azokat a témaköröket, amelyekről a felvételek során az adatközlőink beszélgetnek. Elsőként a 12–13 éves korosztály tagjaival rögzítettünk videó- és hangfelvételeket. Az adatközlő gyermekek a kiválasztott tizenkét témakörrel – *állatok, influenzaszerek, Harry Potter, kirándulás, kreatív hobbi, kutyák, technológia, természet, saját történet / régi emlék, sorozatok, sport, utazás* – készített gondolattérképek alapján beszélgetnek egymással.



1. ábra A Harry Potter témakör gondolattérképe

Egyelőre tíz témakörben rögzítettünk felvételeket, melyekről a rendszerezés után, az Alrite beszédfelismerő program segítségével szövegátiratokat is készítettünk. A videók időtartalmát nem határoztuk meg, a gyerekek addig beszélgethetnek egymással, amíg jónak látják. Így az elkészült felvételek hossza öt és negyvenkét perc között változik. Jelenleg 26 videó és hangfelvétel készült el, amely összesen megközelítőleg 450 percnyi anyagot és 46 ezer szót jelent.

A munkacsoport tervei között szerepel a 12–13 évesek beszéltnyelvi korpuszának folyamatos bővítése, témakörönként legalább tíz videó- és/vagy hangfelvétel rögzítése és átírata. Később szeretnénk a fent felsorolt többi korosztályra is kiterjeszteni a adatgyűjtést.

Az írott nyelvi korpusz építése terveink szerint kétféleképpen fog zajlani. A korosztályok közül elsőként ennél a korpusznál is a 12–13 éveseket választottuk, akik a már fent említett témakörökben fognak adatot szolgáltatni számunkra. Elsőként a fogalmazások összegyűjtésének szervezését kezdtük meg, de tervezzük a kötetlenebb stílusú, a közösségi oldalak beszédmódját reprezentáló megnyilatkozások gyűjtését is.

A közeljövőben a munkacsoport tagjai a már összegyűjtött anyagokon alapuló korpusznyelvészeti kutatásokon, például a pragmatika-diskurzusszervezés vizsgálatán gondolkoznak, valamint a beszélt és írott nyelvi korpuszok egy gyermekek számára készülő tananyag alapjaként is szolgálnak majd. A tananyag öt különböző nyelvtudási szinten öt korosztálynak készül, az életkori sajátosságokat is figyelembe véve. Elsősorban olyan gyermekek részére, akik a magyart származásnyelvként tanulják, de elképzeléseink szerint a magyar mint idegen nyelvi órákon, valamint vegyes nyelvtudási szinteken lévő csoportokban is alkalmazható lesz. A mintamodul hamarosan mindenki számára elérhetővé válik.

A KorSzak gyermeknyelvi pedagógiai korpuszainak különlegessége, hogy az írott nyelvi és a beszélt nyelvi adatgyűjtés ugyanazokban a témakörökben készül, így lehetőség nyílik a kettő összehasonlító vizsgálatára is. A témaköröket a gyermekekkel készült felmérés alapján

választottuk ki, így az számukra is érdekes és releváns. A kötetlenebb stílusok (beszélgetés barátokkal, fórum hozzászólás) miatt a valós, mai nyelvhasználathoz közel álló megnyilatkozásokat tudunk rögzíteni. Korpuszainkat a későbbiek folyamán a Sketch Engine felületén mindenki számára elérhetővé szeretnénk tenni.

2.2. MID Nyelvtanulói Korpusz

A MID Nyelvtanulói Korpusz a magyart idegen nyelvként tanulók által létrehozott szövegek és szóbeli megnyilatkozások (elektronikus adatbázisként tárolt) gyűjteménye. A korpuszt 2020 februárja óta építik a KorSzak Nyelvtanulói Korpusz Munkacsoport tagjai. A nyelvtanulói korpusz két gyűjteményrészről áll:

- (1) a nyelvtanulók írásbeli szövegalkotását rögzítő korpuszból
- (2) és a tanulók szóbeli beszédproduktumait rögzítő korpuszból.

Az első gyűjtemény a magyart idegen nyelvként tanulók írott szövegeit (esszéként beadott házi feladatait) tartalmazza. A jelenleg rendelkezésünkre álló szövegek részben kézírással készült szövegek, melyek eredeti változatát archiváljuk, és a begépelte szövegváltozatot bocsájítjuk rendelkezésre a kutatók számára. A gépelt szövegek tartalmi szempontból hűen tükrözik a diákok által eredetileg elkészített esszéket, és részben a formai sajátosságokat is visszaadják (írásirány, a szöveg tagolása, írásjelek, nagybetűs és kisbetűs írásmódok stb.). Az archivált anyag ezentúl további formai jegyek vizsgálatára is lehetőséget biztosít.

Egy példa a korpuszba felvitt esszészövegekre:

“A vietnami iskolában ugyanazokat a tantárgyakat tanuljuk, mint Magyarországon. Nálunk van háztartási ismeretek, etika és informatika is. Háztartási ismeretekórán házimunkák alapjáról tanultunk. Etikaórán valódi helyzetekről tanultunk. Informatikaórán számítógépes szoftverekkel tanultunk. A kedvenc tantárgyam az angol volt, mert érdekes volt az óra, a tanárom érdekesen magyarázott, és értettem az anyagot. A fizikatanárom szigorú volt, mégis imádtam, mert élveztem az órát. Vietnamban legjobb érdemjegy a tízes, a legrosszabb pedig a nullás. Nálunk reggel hétkor kezdődött a tanítás, és fél tizenkettőkor voltunk vége a tanításnak. Tapasztalataim szerint, ha jól akar tanulni egy nyelvet, akkor ahhoz a nyelvhez szenvedélyre van szüksége, és sok időt kell rá költeni, például zenét hallgatni, filmeket nézni és könyveket olvasni. Amikor a magyar szavakat tanulom, nekem segít az, ha kontextusban tanulom a szavakat. Nálam is jól működik az, ha sokszor leírom, kimondom hangosan, és megpróbálom használni a szót. Mostanában vacsora után vagyok a legfrissebb, akkor tudok koncentrálni.”

(A szöveg kódja: PTENOK_PREP_2019_2020_HF_A2_5F_VN_vietnami_F_19_1_26)

A korpusz második gyűjteménye beszélt nyelvi adatokat kíván gyűjteni a nyelvtanulóktól. A beszélt nyelvi adatok gyűjtését a Pécsi Tudományegyetem Nemzetközi Oktatási Központjában már évek óta végezzük olyan a formában, hogy az évközi és évvégi szintzáró vizsgákról – a hallgatók írásbeli jóváhagyásával – videófelvételeket készítünk. Ezeket a felvételeket a nyelvtanulók egy-egy szóbeli tételként megadott témáról beszélnek, illetve a *MagyarOK* tankönyv egy választott szövegét olvassák fel. Ez utóbbi feladatot a kiejtés mérésére és értékelésére vezettük be, és az így gyűjtött anyag jól alkalmazható a kiejtéssel kapcsolatos vizsgálatokra. A 2020 márciusában bevezetett digitális oktatás számunkra további lehetőségeket biztosított a gyűjtemény gazdagítására, hiszen így teljes tanórákat is felvehettünk. Ezekből a hosszabb felvételekből a tanulói megnyilatkozásokban gazdag részek leiratait fogjuk a korpusz részévé tenni. Így létrejöhét egy olyan egyedülálló anyag, amely vizsgálhatóvá teszi a magyart idegen nyelvként tanulók élőbeszédét különböző nyelvtudási szinteken. Egy ily módon kialakított korpusz kiváló lehetőséget biztosít majd a különböző nyelvészeti, nyelvtanulás-elméleti és -módszertani kutatásokhoz.

Példa egy szóbeli vizsga leiratára B1 szinten (részlet):

Tanár: *Kapcsolatok a valós és virtuális térben. Na, hát ez most szerintem egy nagyon aktuális téma. Hogyan tartod a kapcsolatod a barátaiddal és a családdal most?*

Diák: *Tartom a kapcsolatot a családommal és a barátaimmal a közösségi oldalakon, mint a facebookon, vagy Instagramon, és a közeli (...) a közeli barátaimnak itt, Magyarországon szintén minden nap tanulunk.*

Tanár: *Igen, sokat tanultok. Mondtad, hogy használod a közösségi oldalakat, és többet is felsoroltál. Melyiket mire használod egészen pontosan? Milyen funkciókat használsz ezeknél az oldalaknál?*

Diák: *Sokat használom a közösségi oldalakat, amik az egyik részévé vált. Szinte tartom a kapcsolatot mindenkivel, a közösségi oldalakon keresztül, ezen kívül lelki tornára, e-mailre is. De szoktam használni a közösségi oldalakat a szórakozásra, és megnézem, amely történik a barátaimmal. De sajnos én még pókok vagyok, de nem használom ezeket arra, hogy új és új emberekkel ismerkedjek, vagy kutatok eléggé barátok után. Mert úgy gondolom, hogy azok a kapcsolatok nem igazak a valóság. (...)*

Példa egy óra részletének leiratára B1 szinten:

Tanár: *Ti tudtok palacsintát sütni? Szereted a palacsintát, [név]?*

Diák1: *Nem tudom.*

Tanár: *Ki az, aki tud palacsintát sütni?*

Diák2: *Nagyon szeretek, de nem tudok.*

Diák3: *Aha. Nehéz palacsintát készíteni?*

Tanár: *Nem nehéz. Aha, nem tudod, hogy hogyan lehet sütni. Igen, kell hozzá néhány speciális eszköz, például nem mindegy, hogy milyen a serpenyő. Ha rossz a serpenyő, akkor nem lehet jó palacsintát sütni, mert akkor nem tudod kivenni a serpenyőből, odaragad a serpenyőhöz például. Tehát van néhány fontos lépés, amit be kell tartani. [Név] te szoktál palacsintát készíteni?*

Diák4: *Nem szoktam palacsintát készíteni, de ittem, ittem, és nagyon finom. Csokoládé csokoládéval.*

Tanár: *Aha, csokoládéval. A szír konyhában nincs palacsinta? A szírek nem készítenek palacsintát? [Név], a szírek nem készítenek palacsintát?*

Diák4: *Igen.*

Tanár: *És az más egy kicsit, mint a magyar palacsinta? Más?*

Diák4: *Ugyanaz. Ugyanaz.*

Tanár: *Ugyanaz. Oké. A vietnámi konyhában van palacsinta?*

Diák5: *Igen.*

Tanár: *Igen? És miben különbözik a magyar palacsintától? Más, mint a magyar palacsinta? A vietnámi palacsinta?*

Diák5: *Nem. ettél, nem ettél még.*

Tanár: *Melyiket nem ettél? Magyar palacsintát nem ettél, vagy vietnámit?*

Diák5: *Mind, mind, mind, mind.*

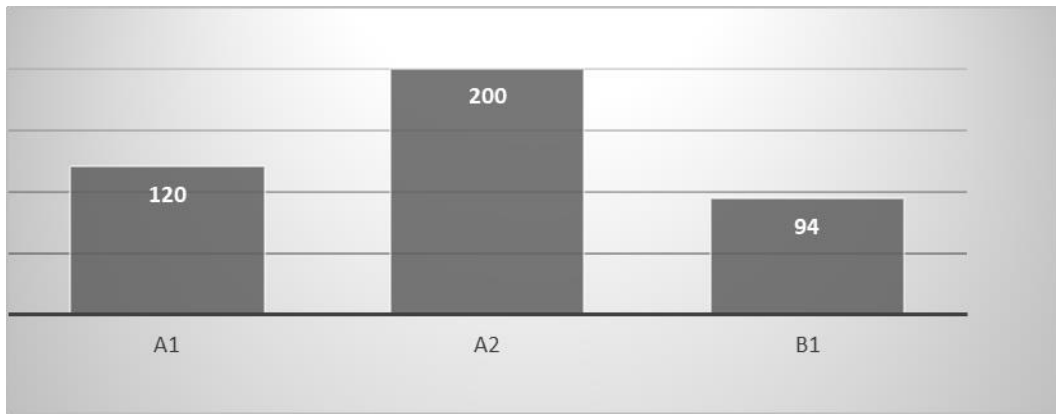
Tanár: *És [név], Mongóliában készítenek palacsintát?*

Diák6: *Igen.*

Tanár: *Igen. És az olyan, mint a magyar, vagy más?*

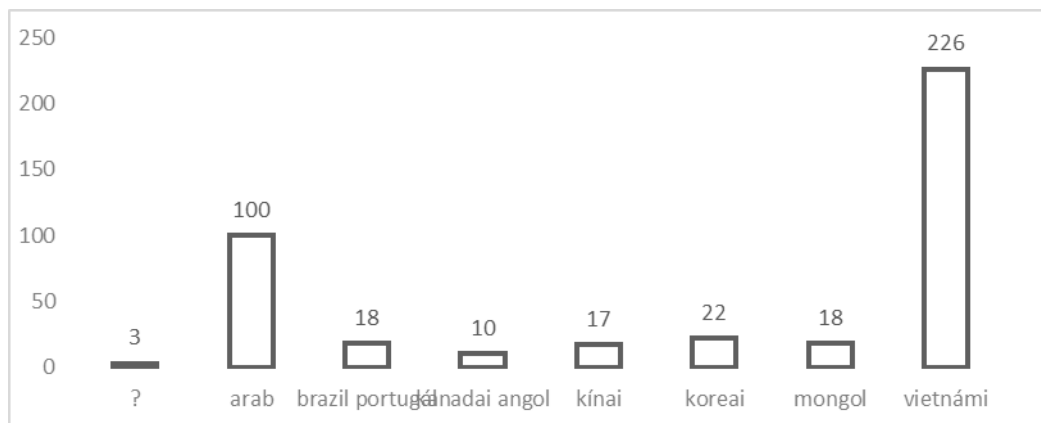
Diák6: *Szerintem kicsit olyan, mint magyar.*

A fenti példák még feldolgozás alatt álló leiratok, annotálásuk, kódolásuk is folyamatban van, így természetesen csak egy kis bepillantást kívánnak nyújtani a folyamatosan alakuló, bővülő korpuszba. A nyelvtanulói korpusz jelenlegi 414 írott szöveggel rendelkezik, A1 szinttől B1 szintig.



2. ábra Az eddig felgyűjtött írott korpusz nyelvtudási szintek eloszlása szerint

A szövegek különböző nemzetiségű nyelvtanulóktól származnak, amint az alábbi diagram mutatja.



3. ábra A korpuszban található anyagok eloszlása a diákok nemzetisége alapján

A beszélt nyelvi adatok gyűjtése és feldolgozása egy későbbi munkafázisban fog megvalósulni, ám már jelenleg is rendelkezünk 25 vizsga leiratával B1 és A1 szinten.

Célunk egy jól kereshető adatbázis létrehozása. A szövegeket megfelelő kódrendszerrel ellátva visszük fel az adatbázisba, így a kutatók számára rendelkezésre fognak állni adatok arról, hogy a szöveg mely oktatói intézményből, milyen kurzusról, melyik évből származik, továbbá arról, hogy a hallgató milyen tananyagból tanult, mi volt az írott esszé témája. Ezenkívül a névtelenül szereplő adatközlők állampolgárságát, első nyelvét, nemét és életkorát is vizsgálhatjuk. A szövegek hibák szerinti annotációja egy következő munkafolyamat része lesz, amelynek során a szövegeket olyan jelölőkkel látjuk el, melyek lehetővé teszik a hibatípusok szerinti keresést.

A MID Nyelvtanulói Korpusz a nyelvészek, hungarológusok, magyar mint idegen nyelv tanításával foglalkozó szakemberek számára gazdag forrásanyag lesz. Nyelvészeti elemzések és az idegennyelv-tanítás módszertanával kapcsolatos elemzések alapjául szolgálhat, a nyelvi elemzéseken keresztül új tudományos eredményekhez juthatunk, ugyanakkor jól alkalmazhatjuk a nyelvtanításban is: feltérképezhetjük vele az egyes nyelvtudási szinteken jellemző nyelvi hibákat és kihívásokat, támpontot adhat korpusz alapú tananyagok kidolgozásához, valamint a nyelvtanítás módszertanát érintő következtetéseket vonhatunk le belőle.

2.3. Az élőnyelvi korpuszt építő munkacsoport

A munkacsoport célja egy olyan élőnyelvi korpusz közreadása, amely segíti a nyelvtanulási és nyelvtanítási folyamatokat, vagyis pedagógiai korpuszokról van szó. A munkacsoport öt alkorpusz létrehozásán dolgozik, amelyek öt különböző szövegtípus produktumaiba engednek betekintést: cset, e-mail, mindennapi beszélgetések, interjúk, prezentációk (TedTalk). Valamennyi korpusz közös jellemzője, hogy jól körülhatárolható interakciókhoz, könnyen definiálható szituációkhoz kötöttek.

A két írott nyelvi alkorpusz – a cset- és az e-mail korpusz – autentikus nyelvi megnyilvánulások gyűjteménye. A csetkorpusz magánbeszélgetéseket tartalmaz, és a rövid üzenetváltások formai és tartalmi sajátosságait hivatott feltérképezni, különös tekintettel az ismétlődő elemekre, amelyeket – gyakoriságuk okán – érdemes a tanításba, tananyagokba is beépíteni. A csetkorpusz annotálása érdekes kérdéseket vet fel, hiszen például a vizuális elemeket (emojikat) is körültekintően kell kódolni.

Az e-mail korpusz a mai mindennapi kommunikáció egy másik lényeges területét rögzíti. Alapvetően hosszabb, egy vagy több levelezőpartnerrel folytatott levélváltások rögzítése a cél, mert ezekben lehet leginkább azonosítani a regiszter- és témaspecifikus nyelvi elemeket. A gyűjteményben hivatalos, félhivatalos és személyes e-mailek is szerepelni fognak, amelyek között a megfelelő annotálás fogja megkönnyíteni a tájékozódást.

A másik három alkorpusz élőbeszédéről készült felvételeket tartalmaz. Az autentikus beszédkorpusz olyan konkrét beszédhelyzetekhez kötődő interakciókat rögzít, ahol a beszélgetéseket semmiféle instrukció nem korlátozza vagy irányítja. Néhány, a nyelvoktatás szempontjából releváns beszédhelyzet kiválasztása után a cél, hogy egy-egy beszédhelyzet elemzéséhez minél nagyobb minta álljon a rendelkezésre. Az eddigi anyaggyűjtések három beszédhelyzetet érintettek: (1) interakciók családtagok és (2) barátok beszélgetései során, valamint (3) munkahelyi környezetben. Továbbá még a tervek között szerepel egy, a szolgáltatások témájához köthető autentikus korpusz felgyűjtése. Jelenleg a gyűjtés folyamata zajlik, illetve a felvételek átíratozási és annotálási feladatai várnak a munkacsoport tagjaira.

Az irányított beszélgetések korpusza olyan tematikus interjúkat tartalmaz, melyek közvetlenül is felhasználhatók a tanítás során. E beszélgetések felvételét is már közvetlen didaktikai szándékok vezették: olyan témákhoz kötött szövegproduktumok rögzítése volt a cél, melyek a nyelvtanulási folyamatban akár ebben a formájukban is felhasználhatók. Az irányított beszélgetések három csoportba sorolhatók:

- (1) Görbe Tamás munkájának eredményeként ennek az alkorpusznak a keretében kerül feldolgozásra a Magyar Hangtár 18 órányi, több, mint négyszáz félautentikus beszélgetést tartalmazó anyaga.
- (2) Olyan félautentikus interjúk kapnak itt még helyet, amelyek tágabb értelemben a Közös Európai Referenciakeret, szűkebben véve pedig a *MagyarOK* tankönyvcsalád témái mentén folynak.
- (3) Valamint készülnek olyan irányítatlan beszélgetések is, amelyek nem előre rögzített tematika mentén haladnak.

Az alkorpuszok körül teljesnek tekinthető a TedTalk korpusz, mely rögzíti az eddig elhangzott összes magyar nyelvű Ted-prezentáció leiratát (500 000 token), és teszi ezt elérhetővé a prezentációk nyelvi eszközeinek vizsgálatához.

A munkacsoport számára az 1–4 alkorpuszok esetén a gyűjtés lesz az elkövetkezendő időszak legfontosabb feladata. Végző célunk egy olyan pedagógiai nagykorpusz létrehozása, amely minden nyelvtanár, nyelvtanuló és nyelvész számára szabadon kereshető. Mint több más pedagógiai korpusz esetében is (vö. Braun 2006, Chambers 2019), a szövegeket nem csak a keresőfelületre töltjük fel, hanem teljességükben is a felhasználók rendelkezésére bocsátjuk.

2.4. A korpuszalapú füzeteket készítő munkacsoport

Ez a munkacsoport már létező, nagy nyelvi adatbázisok, illetve a fent ismertetett korpuszok felhasználásával készít gyakorló füzeteket. A füzetek célja, hogy konkrét feladattípusokon, prezentációs módszereken keresztül adjanak példát arra, miként építhető híd a korpusznyelvészet és a szakmódszertan között, valamint láthatóvá tegyék, miért érdemes korpuszalapú példamondatokkal dolgozni a nyelvtanítás során. A sorozat hiánypótló, hiszen eddig nem sok példa volt a magyar mint idegen nyelv tanításában korpuszok felhasználására.

Első kötetünk a gyakori melléknevek jelentéseit mutatja be és gyakoroltatja A1–B1 szinten, kontextusba ágyazva. A nyelvtanuló már kezdő szinten is számos egyszerű, gyakran használt melléknévvel találkozik, például *jó, új, szép, nagy* stb. Ezek a melléknevek azért is fordulnak elő olyan gyakran, mert sokféle szövegkörnyezetben használhatók, számos jelentésük van. Jelentéseiket azonban a szótárak nem mutatják be részletesen, ráadásul nem is mindig a leggyakoribb jelentésekre helyezik a fő hangsúlyt, valamint a melléknevek használatára csak igen kevés példát kap bennük a nyelvtanuló. A *Gyakori melléknevek kontextusban* munkacímű korpuszalapú füzet – ettől eltérően – az ismert mellékneveket a mindennapokban gyakran használt funkcióban mutatja be.

Munkánk során a természetes nyelvhasználatot helyezzük előtérbe. Amennyire lehetséges, autentikus, illetve félautentikus mondatokkal dolgozunk, háttérbe szorítva vagy kizárva a mesterséges, csak a gyakorlás érdekében a szerzők által megalkotott mondatokat. A köznyelvben előforduló mindennapi kifejezések dominálnak. A gyakorlófüzet bármelyik tananyagot kiegészítheti, ugyanakkor önálló tanulásra is alkalmas.

Forrásaink a Sketch Engine nagy korpusza (huTenTen12) és az ugyanitt megtalálható *MagyarOK* nyílt korpusza, mely mindenki számára szabadon elérhető. Anyagunk bázisa még – a mindennapi kommunikáció leggyakoribb színtereiről – a közösségi oldalakon található informális közléshelyzetek kommentjei és rövid párbeszédei, amelyek a fent ismertetett élőnyelvi korpusz részét is képezik.

Munkamódszerünk a példák eredeti vagy adoptált formájában való beépítése a feladatokba, a válogatásnál már a nyelvész intuícióját is bevonva. Elsőként a korpuszokban végeztünk gyakorisági vizsgálatokat, így választottuk ki a kötetbe kerülő ötven melléknevet. Ezt követte a gyakori kollokációk feltérképezése, illetve a konkrét példamondatok kiválasztása. A kötet feladattípusainak meghatározásakor részben a korpuszalapú tanulás (*data-driven learning*) már létező feladattípusaira támaszkodtunk (Johns 1991, Boulton 2017), részben új feladatokat alkottunk. A műhelymunka része az egyes – egyénileg megírt fejezetek – értékelése és javítása. Nem célunk azonos terjedelmű részek megalkotása, ezt inkább a melléknevek gyakorisága és gyakori jelentéseinek száma fogja meghatározni. A könyvben szereplő kifejezések, kollokációk egy részét angol fordítással is ellátjuk.

A bemutatott melléknevekkel sok lehetőség van gyakorlásra, mind a kiadvány fejezeteiben, mind a digitális formában elérhető feladatok révén. A koncepció része továbbá, hogy a gyakorolt melléknevek folyamatosan, mintegy spirálszerűen építkezve vissza-visszatérnek a teljes kötetben.

A hatékony nyelvtanulás minden formáját bevonva hanganyagot is készítünk az egyes részekhez, amihez nemcsak kitűnően artikuláló, tökéletesen kidolgozott orgániumú színészeket kérnénk fel, hanem képzetlenebb hangú, kevésbé kifinomultan artikuláló személyekkel is rögzítenénk szövegeket, párbeszédeket. Mindezzel a mindennapi életben előforduló valós szituációkat prezentálnánk.

Mostanáig öt fejezet készült el, és 2021 februárjáig további öt fejezetet hozunk létre. 2021 nyarán egy intenzívebb műhelymunka keretében – terveink szerint – a teljes kézirat összeáll. A véglegesítést és a hanganyagok elkészítését 2021 decemberére tervezzük, majd ezt követi a *KorSzak-füzetek* első számának nyomdai kivitelezése.

3. Összegzés

A 2020 februárjában alakult Korpusznyelvészeti és Szakmódszertani Munkacsoport igen ambiciózus célokat tűzött maga elé: amint a neve is mutatja, a csoport munkájával hidat kíván építeni a korpusznyelvészeti eredmények – vagyis a valóban használt nyelv empirikus vizsgálata – és a MID-szakmódszertana közé. Ezt a célt bontottuk kisebb szegmensekre és lépésekre, amikor almunkacsoportokat alakítottunk. A munka az elmúlt hónapokban meg is kezdődött, és a csoportok máris jelentős mennyiségű anyagot gyűjtöttek és készítettek. Elindult az autentikus és félautentikus pedagógiai korpuszok, illetve a nyelvtanulók és gyermekek nyelvi produktumait rendszerező korpuszok építése, valamint a nagy nyelvi korpuszok alapján az újabb tananyagok létrehozása is. Reményeink szerint egyrészt olyan adatbázisokat hozunk létre, amelyeket a kollégák – kutatók, gyakorló nyelvtanárok – is használhatnak munkájukban, másrészt olyan konkrét tananyagokat készítünk, amelyek a korpuszhasználat többletértékét a gyakorlatban is láttatják.

Irodalom

- Braun, Sabine 2010. Getting past Groundhog day. Spoken multimedia corpora for student-centred corpus exploration. In: T. Harris & M. Moleno Jaén (szerk.): *Corpus Linguistics in Language Teaching*. Peter Lang, Bern. 75–98.
- Boulton, Alex 2017. Research timeline: Corpora in language teaching and learning. *Language Teaching* 50/4: 483–506.
- Chambers, Angela 2019. Towards the corpus revolution? Bridging the research–practice gap. *Language Teaching*, 52/4: 460–475.
- Hanks, Patrick 2013. *Lexical Analysis. Norms and Exploitations*. MIT Press, Cambridge, MA.
- Hoey, Michael 2005. *Lexical Priming: A New Theory of Words and Language*. Routledge, London.
- Hunston, Susan (2009). The usefulness of corpus-based descriptions of English for learners: The case of relative frequency. In: K. Aijmer (szerk.): *Corpora and Language Teaching*. John Benjamins, Amsterdam/Philadelphia. 141–157.
- Johns, Tim 1991. Should you be persuaded: Two samples of data-driven learning materials, *English Language Research Journal* 4: 1–16.
- O’Keeffe, Anne – McCarthy, Michael – Carter, Ronald 2007. *From Corpus to Classroom*. Cambridge, Cambridge University Press.

Korpuszok

- Bíró Andrea (2013) Biro Trilingual Corpus. <https://childes.talkbank.org/access/Biling/Biro.html> (2020.11.29.)
- Bodor Péter (1994) Bodor Corpus. <https://childes.talkbank.org/access/Other/Hungarian/Bodor.html> (2020.11.29.)
- Bóna Judit (2017). GABI – Gyermeknyelvi beszédatadtbázis a kutatásban. In: Új utak a gyermeknyelvi kutatásokban, 35–50. http://www.eltereader.hu/media/2017/11/Bona_Gyermeknyelv_READER.pdf (2020.11.29.)
- Görbe Tamás (2015–2020). Magyar Hangtár. <https://rs.cms.hu-berlin.de/magyar/pages/home.php>
- Gyarmathy Dorottya – Neuberger Tilda (2015). Egy hiánypótló adatbázis: a Tini BEA. Beszédkutatás 2015. 209–222. <http://real-j.mtak.hu/2676/1/Beszedkutatás2015.pdf> (2020.11.29.)
- MacWhinney, Brian (1970). MacWhinney Corpus. <https://childes.talkbank.org/access/Other/Hungarian/MacWhinney.html> (2020.11.29.)
- MacWhinney, Brian (2015). CHILDES: Child Language Data Exchange System. <https://www.lti.cs.cmu.edu/projects/spoken-interfaces-and-dialogue-processing/childes-child-language-data-exchange-system> (2020.11.29.)
- Mátyus Kinga – Orosz György (2014). MONYEK – Morfológiailag egyértelműsített óvodai nyelvi korpusz. Beszédkutatás 2014. 237–245. <http://real-j.mtak.hu/974/1/Beszedkutatás2014.pdf> Utoljára megtekintve: 2020.11.29.

Navracsics Judit (1994-1997). Navracsics Bilingual Corpus.

<https://childes.talkbank.org/access/Biling/Navracsics.html> Utoljára megtekintve: 2020.11.29.

Réger Zita (1990). Réger Corpus. <https://childes.talkbank.org/access/Other/Hungarian/Reger.html> Utoljára megtekintve: 2020.11.29.

Sketch Engine. <https://www.sketchengine.eu>

Szita Szilvia – Pelcz Katalin (2020). A *MagyarOK* tankönyvcsalád nyílt korpusza.

<https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Fmagyarok>

Baumann, Tímea – Majoros, Judit – Pelcz, Katalin – Schmidt, Ildikó – Szita, Szilvia – Vermeki, Boglárka

Introduction of the Work Group for Corpus Linguistics and Didactics (*KorSzak*)

The Work Group for Corpus Linguistics and Didactics (*KorSzak*) was created in February 2020. The group aims at building bridges between the domain of corpus linguistics and the methodology of teaching Hungarian as a foreign or second language. The four sub-groups have collected and categorised a great amount of material in the past few months. One group has been working on designing and building fully authentic and semi-authentic corpora, another on collecting children's language, and yet another on compiling a learner corpus of written and spoken linguistic products. The creation of pedagogic materials based on corpus analysis has also begun. *KorSzak* intends to build large pedagogic corpora that can be used by linguists for research purposes and by practitioners in their daily work. The teaching materials will offer concrete examples of how language teaching can be made more effective with the help of corpora.